UNIVERSIDAD DE SAN ANDRES

Seminario del Departamento de Economía

"Panel Data Estimators for Nonseparable Models with Endogenous Regressors"

Rosa Matzkin
(Northwestern University)

Martes 14 de julio de 1998
9:00 hrs
Aula Chica de Planta Baja

Panel Data Estimators for Nonseparable Models with Endogenous Regressors

Joseph Altonji and Rosa Matzkin
Department of Economics
Northwestern University

First Draft: May 1997
Revised April 1998
(preliminary and incomplete)

Abstract

1 Introduction

In this paper we develop estimators for models with nonseparable error terms and endogenous explanatory variables that can be used when the data are clustered in some way. Examples of such data are samples of siblings, panel data on a set of individuals or firms, and data on a set of individuals who grew up in the same neighborhood or attended the same high school. The methods will be applicable to a wide range of topics. Examples include the use of siblings to analyze the effects of teenage pregnancy on the probabilities of being in poverty, of being on welfare, and of ever marrying (Geronimus and Korenman (1992)), the use of children from the same high school to isolate the effects of family background on educational attainment, the use of siblings to study the effects of neighborhood characteristics on high school graduation (Aaronson (forthcoming)), and the use of siblings to study transfers of time and money to and from parents (Altonji, Hayashi and Kotlikoff (1996), Rosenzweig and Wolpin (1994 a and b). The methods cover a wide range of nonlinear models, including binary choice models.

To be more specific, consider the model (1.1)

$$(1.1)\ y_{ik} = m(x_{ik}, \varepsilon_i, u_{ik}),\ i = 1...n,\ k = 1...K_i.$$

where y_{ik} is an outcome of person k from group i , x_{ik} is a $1 \times J_1$ vector of observed variables, ε_i is an error component common to observations from group i, u_{ik} is an error term that is specific to person k of group i, and K is the number of observations in group i. In some applications, the "group" might be a family. In others, it might be a neighborhood, a school, or a firm. In cross section time series data, i might refer to an individual and k to the time period. The function $m(\cdot, \cdot, \cdot)$ may be nonseparable in x_{ik}, ε_i, and u_{ik}. The index k may be an element of x_{ik}, which means that the effect of u_{ik} and ε_i on y_{ik} may depend on sibling order in a family context or age or the time period in a cross section time series context.

To give a simple example, consider Aaronson's (forthcoming) analysis of the effects of neighborhood characteristics on college attendance. In this case i denotes a family and k a specific child. The outcome y_{ik} is 1 if person i, k started college and 0 otherwise and x_{ik} is average neighborhood income while person ik was between the ages of 10 and 16. (We abstract from other elements of x_{ik} such as the income of family i when k is growing up to simplify the exposition). The function $m(x_{ik}, e_i, u_{ik})$ takes on the value 0 or 1. The form of $m(x_{ik}, \varepsilon_i, u_{ik})$ might be

$$\begin{aligned} m(x_{ik}, \varepsilon_i, u_{ik}) &= 1 \text{ if } \Psi(x_{ik}, \varepsilon_i, u_{ik}) = x_{ik}\Gamma + \varepsilon_i + u_{ik} > 0 \\ &= 0 \text{ otherwise.} \end{aligned}$$

or

$$(1.1a)\ m(x_{ik}, \varepsilon_i, u_{ik}) = I(-x_{ik}\Gamma - \varepsilon_i < u_{ik})$$

where the function $I(.)$ is 1 if the inequality is true and 0 if it is false. (We adopt the linear form for the index function for exposition only—our methods are nonparametric and do not require that the index function Ψ be separable in x_{ik}, ε_i, and u_{ik}.)

Let $g(u_{ik}, \varepsilon_i | x_{ik})$ be the density of $(u_{ik}, \varepsilon_i | x_{ik})$. The probability that a person with characteristics x_{ik} attends college is

$$
\begin{aligned}
(1.2) E(y_{ik}|x_{ik}) &= \int_{\varepsilon_i} \int_{u_{ik}} [m(x_{ik}, \varepsilon_i, u_{ik})] g(u_{ik}, \varepsilon_i | x_{ik}) d\varepsilon_i du_{ik} \\
&= \int_{\varepsilon_i} \int_{-\Gamma x_{ik} - \varepsilon_i}^{\infty} g(u_{ik}, \varepsilon_i | x_{ik}) d\varepsilon_i du_{ik}
\end{aligned}
$$

The objective of the analysis is to estimate the expected value of the partial derivative of the probability of attending college with respect to neighborhood income, holding the distribution $g(e_i, u_{ik}|x_{ik})$ constant. Call this derivative $\beta(x_{ik})$, where

$$
(1.3a) \ \beta(x_{ik}) = \int_{\varepsilon_i} \Gamma \cdot g(-\Gamma x_{ik} - \varepsilon_i, \varepsilon_i, |x_{ik}) d\varepsilon_i
$$

in the above binary choice example. The major statistical problem in estimating $\beta(x_{ik})$ arises from the fact that neighborhood income, x_{ik} , is likely to be correlated with unobserved characteristics of families who are clustered in the same neighborhood. Standard parametric methods for binary choice models such as the probit, logit, as well as nonparametric estimators provide biased estimates of $\beta(x_{ik})$ when ε_i is correlated with x_{ik} In those methods, the estimate of $\beta(x_{ik})$ will pick up part of the effect of ε_i on y_{ik}.

Aaronson (forthcoming) attempts to get around this problem by comparing the schooling outcomes of siblings who grew up in different neighborhoods. He assumes that most family background characteristics are the same for siblings conditional on observables such as family income and marital status. He used the linear probability model with family fixed effects and the conditional logit model proposed by Gary Chamberlain (1980, 1984) to do this. Many other authors have used one or both of these methods in other contexts. Unfortunately, the linear probability model is biased in almost all circumstances. Chamberlain's conditional logit model does not estimate the parameter of interest (the population mean at a particular value of x_{it} of the effect of x_{it} on the mean of y for the) and does not use information on groups (e.g., siblings) in which all members have the same value for y_{ik}. There is no suitable estimation method in the literature.

To give a second special case to which our methods apply, consider the model

$$(1.4)\ y_{ik} = m(x_{ik}, \varepsilon_i, u_{ik}) = x_{ik}\beta_1 + H(x_{ik}, \varepsilon_i) + \varepsilon_i + u_{ik}$$

where $E(\varepsilon_i|x_{ik}) \neq 0$ and H is differentiable. Since m is differentiable in x_{ik} the parameter of interest may be written as

$$(1.3)\ \beta(x_{ki}) = \int_{\varepsilon_i} \int_{u_{ki}} [m_{xk}(x_{ki}, \varepsilon_i, u_{ki})] g(u_{ki}, \varepsilon_i | x_{ki}) d\varepsilon_i du_{ik}\ .$$

There is a huge literature that assumes that $H(x_{ik}, \varepsilon_i)$ is 0 and deals with the correlation between ε_i and x_{ik} by controlling for ε_i with a group specific intercept. These "fixed effects" estimators don't work when the impact of x_{ik} on y_{ik} depends on ε_i, and alternatives in the literature require strong assumptions about the form of H and the distributions of the error terms.

In this paper we propose two estimators for these problems as well as a wide class of other panel data models involving nonseparable error terms and endogenous regressors. For simplicity let $K_i = K$ for all groups i, and define x_i to be the vector $[x_{i1}, ..., x_{iK}]'$. Both estimators are based on the assumption that the distribution of u_{ik} and ε_i conditional on $x_{i1}...x_{iK}$ is exchangeable in $[x_{i1}...x_{iK}]$. By "exchangeable" we mean $g(u_{ik}, \varepsilon_i | x_{i1}...x_{iK})$ does not depend on the order in which the x_{ik} are entered into the function $g(u_{ik}, e_i | x_{i1}...x_{i_K})$. That is,

$$\begin{aligned} \textit{Assumption A1.1} \quad &: \quad g(u_{ik}, \varepsilon_i | x_{i1}...x_{i_K}) = g(u_{ik}, \varepsilon_i | x_{ik_1}, x_{ik_2}...x_{ik_K}) \text{ for} \\ k_j \quad &\in \quad \{1, 2.., K\}, k_j \neq k_{j'} \qquad (1) \end{aligned}$$

For example, the assumption implies that $g(u_{ik}, \varepsilon_i | x_{i1}...x_{iK}) = g(u_{ik}, \varepsilon_i | x_{iK}...x_{i1})$. In neighborhood and sibling applications the assumption that the value of the function $g(u_{ik}, \varepsilon_i | x_i)$ is the same regardless of the order in which the x_{ik} are entered into x_i is a natural one provided that the elements of x_{ik} are measured at the same age for each child.

Our first estimator is based on the conditional expectation function $E(y_{ik}|x_i))$, and for this reason we refer to it as the "Regression Estimator". The variable x_{i1} has a direct impact on y_{i1} through the function $m(.)$ and an indirect impact by shifting the distribution of ε_i and u_{i1}. $[x_{i2}..x_{iK}]$ only have an indirect impact through their effects on the distribution of ε_i and u_{i1}. However, exchangeability restricts the distribution of u_{i1} and ε_i to depend on x_i only through a vector of L exchangeable functions $z(x_i)$ of x_i. This implies that $g(u_{ik}, \varepsilon_i|x_{i1}...x_{iK}) = g(u_{ik}, \varepsilon_i|z(x_i))$ and that the mean $E(y_{ik}|x_i)$ can be written as $E(y_{ik}|x_{i1}, z(x_i))$. If the $z(.)$ vector of functions is known and (2) x_{ik} and the elements of the vector $z(x_i)$ vary sufficiently, then the partial derivative $\partial E(y_{ik}|x_{i1}, z(x_i))/\partial x_{i1}$ is identified. Since it is possible to estimate the distribution of $z(x_i)$ conditional on x_{i1}, one can recover the parameter of interest $\beta(x_{i1})$ from the estimator of $\partial E(y_{ik}|x_{i1}, z(x_i))/\partial x_{i1}$by integrating this derivative over the distribution of $z(x_i)$ conditional on x_{ik}.

The second estimator also relies heavily on A1.1 but involves a somewhat different set of assumptions. It does not use or require knowledge of the $z(\cdot)$ functions, but it does require some additional assumptions that we discuss below. The most important is that $m(x_{ik}, \varepsilon_i, u_{ik})$ is of the form $m(x_{ik}, e_{ik})$ and is strictly monotone in e_{ik},where $e_{ik} = \Upsilon(\varepsilon_i, u_{ik})$ and $\Upsilon(.,.)$ is a continous function. The strict monotonicity assumption rules out qualitative choice models but covers many models that take on the form of (1.4). We show that $m(x_{ik}, e_{ik})$ and $g(e_{ik}|x_{ik})$ are identified under exchangeability of $g(\varepsilon_i|x_i)$ from knowledge of the joint distribution of y_{ik} and x_i. With knowledge of $m(x_{ik}, e_{ik})$ and $g(e_{ik}|x_{ik})$ one can estimate the average response $\beta(x_{ik})$ as well as other parameters that characterize the distribution of the response of y_{ik} to a change in x_{ik}.

The basic intuition underlying the second estimator is as follows. Suppose that $K = 2$, with $k = 1, 2$. A shift in x_{i1} alters the distribution of $y_{ik} = m(x_{i1}, e_{ik})$ by shifting $m(x_{i1}, e_{ik})$ for a given value of e_{i1} and by shifting the distribution e_{ik}. A shift in x_{i1} alters the distribution of $m(x_{i2}, e_{ik})$ only by shifting the distribution of e_{ik}. Consequently, one can isolate the direct effect of x_{i1} on the distribution of $m(x_{i1}, e)$ by comparing the change in the distribution of $m(x_{i1}, e_{ik})|\ x_{i1}, x_{i2}$ as x_{i1} changes to the change in the distribution of $m(x_{i2}, e_i)|x_{i2}, x_{i1}$ as x_{i1} changes.

