Endogeneity in Semiparametric Binary Response Models

Richard Blundell[*] and James L. Powell[†]

July 2001
Revised, September 2003

Abstract

This paper develops and implements semiparametric methods for estimating binary response (binary choice) models with continuous endogenous regressors. It extends existing results on semiparametric estimation in single index binary response models to the case of endogenous regressors. It develops a *control function* approach to account for endogeneity in triangular and fully simulataneous binary response models. The proposed estimation method is applied to estimate the income effect in a labor market participation problem using a large micro data set from the British FES. The semiparametric estimator is found to perform well, detecting a significant attenuation bias. The proposed estimator is contrasted to the corresponding Probit and Linear Probability specifications.

JEL: C14, C25, C35, J22.

Key Words: Binary Response, Probit, Endogeneity, Semiparametric Estimation.

Address for correspondence: Richard Blundell, Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK. email: r.blundell@ucl.ac.uk

[*]University College London, Department of Economics, Gower Street, London, WC1E 6BT and Institute for Fiscal Studies. r.blundell@ucl.ac.uk

[†]Econometrics Laboratory, Department of Economics, University of California at Berkeley, Berkeley, CA 94720-3880.

1. Introduction[1]

This aim of this paper is to develop and implement semiparametric methods for estimating binary response models with endogenous regressors. The case of interest here is a single index model for a binary dependent variable with continuous endogenous regressors. Other covariates and the instrumental variables may be discrete. This paper extends the extensive literature dealing with semiparametric estimation in single index binary response models to the continuous endogenous regressor case. It highlights the attractiveness of the *control function* approach, which introduces residuals from the reduced form for the regressors as covariates in the binary response model to account for endogeneity in this framework.

In binary parametric models of this kind the control function approach to the estimation of simultaneous equations models can be linked directly to conditional likelihood and in that setting it has been used extensively to model endogeneity in discrete and limited dependent variable models (see Blundell and Smith 1986, 1989, for example). In semiparametric settings the control function approach has been used to account for endogeneity in the triangular systems in which the endogenous variables are fully observed (e.g., Newey, Powell and Vella 1998; Das, Newey, and Vella 2003). This paper draws on these previous results and develops a semiparametric estimator for the binary single index model. Our focus here is

on the control function approach to nonparametric estimation with endogenous regressors. In nonlinear models this differs from the standard assumption of the instrumental variable (IV) approach – namely, that the instrumental variables are independent of the error term in the equation of interest. In the binary response model the parameters of interest in semi- and nonparametric binary response models are not identified in general under the standard IV assumption (see Blundell and Powell 2003); however, we show that many of these parameters, including the index coefficients and *average structural function* (ASF), are identified through the control function assumptions we consider here. While our general approach is amenable to implementation using any of a number of estimation methods for single index models, this paper focuses on a particular single-index estimator proposed by Ahn, Ichimura, and Powell (1996), which is modified to accommodate endogenous regressors.

The proposed estimator is used to investigate the importance of correcting for the endogeneity of other income in a labor market participation model for a sample of married British men. The participation rate for this group is around 85% and the counterfactual effect we wish to recover is the impact of an exogenous change in other family income. As this variable may be endogenous for the participation decision, the standard parametric and semiparametric estimators for binary response will not recover consistent estimates of this parameter of interest. Instead we use the variation in exogenous benefit income as an instrumental variable to implement the control function approach in this semiparametric setting. We argue that this example is well suited for illustrating the usefulness of our approach for at least three reasons. First, the instrumental variable we have at our disposal exploits the exogenous variation in welfare benefit rules. Second, the sample size

is large. Third, a participation rate of 85% suggests that parametric model results may be sensitive to distributional assumptions which differ in their specification of tail probabilities. We find a strong effect of correcting for endogeneity in this example and show that adjusting for endogeneity using the standard parametric models, the Probit and linear probability models, give a highly misleading picture of the impact on participation of an exogenous change in other income.

As an alternative representation of the simultaneous binary response model, we also consider a framework which corresponds to a "fixed costs of work" representation of the participation decision. This is a fully simultaneous specification in which the binary outcome enters directly in to the model for other income. It corresponds to an economic model in which decision making is over observed outcomes. In this case there is no explicit reduced form for other income since the fixed cost of work is subtracted from income when working. Consequently this specification does not permit a triangular representation and turns out to be a special case of the coherency model framework for dummy endogenous models developed for the simultaneous Probit model by Heckman (1978). Blundell and Smith (1994) considered the control function approach for estimation in the joint normal model. We show that our semiparametric framework is equally well suited to this fully simultaneous specification of the binary response model with endogenous regressors.

The layout of the paper is as follows. In section 2 we present the model specification and discuss our approach to identification of the binary response coefficients and the average structural function. Section 3 defines and motivates the proposed estimators and derives their statistical properties. In section 4 we describe the data on participation and the empirical implementation of the approach to the

3

correction of endogeneity of other income in the participation decision. The correction for endogeneity is found to be important and the estimated effect is shown to be strongly biased when inappropriate parametric distributional assumptions are imposed. Section 5 develops the implementation of our approach to the fully simultaneous coherency model which allows for fixed costs of participation. Section 6 concludes.

2. Model Specification

In this paper we consider the binary response model

$$y_{1i} = 1\{y_{1i}^* > 0\}, \tag{2.1}$$

where the latent variable y_1^* is assumed to be generated from a linear model of the form

$$y_{1i}^* = \mathbf{x}_i \boldsymbol{\beta}_0 + u_i, \tag{2.2}$$

where $\mathbf{x}_i$ is a (row) vector of explanatory variables for observations $i = 1, ..., n$, u_i is an unobservable scalar error term, and the conformable column vector $\boldsymbol{\beta}_0$ of unknown regression ("index") coefficients is defined up to some scalar normalization (possibly involving the distribution of u_i). If u_i were assumed to be independent of $\mathbf{x}_i$, with (possibly unknown) distribution function $\Pr\{-u_i \leq \lambda\} = G(\lambda)$, the binary variable y_{1i} would satisfy a "single index regression" model of the form

$$E[y_{1i} \mid \mathbf{x}_i] = G(\mathbf{x}_i \boldsymbol{\beta}_0), \tag{2.3}$$

and the coefficients $\boldsymbol{\beta}_0$ could be estimated using standard single-index estimation procedures.

4

In some settings, though, the assumption of independence of the error term u_i and the regressors $\mathbf{x}_i$ would be suspect, if some components of $\mathbf{x}_i$ (denoted $\mathbf{y}_{2i}$ here) were determined jointly with the latent variable y_{1i}^*, as in the usual simultaneous equations framework. That is, endogeneity in some components of $\mathbf{x}$ might be accommodated through the following recursive structural form:

$$
\begin{aligned}
y_{1i} &= 1\{\mathbf{x}_i\boldsymbol{\beta}_0 + u_i > 0\} \\
&= 1\{\mathbf{z}_{1i}\boldsymbol{\beta}_1 + \mathbf{y}_{2i}\boldsymbol{\beta}_2 + u_i > 0\},
\end{aligned}
\tag{2.4}
$$

where $\mathbf{x}_i \equiv (\mathbf{z}_{1i}, \mathbf{y}_{2i})$ is of dimension $(1 \times (p+q))$, the q-vector $\mathbf{y}_{2i}$ is assumed to be determined by the reduced form

$$
\begin{aligned}
\mathbf{y}_{2i} &= E[\mathbf{y}_{2i}|\mathbf{z}_i] + \mathbf{v}_i \\
&\equiv \boldsymbol{\pi}(\mathbf{z}_i) + \mathbf{v}_i,
\end{aligned}
\tag{2.5}
$$

and the vector of instruments

$$
\mathbf{z}_i \equiv (\mathbf{z}_{1i}, \mathbf{z}_{2i})
\tag{2.6}
$$

is of dimension $1 \times (p+m)$, with $m \geq q$. (Here $m - q$ describes the degree of overidentification.) By construction, the reduced form error terms $\mathbf{v}_i$ have

$$
E(\mathbf{v}_i|\mathbf{z}_i) = \mathbf{0},
\tag{2.7}
$$

though alternative centering assumptions (e.g., conditional median zero) would be compatible with the approach taken here.

If the joint distribution of the structural error term u_i and reduced form error terms $\mathbf{v}_i$ were parametrically specified (as, say, Gaussian and independent of $\mathbf{z}_i$), and if the reduced-form regression function $\pi(\mathbf{z}_i)$ were also parametrized, then

maximum likelihood estimation could be applied to obtain consistent estimators of $\boldsymbol{\beta}_0$, $\boldsymbol{\pi}(\cdot)$, and the unknown parameters of the joint distribution function of the errors. To be specific, assuming a joint normal distribution for *the* error terms and a particular normalization for $Var(u)$, we have

$$
\begin{aligned}
E(y_{1i}|\mathbf{x}_i, \mathbf{v}_i) &= \Pr[u_i > -\mathbf{x}_i\boldsymbol{\beta}|\mathbf{v}_i] \\
&= \Phi(\mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\rho}\mathbf{v}_i),
\end{aligned}
\tag{2.8}
$$

where $\boldsymbol{\rho}$ is the vector of population regression coefficients of u_i on $\mathbf{v}_i$. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ can be estimated directly from the conditional likelihood for y_{1i} given $\mathbf{x}_i$ and $\mathbf{v}_i$. Blundell and Smith (1986) show that, unlike in the linear model case, this "control function" approach is asymptotically more efficient in discrete and censored normal models than an alternative two-stage estimation approach.

In a semiparametric setting, where the joint error distribution and reduced form regression functions are not parametrically specified, one possible "two stage" estimation approach would insert the reduced form for y_{2i} into the structural model (2.4), yielding

$$
y_{1i} = 1\{\mathbf{z}_{1i}\boldsymbol{\beta}_1 + \boldsymbol{\pi}(\mathbf{z}_i)\boldsymbol{\beta}_2 + u_i + \mathbf{v}_i\boldsymbol{\beta}_0 > 0\},
\tag{2.9}
$$

which would yield a single index representation

$$
E[y_{1i}|\mathbf{z}_i] = H\left(\boldsymbol{\pi}(\mathbf{z}_i)\boldsymbol{\beta}_0\right),
\tag{2.10}
$$

assuming the composite error term $u_i + \mathbf{v}_i\boldsymbol{\beta}_0$ is independent of $\mathbf{z}_i$ with c.d.f. $H(-\lambda)$. While a two-stage estimation approach (using a nonparametric estimator of $\boldsymbol{\pi}(\mathbf{z}_i)$ in the first stage) could be used to consistently estimate the parameters $\boldsymbol{\beta}_0$, the assumption of independence of the instruments $\mathbf{z}_i$ and the composite error

6

term $u_i + \mathbf{v}_i\boldsymbol{\beta}_0$ might be difficult to maintain, particularly if the reduced form error terms $\mathbf{v}_i$ do not appear to be independent of the instruments (as might be revealed by standard tests for heteroskedasticity, etc.). Moreover, this 'fitted value' approach does not easily yield an estimator of the marginal c.d.f. $G(\cdot)$ of the error terms $-u_i$, which would be needed to evaluate the effect on response probabilities of an exogenous shift in the regressors $\mathbf{x}_i$.

An alternative approach to estimation of the components of this model, adopted here, uses estimates of the reduced form error terms $\mathbf{v}_i$ as "control variables" for the endogeneity of the regressors in the original structural equation. The key identifying assumptions for estimation of the unknown coefficients $\boldsymbol{\beta}$ and the distribution function of the error term u_i is a distributional exclusion restriction, which requires that the dependence of the structural error term u_i on the vector of regressors $\mathbf{x}_i$ and instrumental variables $\mathbf{z}_i$ is completely characterized by the reduced form error vector $\mathbf{v}_i$: that is,

$$u_i|\mathbf{x}_i, \mathbf{z}_i \quad \sim \quad u_i|\mathbf{x}_i, \mathbf{v}_i \tag{2.11}$$

$$\sim \quad u_i|\mathbf{v}_i, \tag{2.12}$$

where the tilde symbol denotes equality of conditional distributions. Under this last condition, the conditional expectation of the binary variable y_{1i} given the regressors $\mathbf{x}_i$ and reduced form errors $\mathbf{v}_i$ takes the form

$$
\begin{aligned}
E[y_{1i}|\mathbf{x}_i, \mathbf{v}_i] \quad &= \quad \Pr[-u_i \leq \mathbf{x}_i\boldsymbol{\beta}_0|\mathbf{x}_i, \mathbf{v}_i] \\
&= \quad F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i),
\end{aligned}
\tag{2.13}
$$

where $F(\cdot, \mathbf{v}_i)$ is the conditional c.d.f. of $-u_i$ given $\mathbf{v}_i$. Thus, y_{1i} can be characterized by a "multiple index regression" model, with conditional distribution, given

7

$\mathbf{x}_i$ and $\mathbf{v}_i$, that depends upon $\mathbf{x}_i$ only through the single index $\mathbf{x}_i\boldsymbol{\beta}_0$. As in the single index regression model, it is the dimensionality reduction in the conditional expectation of y_{1i} given $\mathbf{x}_i$ and $\mathbf{v}_i$ that can be exploited to obtain a $\sqrt{n}$-consistent estimator of the index coefficients $\boldsymbol{\beta}_0$.

Our approach to identification and estimation of the unknown regression coefficients $\boldsymbol{\beta}_0$ uses an extension of the Ahn, Ichimura and Powell (1996) "matching" estimator of $\boldsymbol{\beta}_0$ for the single index model without endogeneity. This approach, adapted to the present context, assumes both the monotonicity and continuity of $F(\mathbf{x}_i\boldsymbol{\beta}, \mathbf{v}_i)$ in its first argument. Since the structural index model is related to the conditional mean of y_{1i} given $\mathbf{w}_i = (\mathbf{x}_i, \mathbf{v}_i)$ by the relation

$$E[y_{1i}|\mathbf{x}_i, \mathbf{v}_i] \equiv g(\mathbf{w}_i) = F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i), \tag{2.14}$$

invertibility of $F(\cdot)$ in its first argument implies

$$\mathbf{x}_i\boldsymbol{\beta}_0 - \psi(g(\mathbf{w}_i), \mathbf{v}_i) = 0, \qquad \text{(w.p.1)}, \tag{2.15}$$

where $\psi(\cdot, \mathbf{v}) \equiv F^{-1}(\cdot, \mathbf{v})$, i.e.,

$$F(\psi(g, \mathbf{v}), \mathbf{v}) = g. \tag{2.16}$$

If two observations (with subscripts i and j) have identical conditional means (i.e., $g(\mathbf{w}_i) = g(\mathbf{w}_j)$) and identical reduced form error terms ($\mathbf{v}_i = \mathbf{v}_j$), it follows that their indices $\mathbf{x}_i\boldsymbol{\beta}_0$ and $\mathbf{x}_j\boldsymbol{\beta}_0$ are also identical:

$$\begin{aligned}
(\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\beta}_0 &= \boldsymbol{\psi}(g(\mathbf{w}_i), \mathbf{v}_i) - \boldsymbol{\psi}(g(\mathbf{w}_j), \mathbf{v}_j) \\
&= 0 \qquad \text{if} \tag{2.17} \\
g(\mathbf{w}_i) &= g(\mathbf{w}_j) \qquad \text{and} \qquad \mathbf{v}_i = \mathbf{v}_j. \tag{2.18}
\end{aligned}$$

So, for any nonnegative function $\omega(\mathbf{w}_i, \mathbf{w}_j)$ of the conditioning variables $\mathbf{w}_i$ and $\mathbf{w}_j$, it follows that

$$
\begin{aligned}
0 &= E[\omega(\mathbf{w}_i, \mathbf{w}_j) \cdot ((\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\beta}_0)^2 \mid g(\mathbf{w}_i) = g(\mathbf{w}_j), \mathbf{v}_i = \mathbf{v}_j] \\
&\equiv \boldsymbol{\beta}_0' \boldsymbol{\Sigma}_w \boldsymbol{\beta}_0, \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (2.19)
\end{aligned}
$$

where

$$
\boldsymbol{\Sigma}_w \equiv E[\omega(\mathbf{w}_i, \mathbf{w}_j) \cdot (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) \mid g(\mathbf{w}_i) = g(\mathbf{w}_j), \mathbf{v}_i = \mathbf{v}_j]. \quad (2.20)
$$

That is, the nonnegative-definite matrix $\boldsymbol{\Sigma}_w$ is singular, and the unknown parameter vector $\boldsymbol{\beta}_0$ is the eigenvector (with an appropriate normalization) corresponding to the zero eigenvalue of $\boldsymbol{\Sigma}_w$. Under the identifying assumption that $\boldsymbol{\Sigma}_w$ has rank $p + q - 1 = \dim(\mathbf{x}) - 1$ – which requires that any nontrivial linear combination $(\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\lambda}$ of the difference in regressors has nonzero conditional variance when $\boldsymbol{\lambda} \neq \boldsymbol{\beta}_0$ – the parameter vector $\boldsymbol{\beta}_0$ as the eigenvector corresponding to the unique zero eigenvalue of $\boldsymbol{\Sigma}_w$. We construct an estimator of $\boldsymbol{\beta}_0$ as the eigenvector corresponding to the smallest eigenvalue (in magnitude) of a sample analogue to the $\boldsymbol{\Sigma}_w$ matrix, for a convenient choice of the weighting function $\omega(\cdot)$.