The theoretical literature on estimating nonseparable panel data models when the regressors are correlated with the error term is relatively small. (See Powell (1994)). An exception is the recent independent paper by Abrevaya (1997) which deals with generalized regression models with fixed effects. Abrevaya's approach permits estimation of slope parameters up to scale but, in contrast to our approaches, does not permit estimation of the partial effects of x_{ik} on mean of y_{ik}. In the case of qualitative response models we have already mentioned the linear probability model with fixed effects and the conditional logit model. The conditional logit model and the other "fixed effects" approaches that we are aware of are restricted to specifications that take on the additively separable form of (1.1a). [1] The fixed effects probit model is sometimes used to estimate Γ up to scale. It is well known that the fixed effects probit model is inconsistent when the group size is fixed, but Heckman (1981) provides monte carlo evidence suggesting that the bias is small when K is on the order of 10.[2]

Manski (1987) provides a way to estimate Γ up to scale under more general assumptions than the conditional logit. He places no restrictions on the distribution of ε_i and assumes that the distribution of $u_{i1}|\varepsilon_i, x_{i1}, x_{i2}$, is the same as the distribution of $u_{i2}|\varepsilon_i, x_{i1}, x_{i2}$. He proposes a maximum score estimator that exploits that fact that $sgn(E((y_{i2} - y_{i1}) \,|x_{i1}, x_{i2}) = sgn(x_{i2}\Gamma - x_{i1}\Gamma)$ where $sgn(.)$ is -1 if the argument is negative and 1 if it is positive. In contrast to Manski's estimator, our approach requires appriori information about the distribution of $\varepsilon_i|x_{i1}, x_{i2}$. However, it permits us to estimate the partial effect of x_{ik} on the probability that y_{ik} is 1 as well as the parameter vector Γ up to scale. Furthermore, our "regression" approach can handle qualitative choice models cases in which x_{ik} and the error components interact in arbitrary ways while the other approaches in the literature cannot.

We should point out however that the estimators in there current form cannot accomodate dynamics in the model, which are addressed in recent papers by Honore and Kyriazidou (1997) and Kyriazidou (1997).

The conditional logit and the fixed effects probit estimators may be thought of as parametric "fixed effects" approaches. In addition, Chamberlain (1984) discusses parametric random effects approaches to estimating Γ up to scale in (1.3a). This approach has been used in a number of applications, including Jakubsen (19??). Assume that u_{ik} is normal, identically

[1]The most common method in empirical studies is the linear probability model with fixed effects, which forces one to maintain that the probability of y is the sum of e_i and a function of x_{ik}.

[2]Heckman and MaCurdy (1980) apply this estimator as well as the fixed effects Tobit estimator to the analysis of life cycle labor supply. Note that one can recover an estimate of the partial effect of x_{ik} on the probability that y_{ik} is 1 from the probit coefficients and the distribution of ε_i given x_{ik} . However, the MLE estimates of ε_i are unbounded when y is the same for all group members, so one cannot obtain an estimate of the distribution of $\varepsilon_i|x_{ik}$ without making assumptions about this distribution. The same is true in the case of the conditional logit.

distributed across k,and independent of x_i and that ε_i is the sum of a function of $f(x_i;\theta)$ plus a normally distributed error term that is independent of x_i. Then one can estimate Γ up to scale by adding $f(x_i;\theta)$ to a probit model for each k and jointly estimating θ and Γ while imposing cross restrictions across the models for each k. One may also recover the partial effect of x_{ik} on the probability y_{ik} that 1. The main disadvantages of this approach relative to ours is that it requires the assumption that u_{ik} and $\varepsilon_i|x_i$ are normal and the assumption of additive separability, as in (1.3a)

In the case of continuous variables, the incidental parameters problem limits the utility of parametric "fixed effects" approaches for models such as (1.4). In special cases, parametric random effects approaches may be available. GMM is often used to estimate the parameters of nonseparable models and it may be possible in some cases to estimate elements of β_1 or some parameters of the $H(x_{ik},\varepsilon_i;\theta)$ when that function is parametric. However, there are many cases in which this method cannot be used to estimate the partial effect of x_{ik} on the mean of y_{ik}.[3]

The paper continues in section 2, where we present the "Regression" estimator based on $E(y_{ik}|x_i)$). In section 3 we discuss a nonparametric version of the estimator and analyze its asymptotic distribution. In section 4 we discuss an extension of the estimator to the case in which x_{ik} is correlated with u_{ik} conditional on $x(x_i)$ but an instrumental variable is available. This estimator provides an alternative to the "fixed effects-IV" linear probability model that is sometimes used in applied studies even though it is inconsistent. In section 5 we derive the second estimator and provide results on its asymptotic properties. In section 6 we present some very encouraging monte carlo evidence on the performance of the "Regression Estimator. In section we provide some concluding remarks.

[3]Thus far, neither of our estimators cover other limited dependent variables models such as the censored regression models or sample selection models. Honore (1992) provides a fixed effects estimator for the limited dependent variables case. Kyriazidou (1997) uses an exchangeability assumption that is similar to ours in her work on panel data sample selection models. The approaches in both of these papers are based on differencing the observations in clever ways and are quite distinct from our approaches.

2 An Estimator Based on $E(y_{ik}|x_i)$

In this section we present our regression based estimator. The estimator uses exchangeable functions $z^1(x_{i1}...x_{iK}), z^2(x_{i1}...x_{iK}), ..., z^L(x_{i1}...x_{iK})$ of $(x_{i1}...x_{iK})$ satisfying that property that for all $x_{i1}...x_{iK}$, $g(u_{ik}, \varepsilon_i|x_{i1}...x_{iK}) = g(u_{ik}, e_i|z_i^1, ..., z_i^L)$. By exchangeable we mean that the functions are invariant to the order in which the elements of $(x_{i1}...x_{iK})$ enter the function. For example, z_i^1 might be the mean of $x_{i1}..x_{iK}$ for family i and z_i^2 might be the average over k of $(x_{ik} - z_{i1})^2$. As we noted in the introduction, assumption (A1.1) that $g(u_{ik}, \varepsilon_i|x_{i1}...x_{iK})$ is exchangeable in $x_{i1}...x_{iK}$ means that without loss of generality we can $g(u_{ik}, \varepsilon_i|x_{i1}...x_{iK})$ as $g(u_{ik}, e_i|z_i^1, ..., z_i^L)$, where $z_i^1, ..., z_i^L$ are symmetric functions $z^1(x_{i1}...x_{iK}), z^2(x_{i1}...x_{iK}), ..., z^L(x_{i1}...x_{iK})$ of $(x_{i1}...x_{iK})$. Let z_i be the vector of z_i^ℓ variables for family i. The first estimator requires the following assumptions in addition to (A*1.1*)

Assumption 2.1. The functions that define z_i in terms of x_i are known.

Assumption 2.2. The distributions of each element of the vector x_{ik}, z_i^1,...., z_i^L conditional on the other elements of the vector are nondegenerate.

With these assumptions one may estimate $E(y_{ik}|x_{ik}, z_i)$ nonparametrically. (As we discuss below, Assumption 2.2 can be weakened appriori information about the functional form of $E(y_{ik}|x_{ik}, z_i)$ is available.) The Regression estimator of $\beta(x_{ik})$ is based on the conditional expectation function $E(y_{ik}|x_{ik}, z_i)$. Suppressing the i subscript where it is not needed for clarity and setting k to 1 for concreteness, this function is

$$E(y_1|x_1, z) = \int_\varepsilon \int_{u_1} m(x_1, \varepsilon, u_1) g(u_1, \varepsilon|x_1, z) d\varepsilon, du_1 \qquad (2)$$

The idea of the estimator is to recover $\beta(x_1)$ from

$$(2.2)\ E_{x_1}(y_1|x_1, z) \equiv \frac{\partial}{\partial x_1} E(y_1|x_1, z)$$

and $h(z|x_1)$, the conditional distribution of $z|x_1$, The distribution $h(z|x_1)$ can be estimated from the observations on x_1 for the cross section of groups i. The derivative with respect to x_1 is :

$$(2.3)\ E_{x_1}(y_1|x_1, z) = \int_{u_1} \int_\varepsilon m_{x1}(x_1, \varepsilon, u_1) g(u_1, \varepsilon|x_1, z) d\varepsilon\, du_1$$

because

$$\int_{u_1}\int_{\varepsilon} m(x_1,\varepsilon,u_1)\; g_{x1}(u_1,\varepsilon|x_1,z)d\varepsilon\, du_1 = 0$$

The form of (2.3) is analogous to (1.3a) in the binary choice case. In this case

$$(2.4)\ E_{x_1}(y_1|x_1,z) \ = \int_{\varepsilon} \ \Gamma \cdot g(-\Gamma x_1 - \varepsilon, \varepsilon|x_1,z)d\varepsilon.$$

Note that $E_{x_1}(y_1|x_1,z)$ differs from $\beta(x_1)$ because the distribution of u_1 and ε is conditioned on both x_1 and z_i. (See the right hand side of (2.3) or (2.4).) However, one may integrate out z to obtain $\beta(x_1)$ from $E_{x_1}(y_1|x_1,z)$. To see how to do this, note first that

$$(2.5)\ g(u_1,\varepsilon|x_1) = \int_z g(u_1,\varepsilon|x_1,z)h(z|x_1)dz$$

where $h(z|x_1)$ is the conditional density of z given x_1. Multiply both sides of (2.3) by $h(z|x_1)$ and integrate over the range of z. This yields (2.6)

$$(2.6)\ \int_z E_{x_1}(y_1|x_1,z)h(z|x_1) = \int_z\int_{u_1}\int_{\varepsilon} m_{x1}(x_1,\varepsilon,u_1)g(u_1,\varepsilon|x_1,z)h(z|x_1)\, d\varepsilon\, du_1\, dz$$

.

Re-arranging the order of integration on the right hand side of the equality and using (2.5) establishes that the right hand side is $\beta(x_1)$, the function we would like to estimate. That is,

$$(2.7)\ \beta(x_1) = \int_z E_{x_1}(y_1|x_1,z)h(z|x_1)dz$$

The above equation forms the basis of our first estimator. The estimator is obtained by substituting parametric or nonparametric estimators of the components of the right hand side of (2.7) into the equation.[4] In the next section we provide the asymptotic distribution theory for a nonparametric approach in which kernel estimators of $E_{x_1}(y_1|x_1,z)$ and $h(z|x_1)$ are used.

[4]When K, the number of observations per group, differs across i, one could do the estimation for each group size and then combine the estimates.

The key assumption, aside from exchangeability, is that $E(y_1|x_1, z)$ is identified conditional on prior information about how x_1 and z enter the expectation function. The expectation function is identified nonparametrically only if x_1 varies conditional on z. The conditions for identification involve trade-offs among (a) the size of the panel K, (b) the number L of elements in z, and (c) parametric or nonparametric restrictions on $E(y_1|x_1, z)$. For example, consider the case in x_k is a scalar and $K = 2$ (i.e.e, there are observations on 2 children per family). In this case, we conjecture that variables that are continuous exchangeable functions of the elements of $x_i = \{x_{i1}, x_{i2}\}$ may be approximated arbitrarily closely be functions of $z_i^1 = x_{i1} + x_{i2}$ and $z_i^2 = |x_{i1} - x_{i2}|$.[5]If the conjecture is correct, then conditioning the distribution of $(\varepsilon_i,\ u_{ik})$ on z_{i1} and z_{i2} is general enough. However, $z_i^2 = |2x_{i2} - z_{i1}|$. This dependence among z_i^1, z_i^2, and x_{i1} means that

$$E(y_1|x_1, z_1, z_2)$$

is not identified nonparametrically when K is 2. However, if the function

$$E(y_1|x_1, z_1, z_2)$$

is a low order polynomial in the three variables or z_2 does not enter at all, then it may be identified. It is also possible that the function will be identified over