In addition to the vector of regression coefficients $\boldsymbol{\beta}_0$ (defined up to some scale normalization), the other key parameter of interest is the marginal probability distribution function of the structural errors $-u_i$,

$$
G(\lambda) = \Pr[-u_i \leq \lambda]; \quad\quad\quad\quad\quad\quad\quad\quad (2.21)
$$

when $\lambda = \mathbf{x}\boldsymbol{\beta}_0$, we define this function $G(\mathbf{x}\boldsymbol{\beta}_0)$ to be the *Average Structural Function (ASF)* since it represents the response probability for an exogenously-determined setting of the regressors $\mathbf{x}$ (analogous to the 'structural demand function' in a simultaneous equations model), and its derivative with respect to $\mathbf{x}$

is the marginal response to an exogenous shift in $\mathbf{x}$. The marginal distribution function $G(\lambda)$ of $-u_i$ can be identified as the "partial mean" of this conditional distribution function $F(\mathbf{x}\boldsymbol{\beta}_0, \mathbf{v}_i)$, holding the index $\mathbf{x}\boldsymbol{\beta}_0$ fixed and averaging over the marginal distribution of the reduced form errors $\mathbf{v}_i$:

$$G(\mathbf{x}\boldsymbol{\beta}_0) = \int F(\mathbf{x}\boldsymbol{\beta}_0, \mathbf{v})dF_{\mathbf{v}}. \tag{2.22}$$

And, given a particular marginal distribution $F_{\mathbf{x}}^*$ for the regressors $\mathbf{x}$ of interest (including possibly the observed marginal distribution), the average response probability for exogenous regressors with that distribution would be

$$\gamma^* \equiv \int G(\mathbf{x}\boldsymbol{\beta}_0)dF_{\mathbf{x}}^*, \tag{2.23}$$

which may be of interest for policy analysis (Stock 1989).

3. Estimation

3.1. The semiparametric estimation approach

The approach adopted here for estimation of the parameters of interest in this model follows three main steps. The first step uses nonparametric regression methods to estimate the error term $\mathbf{v}_i$ in the reduced form, as well as the unrestricted conditional mean of y_{1i} given $\mathbf{x}_i$ and $\mathbf{v}_i$,

$$
\begin{aligned}
E[y_{1i}|\mathbf{z}_{1i}, \mathbf{y}_{2i}, \mathbf{v}_i] &\equiv E[\mathbf{y}_{1i}|\mathbf{x}_i, \mathbf{v}_i] && (3.1) \\
&\equiv E[\mathbf{y}_{1i}|\mathbf{w}_i] \\
&\equiv g(\mathbf{w}_i)
\end{aligned}
$$

where $\mathbf{w}_i$ is the $1 \times (p + 2q)$ vector

$$\mathbf{w}_i \equiv (\mathbf{x}_i, \mathbf{v}_i). \tag{3.2}$$

10

This step can be viewed as an intermediate 'structural estimation' step, which imposes the first exclusion restriction of (2.11) and (2.12) but not the second.

The second step imposes the linear index assumptions on the unrestricted conditional mean

$$
\begin{aligned}
E[\mathbf{y}_{1i}|\mathbf{z}_{1i}, \mathbf{y}_{2i}, \mathbf{v}_i] &\equiv E[\mathbf{y}_{1i}|\mathbf{x}_i, \mathbf{v}_i] \qquad\qquad (3.3) \\
&= E[\mathbf{y}_{1i}|\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i] \\
&\equiv F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i)
\end{aligned}
$$

to obtain an estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$.

The final step recovers an estimator of $G(\mathbf{x}\boldsymbol{\beta}_0) \equiv \int F(\mathbf{x}\boldsymbol{\beta}_0, \mathbf{v})dF_{\mathbf{v}}$. It is obtained by computing a sample average of $\hat{F}(\lambda, \mathbf{v}_i)$ over the observations, holding λ fixed and substituting the estimates $\hat{\mathbf{v}}_i$ for $\mathbf{v}_i$, an application of "partial mean" estimation (Newey 1994).

3.2. Implementation of the estimation approach

Let $\{(\mathbf{y}_i, \mathbf{z}_i)\}_{i=1}^n$ be a random sample of observations from the model described above. In estimation of the conditional mean of y_{1i}, $\mathbf{v}_i$ is replaced by the residual from the nonparametric regression of $\mathbf{y}_{2i}$ on $\mathbf{z}_i$, so that in place of (3.2) we have

$$
\begin{aligned}
\widehat{\mathbf{w}}_i &= (\mathbf{x}_i, \mathbf{y}_{2i} - \widehat{\boldsymbol{\pi}}(\mathbf{z}_i)) \qquad\qquad (3.4) \\
&\equiv (\mathbf{x}_i, \hat{\mathbf{v}}_i),
\end{aligned}
$$

where $\widehat{\boldsymbol{\pi}}(\mathbf{z}_i)$ is the unrestricted Nadaraya-Watson kernel regression estimator for the mean of $\mathbf{x}_i$ given $\mathbf{z}_i$.

Estimation of the conditional mean function in step 1 then replaces the population moments $g(\mathbf{w}_i) = E[y_{1i}|\mathbf{w}_i]$ by the unrestricted Nadaraya-Watson kernel

11

regression estimator

$$\widehat{g}(\widehat{\mathbf{w}}_i) \equiv \widehat{f}(\widehat{\mathbf{w}}_i)^{-1}\widehat{r}(\widehat{\mathbf{w}}_i), \tag{3.5}$$

with

$$\widehat{r}(\widehat{\mathbf{w}}_i) \equiv \frac{1}{n}\sum_j K_w\left(\widehat{\mathbf{w}}_i - \widehat{\mathbf{w}}_j\right)y_{1j}, \tag{3.6}$$

$$\widehat{f}(\widehat{\mathbf{w}}_i) \equiv \frac{1}{n}\sum_j K_w\left(\widehat{\mathbf{w}}_i - \widehat{\mathbf{w}}_j\right), \tag{3.7}$$

where $K_w(\boldsymbol{\zeta}) = h_w^{-(p+2q)}K(\boldsymbol{\zeta}/h_w)$ for bandwidth h_n satisfying $h_w \to 0$ and $nh_w^{p+2q} \to \infty$ as $n \to \infty$, and some kernel function $K : R^{p+2q} \to R^+$ that satisfies standard conditions like $\int K(\boldsymbol{\zeta})d\boldsymbol{\zeta} = 1$.

Given the preliminary nonparametric estimators $\widehat{\mathbf{v}}_i$ and $\widehat{g}(\widehat{\mathbf{w}}_i)$ of $\mathbf{v}_i$ and $g(\mathbf{w}_i)$ defined above, and assuming smoothness (continuity and differentiability) of the inverse function $\psi(\cdot)$ in (2.16), a consistent estimator of $\boldsymbol{\Sigma}_w$ for a particular weighting function $\omega(\mathbf{w}_i, \mathbf{w}_j)$ can be obtained by a 'pairwise differencing' or 'matching' approach which takes a weighted average of outer products of the differences $(\mathbf{x}_i - \mathbf{x}_j)$ in the $\binom{n}{2}$ distinct pairs of regressors, with weights that tend to zero as the magnitudes of the differences $|\widehat{g}(\widehat{\mathbf{w}}_i) - \widehat{g}(\widehat{\mathbf{w}}_j)|$ and $|\mathbf{v}_i - \mathbf{v}_j|$ increase. As in Ahn, Ichimura, and Powell (1996), the estimator of $\boldsymbol{\Sigma}_w$ is defined as

$$\widehat{\mathbf{S}}_w \equiv \binom{n}{2}^{-1}\sum_{i<j}\widehat{\omega}_{ij}(\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j), \tag{3.8}$$

for

$$\widehat{\omega}_{ij} \equiv h_n^{-q+1}K_g\left(\frac{\widehat{g}(\widehat{\mathbf{w}}_i) - \widehat{g}(\widehat{\mathbf{w}}_j)}{h_n}\right)K_v\left(\frac{\widehat{\mathbf{v}}_i - \widehat{\mathbf{v}}_j}{h_n}\right)t_it_j, \tag{3.9}$$

where K_g and K_v are kernel functions analogous to the kernel function K_w defined above, and where t_i and t_j are 'trimming' terms which equal zero when

12

the conditioning variables $\mathbf{z}_i$ and $\widehat{\mathbf{w}}_i$ lie outside a compact set. As the sample size n increases, and the bandwidth h_n shrinks to zero, the weighting term $\widehat{\omega}_{ij}$ tends to zero, except for pairs of observations with $\widehat{g}(\widehat{\mathbf{w}}_i) \simeq \widehat{g}(\widehat{\mathbf{w}}_j)$ and $\widehat{\mathbf{v}}_i \simeq \widehat{\mathbf{v}}_j$.