[5](Preliminary) A sketch of a proof of this conjecture follows. Let $z_i^3 = x_{i1}x_{i2}$. $z_i^3 = .25\{(z_i^1)^2 - (z_i^2)^2\}$. Consider functions of the form $h_1(\Lambda_1(x_1) + \Lambda_1(x_2))$. Assuming $\Lambda_1(x_I)$ is continuous, then it can be approximated arbitrarily closely by a polynomial in $\Lambda_1(x_1)$. Consequently, $h_1(\Lambda(x_1) + \Lambda(x_2)) \simeq h_1(a_0 + a_1(x_1 + x_2) + a_2(x_1^2 + x_2^2) + a_3(x_1^3 + x_2^3) + ...)$ where we have stopped at three terms for simplicity. It is tedious but straightforward to show that one can express terms $x_1^j + x_2^j$as a linear combination of powers of z_i^1, powers of z_i^3, and products of powers of the two. Thus, functions of the form $h_1(\Lambda_1(x_1) + \Lambda_1(x_2))$ can be approximated arbitrarily closely by functions of z_i^1 and z_i^2. Functions of the form $h_2(|\Lambda_2(x_1) - \Lambda_2(x_2)|) \simeq h(|a_1(x_1 - x_2) + a_2(x_1^2 - x_2^2) + a_3(x_1^3 - x_2^3) + ...|)$. The sum of the first three terms inside the absolute value sign is equal $|(x_1 - x_2)| \cdot |a_1 + a_2(x_1 + x_2) + a_3(x_1^2 + x_2^2) + a_3(x_1 x_2)|$. The first term in the product is z_i^2. The components of the second term may be expressed as functions of z_i^1, z_i^2 and powers of z_i^3. Exchangeable functions of the form $h_3(\Lambda_3(x_1) \cdot \Lambda_3(x_2))$ can be expressed as functions of the three z variables in similar fashion by working with a polynomial approximation to $\Lambda_3(x_1) \cdot \Lambda_3(x_2)$. We conjecture that any continuous exchangeable function of x_1 and x_2 must be a composite function of the terms involving $h_1(\Lambda_1(x_1) + \Lambda_1(x_2))$, $h_2(|\Lambda_2(x_1) - \Lambda_2(x_2)|)$, and $h_3(\Lambda_3(x_1) \cdot \Lambda_3(x_2))$. If this is true, then we conclude that any continous exchangeable function of x_1 and x_2 may be written approximated arbitrarily closely by a function of z_i^1 and z_i^2. We do not know if there is a useful generalization of this result when K > 2.

some ranges of x where there is variation in x conditional on z but not others. Note that when K is greater than 2, it is possible to test restrictions on the dimensionality of z. For example, when $K = 4$ one can test the hypothesis that the distribution of $(\varepsilon, u_k)|x_i$ depends only on $z_{i1} = (x_{i1}+x_{i2}+x_3+x_k)/4$ and $\sum |x_{ik} - z_{i1}|$. Finally, note that it is necessary to identify the effects of $z_1$and z_2 on the mean of y_1 conditional on x_1 only over the range in which the conditional density $h(z_1, z_2)$ is positive since it is only these values that enter into (2.7).

2.1 Discussion

One very attractive feature of the estimator in the binary choice case compared to the conditional logit or fixed effects probit estimators is that it can utilize groups in which y_{ik} is either 1 or 0 for all k. It is quite common panel data applications, particulary for rare events, that all group members have the same value for y_{ik}. [6]

A second very attractive feature of the regression estimator is that it only requires that data on the dependent variable y_{ik} be available for one member of group i, although data on x_{ik} must be available for at least 2 members of group i. In contrast, the conditional logit estimator and standard fixed effects estimators require data on y_{ik} and x_{ik} for at least 2 group members. Consequently, data on children as young as 16 can be included in studies of the effects of neighborhood characteristics during childhood on outcomes that occur later in life, such as college graduation or marriage. This will substantially increase the sample sizes for sibling studies. To provide a bit of intuition for why one only needs data on y_{ik} for one member of group i as well as the intuition underlying the Regression estimator, it is helpful to relate it to other panel data estimators for the standard separable case. Note that the standard linear regression model with an additive family fixed effect is a special case of our model (1.1). Consider the model

$$(2.8)\ y_{ik} = x_{ik}\beta + \varepsilon_i + u_{ik}$$

[6] The regression estimator can also be applied to multinomial models. A simple way to do this is to treat each outcome as a seperate 0-1 variable, and estimate $E_{x_{ik}}(y_{ik}|x_{ik}, z_i)$, impose the adding up constraint, and integrate out z_i.

. Chamberlain (1984) and Mundlak (1978) point out that the parameter β, which is the effect of $x_{ik} on y_{ik}$ holding ε_i constant, may be estimated by using the decomposition of $\varepsilon_i$into its least squares linear projection on the elements of x_i and the orthogonal error term v_i to eliminate ε_i from the above equation, and using the K observations on group i to estimate the system of equations

$$(2.9)\ y_{ik} = x_{ik}\beta + x_i\lambda + u_{ik} + v_i,\ \ k = 1...K$$

with cross equation restrictions imposed. This does not require the assumption of exchangeability. The assumption of exchangeability places restrictions on the coefficient vector λ summarizing the relationship between ε_i and the elements of x_i. In this case our regression based estimation procedure would amount to running the regression

$$(2.10)\ y_{ik} = x_{ik}\beta + f(z_i; \beta_1) + u_{ik} + v_i$$

where z_i is a vector of exchangeable functions of x_i and $f(.; \beta_1)$ is a function with parameter vector β_1, such as a polynomial. In the above model $m_{x_{ik}}(x_{ik}, u_{ik}, \varepsilon_i)$, the partial derivative of y_{ik} with respect to x_{ik} is a constant β, so it is not necessary to integrate out over the distribution $h(z_i|x_{ik})$). A special case is when z_i only contains z_i^1, the sum of the elements of x_i and $f(.;.)$ is linear. In this case, (2.10) is equivalent to (2.9) with the restriction that the elements of λ are all equal. With these restrictions one does not need to have data on all of the y_{ik} to identify β from (2.8).

3 Asymptotic Properties of the Regression Estimator in the Nonparametric Case

The nonparametric version of the estimator introduced in Section 2 is given by

$\hat{\beta}(x) = \int \frac{\partial}{\partial x}\hat{E}(y|x,z)\ \hat{h}(z|x)\ dz,$

where $\hat{E}(y|x,z)$ is a kernel estimator of the conditional expectation of Y given (X, Z) and $\hat{h}(z|x)$ is a kernel estimator of the conditional pdf of

Z given X. We suppress the i subscript and k subscripts on y_{ik} and x_{ik} and the i subscript on z_i.

To derive the asymptotic properties of $\hat{\beta}(x)$, we note that $\hat{\beta}(x)$ is a functional of the kernel estimator, $\hat{F}(y,x,z)$, for the joint cdf $F(y,x,z)$ of (Y,X,Z). Hence, a "delta-method" such as the ones described in Newey (1994) or Ait-Sahalia (1992) can be used to derive its asymptotic distribution.

Let K_1 denote the dimension of x and K_2 denote the dimension of z.

Let $d = 1 + K_1 + K_2$, and let f denote the pdf of (y,x,z). Then,

$\hat{F}(y,x,z) = \int_{-\infty}^{y} \int_{-\infty}^{x} \int_{-\infty}^{z} \hat{f}_N(t_y,t_x,t_z)\, dt_y\, dt_x\, dt_z$

where

$\hat{f}_N(t_y,t_x,t_z) = \frac{1}{N\sigma_N^d} \sum_{i=1}^{N} K(\frac{u_y-t_y}{\sigma}, \frac{u_x-t_x}{\sigma}, \frac{u_z-t_z}{\sigma})$ for all $(t_y,t_x,t_z) \in R^d$.

The following assumptions will be needed:

ASSUMPTION 1: The sequence $\{y_i, x_i, z_i\}$ is a strictly stationary $\beta-$mixing sequence satisfying $k^{\delta}\beta_k \to 0$ as $k \to \infty$, for some $\delta > 1$.

ASSUMPTION 2: $f(y,x,z)$ has a compact support $\Theta \subset R^d$, $f(\partial\Theta) = 0$ where $\partial\Theta$ denotes the boundary of Θ, and $f(y,x,z)$ is continuously differentiable up to order g where $g \geq 2d$.

ASSUMPTION 3: The kernel function $K(\cdot,\cdot,\cdot)$ is an even function, integrates to 1, is of order r where r is an even integer satisfying $d/2 < r < g$, is continuously differentiable up to order g and its derivatives of order up to g are in $L^2(R^d)$. I.e.,

(i) $K(-y,-w,-z) = K(y,w,z)$ for all $(y,w,z) \in R^d$,

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K(y, w, z)\, dy\, dw\, dz = 1$,

(iii) for some even integer r such that $d/2 < r < g$,

(a) $\forall \lambda \in N^d$ such that $\sum_{k=1}^{d} \lambda_k \in \{1, 2, ..., r-1\}$

$\int_{-\infty}^{\infty} t_1^{\lambda_1} t_2^{\lambda_2} \cdots t_d^{\lambda_d}\, K(t)\, dt = 0, \qquad$ where $t = (t_1, ..., t_d)$

(b) $\exists \lambda \in N^d$ such that $\sum_{k=1}^{d} \lambda_k = r$ and

$\int_{-\infty}^{\infty} t_1^{\lambda_1} t_2^{\lambda_2} \cdots t_d^{\lambda_d}\, K(t)\, dt \neq 0$, and

(c) $\int_{-\infty}^{\infty} \|t\|^r \; |K(t)|\; dt < \infty$, and

(iv) $K(\cdot, \cdot, \cdot)$ is continuously differentiable on R^d up to order g and its derivatives of order up to g are in $L^2(R^d)$.

ASSUMPTION 4: As $N \to \infty$, $\sigma_N \to 0$, $N\sigma_N^d \to \infty$, $\sqrt{N}\sigma_N^r \to 0$, and $\sqrt{N}\sigma_N^{2m} \to \infty$ where m is an integer such that $m < r/2$ and $m + r \leq g + d$.

ASSUMPTION 5: $\infty > f(x) > 0$, and $\int\int f_x^2(y,x,z)\, dydz$, $\int \frac{\left(\int f_x(y,x,z)\, dy\right)^2}{f(x,z)^8} dz$, $\int \frac{\left(\int y\, f(y,x,z)\, dy\right)^2}{f(x,z)^8} dz$, $\int \frac{\left(\int f_x(y,x,z)\, dy\right)^2 \left(\int y\, f(y,x,z)\, dy\right)^2}{f(x,z)^{10}} dz$, $\int \frac{1}{f(x,z)^2} dz$, $\int y\, f_x(y,x,z)\, dydz$, $\int \frac{\left|\int f_x(y,x,z)\, dy\right|}{f(x,z)^2} dz$, $\int \frac{\left|\int y\, f(y,x,z)\, dy\right|}{f(x,z)} dz$, $\int \frac{\left|\int f_x(y,x,z)\, dy\right| \left|\int y\, f(y,x,z)\, dy\right|}{f(x,z)} dz$, are all bounded.

Theorem 1 *1: If Assumptions 1-5 are satisfied, then $\hat{\beta}(x)$ is a consistent estimator of $\beta(x)$ and*

$$\sqrt{N}\, \sigma_N^{(K_1/2)+1} \left(\hat{\beta}(x) - \beta(x)\right) \to N(0, V) \text{ in distribution}$$

where

$$V = \{\int Var\,(y|x,z)\ f(x,z)\ dz\} \left\{\int \left[\frac{\partial}{\partial x}\left(\int\int K(y,x,z)dydz\right)\right]^2 dx\right\}\ \frac{1}{f(x)^2}$$

PROOF: See the Appendix.

4 An Extension: Correlation between u_{ik} and x_{ik} conditional on z_i.

In some applications, x_{ik} will be correlated with the idiosyncratic error component u_{ik} even after one conditions on z_i. It is common in panel data applications involving continuous dependent variables with additive error terms such as (2.8) to use an instrumental variable approach to deal with this problem while at the same time adding group specific intercepts to control for ε_i or to use the class of estimators discussed in Hausman and Taylor (1982). Unfortunately, this approach is not available in the case of nonseparable models. Here we extend the Regression estimator to handle correlation between u_{ik} and x_{ik} when an instrumental variable A_{ik} is available. We modify the above model by dropping A1.1 and replacing it with the following assumptions.

(A1.1a) $x_{ik} = \Upsilon(A_{ik}) + \xi_{ik}$, where $\Upsilon(A_{ik})$ is independent ξ_{ik}

(A1.1b) A_{ik} is independent of ε_i, u_{ik} conditional on z_i, ξ_{ik}.

Note that the correlation of x_{ik} with u_{ik} comes from ξ_{ik}. Since x_{ik} and A_{ik} are both observed, it possible to consistently estimate $\Upsilon(A_{ik})$ and ξ_{ik}, particularly if $\Upsilon(A_{ik})$ has a finite number of parameters. Given these facts, we modify the approach to estimation underlying (2.7) by working with the $E_{x_{ik}}(y_{ik}|x_{ik}, z_i, \xi_{ik})$ rather than $E_{x_{ik}}(y_{ik}|x_{ik}, z_i)$.

Suppressing the i subscript and set k to 1.

$$\text{(i) } E(y_1|x_1, z, \xi_1) = \int_{u_1}\int_{\varepsilon} m(x_1, \varepsilon, u_1) g(u_1, \varepsilon|x_1, z, \xi_1) d\varepsilon\, du_1$$

Since $\Upsilon(A_{ik})$ is independent of (ε, u_k),

$$\text{(ii) } \frac{\partial g(u_1, \varepsilon|x_1, z, \xi_1)}{\partial x_1} = \frac{\partial g(u_1, \varepsilon|x_1, z, \xi_1)}{\partial \Upsilon(A_1)} = 0$$

Using (i) and (ii) leads to

$$\text{(iii)}\ E_{x_i}(y_1|x_1, z, \xi_1) = \int\limits_{u_1}\int\limits_{\varepsilon} m_{x_1}(x_1, \varepsilon, u_1) g(u_1, \varepsilon|x_1, z, \xi_1) d\varepsilon\, du_1$$

$E_{x_1}(y_1|x_1, z)$ differs from $\beta(x_1)$ because the distribution of u_1 and ε is conditioned on z and $\xi_1$as well as x_1. However, one may integrate out z and ξ_1 to obtain $\beta(x_1)$ from $E_{x_1}(y_1|x_1, z)$. To see how to do this, note first that

$$\text{(iv)}\ g(u_1, \varepsilon|x_1) = \int\limits_{z}\int\limits_{\xi_1} g(u_1, \varepsilon|x_1, z) h(z, \xi_1|x_1) dz$$

where $h(z, \xi_1|x_1)$ is the conditional density of z given x_1. Following the approach above, multiply both sides of (iii) by $h(z, \xi_1|x_1)$ and integrate over the range of z and ξ_1. This yields (v)

$$\text{(v)}\ \int\limits_{z}\int\limits_{\xi_1} E_{x_1}(y_1|x_1, z, \xi_1) h(z, \xi_1|x_1) d\xi dz$$

$$= \int\limits_{z}\int\limits_{\xi_1}\int\limits_{u_1}\int\limits_{\varepsilon} m_{x1}(x_1, \varepsilon, u_1) g(u_1, \varepsilon|x_1, z) h(z|x_1)\, d\varepsilon\, du_1\, d\xi dz$$

.

Using (iv) and re-arranging the order of integration on the right hand side of the equality establishes that the right hand side is $\beta(x_1)$, the function we would like to estimate.

The key assumption is (A1.1b), but this assumption is very unlikely to hold when ξ_{ik} is correlated with ε_i and (A1.1a) does not hold. Altonji and Ichimura (1997) take a similar approach to treatment of endogenous explanatory variables in the context of nonseparable linear dependent variables models. As they point out, the conditions that A_{ik} is independent of ξ_{ik} and independent of ε_i and u_{ik} are much stronger than the usual conditions on instrumental variables for IV estimators. However, it should be kept in mind that IV estimators of partial derivatives are inconsistent in models such as (1.4), where slope coefficients are random and correlated with the endogenous variable.

5 Estimating the effect of x_{i1} on y_{i1} from the joint distribution of y_i and x_i.

Our second estimator uses the entire distribution of y_{i1} given x_i rather than just the conditional expectation function . We show that under certain assumptions it is possible to identify $m(x_{ik}, \varepsilon_i, u_{ik})$ and $g(u_{ik}, \varepsilon_i|x_i)$ from the joint distribution of $y_{ik}|x_i$ and the distribution of $y_{ik}|x_{ik}$. Consequently, various functions of $m(x_{ik}, \varepsilon_i, u_{ik})$ and $g(u_{ik}, \varepsilon_i|x_i)$, including averages such as $\beta(x_{i1})$, are identified. Our proof of identification is a constructive proof and, hence, it provides a way of estimating $m(x_{ik}, \varepsilon_i, u_{ik})$ and $g(\varepsilon_i, u_{ik}|xi)$ from a nonparametric estimator for the joint distribution of y_i and x_i.

To simplify the notation we will sometimes suppress both the k and i subscripts and use y to refer to y_{i1} , x to refer to x_{i1} , and u to refer to u_{i1}.The model underlying the second approach to estimation is described by the following assumptions:

(A5.1) There exists a function $\Upsilon(\varepsilon, u)$ such that $y = m(x, e)$, where $e = \Upsilon(\varepsilon_i, u_i)$

Let $q(e|x_1, x_2)$ denote the conditional density of given (x_1, x_2).

(A5.2) $\forall w, w' \quad q(e|w, w') = q(e|w', w)$.

(A5.3) $\forall x \quad m(x, \cdot)$ is strictly increasing in e.

(A5.4) $q(e|w, w')$ is strictly positive everywhere.

Assumption (A5.1) states that $m(\cdot.\cdot)$ is weakly separable in x and a function $\Upsilon(\varepsilon, u)$ of ε and u. We did not need this restriction for the regression estimator, but it is also assumed to hold in most nonlinear panel data models in the literature, such as the probit and logit binary choice models. Assumption (A5.2) states that the conditional distribution of e is exchangeable in x_1 and x_2. If, for example, $e = \Upsilon(\varepsilon, u) = \varepsilon + u$, the conditional distribution of ε is exchangeable in x_1 and x_2, and u is distributed independently of x and

ε, then, Assumption (A5.2) is satisfied. [7]

The strict monotonicity assumption (A5.3) is not required for the regression estimator based on (2.7) but it seems to be critical for the identification of $m(x_{ik}, e_{ik})$ and $g(e_{ik}|x_{ik})$. As we have noted in the introduction, strict monotonicity of m in e restricts the analysis rules out qualitative choice models.[8] On the other hand, this second estimator has a number of advantages over the regression estimator. First, it does not require the knowledge of or the use of the z functions, which are required for the regression estimator. Second, it does not require that x and the relevant z functions vary independently for nonparametric identification. Finally, it permits one to estimate $m(x_{ik}, e_{ik})$ and $g(e_{ik}|x_{ik})$ and various functions of them that include but are not limited to $\beta(x_{ik})$. Thus the two approaches have different strengths and weaknesses and are complementary.

We adopt the following normalization:

$(A5.5)\ \ m(0, e) = e.$

This is innocuous because the assumption of strict monotonicity of m in e implies that, given any function $m(\cdot,\cdot)$, one can define a new function $m'(\cdot,\cdot)$ by $m'(x,\tilde{e}) = m(x, m^{-1}(0,\tilde{e}))$, where $m^{-1}(0,\cdot)$ denotes the inverse function of m with respect to e, when $x = 0$. From the definition of m' it follows that for all $\tilde{e}$, $m'(0,\tilde{e}) = \tilde{e}$. Moreover, since for all x and all $\tilde{e}$ and e such that $m'(x,\tilde{e}) = m(x,e)$ it is the case that $\frac{\partial}{\partial x}m'(x,\tilde{e}) = \frac{\partial}{\partial x}m(x,e)$, it follows from Brown and Matzkin (1996) that m' and m are observationally

[7] To see this, let $g(\varepsilon|w,w')$ and $s(u)$ denote, respectively, the conditional pdf of ε and the pdf of u, and note that

$$
\begin{aligned}
(5.2)\ \ \forall w, w'\ \ q(e|w,w') &= \int s(e-\varepsilon|w,w')\, g(\varepsilon|w,w')\, d\varepsilon \\
&= \int s(e-\varepsilon)\, g(\varepsilon|w,w')\, d\varepsilon \\
&= \int s(e-\varepsilon|w',w)\, g(\varepsilon|w',w)\, d\varepsilon \\
&= q(e|w',w).
\end{aligned}
$$

[8] The sign of the effect of e on m can depend on x_1 provided that the analyst knows the values of x_1 at which the sign switches. For example, in the case of the model $m(x_1,e) = \alpha_1 x_1 + \alpha_2 x_1 e$ the sign of the effect of e depends on the sign of x_1.

equivalent.

Theorem 2 *Under A5.1-A5.5 $m(x,e)$ is identified from the distribution of $y|x_1, x_2$.*

Proof. : Equation (5.2) implies that
$\forall w, w' \ \ \Pr(e \leq \eta | w', w) = \Pr(e \leq \eta | w, w')$.

It follows by (A5.3) that

$\forall w', w$, $\forall e$,and $\forall \eta$, $\Pr(m(w',e) \leq m(w',\eta)|w',w) = \Pr(m(w,e) \leq m(w,\eta)|w,w')$,
or

(5.3) $\Pr(y \leq m(w',\eta)|w',w) = \Pr(y \leq m(w,\eta)|w,w')$.
In other words,

$F_{y|w',w}(m(w',\eta)) = F_{y|w,w'}(m(w,\eta))$ where $F_{y|w',w}(\cdot)$ is the CDF of y conditional on $x_1 = w'$ and $x_2 = w$. In particular,

(5.4) $F_{y|x,0}(m(x,e)) = F_{y|0,x}(m(0,e)) = F_{y|0,x}(e))$,

where the last equality follows from (A5.5).

Let $f(y|w',w)$ be the pdf of y conditional on $x_1 = w', x_2 = w$. Since $f(y|w',w) = g(m^{-1}(w',y)) \frac{\partial m^{-1}(w',y)}{\partial y}$, it follows from (A5.3) and (A5.4) that

$F_{y|w',w}(\cdot)$ is strictly increasing. Hence, by (A5.4)

(5.5) $m(x,e)) = F^{-1}_{y|x,0}(F_{y|0,x}(e))$.

This shows that the function m is identified. ■

The basic principle underlying identification is quite simple. The assumption of exchangeability implies that the CDF of $e|x,0$ is the same as the CDF of $e|0,x$. When one changes x, the distribution of $m(0,e)$ changes only because of the change in the distribution of $e|0,x$, while the distribution of $m(x,e)$ changes both because of the identical change in the distribution of $e|x,0$ and because x has a direct effect on $m(x,0)$ for each value of e. Consequently, one can isolate the direct effect of x on the distribution of $m(x,e)$ by comparing the change in the distribution of $m(x,e)|\ x,0$ as x changes to the change in the distribution of $m(0,e)|0,x$ as x changes. Our exchangeability and strong monotonicity assumptions imply that $F_{y|x,0}(m(x,e)) = F_{y|0,x}(m(0,e))$, which allows one to pin down $m(x,e)$ subject to the normalization $m(0,x) = e$. The mechanics are roughly as follows. Find the CDF of $y|0,x$ which, is the CDF of $e|0,x$. Then find the CDF of $y|x,0$ For each value of e, $m(x,e)$ is the value of y at which the 2 CDFs are equal

5.1 Identification of $g(e|x)$

Next we derive an expression for $g(e|x)$ in terms of the pdf and CDF of $y|x_1,x_2$ and of $y|x$, both of which are identified from data on y, x_1, and x_2.

Note first that

(5.6) $g(e|x) = f_{y|x}(m(x,e))\ \frac{\partial m(x,e)}{\partial e}$

where $f_{y|x}(m(x,e))$ is the pdf of $y|x$ and we recall that x is x_{i1}. Since $g(e|x,0) = g(e|0,x)$, it follows that

$$f_{y|x,0}(m(x,e))\frac{\partial m(x,e)}{\partial e} = f_{y|0,x}(m(0,e))\frac{\partial m(0,e)}{\partial e}, \text{ i.e.}$$

(5.7) $f_{y|x,0}(m(x,e))\frac{\partial m(x,e)}{\partial e} = f_{y|0,x}(e)$. Hence,

(5.8) $\frac{\partial m(x,e)}{\partial e} = \frac{f_{y|0,x}(e)}{f_{y|x,0}(m(x,e))}$. Or, using (5.5),

(5.9) $\frac{\partial m(x,e)}{\partial e} = \frac{f_{y|0,x}(e)}{f_{y|x,0}(F^{-1}_{y|x,0}(F_{y|0,x}(e)))}$.

Using (5.5). (5.6) and (5.9), we obtain

$$(5.10) \quad g(e|x) = f_{y|x}(m(x,e)) \, \frac{f_{y|0,x}(e)}{f_{y|x,0}(m(x,e))} = f_{y|x}(F^{-1}_{y|x,0}(F_{y|0,x}(e))) \, \frac{f_{y|0,x}(e)}{f_{y|x,0}(F^{-1}_{y|x,0}(F_{y|0,x}(e)))}.$$

Equations (5.5) and (5.10) provide a way to estimate $m(x,e)$ and $g(e|x)$ provided that one can estimate the pdf and CDF of $y|x_1, x_2$ and the pdf of $y|x$. With knowledge of $m(x,e)$ and $g(e|x)$, it is of course possible to obtain a number of functions summarising the effect of a change in x on the distribution of y holding the distribution of e constant. Consider for example, the average partial derivative $\beta(x_1) \equiv \beta(x)$ defined in (1.?). We now provide an expression that can serve as the basis for an estimator of $\beta(x)$.