With this estimator of the matrix $\boldsymbol{\Sigma}_w$, a corresponding estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ can be defined as the eigenvector corresponding to the eigenvalue $\hat{\eta}$ of $\hat{\mathbf{S}}_w$ that is closest to zero in magnitude. (Because the kernel functions K_g and K_v will be permitted to become negative for some values of their arguments, we choose the eigenvalue closest to zero, rather than the minimum of the eigenvalues.) A convenient normalization of this eigenvector sets the first component to unity; that is, we normalize

$$\boldsymbol{\beta}_0 = \begin{pmatrix} 1 \\ -\boldsymbol{\theta}_0 \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 \\ -\hat{\boldsymbol{\theta}} \end{pmatrix}, \tag{3.10}$$

and, partitioning $\hat{S}$ conformably as

$$\hat{\mathbf{S}}_w = \begin{bmatrix} \hat{\mathbf{S}}_{11} & \hat{\mathbf{S}}_{12} \\ \hat{\mathbf{S}}_{21} & \hat{\mathbf{S}}_{22} \end{bmatrix}, \tag{3.11}$$

the estimator of the subvector $\boldsymbol{\theta}_0$ can be written as

$$\hat{\boldsymbol{\theta}} = [\hat{\mathbf{S}}_{22} - \hat{\eta}\mathbf{I}]^{-1}\hat{\mathbf{S}}_{21}, \tag{3.12}$$

where

$$\hat{\eta} \equiv \arg\min_{\alpha:\|\alpha\|=1} |\alpha'\hat{\mathbf{S}}_w\alpha|. \tag{3.13}$$

Using the estimator $\widehat{\boldsymbol{\beta}}$, we can estimate the index as $\widehat{\lambda} = \mathbf{x}\widehat{\boldsymbol{\beta}}$ for any value of $\mathbf{x}$, which we can then use to estimate the average structural function $G(\mathbf{x}\boldsymbol{\beta}_0)$, the marginal probability that $y_{1i} = 1$ given an exogenous $\mathbf{x}$. The joint probability function $F(\mathbf{x}\boldsymbol{\beta}_0, \mathbf{v})$ can be directly estimated through a nonparametric (kernel) regression of y_{1i} on $\mathbf{x}_i\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{v}}_i$. The approach then estimates the average structural

function $G(\mathbf{x}\boldsymbol{\beta}_0)$ from a sample average of $\widehat{F}(\mathbf{x}\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{v}}_i)$ over $\widehat{\mathbf{v}}_i$,

$$\widehat{G}(\mathbf{x}\widehat{\boldsymbol{\beta}}) = \sum_{i=1}^{n} \widehat{F}(\mathbf{x}\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{v}}_i)\tau_i, \tag{3.14}$$

where $\tau_i = \tau(\widehat{\mathbf{v}}_i, n)$ is some 'trimming' term which downweights observations for which $\widehat{F}(\cdot)$ is imprecisely estimated.

3.3. Alternative assumptions and estimators

Given the assumptions imposed in section 2 above, several variations on the estimation strategy outlined in the previous section might be adopted. For example, to obtain the first-stage estimates of the conditional mean function $g(\mathbf{w}_i) = E[y_{1i}|\mathbf{w}_i]$ and residuals $\mathbf{v}_i = \mathbf{x}_i - E[\mathbf{x}_i|\mathbf{z}_i]$, more sophisticated nonparametric regression methods like local polynomial regression might be used instead of the simpler kernel estimators adopted here. And, given a $\sqrt{n}$-consistent estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, an alternative to the "partial mean" estimator of the average structural function $\widehat{G}(\lambda)$ could be based upon "marginal integration" of $\widehat{F}(\lambda, \mathbf{v})$, weighting by the estimated joint density function $\widehat{f}_{\mathbf{v}}(\mathbf{v})$ for $\mathbf{v}_i$,

$$\widetilde{G}(\lambda) = \int \widehat{F}(\lambda, \mathbf{v})\widehat{f}_{\mathbf{v}}(\mathbf{v})d\mathbf{v}, \tag{3.15}$$

as discussed by Linton and Nielson (1995) and Tjosthiem and Auestad (1994).

For estimation of $\boldsymbol{\beta}_0$, the index coefficients, the multi-index restriction $E[y_{1i}|\mathbf{x}_i, \mathbf{v}_i] = F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i)$ can be exploited with modification of most existing single-index estimation methods. The "average derivative" (Härdle and Stoker 1989) or "weighted average derivative" (Powell, Stock, and Stoker 1989) estimation methods for single index models could be adapted to estimation of $\boldsymbol{\beta}_0$ here, using the fact that

$$E\left[w(\mathbf{x}_i, \mathbf{v}_i)\cdot\frac{\partial E[y_{1i}|\mathbf{x}_i, \mathbf{v}_i]}{\partial \mathbf{x}}\right] = \boldsymbol{\beta}_0\cdot E\left[w(\mathbf{x}_i, \mathbf{v}_i)\cdot\frac{\partial F(\mathbf{x}_i\boldsymbol{\beta}, \mathbf{v}_i)}{\partial(\mathbf{x}\boldsymbol{\beta})}\right] \propto \boldsymbol{\beta}_0. \tag{3.16}$$

14

With these estimators, as with the recent proposal by Hristache, Juditsky, and Spokoiny 2001 (which combines average derivative and local linear regression estimation), the function $F(\cdot)$ need not be assumed to be monotonic in $\mathbf{x}_i\boldsymbol{\beta}_0$, but all components of $\mathbf{x}_i$ must be continuously distributed, which is uncommon in empirical applications, including the one presented herein. Other possibilities include generalizations of the single-index regression estimator of Ichimura (1993) or semiparametric binary response estimator of Klein and Spady (1993); these would fit the nonlinear regression function $E[y_{1i}|\mathbf{x}_i, \mathbf{v}_i] = F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i)$ iteratively by either nonlinear least squares or binary response maximum likelihood, using nonparametric regression to estimate $F(\mathbf{x}_i\mathbf{b}, \mathbf{v}_i)$ as a function of $\mathbf{b}$. Like the average derivative estimators, these estimators would not need the monotonicity requirement, can accommodate discontinuous regressors, involves lower-dimensional nonparametric regression, and can be more statistically efficient. However, while each nonparametric regression step involves a lower-dimensional regression (to estimate $E[y_{1_i}|\mathbf{x}_i\mathbf{b}, \mathbf{v}_i]$ rather than $E[y_{1i}|\mathbf{x}_i, \mathbf{v}_i]$), iteration between estimation of F and minimization or maximization over $\mathbf{b}$ makes these estimators less computationally tractable than the present estimator. Yet another candidate would be a "local probit" estimator which estimated the reparametrised function

$$F(\mathbf{x}\boldsymbol{\beta}_0, \mathbf{v}) \equiv \Phi(\mathbf{x}\boldsymbol{\beta}(\mathbf{x}, \mathbf{v}) + \mathbf{v}\boldsymbol{\rho}(\mathbf{x}, \mathbf{v})) \tag{3.17}$$

in a neighborhood of each value of $\mathbf{x}_i$ and $\widehat{\mathbf{v}}_i$ (where Φ is the standard normal cumulative) by maximizing a local likelihood

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\rho}; \mathbf{w}) \equiv \sum_{i=1}^{n} K_w\left(\mathbf{w} - \widehat{\mathbf{w}}_i\right) \left[y_{1i}\ln(\Phi(\mathbf{x}\boldsymbol{\beta} + \mathbf{v}\boldsymbol{\rho})) + (1 - y_{1i})\ln(1 - \Phi(\mathbf{x}\boldsymbol{\beta} + \mathbf{v}\boldsymbol{\rho}))\right];$$
$$\tag{3.18}$$

15

given local estimators $\widehat{\boldsymbol{\beta}}(\widehat{\mathbf{w}}) = \widehat{\beta}(\mathbf{x}, \widehat{\mathbf{v}})$ of the slope coefficients, the index coefficients $\boldsymbol{\beta}_0$ could be recovered by integration,

$$\widetilde{\boldsymbol{\beta}} = \int \widehat{\boldsymbol{\beta}}(\mathbf{x}, \mathbf{v}) d\widehat{F}(\mathbf{x}, \mathbf{v}), \tag{3.19}$$

and the parametric might model could be tested as a special case by testing the constancy of the estimated $\boldsymbol{\rho}(\mathbf{x}, \mathbf{v})$ function.[2]

All of these alternative index coefficient estimators, like the ones proposed here, are based upon the multi-index regression function $E[y_{1i}|\mathbf{x}_i, \mathbf{v}_i] = F(\mathbf{x}_i \boldsymbol{\beta}_0, \mathbf{v}_i)$, which in turn are derived from the distributional exclusion restrictions (2.11) and (2.12), which are the basis for the control function approach; thus, they all share in some restrictive aspects of those assumptions. First, as for single-index estimators of binary response models without endogeneity, the form of permissible heteroskedasticity in the structural error u_i is limited to functions of $\mathbf{x}_i \boldsymbol{\beta}_0$ and $\mathbf{v}_i$, ruling out, e.g., regressors with random coefficients. Also, location of the error term u_i is not identified, though this is only because no location restriction on u_i is imposed by the exclusion restrictions, and location parameters for u_i could be recovered directly from the ASF $G(\lambda)$.

More fundamentally, in contrast to the linear model, consistency of the control function approach for the binary response and other nonlinear models requires a correct, or at least complete, specification of the vector $\mathbf{z}_i$ of instruments for the first-stage nonparametric regression. In the linear model with endogenous regressors, least-squares regression of the dependent variable on the endogenous variables and first-stage residuals yields the two-stage least squares estimators as the estimated regression coefficients, so any set of valid instruments satisfying

---

[2]We are grateful to a referee for suggesting this possible extension.

the relevant rank condition will yield consistent estimators; in general, though, the conditional independence restrictions (2.11) and (2.12) will no longer hold in general if $\mathbf{z}_i$ is replaced by a subvector, so "correct" specification of the first stage is essential. Also, as discussed in more detail by Blundell and Powell (2003), estimation of the multi-index regression function $F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i)$ and its partial mean, the ASF $G(\mathbf{x}_i\boldsymbol{\beta}_0)$, requires the first-stage residual $\mathbf{v}_i$ (and thus the endogenous regressors) to be continuously distributed, which holds for our empirical application but not more generally.

If the primary interest is in the index coefficients $\boldsymbol{\beta}_0$, rather than the ASF $G(\mathbf{x}\boldsymbol{\beta}_0)$, it may be identified and consistently estimated under weaker conditions than the conditional independence restrictions (2.11) and (2.12). Consider, for example, replacing the distributional exclusion restrictions with the corresponding median exclusion restrictions

$$
\begin{aligned}
med\{u_i|\mathbf{x}_i, \mathbf{z}_i\} &= med\{u_i|\mathbf{x}_i, \mathbf{v}_i\} \\
&= med\{u_i|\mathbf{v}_i\} \quad\quad\quad (3.20) \\
&\equiv \eta(\mathbf{v}_i),
\end{aligned}
$$

which restrict only the 50th percentiles of the corresponding conditional distributions. Under these restrictions,

$$
med\{y_{1i}|\mathbf{x}_i, \mathbf{v}_i\} = 1\{\mathbf{x}_i\boldsymbol{\beta}_0 + \eta(\mathbf{v}_i)\}, \quad\quad\quad (3.21)
$$

and estimation of $\boldsymbol{\beta}_0$ might be based upon a generalization of Manski's (1975, 1985) maximum score estimation to semilinear regression models.[3] Identification

---

[3]We are grateful to a referee for suggesting this possible extension, and for the local probit estimation approach.

of the ASF, though, would require strengthening of the median exclusion restrictions to independence restrictions.

When the regressor vector $\mathbf{x}_i$ includes a (scalar) exogenous variable z_{0i} which is continuously distributed with a large support, it is possible to derive $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}$ under substantially weaker versions of the exclusion restrictions (2.11) and (2.12), using the approach proposed by Lewbel (2000) and its variants (Lewbel 1998, Honoré and Lewbel 2001). Writing the vector of regressors as $\mathbf{x}_i \equiv (z_{0i}, \mathbf{z}_{1i}, \mathbf{y}_{2i})$, the distributional exclusion restrictions can be relaxed to the form

$$
\begin{align}
u_i | \mathbf{x}_i, \mathbf{z}_i \quad &\sim \quad u_i | \mathbf{z}_{1i}, \mathbf{y}_{2i}, \mathbf{v}_i \tag{3.22} \\
&\sim \quad u | \mathbf{z}_{1i}, \boldsymbol{\pi}(\mathbf{z}_i), \mathbf{y}_{2i} \tag{3.23}
\end{align}
$$

which is supplemented by the unconditional moment restrictions

$$
\begin{align}
E[\mathbf{z}_{1i} u_i] &= \mathbf{0}, \tag{3.24} \\
E[\boldsymbol{\pi}(\mathbf{z}_i) u_i] &= \mathbf{0}, \tag{3.25}
\end{align}
$$

where $\boldsymbol{\pi}(\mathbf{z}_i) \equiv E[\mathbf{y}_{2i} | \mathbf{z}_i]$. Under this combination of conditional independence and unconditional moment restrictions, Lewbel (2000) shows that a linear least-squares regression of $[y_{1i} - 1\{z_{0i} > 0\}] / f(z_{0i} | \mathbf{z}_{1i}, \boldsymbol{\pi}(\mathbf{z}_i), \mathbf{y}_{2i})$ on $\mathbf{z}_{1i}$ and $\boldsymbol{\pi}(\mathbf{z}_i)$, where $f(\cdot)$ is the conditional density of the "special regressor" z_{0i} (which must be replaced by a consistent estimator in practice), yields a consistent estimator of the "normalized" coefficients $\boldsymbol{\theta}_0$ from (3.10). Additional restrictions on z_{0i} (e.g., its independence from $\boldsymbol{\pi}(\mathbf{z}_i)$, or from u_i and the other regressors) permit more general forms of heteroskedasticity and an incomplete specification of the set of instruments in the first stage regression. And, while Lewbel's approach requires a continuously-

18

distributed, exogenous regressor in the binary response equation (which is not available in our empirical example), it would accommodate discrete endogenous regressors, unlike the control function approach. Hence, the two approaches to estimation of $\boldsymbol{\beta}_0$ may be viewed as complementary, depending upon whether the endogenous or exogenous regressors are continuously distributed,[4] and either might be combined with the "partial mean" approach to estimation of the ASF $G(\mathbf{x}\boldsymbol{\beta}_0)$, which represents the response probability for an exogenously-specified regression vector $\mathbf{x}$. Our choice of the Ahn, Ichimura, and Powell (1996) generalization is motivated by its applicability to our empirical problem and its computational simplicity compared to competing procedures.

3.4. Large-sample properties of the proposed estimators

The objectives of the asymptotic theory for the estimators proposed here are, first, demonstration of consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ and, second, demonstration of (pointwise) consistency of the marginal distribution function estimator $\widehat{G}(\lambda)$. Derivation of the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ will be similar to that for the 'monotone single index' estimator proposed by Ahn, Ichimura and Powell (1996); the present derivation is more complex, both in analysis of the first stage estimator of $g(\mathbf{w}_i)$ (since the estimated residuals $\hat{\mathbf{v}}_i$ are used as covariates in a nonparametric regression step) and in the second stage estimator $\hat{\mathbf{S}}_w$ of $\boldsymbol{\Sigma}_w$, which must condition on the reduced-form residuals as well as the preliminary estimators $\hat{g}_i \cong \hat{g}(\tilde{\mathbf{w}}_i)$ and $\hat{\mathbf{v}}_i \equiv \mathbf{y}_{2i} - \hat{\boldsymbol{\pi}}(\mathbf{z}_i) \equiv \mathbf{y}_{2i} - \hat{\boldsymbol{\pi}}_i$.

Demonstration of consistency of the estimator $\boldsymbol{\beta}_0$ follows from consistency

---

[4]Identification of the ASF and index coefficients is problematic in nonlinear models with endogenous regressors without continuous components – see Blundell and Powell (2003).

19

of the matrix $\tilde{\mathbf{S}}_w$ for $\boldsymbol{\Sigma}_w$, plus an identification condition. Making a first-order Taylor's series expansion around the true values $g_i \equiv g(\mathbf{w}_i)$ and $\mathbf{v}_i$, the matrix $\hat{\mathbf{S}}_w$ can be decomposed as

$$\hat{\mathbf{S}}_w = \mathbf{S}_0 + \mathbf{S}_1, \tag{3.26}$$

where

$$\mathbf{S}_l \equiv \binom{n}{2}^{-1} \sum_{i<j} \omega_{ij}^l (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j), \qquad l = 0, 1, \tag{3.27}$$

$$\boldsymbol{\omega}_{ij}^0 \equiv h_n^{-(q+1)} K_g \left(\frac{g_i - g_j}{h_n} \right) K_v \left(\frac{\mathbf{v}_i - \mathbf{v}_j}{h_n} \right) t_i t_j, \tag{3.28}$$

and

$$\begin{aligned}
\omega_{ij}^1 &\equiv h_n^{-(q+2)} K_g^{(1)} \left(\frac{g_{ij}^*}{h_n} \right) K_v \left(\frac{\mathbf{v}_{ij}^*}{h_n} \right) (\hat{g}_i - \dot{g}_i - \hat{g}_j + g_j) t_i t_j \\
&\quad - h_n^{-(q+2)} K_g \left(\frac{g_{ij}^*}{h_n} \right) K_v^{(1)} \left(\frac{\mathbf{v}_{ij}^*}{h_n} \right) (\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i - \hat{\boldsymbol{\pi}}_j - \boldsymbol{\pi}_j)' t_i t_j. \tag{3.29}
\end{aligned}$$

In these expressions, the superscript '(1)' denotes the (row) vectors of first derivatives of the respective kernel functions, while g_{ij}^* and $\mathbf{v}_{ij}^*$ denote intermediate value terms.