Differentiating (5.5),

$$\begin{aligned}
(5.10) \quad \frac{\partial m(x,e)}{\partial x} &= \frac{\partial F^{-1}_{y|x,0}(F_{y|0,x}(e))}{\partial y} \, \frac{\partial F_{y|0,x}(e))}{\partial x} + \frac{\partial F^{-1}_{y|x,0}(F_{y|0,x}(e))}{\partial x} \\
&= \frac{\frac{\partial F_{y|0,x}(e))}{\partial x}}{\frac{\partial F_{y|x,0}(F^{-1}_{y|x,0}(F_{y|0,x}(e))}{\partial y}} + \frac{\partial F^{-1}_{y|x,0}(F_{y|0,x}(e))}{\partial x} \\
&= \frac{\frac{\partial F_{y|0,x}(e))}{\partial x}}{\frac{\partial F_{y|x,0}(m(x,e))}{\partial y}} + \frac{\partial F^{-1}_{y|x,0}(F_{y|0,x}(e))}{\partial x}.
\end{aligned}$$

Hence,

$$(5.11) \quad \int \frac{\partial m(x,e)}{\partial x} \, g(e|x) \; de$$

$$= \int \left[\frac{\frac{\partial F_{y|0,x}(e))}{\partial x}}{\frac{\partial F_{y|x,0}(m(x,e))}{\partial y}} + \frac{\partial F^{-1}_{y|x,0}(F_{y|0,x}(e))}{\partial x} \right] f_{y|x}(m(x,e)) \, \frac{f_{y|0,x}(e)}{f_{y|x,0}(m(x,e))} \, de$$

$$= \int \left[\frac{\frac{\partial F_{y|0,x}(e))}{\partial x}}{\frac{\partial F_{y|x,0}(F^{-1}_{y|x,0}(F_{y|0,x}(e)))}{\partial y}} + \frac{\partial F^{-1}_{y|x,0}(F_{y|0,x}(e))}{\partial x} \right] f_{y|x}(F^{-1}_{y|x,0}(F_{y|0,x}(e))) \, \frac{f_{y|0,x}(e)}{f_{y|x,0}(F^{-1}_{y|x,0}(F_{y|0,x}(e)))} \, de.$$

5.2 Estimation and Asymptotic Properties

We now proceed to define a particular estimator for $m(x,e)$. Let K denote the dimension of x. Let $d = 1 + 2K$, and let f denote the pdf of (y, x_1, x_2).

Then, the kernel estimator for the unconditional and conditional pdf's and cdf's are:

$$\hat{f}_N(t_y, t_1, t_2) = \frac{1}{N\sigma_N^d} \sum_{i=1}^{N} K\left(\frac{u_y - t_y}{\sigma}, \frac{u_x - t_1}{\sigma}, \frac{u_x - t_2}{\sigma}\right) \quad \text{for all } (t_y, t_1, t_2) \in R^d,$$

$$\hat{F}(y, x_1, x_2) = \int_{-\infty}^{y} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \hat{f}_N(t_y, t_1, t_2)\, dt_y\, dt_1\, dt_2,$$

$$\hat{F}_{y|x_1,x_2}(y|x_1, x_2) = \frac{\int_{-\infty}^{y} \hat{f}_N(t_y, x_1, x_2)\, dt_y}{\int_{-\infty}^{\infty} \hat{f}_N(t_y, x_1, x_2)\, dt_y}, \text{ and}$$

$$\hat{f}_N(y|x_1, x_2) = \frac{\hat{f}_N(y, x_1, x_2)}{\int_{-\infty}^{\infty} \hat{f}_N(t_y, x_1, x_2)\, dt_y}.$$

The estimator for $m(x, e))$ can then be defined by

$$\hat{m}(x, e) = \hat{F}^{-1}_{y|x,0}(\tilde{F}_{y|0,x}(e)),$$

where

$$\tilde{F}_{y|0,x}(e) = \left\{ \begin{array}{ll} t^u & if\ \hat{F}_{y|0,x}(e) \geq t^u \\ \hat{F}_{y|0,x}(e) & if\ t_l < \hat{F}_{y|0,x}(e) < t^u \\ t_l & if\ \hat{F}_{y|0,x}(e) \leq t_l \end{array} \right\},$$

$t^u = \sup_e\{\hat{F}_{y|x,0}(e)\}$ and $t_l = \inf_e \hat{F}_{y|x,0}(e)\}$.

Since we do not restrict $\hat{F}_{y|x,0}$ and $\hat{F}_{y|0,x}$ to be strictly increasing, $\hat{m}(x, e)$ need not be a singleton. To measure the distance between $\hat{m}(x, e)$ and $m(x, e)$ we will use the metric ρ defined by

$$\rho(m, m') = \sup_{x,e}\left\{\max\left\{\sup_{n\in m(x,e)} \inf_{n'\in m'(x,e)} |n - n'|, \sup_{n'\in m'(x,e)} \inf_{n\in m(x,e)} |n - n'|\right\}\right\}$$

for any functions m and m'. To establish the asymptotic properties of this estimator, we make the following assumptions:

ASSUMPTION 1': The sequence $\{y_i, x_{1i}, x_{i2}\}$ is a strictly stationary $\beta-$mixing sequence satisfying $k^\delta\, \beta_k \to 0$ as $k \to \infty$, for some $\delta > 1$.

ASSUMPTION 2': $f(y, x_1, x_2)$ has a compact support $\Theta \subset R^d$, $f(\partial\Theta) = 0$ where $\partial\Theta$ denotes the boundary of Θ, and $f(y, x, z)$ is continuously differentiable up to order g where $g \geq 2d$.

ASSUMPTION 3': The kernel function $K(\cdot, \cdot, \cdot)$ is an even function, integrates to 1, is of order r where r is an even integer satisfying $d/2 < r < g$, is continuously differentiable up to order g and its derivatives of order up to g are in $L^2(R^d)$.

ASSUMPTION 4': As $N \to \infty$, $\sigma_N \to 0$, $N\sigma_N^d \to \infty$, $\sqrt{N}\sigma_N^r \to 0$, and $\sqrt{N}\sigma_N^{2m} \to \infty$ where m is an integer such that $m < r/2$ and $m + r \leq g + d$.

ASSUMPTION 5': $f(0, x)$, $f(x, 0)$, and $f(F_{y|x,0}^{-1}(F_{y|0,x}(e)), x, 0)$ are bounded from above and, bounded away from zero from below. $0 < F_{y|0,x}(e)) < 1$ and $0 < F_{y|x,0}(t)) < 1$, where $t = F_{y|x,0}^{-1}(F_{y|0,x}(e))$.

Theorem 3 : *If Assumptions (A5.1)-(A5.5) and 1'-5' are satisfied, then $\hat{m}(x, e)$ is a consistent estimator of $m(x, e)$, with respect to ρ, and*

$$\sqrt{N}\, \sigma_N^K \left(\hat{m}(x, \varepsilon) - m(x, e)\right) \to N(0, \tilde{V}) \text{ in distribution}$$

where

$$\tilde{V} = \frac{\left\{\int\int\left[\int K(y,x,z)dy\right]^2 dt_1 dt_2\right\} \left\{\left[F_{y|0,x}(e)\left(1-(F_{y|0,x}(e)\right)\right]+\left[F_{y|x,0}(m(x,e))\left(1-F_{y|0,x}(m(x,e))\right)\right]\right\}}{f_{y|0,x}(m(x,e))^2\, f(0,x)^3,}.$$

Proof: See Appendix.

In a future draft we will propose a specific estimator for g and analyze its asymptotic properties.

6 Monte Carlo Evidence

We have performed a limited monte carlo analysis of our Regression estimator, which we refer to in the tables as AM-1. Our first set of experiments involves cases in which y is a continuous variable. The cases are nested in the following model:

Model 1

$$\begin{aligned} y_{ik} &= m(x_{ik}, \varepsilon_i, u_{ik}) = b_0 + b_1 x_{ik} + \gamma x_{ik}\eta_i + \theta_\eta \eta_i + \theta_\varepsilon \varepsilon_i + u_{ik} \ ; \ k = 1,2 \ ; \ i = 1,2,...n \\ x_{ik} &= x_i + \tilde{x}_{ik} \\ \varepsilon_i &= \theta_{\varepsilon x} x_i + \tilde{\varepsilon}_i \\ \eta_\iota &= \theta_{\eta x} x_i + \tilde{\eta}_{ik} \\ & x_i \sim N(0,1); \ \tilde{x}_{ik} N(0,1); \ \tilde{\varepsilon}_i \sim N(0,1); \ \tilde{\eta} \sim N(0,1); u_{ik} \sim N(0,1) \ . \end{aligned}$$

The random variables x_i, $\tilde{x}_{ik}$, $\tilde{\varepsilon}_i$, and u_{ik} are i.i.d. and mutually independent. Model 1 is a special case of the model in (1.4) that we used to motivate that paper. In applying AM-1 we defined z_i to $(x_{i1} + x_{i2})/2$. Because of the linear relationship among the stochastic components and the fact that $\tilde{\varepsilon}_{ik}$, x_i, and $\tilde{x}_{ik}$ are all normally distributed, $g(\eta_i, \varepsilon_i, u_{ik}|x_{i1}, x_{i2}) = g(\eta_i, \varepsilon_i, u_{ik}|(x_{i1} + x_{i2})/2)$, so z_i may be restricted to the mean of x_{ik} group i. In practice, the researcher will not know the distribution of the random components, and so it will be necessary to experiment with additional symmetric functions of x_{i1} and x_{i2}. It seems sensible to us begin with the case in which we have the right conditioning variables. For this design $\beta(x) = b_1 + \gamma E(\eta_i|x_{ik}) = b_1 + \gamma\theta_{\eta x}\theta_{x_i x_{ik}} x_{ik}$ where $\theta_{x_i x_{ik}}$ is the population coefficient of the regression of x_i on x_{ik}.

We implement AM-1 using 4 different approaches. The first, which we refer to as AM-1(poly/llr) in the table, is to approximate $E(y_{ik}|x_{ik}, z_i)$ as a third order polynomial in x_{ik} and z_i with interactions up to the third order. We use OLS to estimate the coefficients of the polynomial. We use local linear regression to estimate $h(z_i|x_{ik})$. The second, called A-M(poly/ker) in the table, combines the polynomial approximation to $E(y_{ik}|x_{ik}, z_i)$ with a kernel regression estimator of $h(z_i|x_{ik})$. AM-1(llr/llr) uses local linear regression to estimate both $E(y_{ik}|x_{ik}, z_i)$ and $h(z_i|x_{ik})$, while AM-1(ker/ker) uses kernel regression to estimate both. We also report of $\beta(x)$ based on applying OLS to $y_{ik} = b_0 + b_1 x_{ik} + \gamma_2 x_{ik}\varepsilon_i + \varepsilon_i + u_{ik}$ both with and without group specific intercepts or "fixed effects". We set n to 1,500. The monte carlo results are based on 750 replications. The numbers in parentheses are the standard deviations of the estimators over the 750 replications.

We consider 3 cases of Model 1. In all cases we set b_0 to $0, b_2$ to 2, and evaluate the estimates at $x_{it} = 0.5$. In Case 1 γ is 0, θ_ε is 0, and $\theta_{\eta x}$ is 1. Case 1 corresponds to the classic case in which the OLS-fixed effects estimator is used in the literature, and it is unbiased. The fact that $\gamma = 0$ implies that $\beta(x) = b_1 = 2$. In this design x_{ik} is correlated with η_i so OLS is inconsistent. Given that $\gamma = 0$ and $\theta_{\eta x} = 0$ the mean of the OLS estimator is $b_1 + \theta_{\eta x}\theta_{x_i x_{ik}} = 2.5$ where $\theta_{x_i x_{ik}}$ is the population coefficient of the regression of x_i on x_{ik}.

The results of the monte carlo simulations are reported in Table 1. The mean of all 4 versions of the AM-1 estimator are very close to the true value of 1, as is the mean of the OLS-fixed effect estimator. Perhaps surprisingly, the sampling error of all 4 versions of the AM-1 estimator are smaller than that of the OLS-fixed estimator. In constrast, OLS is biased, with a mean of 2.5. The results establish that the regression based estimator performs well in the standard case in which OLS-fixed effects can be employed.

In Case 2 we use the same parameter values as Case 1 except that we change γ from 0 to 1. For this design and parameter values, $\beta(.5) = 2.25$. Controlling for group specific intercepts does not control for the $x_{it}\eta_i$. All versions of the AM-1 estimator have a mean that is very close to 2.25. In contrast, OLS is badly biased with a mean of 3.49, and OLS-fixed effects also suffers from a substantial bias, with a mean of 2.5. The standard deviation of the various versions of the AM-1 estimator range from .127 to .107, which is somewhat larger than the value of .101 for OLS-fixed effects. However, the AM-1 estimator clearly dominates on mean squared error grounds.

Case 3 is the same as Case 2 except θ_η is set to 0, θ_ε is set to 1, and $\theta_{\varepsilon x}$ is set to -1. In this design x_{it} is correlated with the random slope parameter η_i and is negatively correlated with a separate random intercept term ε_i, while in Case 2 η_i is both a random slope and a random intercept. The true value of $\beta(.5)$ is 2.25. OLS is badly downward biased, with a mean of 1.5. The mean of the OLS-fixed effects estimator is 2.00. In contrast, the AM-1 estimator is essentially unbiased regardless of how we estimate the conditional mean of y and the conditional density of z. Overall, the results for the AM-1 estimator in the continuous dependent variable case are very encouraging. In a future draft, we will evaluate the performance of our second estimator using a similar design.

We now turn to analysis of a binary choice model. The model is of the form

$$
\begin{aligned}
& \text{Model 2} \\
y_{ik} &= I(m(x_{ik}, \varepsilon_i, u_{ik})) = I(b_0 + b_1 x_{ik} + \gamma x_{ik}\eta_i + \theta_\eta \eta_i + u_{ik}) \ ; \ k = 1, 2; \ i = 1, 2, ...n \\
x_{ik} &= x_i + \tilde{x}_{ik} \\
\eta_\iota &= \theta_{\eta x} x_i + \tilde{\eta}_{ik} \\
& x_i \tilde{\ } N(0, \sigma_x^2); \ \tilde{x}_{ik} \tilde{\ } N(0, \sigma_{\tilde{x}}^2); \ \tilde{\eta} \tilde{\ } N(0, \sigma_{\tilde{\eta}}^2); u_{ik} \tilde{\ } N(0, 1.5)
\end{aligned}
$$

For this design we also set z_i to $(x_{i1} + x_{i2})/2$ for reasons discussed above. We used two approaches to implementing the AM-1 estimator. In the column labelled AM-1(poly/llr) we take a parametric approach and assume that $E(y_{ik}|x_{ik}, z_i)$ is well approximated by a probit model with an index consisting of a third order polynomial in x_{ik} and z_i with interactions up the third order, and we use local linear regression to estimate $h(z_i|x_{ik})$. In the column labelled AM-1(ker/ker) we use kernel regression to estimate both functions. We also report estimates of $\beta(x)$ based on a probit model involving a third order polynomial in x_{ik}. As before, we set n to 1,500 and perform 750 replications. In all the experiments, b_0 is 0. The results are in Table 2.

In Case 1, which we do not report in the table, $\gamma = 0$ and $\theta_{\eta x}$ is 0, and $b_1 = 0$. We also set $\sigma_{\tilde{x}}^2$ to 1.5. In this case the probit estimator is consistent and $\beta(x) = 0$ for all x. The mean of both the AM-1 estimator and the probit are essentially 0. In Case 2, $\gamma = 0$, but $\theta_{\eta x}$ is 1. We also set $\theta_\eta = 1$, $\sigma_x^2 = 1.5$, $\sigma_{\tilde{x}}^2 = 1.5$, and $\sigma_{\tilde{\eta}}^2 = .5$. This design is essentially a probit model with a group specific error component (η) that is correlated with x_{ik}. We report estimates of $\beta(x)$ at $x = -2$, $x = -1$, $x = 0$, $x = 1$, and $x = 2$. Not surprisingly, the probit estimator suffers from a strong positive bias at each value x. The estimates range between .101 and .118. Since the estimated $\beta(x)$ is almost constant across value of x the probit estimates imply that an increase of x from -2 to 2 would increase the probability that y is positive by about .4, while the true effect is 0. In contrast, the AM-1 estimator does very well. In the case of AM-1(ker/ker) the means of the estimator at the various X values lie between -.001 to .001. The standard deviations are larger than those for the probit but are still quite small.

In case 3 η_i is correlated with x_{ik} and enters as both a random intercept and a random slope. The design is the same as Case 2 except that we change the random slope parameter γ from 0 to 1. In this case, $\beta(x)$ varies with x_{ik} for two reasons. The first is that the mean $\gamma E(\eta_i|x_{ik}) = 1 \cdot .75 \cdot x_{ik}$ of the random slope obviously shifts with x_{ik}, and the second is that the value η influences the likelihood that a given change in $b_1 x_{ik}$ will lead the index $b_0 + b_1 x_{ik} + \gamma x_{ik}\eta_i + \theta_\eta \eta_i + u_{ik}$ to exceed 0. $\beta(x)$ is _ , __, and __ when x_{ik} is -2, 0, and 2 respectively. Once again, the probit estimator is seriously biased at some values of x. On the other hand, A-M does quite well, with a mean of __, __, and __ when x_{ik} is -2, 0, and 2 respectively. These result are very encouraging. They suggest that the regression based estimator does provide a viable way to estimate qualitative choice models with random errors that interact with and are correlated with the explanatory variables in the model.

The monte carlo evidence is obviously limited by the small set of designs and parameter values we have chosen. However, we find the results to date to be quite encouraging.

7 Conclusion (preliminary and incomplete)

There has been an explosion of empirical studies that use variation among members of a panel to try to deal with endogeneity of explanatory variables. In this paper we provide two estimators for models with nonseparable error terms and endogenous explanatory variables that can be used with panel data. One important class of such models are qualitative choice models with group error components that are correlated with the regressors. Another set of examples consists of random coefficients models in which a group specific random coefficient that is correlated with the regressors. The applied econometrician does not have good options in the literature to estimate such models, except in special cases. We believe that our estimators may provide attractive options in a wide range of situations.

There is a long research agenda beyond the obvious tasks of completing the asymptotic theory and providing some initial monte carlo evidence for the second estimator. First, detailed monte carlo studies of both estimators are needed, particularly in a multivariate context. It is important to try different distributions of the error terms and the explanatory variables. Second, it is important to gain some experience with the estimators using real data.

With this objective in mind we are in the midst of using the regression based estimator to estimate how the probabilities of time and money transfers between married couples and the parents of the husband and of the wife are affected by the relative incomes, health status, and other characteristics factors of the two sets of parents. We will compare our results to those of Altonji, Hayashi and Kotlikoff (1996), who have applied fixed effects linear probability models and conditional logit models to this problem. There are many practical issues that will arise in implementing the estimators that are best worked out in the context of real world applications.

8 Appendix: Proof of Theorems 1 and 3

PROOF OF THEOREM 1:

Define the functional $\Phi(\cdot)$ by:

$$\Phi(F) = \int \frac{\partial}{\partial x} E\left(y|x,z\right)\ h(z|x)\, dz.$$

Then,

$$\beta(x) = \Phi(F) \quad \text{and} \quad \hat{\beta}(x) = \Phi(\hat{F})$$

where $\hat{F}$ is the kernel estimator for the cdf F.

Note that

$$\begin{aligned}
\Phi(F) &= \int \frac{\partial}{\partial x} E\left(y|x,z\right)\ h(z|x)\, dz \\
&= \int \frac{\partial}{\partial x}\left[\frac{\int y\, f(y,x,z)\, dy}{\int f(y,x,z)\, dy}\right]\left[\frac{\int f(y,x,z)\, dy}{\int f(y,x,z)\, dy\, dz}\right] dz \\
&= \int \frac{\int y\, f_x(y,x,z)\, dy}{\int f(y,x,z)\, dy\, dz}\, dz - \int \frac{\int f_x(y,x,z)\, dy}{\int f(y,x,z)\, dy}\frac{\int y\, f(y,x,z)\, dy}{\int f(y,x,z)\, dy\, dz}\, dz \\
&= \int \frac{\int y\, f_x\, dy}{f(x)}\, dz - \int \frac{\int f_x\, dy}{f(x,z)}\frac{\int y\, f\, dy}{f(x)}\, dz
\end{aligned}$$

and for any H,

$$\begin{aligned}
\Phi(F+H) - \Phi(F) = &\left[\int \frac{\int y\,(f_x + h_x)\, dy}{f(x)+h(x)}\, dz - \int \frac{\int y\, f_x\, dy}{f(x)}\, dz\right] \\
&- \left[\int \frac{\int (f_x + h_h)\, dy}{(f(x,z)+h(x,z))}\frac{\int y\,(f+h)\, dy}{(f(x)+h(x)}\, dz - \int \frac{\int f_x\, dy}{f(x,z)}\frac{\int y\, f\, dy}{f(x)}\, dz\right]
\end{aligned}$$

$$= \int \frac{\left(\int y\, h_x\, dy\right) f(x) - \left(\int y\, f_x\, dy\right) h(x)}{f(x)^2}\, dz$$

$$- \int \frac{\int f_x dy \int y h dy\, f(x,z)\, f(x) + \int h_x dy \int y f dy\, f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2}\, dz$$

$$+ \int \frac{\int f_x dy \int y f dy\, f(x,z)\, h(x) + \int f_x dy \int y f dy\, h(x,z)\, f(x)}{f(x,z)^2\, f(x)^2}\, dz$$

$$+ \int \frac{\left(\int y\, h_x\, dy\right) f(x) - \left(\int y\, f_x\, dy\right) h(x)}{f(x)^2} \left[\frac{1}{f(x)^2 + f(x)h(x)} - \frac{1}{f(x)^2}\right] dz$$

$$- \int \frac{\int f_x dy \int y h dy\, f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)} - \frac{1}{f(x,z)^2\, f(x)^2}\right] dz$$

$$- \int \frac{\int h_x dy \int y f dy\, f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)} - \frac{1}{f(x,z)^2\, f(x)^2}\right] dz$$

$$+ \int \frac{\int f_x dy \int y f dy\, f(x,z)\, h(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)} - \frac{1}{f(x,z)^2\, f(x)^2}\right] dz$$

$$+ \int \frac{\int f_x dy \int y f dy\, h(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)} - \frac{1}{f(x,z)^2\, f(x)^2}\right] dz$$

$$- \int \frac{\int h_x dy \int y h dy\, f(x,z)\, f(x) - \int f_x dy \int y f dy\, h(x,z)\, h(x)}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)}$$

Let

$$\Phi'(F,H) = \int \frac{\left(\int y\, h_x\, dy\right) f(x) - \left(\int y\, f_x\, dy\right) h(x)}{f(x)^2}\, dz$$

$$- \int \frac{\int f_x dy \int y h dy\, f(x,z)\, f(x) + \int h_x dy \int y f dy\, f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2}\, dz$$

$$+ \int \frac{\int f_x dy \int y f dy\, f(x,z)\, h(x) + \int f_x dy \int y f dy\, h(x,z)\, f(x)}{f(x,z)^2\, f(x)^2}\, dz\,,$$

and

$$R(F,H) = \int \frac{\left(\int y\, h_x\, dy\right) f(x) - \left(\int y\, f_x\, dy\right) h(x)}{f(x)^2} \left[\frac{1}{f(x)^2 + f(x)h(x)} - \frac{1}{f(x)^2}\right] dz$$

$$- \int \frac{\int f_x dy \int y h dy\, f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)} - \frac{1}{f(x,z)^2\, f(x)^2}\right] dz$$

$$- \int \frac{\int h_x dy \int y f dy\, f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)} - \frac{1}{f(x,z)^2\, f(x)^2}\right] dz$$

$$+ \int \frac{\int f_x dy \int y f dy\, f(x,z)\, h(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x)) f(x,z) f(x)} - \frac{1}{f(x,z)^2\, f(x)^2}\right] dz$$

$$+ \int \frac{\int f_x dy \int y f dy\, h(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{1}{(f(x,z)+h(x,z))(f(x)+h(x))f(x,z)f(x)} - \frac{1}{f(x,z)^2\, f(x)^2} \right] dz$$

$$- \int \frac{\int h_x dy \int y h dy\, f(x,z)\, f(x) - \int f_x dy \int y f dy\, h(x,z)\, h(x)}{(f(x,z)+h(x,z))(f(x)+h(x))f(x,z)f(x)}\, dz$$