The leading-order term $\mathbf{S}_0$ is the simplest to analyze, since it is in the form of a second-order U-statistic (with kernel depending upon the sample size n). It is straightforward to show that the first two moments of the summand in $\mathbf{S}_0$ are of order h_n^{q+1} (for q the dimension of $\mathbf{y}_{2i}$, the vector of endogenous regressors), provided the regressors $\mathbf{x}_i$ have four moments and the trimming terms t_i and kernel functions $K_g(\cdot)$ and $K_v(\cdot)$ are bounded. Thus, by Lemma 3.1 of Powell, Stock and Stoker (1989), if

$$h_n = o(1), \qquad h_n^{-(q+1)} = o(n), \tag{3.30}$$

20

then

$$\mathbf{S}_0 - E[\mathbf{S}_0] = o_p(1). \tag{3.31}$$

Furthermore, with continuity of the underlying conditional expectation and density functions, it can be shown that, as $n \to \infty$ (and $h_n \to 0$),

$$E[\mathbf{S}_0] \to \mathbf{\Sigma}_w \equiv E[2f(g_i, \mathbf{v}_i) \cdot (\tau_i \boldsymbol{\nu}_i - \boldsymbol{\mu}_i' \boldsymbol{\mu}_i) \mid g_i = g_j, \mathbf{v}_i = \mathbf{v}_j], \tag{3.32}$$

where $f(g, \mathbf{v})$ is the joint density function of $g_i \equiv g(\mathbf{w}_i)$ and $\mathbf{v}_i$,, and where

$$\begin{aligned}
\tau_i &\equiv E[t_i | g_i, \mathbf{v}_i], \\
\boldsymbol{\mu}_i &\equiv E[t_i \mathbf{x}_i | g_i, \mathbf{v}_i], \text{ and} \\
\boldsymbol{\nu}_i &\equiv E[t_i \mathbf{x}_i' \mathbf{x}_i | g_i, \mathbf{v}_i]. \tag{3.33}
\end{aligned}$$

If the preliminary estimators $\hat{g}_i$ and $\hat{\pi}_i$ converge at a sufficiently fast rate, and if the levels and derivatives of the kernel functions $K_g(\cdot)$ and $K_v(\cdot)$ are uniformly bounded, the term $\mathbf{S}_1$ will be asymptotically negligible. More precisely, assuming

$$\max_i t_i [|\hat{g}_i - g_i| + ||\hat{\boldsymbol{\pi}}_i - \pi_i||] = o_p(h_n^{q+2}) \tag{3.34}$$

and the levels and derivatives of the kernels are bounded,

$$||K_g^{(l)}|| + ||K_v^{(l)}|| < C, \qquad j = 0, 1, \text{ some } C, \tag{3.35}$$

then it is simple to show that

$$\mathbf{S}_1 = o_p(1). \tag{3.36}$$

Primitive conditions for the uniform convergence rates given in(3.34) of the preliminary nonparametric estimators can be found in, e.g., Ahn (1993); these conditions involve high-order differentiability of the functions being estimated, 'higher-order

21

bias-reducing' kernel functions, the dimensionality of the vectors $\mathbf{x}_i$, $\mathbf{v}_i$, and $\mathbf{z}_i$, and particular convergence rates for the bandwidths of the first-stage nonparametric estimators.

Thus, if these conditions hold, it follows that

$$\hat{\mathbf{S}}_w = \mathbf{\Sigma}_w + o_p(1). \tag{3.37}$$

Moreover, by the identity (2.15), it follows that

$$
\begin{aligned}
(\tau_i \boldsymbol{\nu}_i - \boldsymbol{\mu}_i' \boldsymbol{\mu}_i)\beta_0 &= \tau_i E[t_i \mathbf{x}_i' \mathbf{x}_i \beta_0 \mid g_i, \mathbf{v}_i] - \boldsymbol{\mu}_i' E[t_i \mathbf{x}_i \beta_0 \mid g_i, \mathbf{v}_i] \\
&= \tau_i E[t_i \mathbf{x}_i' \mid g_i, \mathbf{v}_i]\psi(g_i, \mathbf{v}_i) - \boldsymbol{\mu}_i' E[t_i \mid g_i, \mathbf{v}_i]\psi(g_i, \mathbf{v}_i) \\
&= 0, \tag{3.38}
\end{aligned}
$$

so that the limit matrix $\mathbf{\Sigma}_w$ is indeed singular with

$$\mathbf{\Sigma}_w \boldsymbol{\beta}_0 = 0, \tag{3.39}$$

as required. The consistency argument is completed with an assumption that $\boldsymbol{\beta}_0$ is the unique nontrivial solution to the system of homogeneous linear equations (3.39), i.e.,

$$\mathbf{\Sigma}_w \boldsymbol{\lambda} = 0, \ \boldsymbol{\lambda} \neq \boldsymbol{\beta}_0 \implies \boldsymbol{\lambda} = \mathbf{0}. \tag{3.40}$$

This in turn implies a 'full rank' condition for the conditional expectations $\pi_i = E[\mathbf{y}_{2i}|\mathbf{z}_i]$, since (3.40)

$$
\begin{aligned}
\Pr\{(\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\lambda} \ &\neq \ 0|\mathbf{v}_i = \mathbf{v}_j, (\mathbf{x}_i - \mathbf{x}_j)\boldsymbol{\beta}_0 = 0, \boldsymbol{\lambda} \neq \boldsymbol{\beta}_0, \boldsymbol{\lambda} \neq \mathbf{0}\} \\
&= \ \Pr\{(\boldsymbol{\pi}_i - \boldsymbol{\pi}_j)\boldsymbol{\lambda} \neq 0|\mathbf{v}_i = \mathbf{v}_j, (\boldsymbol{\pi}_i - \boldsymbol{\pi}_j)\boldsymbol{\beta}_0 = 0, \boldsymbol{\lambda} \neq \boldsymbol{\beta}_0, \boldsymbol{\lambda} \neq \mathbf{0}\} \\
&> \ 0, \tag{3.41}
\end{aligned}
$$

22

where the equality follows from the identity $\mathbf{x}_i \equiv \boldsymbol{\pi}_i + \mathbf{v}_i$.

Since eigenvalues and (normalized) eigenvectors are continuous functions of their matrix arguments, the assumptions and calculations above yield weak consistency of the estimator $\hat{\boldsymbol{\beta}}$, i.e.

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + o_p(1). \tag{3.42}$$

Establishing the $\sqrt{n}$-consistency and asymptotic normality of $\hat{\boldsymbol{\beta}}$ would require a more refined asymptotic argument, based upon a second-order Taylor's series expansion of the matrix $\hat{\mathbf{S}}_w$,

$$\hat{\mathbf{S}}_w = \mathbf{S}_0 + \mathbf{S}_1 + \mathbf{S}_2, \tag{3.43}$$

where, as before,

$$\mathbf{S}_l \equiv \binom{n}{2}^{-1} \sum_{i<j} \omega_{ij}^l (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j), \qquad l = 0, 1, 2, \tag{3.44}$$

with

$$\boldsymbol{\omega}_{ij}^0 \equiv h_n^{-(q+1)} K_g \left(\frac{g_i - g_j}{h_n} \right) K_v \left(\frac{\mathbf{v}_i - \mathbf{v}_j}{h_n} \right) t_i t_j, \tag{3.45}$$

as before, but now

$$\begin{aligned}
\omega_{ij}^1 \equiv\ & h_n^{-(q+2)} K_g^{(1)} \left(\frac{g_i - g_j}{h_n} \right) K_v \left(\frac{\mathbf{v}_i - \mathbf{v}_j}{h_n} \right) (\hat{g}_i - \dot{g}_i - \hat{g}_j + g_j) t_i t_j \\
& - h_n^{-(q+2)} K_g \left(\frac{g_i - g_j}{h_n} \right) K_v^{(1)} \left(\frac{\mathbf{v}_i - \mathbf{v}_j}{h_n} \right) (\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i - \hat{\boldsymbol{\pi}}_j + \boldsymbol{\pi}_j)' t_i t_j \tag{3.46}
\end{aligned}$$

and

$$
\begin{aligned}
\boldsymbol{\omega}_{ij}^2 \;\equiv\; & \frac{h_n^{-(q+3)}}{2} K_g^{(2)}\left(\frac{g_{ij}^*}{h_n}\right) K_v\left(\frac{\mathbf{v}_{ij}^*}{h_n}\right)(\hat{g}_i - \dot{g}_i - \hat{g}_j + g_j)^2 t_i t_j \\
& - h_n^{-(q+3)}(\hat{g}_i - g_i - \hat{g}_j + g_j) K_g^{(1)}\left(\frac{g_{ij}^*}{h_n}\right) K_v^{(1)}\left(\frac{\mathbf{v}_{ij}^*}{h_n}\right) \\
& \qquad \cdot (\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i - \hat{\boldsymbol{\pi}}_j + \boldsymbol{\pi}_j)' t_i t_j \qquad\qquad\qquad\qquad\qquad (3.47) \\
& + \frac{h_n^{-(q+3)}}{2}(\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i - \hat{\boldsymbol{\pi}}_j + \boldsymbol{\pi}_j) K_g\left(\frac{g_{ij}^*}{h_n}\right) K_v^{(2)}\left(\frac{\mathbf{v}_{ij}^*}{h_n}\right) \\
& \qquad \cdot (\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i - \hat{\boldsymbol{\pi}}_j + \boldsymbol{\pi}_j)' t_i t_j. \qquad\qquad\qquad\qquad (3.48)
\end{aligned}
$$