For any H, let $\|H\|$ denote the supremum over the values and derivatives up to order $2(K_1 + K_2 + 1)$ of H. Then, letting $R_k(F, H)$ denote the $k-th$ coordinate of $R(F, H)$, for $k = 1, ..., K$, and h_x and f_x denote the $k-th$ coordinate of h_x and f_x, it follows that

$$|R_k(F,H)| \leq \frac{\int\int |y\, h_x|\, dydz\;\; f(x) \; + \left|\int\int y\, f_x\, dydz\right| h(x)}{f(x)^2} \left[\frac{h(x)}{f(x)^3}\right]$$

$$+ \int \frac{\left|\int f_x\, dydz\right| \int |y\, h|\, dy\, f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left|\int h_x\, dy\right| \left|\int y\, f\, dy\right|\; f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left|\int f_x\, dy\right| \left|\int y\, f\, dy\right|\, f(x,z)\, h(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left|\int f_x\, dy\right| \left|\int y\, f\, dy\right|\, h(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left|\int h_x\, dy\right| \int |y\, h|\, dy\, f(x,z)\, f(x) + \left|\int f_x\, dy\right| \left|\int y\, f\, dy\right|\, h(x,z)\, h(x)}{f(x,z)^2\, f(x)^2}\, dz$$

$$\leq \frac{\left(\int\int y^2\, dydz\right)^{1/2} \left(\int h_x^2\, dy\right)^{1/2} f(x) \; + \left|\int\int y\, f_x\, dydz\right| h(x)}{f(x)^2} \left[\frac{h(x)}{f(x)^3}\right]$$

$$+ \int \frac{\left|\int f_x\, dy\right| \left(\int y^2\, dy\right)^{1/2} \left(\int h^2\, dy\right)^{1/2} f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left(\int h_x^2\, dy\right)^{1/2} \left|\int y\, f\, dy\right|\; f(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left|\int f_x\, dy\right| \left|\int y\, f\, dy\right|\, f(x,z)\, h(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left|\int f_x\, dy\right| \left|\int y\, f\, dy\right|\, h(x,z)\, f(x)}{f(x,z)^2\, f(x)^2} \left[\frac{f(x,z)h(x) + h(x,z)f(x) + h(x,z)h(x)}{f(x,z)^3\, f(x)^3}\right] dz$$

$$+ \int \frac{\left(\int h_x^2\, dy\right)^{1/2} \left(\int y^2\, dy\right)^{1/2} \left(\int h^2\, dy\right)^{1/2} f(x,z)\, f(x) + \left|\int f_x\, dy\right| \left|\int y\, f\, dy\right|\, h(x,z)\, h(x)}{f(x,z)^2\, f(x)^2}\, dz$$

$$\leq \ \|H\|^2 [A]$$

where $$[A] = \frac{\left(\int\int y^2 \, dydz\right)^{1/2}}{f(x)^4} + \frac{\left(\int\int y^2 \, dydz\right)^{1/2}\left(\int\int f_x^2 \, dydz\right)^{1/2}}{f(x)^5}$$

$$+ \left[\int \frac{\left|\int f_x \, dy\right|^2 \left(\int y^2 \, dy\right)}{f(x,z)^6 \, f(x)^8} dz\right]^{1/2} + \left[\int \frac{\left|\int f_x \, dy\right|^2 \left(\int y^2 \, dy\right)}{f(x,z)^8 \, f(x)^6} dz\right]^{1/2}$$

$$+ \|H\| \left[\int \frac{\left|\int f_x \, dy\right|^2 \left(\int y^2 \, dy\right)}{f(x,z)^8 \, f(x)^8} dz\right]^{1/2} + \left[\int \frac{\left|\int y \, f \, dy\right|^2}{f(x,z)^6 \, f(x)^8} dz\right]^{1/2}$$

$$+ \left[\int \frac{\left|\int y \, f \, dy\right|^2}{f(x,z)^8 \, f(x)^6} dz\right]^{1/2} + \|H\| \left[\int \frac{\left|\int y \, f \, dy\right|^2}{f(x,z)^8 \, f(x)^8} dz\right]^{1/2}$$

$$+ \left[\int \frac{\left|\int f_x \, dy\right|^2 \left|\int y \, f \, dy\right|^2}{f(x,z)^6 \, f(x)^{10}} dz\right]^{1/2} + \left[\int \frac{\left|\int f_x \, dy\right|^2 \left|\int y \, f \, dy\right|^2}{f(x,z)^8 \, f(x)^8} dz\right]^{1/2}$$

$$+ \|H\| \left[\int \frac{\left|\int f_x \, dy\right|^2 \left|\int y \, f \, dy\right|^2}{f(x,z)^8 \, f(x)^{10}} dz\right]^{1/2} + \left[\int \frac{\left|\int f_x \, dy\right|^2 \left|\int y \, f \, dy\right|^2}{f(x,z)^8 \, f(x)^8} dz\right]^{1/2}$$

$$+ \left[\int \frac{\left|\int f_x \, dy\right|^2 \left|\int y \, f \, dy\right|^2}{f(x,z)^{10} \, f(x)^6} dz\right]^{1/2} + \|H\| \left[\int \frac{\left|\int f_x \, dy\right|^2 \left|\int y \, f \, dy\right|^2}{f(x,z)^{10} \, f(x)^8} dz\right]^{1/2}$$

$$+ \left[\int \frac{\left(\int y^2 \, dy\right)}{f(x,z)^2 \, f(x)^2} dz\right]^{1/2} + \left[\int \frac{\left|\int f_x \, dy\right|^2 \left|\int y \, f \, dy\right|^2}{f(x,z)^4 \, f(x)^4} dz\right]^{1/2}$$

Since by Assumptions 2 and 5, all the terms in brackets are finite, and $\|H\|$ is bounded on a neighborhood of F, it follows that

$[A] \leq C(F)$ for some constant $C(F)$.

Hence,

$(^*)\ \|R(F, H)\| \leq C(F) \ \|H\|^2.$

Also, for each coordinate k of $\Phi'(F, H)$

$$\left|\Phi_k'(F, H)\right| \leq \frac{\left(\int\int y^2 dydz\right)^{1/2}\left(\int\int h_z^2 dydz\right)^{1/2}}{f(x)} + \frac{\left|\int\int y \, f_x \, dydz\right| h(x)}{f(x)^2}$$

$$+ \int \frac{\left|\int f_x dy\right| \left(\int y^2 dy\right)^{1/2} \left(\int h^2 dy\right)^{1/2}}{f(x,z)^2\, f(x)^2} dz + \int \frac{\left(\int h_x^2 dy\right)^{1/2} \left|\int y f dy\right|}{f(x,z)\, f(x)}\, dz$$

$$+ \int \frac{\left|\int f_x dy\right| \left|\int y f dy\right| h(x)}{f(x,z)\, f(x)^2}\, dz + \left(\int \frac{\left(\int f_x dy\right)^2 \left(\int y f dy\right)^2}{f(x,z)^4\, f(x)^2} dz\right)^{1/2} \left(\int h(x,z)^2 dz\right)^{1/2}$$

$$\leq \|H\|\ [B],$$

where

$$[B] = \frac{\left(\int\int y^2 dy dz\right)^{1/2}}{f(x)} + \frac{\left|\int\int y\, f_x\, dy dz\right|}{f(x)^2}$$

$$+ \int \frac{\left|\int f_x dy\right| \left(\int y^2 dy\right)^{1/2}}{f(x,z)^2\, f(x)^2} dz + \int \frac{\left|\int y f dy\right|}{f(x,z)\, f(x)}\, dz$$

$$+ \int \frac{\left|\int f_x dy\right| \left|\int y f dy\right|}{f(x,z)\, f(x)^2} dz + \left(\int \frac{\left(\int f_x dy\right)^2 \left(\int y f dy\right)^2}{f(x,z)^4\, f(x)^2} dz\right)^{1/2}.$$

is, by Assumptions 2 and 5, bounded. Let $D(F)$ be such that $[B] \leq D(F)$.

Then,

$$(^{**})\ \left|\Phi_k'(F,H)\right| \leq \|H\|\ D(F),$$

¿From (*) and (**) it follows that Φ' is the Hadamard derivative of Φ at F.

Since, as we have just shown,

$$|\Phi(F+H) - \Phi(F)| \leq D(F)\ \|H\| + C(F)\ \|H\|^2$$

for any bounded H, it follows that

$$\left|\Phi(\hat{F}) - \Phi(F)\right| \leq D(F)\ \left\|\hat{F} - F\right\| + C(F)\ \left\|\hat{F} - F\right\|^2$$

Hence, since by Assumptions 1-4, $\hat{F}$ converges in probability to F in the supremum norm, it follows that

$$\left|\hat{\beta}(x) - \beta(x)\right| = \left|\Phi(\hat{F}) - \Phi(F)\right|$$

converges in probability to 0. Hence,

$\hat{\beta}(x)$ is a consistent estimator for $\beta(x)$.

Also,

$$\Phi'(F,H) = \int \frac{\left(\int y\, h_x\, dy\right) f(x) - \left(\int y\, f_x\, dy\right) h(x)}{f(x)^2}\, dz$$

$$- \int \frac{\int f_x dy\ \int y h dy\ f(x,z)\ f(x) + \int h_x dy\ \int y f dy\ f(x,z)\ f(x)}{f(x,z)^2\ f(x)^2}\, dz$$

$$+ \int \frac{\int f_x dy\ \int y f dy\ f(x,z)\ h(x) + \int f_x dy\ \int y f dy\ h(x,z)\ f(x)}{f(x,z)^2\ f(x)^2}\, dz\,,$$

$$= - \int\int\int \frac{\left(\int\int\int y\ 1_x^{(1)}(y,w,z)\right)}{f(x)}\, dH(y,w,z)$$

$$- \int\int\int \left(\frac{\left(\int\int y\ f_x(y,x,z)\ dydz\right)\ 1_x(y,w,z)}{f(x)^2}\right) dH(y,w,z)$$

$$- \int\int\int \frac{\left(\int f_x(y,x,z)\ dy\right)\ 1_x(y,w,z)\ y}{f(x,z)\ f(x)}\, dH(y,w,z)$$

$$+ \int\int\int \frac{1_x^{(1)}(y,w,z)\ \left(\int y\ f(y,x,z)\ dy\right)}{f(x,z)\ f(x)}\, dH(y,w,z)$$

$$+ \int\int\int \left(\int \frac{\left(\int f_x(y,x,z)\ dy\right)\ (y\ f(y,x,z)\ dy) 1_x(w)}{f(x,z)}\, dz\right)\ 1_x(y,w,z)\ dH(y,w,z)$$

$$+ \int\int\int \frac{\left(\int f_x(y,x,z)\ dy\right)\ \left(\int y\ f(y,x,z)\ dy\right)\ 1_x(y,w,z)}{f(x,z)^2\ f(x)}\, dH(y,w,z)$$

$$= \int\int\int \left[\frac{\left(\int y\ f(y,x,z)\ dy\right)}{f(x,z)\ f(x)} - \frac{y}{f(x)}\right] 1_x^{(1)}(y,w,z)\ dH(y,w,z)$$

$$- \int\int\int \left[\frac{\left(\int\int y\ f_x(y,x,z)\ dydz\right)}{f(x)^2} + \frac{\left(\int f_x(y,x,z)\ dy\right)\ y}{f(x,z)\ f(x)}\right] 1_x(y,w,z)\ dH(y,w,z)$$

$$+ \int\int\int \left[\int \frac{\left(\int f_x(y,x,z)\ dy\right)\ \left(\int y\ f(y,w,z)\ dy\right)}{f(x,z)\ f(x)^2}\right] 1_x(y,w,z)\ dH(y,w,z)$$

$$+ \quad \int\int\int \left[\int \frac{\left(\int f_x(y,x,z)\ dy\right)\left(\int y\ f(y,w,z)\ dy\right)}{f(x,z)^2\ f(x)} \right] 1_x(y,w,z)\ dH(y,w,z).$$

where for any (y,z), $1_x(y,w,z)$ and $1_x^{(1)}(y,w,z)$ denote, respectively, the Dirac function and the derivative of the Dirac function with respect to x at the point (y,x,z).

It follows by Assumptions 1-4 and Theorem 3 in Ait-Sahalia (1992) that

$\sqrt{N}\ \sigma_N^{(K_1/2)+1}\left(\Phi(\hat{F}) - \Phi(F)\right) = \sqrt{N}\ \sigma_N^{(K_1/2)+1}\left(\hat{\beta}(x) - \beta(x)\right) \rightarrow N(0,V)$ in distribution

where

$$V = \left\{ \int\int \left[y - \frac{\left(\int y\ f(y,x,z)\ dy\right)}{f(x,z)} \right]^2 f(y,x,z)\ dydz \right\} \left\{ \int \left[\frac{\partial}{\partial x} \left(\int\int K(y,x,z) dydz\right) \right]^2 dx \right\} \frac{1}{f(x)^2}$$

$$= \left\{ \int Var\,(y|x,z)\ \ f(x,z)\ dz \right\} \left\{ \int \left[\frac{\partial}{\partial x} \left(\int\int K(y,x,z) dydz\right) \right]^2 dx \right\} \ \frac{1}{f(x)^2}$$

■

PROOF OF THEOREM 2 (preliminary and incomplete). The proof follows using arguments similar to those used in Matzkin and Newey (1993), Matzkin (1997), and in the proof of Theorem 1. In particular, to derive the asymptotic distribution, one can show that the Hadamard derivative of the functional:

$$\Phi(F) = F_{y|x,0}^{-1}\left(F_{y|0,x}(e)\right)$$

is given by

$$\Phi'(F,H) = \frac{1}{f_{y|x,0}(m(x,e))}\left(L_1 + L_2\right)$$

where

$$L_1 = \frac{\int_{-\infty}^{e} h(s,0,x) ds}{\int_{-\infty}^{\infty} f(s,0,x) ds} \ - \ \frac{\int_{-\infty}^{e} f(s,0,x) ds}{\left(\int_{-\infty}^{\infty} f(s,0,x) ds\right)^2} \int_{-\infty}^{\infty} h(s,0,x) ds$$

$$= \int \left[\frac{1[s \le e]\ f(0,x) - \int_{-\infty}^{e} f(s,0,x)ds}{f(m(x,e),0,x)\ f(0,x)^2} \right] 1_{s,0,x}(s,w,z)\ h(s,w,z)\ ds\ dw\ dz$$

and

$$L_2 = \frac{\int_{-\infty}^{m(x,e)} f(s,x,0)ds}{\left(\int_{-\infty}^{\infty} f(s,x,0)ds\right)^2} \int_{-\infty}^{\infty} h(s,x,0)ds \ - \ \frac{\int_{-\infty}^{m(x,e)} h(s,x,0)ds}{\int_{-\infty}^{\infty} f(s,x,0)ds} -$$

$$= \int \left[\frac{\int_{-\infty}^{m(x,e)} f(s,0,x)ds - 1[s \le m(x,e)]\ f(x,0)}{f(m(x,e),0,x)\ f(0,x)^2} \right] 1_{s,x,0}(s,w,z)\ h(s,w,z)\ ds\ dw\ dz$$

By Ait-Sahalia (1992), it then follows that

$\sqrt{N}\ \sigma_N^K\left(\hat{m}(x,e) - m(x,e)\right) = \sqrt{N}\ \sigma_N^K\left(\Phi(\hat{F}) - \Phi(F)\right) \rightarrow N(0,\tilde{V})$ in distribution

where

$$\tilde{V} = \left\{ \int \int \left[\int K(y,w,z)dy\right]^2 dwdz \right\}\ L$$

and $L = \int \left[\frac{1[s \le e, w=0, z=x]\ f(0,x) - \int_{-\infty}^{e} f(s,0,x)ds}{f(m(x,e),0,x)\ f(0,x)^2} \right]^2 f(s,0,x)\ ds$

$$+ \int \left[\frac{\int_{-\infty}^{m(x,e)} f(s,0,x)ds - 1[s \le m(x,e), w=x, z=0]\ f(x,0)}{f(m(x,e),0,x)\ f(0,x)^2} \right] f(s,x,0)\ ds$$

$$= \frac{\left\{ \left[F_{y|0,x}(e)\left(1-(F_{y|0,x}(e)\right)\right] + \left[F_{y|x,0}(m(x,e))\left(1-F_{y|0,x}(m(x,e))\right)\right] \right\}}{f_{y|0,x}(m(x,e))^2\ f(0,x)^3}$$

■

9 References (incomplete)

Aaronson, D. (forthcoming)"Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes", *Journal of Human Resources.*

Abrevaya, J., (1997) "Rank Estimation of a Generalized Fixed Effects Regression Model", Unpublished Paper, University of Chicago.

Ait-Sahalia, Y. (1992) "The Delta and Bootstrap Methods for Nonlinear Functionals of Nonparametric Kernel Estimators Based on Dependent Multivariate Data," mimeo.

Altonji, J.G. and H. Ichimura, (1997) "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables", mimeo.

Brown, D.J. and R.L. Matzkin, (1996) "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," mimeo, Northwestern University.

Chamberlain, G. (1080) "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47 225-238.

Chamberlain, G. (1984), "Panel Data" in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics*, Volume II, Amsterdam: North Holland.

Geronimus, A.T. and S. Korenman, (1992) "The socioeconomic consequences of teen childbearing reconsidered", *Quarterly Journal of Economics,* v107 n4 p1187-1228.

Hausman, J.A., and W.E. Taylor, "Panel Data and Unobservable Individual Effects", *Econometrica* 49, pp. 1377-1398.

Heckman, J.J. "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating A Discrete Time-Discrete Data Stochastic Process and Some Monte Carlo Evidence.", in C. Manski and D. McFadden (eds.) *Structural Analysis of Discrete Data With Econometric Applications.* M.I.T. Press

Heckman, J.J. and T. E. MaCurdy, "A Life Cycle Model of Female Labour Supply." Review of Economic Studies (1980), pp. 47-74.

Honoré, B. E., "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica,* 60, pp. 533-565.

Honoré, B. E. and E. Kyriazidou, (1997) "Panel Data Discrete Choice Models with Lagged Dependent Variables", unpublished paper, Department of Economics, Princeton University.

Horowitz J.L. (1992) "A Smoothed Maximum Score Estimator for the Binary Response Model" *Econometrica* 60 :505-531.

Kyriazidou, E. (1997) "Estimation of a Panel Data Sample Selection Model", *Econometrica,* forthcoming (November)

Kyriazidou, E. (1995), *Essays in Estimation and Testing of Econometric Models.* Ph. D. thesis, Department of Economics, Northwestern University

Manski, C., (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, pp. 357-362.

Matzkin, R.L. (1997) "Nonparametric Estimation of Random Functions in Economic Models," mimeo.

Matzkin, R.L. and W. Newey (1993) "Kernel Estimation of Nonparametric Limited Dependnet Variable Models,", mimeo.

Mundlak, Y. (1978)"On the Pooling of Time Series and Cross Section Data", *Econometrica* 46 pp. 69-85

W. Newey (1994) "Kernel Estimation of Partial Means and a General Variance Estimator", *Econometric Theory*, 10, 233-253.

Powell, J.L, (1994) "Estimation of Semiparametric Models", in *Handbook of Econometrics*, Volume IV, Edited by R.F. Engle and D.L. McFadden, Elsevier Science B.V.

Rosenzweig, M. R. and K. I. Wolpin, (1994) "Parental and public transfers to young women and their children", *American Economic Review*, v84 n5 p1195-2018.

Rosenzweig, M. R. and K. I. Wolpin, (1994) "Inequality among young adult siblings, public assistance programs, and intergenerational living arrangements", *Journal of Human Resources* Fall v29 n4 pp110-25.

Table 1: *Monte Carlo Simulations of AM-1. Continuous Dependent Variable Case*
(750 replications). Slopes evaluated at x_{ik}=0.5

$y_{it} = b_0 + b_1 x_{ik} + \gamma x_{ik}\eta_i + \theta_\eta \eta_i + \theta_\varepsilon \varepsilon_i + u_{ik}$: $k = 1,2$; $i = 1,2,...1500$

$x_{ik} = x_i + \tilde{x}_{ik}$; $\varepsilon_i = \theta_{\varepsilon x} x_i + \tilde{\varepsilon}_i$; $\eta_i = 1x_i + \tilde{\eta}_{ik}$;

$x_i \sim N(0,1)$; $\tilde{x}_{ik} \sim N(0,1)$; $\tilde{\varepsilon}_i \sim N(0,1)$; $\tilde{\eta}_{ik} \sim N(0,1)$; $u_{ik} \sim N(0,1)$

Case 1: $y_{it} = 0 + 2x_{ik} + 0x_{ik}\eta_i + 1\eta_i + 0\varepsilon + u_{ik}$; $\theta_{\eta x}$=1

		Estimation Method					
	True value	AM-1 (poly/llr)	AM-1 (poly/ker)	AM-1 (llr/llr)	AM-1 (ker/ker)	OLS	OLS Fixed Effects
Mean of $\hat{\beta}(0.5)$	2.0	2.00 (0.056)	2.00 (0.043)	2.00 (0.078)	2.00 (0.043)	2.50 (0.072)	2.00 (0.092)

Case 2: $y_{it} = 0 + 2x_{ik} + 1x_{ik}\eta_i + 1\eta_i + 0\varepsilon + u_{ik}$:

Mean of $\hat{\beta}(0.5)$	2.25	2.25 (0.127)	2.25 (0.1[illegible]2)	2.23 (0.107)	2.25 (0.109)	2.50 (0.072)	2.00 (0.092)

Case 3: $y_{it} = 0 + 2x_{ik} + 1x_{ik}\eta_i + 1\eta_i + 0\varepsilon + u_{ik}$:

Mean of $\hat{\beta}(0.5)$	2.25	2.25 (0.107)	2.25 (0.1[illegible])	2.24 (0.90)	2.25 (0.010)	1.51 (0.074)	2.01 (0.069)

Notes: AM-1(poly/llr) uses a third order polynomial in x_{ik} and z_i to approximate $E(y_{ik}|x_{ik},z_i)$ and a local linear regression to approximate $h(z_i|x_{ik})$. AM-1(poly/ker) uses a third order polynomial in x_{ik} and z_i to approximate [illegible] AM-1(poly/ker) uses kernel regression to estimate $h(z_i|x_{ik})$. AM-1(llr/llr) uses local linear regression to estimate for both parts of the estimation, and AM-1(ker/ker) uses kernel regression. OLS is OLS regression applied to a model relating y_{ik} to a constant and a cubic in x_{ik}. OLS-Fixed Effects is OLS applied to a model that contains a separate intercept for each group i. Technical details are given in footnote ___.

Table 2: Monte Carlo Simulations of AM-1 in the Case of a Binary Dependent Variable (750 replications).

$y_{it} = I(b_0 + b_1 x_{ik} + \gamma x_{ik}\eta_i + \theta_\eta \eta_i + u_{ik})$; $k = 1,2$; $i = 1,2,\ldots 1500$

$x_{ik} = x_i + \tilde{x}_{ik}$; $\eta_i = \theta_{\eta x} x_i + \tilde{\eta}_{ik}$;

$x_i \sim N(0,\sigma_x^{\ 2})$; $\tilde{x}_{ik} \sim N(0,\sigma_{\tilde{x}}^{\ 2})$; $\tilde{\eta}_{it} \sim N(0.05)$; $u_{ik} \sim N(0,1)$

Case 1: $y_{it} = I(0 + 0x_{ik} + 0x_{ik}\eta_i + 1\eta_i + 0\varepsilon_i + u_{ik})$; $\theta_{\eta x}=1$; $\sigma_x^2=1.5$; $\sigma_{\tilde{x}}^2= 1.5$; $\sigma_{\tilde{\eta}}^2=0.5$;

		Estimation Method		
	True value	AM-1 (poly. lir)	AM-1 (kernel)	Probit
Mean (sd) of $\hat{\beta}(2.0)$	0.0	0.002 (0.022)	-0.001 (0.025)	0.101 (0.011)
Mean (sd) of $\hat{\beta}(1.0)$	0.0	0.001 (0.015)	-0.001 (0.022)	0.114 (0.009)
Mean (sd) of $\hat{\beta}(0.0)$	0.0	0.000 (0.011)	0.001 (0.021)	0.118 (0.074)
Mean (sd) of $\hat{\beta}(-1.0)$	0.0	0.000 (0.015)	0.001 (0.021)	0.114 (0.009)
Mean (sd) of $\hat{\beta}(-2.0)$	0.0	0.001 (0.021)	-0.001 (0.025)	0.101 (0.011)