Since demonstration of $\sqrt{n}$-consistency of $\hat{\boldsymbol{\beta}}$ involves normalization of each of these terms by the factor $\sqrt{n}$, the regularity conditions would have to be strengthened substantially, with considerably more smoothness of the unknown density and expectation functions, a higher rate of convergence of the preliminary nonparametric estimators, bias-reducing kernels of yet higher order, and a more restricted rate of convergence of the second-step bandwidth to zero. Still, with appropriate modification of the conditions and calculations in Ahn, Ichimura, and Powell (1996), the terms $\mathbf{S}_0$ and $\mathbf{S}_2$, when postmultiplied by the true parameter vector β_0, and appropriately normalized, should be asymptotically negligible,

$$
\sqrt{n}\mathbf{S}_0\boldsymbol{\beta}_0 = o_p(1) = \sqrt{n}\mathbf{S}_2\boldsymbol{\beta}_0, \qquad\qquad (3.49)
$$

while the first-order component matrix $\mathbf{S}_1$ should satisfy an asymptotic linearity relation of the form

$$
\begin{aligned}
\sqrt{n}\mathbf{S}_1\boldsymbol{\beta}_0 \;=\; & \sqrt{n}\hat{\mathbf{S}}_w\boldsymbol{\beta}_0 + o_p(1) \qquad\qquad\qquad\qquad (3.50) \\
\;=\; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{e}_{i1} + \mathbf{e}_{i2}) + o_p(1), \qquad\qquad (3.51)
\end{aligned}
$$

24

with

$$\mathbf{e}_{1i} \equiv 2\tau_i f(g_i, \mathbf{v}_i) \cdot (\tau_i \mathbf{x}_i - \boldsymbol{\mu}_i)' \cdot \frac{\partial \psi(g(\mathbf{w}_i), \mathbf{v}_i)}{\partial g} \cdot (y_{1i} - g(\mathbf{w}_i)) \tag{3.52}$$

being the component of the influence function for $\hat{\mathbf{S}}_w \boldsymbol{\beta}_0$ due to the nonparametric estimation of $g_i \equiv E[y_{1i} | \mathbf{x}_i, \mathbf{v}_i]$ and

$$\mathbf{e}_{2i} \equiv -E\left[2\tau_i f(g_i, \mathbf{v}_i) \cdot (\tau_i \mathbf{x}_i - \boldsymbol{\mu}_i)' \cdot \frac{\partial \psi(g(\mathbf{w}_i), \mathbf{v}_i)}{\partial \mathbf{v}'} | \mathbf{z}_i \right] \cdot \mathbf{v}_i \tag{3.53}$$

being the influence function component accounting for the first-stage nonparametric estimation of $\mathbf{v}_i \equiv \mathbf{y}_{2i} - E[y_{2i} | \mathbf{z}_i]$.

Given the validity of (3.51), the same argument as given for (3.38) would yield

$$\sqrt{n} \boldsymbol{\beta}_0' \hat{\mathbf{S}}_w \boldsymbol{\beta}_0 = o_p(1), \tag{3.54}$$

from which the $\sqrt{n}$-consistency and asymptotic normal distribution of the lower $(p+q-1)$-dimensional subvector $\widehat{\boldsymbol{\theta}}$ of $\widehat{\boldsymbol{\beta}}$,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_{22}^{-1} \mathbf{V}_{22} \boldsymbol{\Sigma}_{22}^{-1}), \tag{3.55}$$

will follow from the same argument as in Ahn, Ichimura, and Powell (1996), with

$$\mathbf{V} \equiv E[(\mathbf{e}_{1i} + \mathbf{e}_{i2})(\mathbf{e}_{1i} + \mathbf{e}_{2i})'] \tag{3.56}$$

and with $\boldsymbol{\Sigma}_{22}$ and $\mathbf{V}_{22}$ the lower $(p+q-1) \times (p+q-1)$ diagonal submatrices of $\boldsymbol{\Sigma}_w$ and $\mathbf{V}$, as in (3.11).

Demonstration of consistency and asymptotic normality of the estimator $\widehat{G}(\mathbf{x}\widehat{\boldsymbol{\beta}})$ of the average structural function $G(\mathbf{x}\boldsymbol{\beta})$ would require a substantial extension of the large-sample theory for "partial means" of kernel regression estimators given

by Newey (1994). First, while Newey's results assumed the trimming term τ_i in (3.14) to be independent of the sample size (so that a positive fraction of observations is trimmed in the limit, to help bound the denominator of the kernel regression away from zero), it would be important to extend the results to permit $\tau_i = \tau_i(\widehat{\mathbf{v}}_i, n)$ to tend to unity as $n \to \infty$ for all i. This would ensure that the probability limit of $\widehat{G}(\mathbf{x}\widehat{\boldsymbol{\beta}})$ is an expectation, and not a truncated expectation, of $F(\mathbf{x}\boldsymbol{\beta}, \mathbf{v}_i)$ over the marginal distribution of $\mathbf{v}_i$.[5]

Another important extension of Newey's partial-mean results would account for the use of estimated regressors in the nonparametric estimation of the nonlinear regression function $F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i)$. While these results would directly apply to kernel regressions of y_{1i} on the true index $\mathbf{x}_i\boldsymbol{\beta}_0$ and control variable $\mathbf{v}_i$, they do not account for the added imprecision due to the semiparametric estimator $\widehat{\boldsymbol{\beta}}$ of the index coefficients, nor of the nonparametric estimator $\widehat{\mathbf{v}}_i$ of the first-stage residuals. While the "delta method" approach taken by Newey could, in principle, be used to derive variance estimators in this case, it is difficult to obtain analytic formulae using this approach.

Though a rigorous derivation of this extension is beyond the scope of the present paper, a preliminary analysis suggests that the qualitative rate-of-convergence results for partial means in Newey (1994) will carry over to the case where the regressors are estimated: that is, the convergence rate of $\widehat{G}(\mathbf{x}\widehat{\boldsymbol{\beta}})$ will be the same as that for a one-dimensional kernel regression estimator under appropriate conditions on the bandwidth and trimming terms, and "full means" – weighted integrals of $\widehat{G}(\mathbf{x}\widehat{\boldsymbol{\beta}})$ over all components of $\mathbf{x}$, as in (2.23) – will be $\sqrt{n}$-consistent (and as-

---

[5]In our empirical application, we follow the common, if theoretically questionable, practice of setting the trimming term τ_i equal to one for all observations.

ymptotically normal). Moreover, based upon formal calculations analogous to those in Ahn and Powell (1993) and Ahn, Ichimura, and Powell (1996), we conjecture that the asymptotic distributions for the partial means and full means using the nonparametric estimator $\widehat{F}(\mathbf{x}\widehat{\boldsymbol{\beta}},\widehat{\mathbf{v}}_i)$ based upon a kernel regression of y_{1i} on $\mathbf{x}_i\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{v}}_i$ will have the same asymptotic distribution as the analogous partial and full means for the nonparametric estimator $\widetilde{F}(\mathbf{x}\boldsymbol{\beta}_0, \mathbf{v}_i)$ based on a regression of y_i^+ on $\mathbf{x}_i\boldsymbol{\beta}_0$ and $\mathbf{v}_i$, for

$$y_i^+ = y_{1i} - 2\frac{\partial F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i)}{\partial(\mathbf{x}\boldsymbol{\beta})}\mathbf{x}_i(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+2\frac{\partial F(\mathbf{x}_i\boldsymbol{\beta}_0, \mathbf{v}_i)}{\partial\mathbf{v}}\mathbf{v}_i, \qquad (3.57)$$

where the second and third terms account for the preliminary estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{v}}_i$, respectively. If this conjecture could be verified, then the asymptotic distributions of the average structural function estimator $\widehat{G}$ and weighted averages of it could be obtained directly from Theorems 4.1 and 4.2 of Newey (1994), assuming these can be shown to hold when the trimming term $\tau_i \to 1$ and inserting the asymptotic linearity approximation for $\widehat{\boldsymbol{\beta}}$ implicit in (3.55). Consistent estimates of the corresponding asymptotic covariance matrices would still need to be developed, though, to make such results useful in practice.

An alternative approach to inference, used in the empirical application below, can be based upon bootstrap estimates of the sampling distribution of $\widehat{\boldsymbol{\beta}}$ and $\widehat{G}$. Like much of the asymptotic theory outlined above, the theoretical validity of the bootstrap remains to be established in the present context, but there is no prior reason to suspect it would yield misleading inferences when applied to the present problem.

4. A Empirical Investigation: The Income Effect on Labour Market Participation

4.1. The Data

In this empirical application we consider the work participation decision by men without college education in a sample of British families with children. Employment in this group in Britain is surprisingly low. More than 12% do not work and this approaches 20% for those with lower levels of education. Consequently this group is subject to much policy discussion. We model the participation decision (y_1), see (2.4), in terms of a binary response framework that controls for market wage opportunities and the level of other income sources in the family. Educational level (z_1) is used as a proxy for market opportunities and is treated as exogenous for participation. But other income (y_2), which includes the earned income of the spouse, is allowed to be endogenous for the participation decision.

As an instrument (z_{21}) for other family income we use a welfare benefit entitlement variable. This instrument measures the transfer income the family would receive if neither spouse was working and is computed using a benefit simulation routine designed for the evaluation of welfare benefits for households in the British data used here. The value of this variable just depends on the local welfare benefit rules, the demographic structure of the family, the geographic location and housing costs.[6] As there are no earnings-related benefits in operation in Britain over this period under study, we may be willing to assume it is exogenous for the participation decision. Moreover, although this benefit entitlement variable will be a determinant of the reduced form for participation and other income, for the

---

[6] See Blundell, Reed and Stoker (2003) for more details on this instrument.

structural model below, it should not enter the participation decision conditional on the inclusion of other income variable. Another instrumental variable will be the education level of the spouse (z_{22}).

The sample of married couples is drawn from the British Family Expenditure Survey (FES). The FES is a repeated continuous cross-sectional survey of households which provides consistently defined micro data on family incomes, employment status and education, consumption and demographic structure. We consider the period 1985-1990. The sample is further selected according to the gender, educational attainment and date of birth cohort of the head of household. We choose male head of households, born between 1945 and 1954 and who did not receive college education. We also choose a sample from the North West region of Britain. These selections are primarily to focus on the income and education variables.

For the purposes of modeling, the participating group consists of employees; the non-participating group includes individuals categorized as searching for work as well as the unoccupied. The measure of education used in our study is the age at which the individual left full-time education. Individuals in our sample are classified in two groups; those who left full-time education at age 16 or lower (the lower education base group) and those who left aged 17 or 18. Those who left aged 19 or over are excluded from this sample.

Our measure of exogenous benefit income is constructed for each family as follows: a tax and benefit simulation model[7] is used to construct a simulated budget constraint for each individual family given information about age, location,

---

[7]The IFS tax and benefit simulation model TAXBEN, designed for the British Family Expenditure Survey data used in this paper, www.ifs.org.uk.

benefit eligibility etc. The measure of out-of-work income is largely comprised of income from state benefits; only small amounts of investment income are recorded. State benefits include eligible unemployment benefits[8], housing benefits, child benefits and certain other allowances. Since our measure of out-of-work income will serve to identify the structural participation equation, it is important that variation in the components of out-of-work income over the sample are exogenous for the decision to work. In the UK, the level of benefits which individuals receive out of work varies with age, time, household size and (in the case of Housing Benefit) by region. Housing benefit varies systematically with time, location and cohort.

Table 4.1: Descriptive Statistics

Variable		Mean	Std Dev.
Work	(y_1)	.871	.387
education > 16	(z_1)	.196	.396
ln(other income)	(y_2)	5.016	.434
ln(benefit income)	(z_{21})	3.314	.289
education (spouse)	(z_{22})	.204	.403
age		39.191	10.256

Notes to Table 4.1: number of observations, 1606. The income and benefit income variables are measured in log £'s per week. The education dummy (education>16) is a binary indicator that takes the value unity if the individual stayed on a school after the minimum school leaving age of 16.

---

[8]Unemployment benefit included an earnings-related supplement up to 1979, but this was abolished in 1980 and does not therefore impact on our benefit entitlement measure.
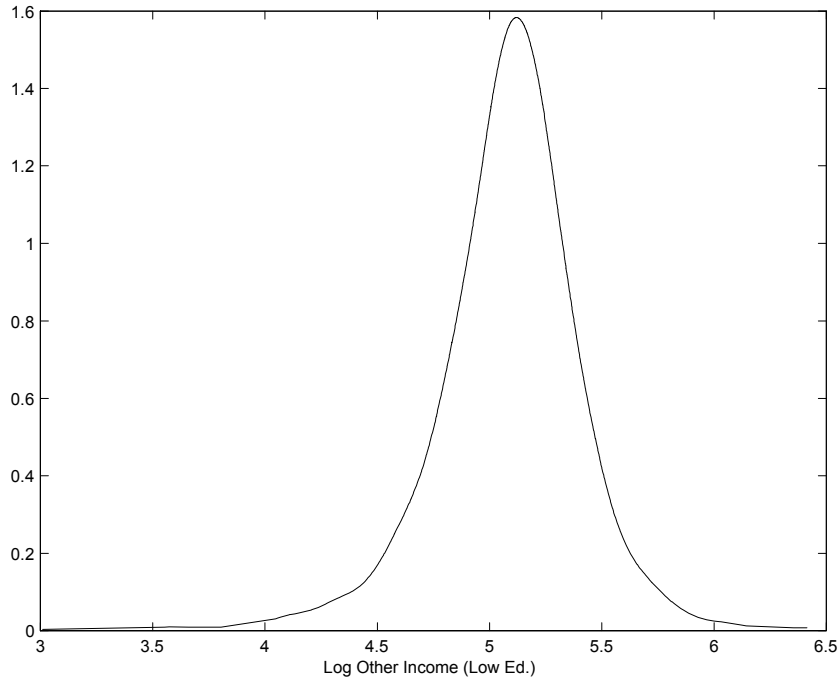
Figure 1: Density of Log Other Income: Low Income Subsample

After making the sample selections described above, our sample contains 1606 observations. A brief summary of the data is provided in Table 4.1. The 87.1% employment figure for men in this sample is reduced to less than 82% for the lower education group that makes up more than 75% of our sample. As mentioned above, this lower education group refers to those who left formal schooling at 16 years of age or before and will be the group on which we focus in much of this empirical application. The kernel density estimate of log other income for the low education sub-sample is given in Figure 1.

4.2. A Model of Participation in Work

To motivate the model specification, suppose that observed participation is described by a simple threshold model of labor supply. In this model the desired supply of hours of work for individual i can be written

$$h_i^* = \delta_0 + \mathbf{z}_{1i}\boldsymbol{\delta}_1 + \ln w_i \delta_2 + \ln \mu_i \delta_3 + \zeta_i, \tag{4.1}$$

where $\mathbf{z}_{1i}$ includes various observable social demographic variables, $\ln w_i$ is the log hourly wage, $\ln \mu_i$ is the log of 'virtual' other income, and ζ_i is some unobservable heterogeneity. As $\ln w_i$ is unobserved for nonparticipants we replace it in (4.1) by the wage equation

$$\ln w_i = \theta_0 + \mathbf{z}_{1i}\boldsymbol{\theta}_1 + \omega_i \tag{4.2}$$

where $\mathbf{z}_{1i}$ is now defined to include the education level for individual i as well as other determinants of the market wage. Labor supply (4.1) becomes

$$h_i^* = \phi_0 + \mathbf{z}_{1i}\boldsymbol{\phi}_1 + \ln \mu_i \phi_2 + \nu_i. \tag{4.3}$$

Participation in work occurs according to the binary indicator

$$y_{1i} = 1\{h_i^* > h_i^0\} \tag{4.4}$$

where

$$h_i^0 = \gamma_0 + \mathbf{z}_{1i}\boldsymbol{\gamma}_1 + \xi_i \tag{4.5}$$

is some measure of reservation or threshold hours of work.

Combining these equations, the binary response model for participation is now described by

$$
\begin{aligned}
y_{1i} &= 1\{\phi_0 + \mathbf{z}_{1i}\boldsymbol{\phi}_1 + \ln \mu_i \phi_2 + \nu_i > \gamma_0 + \mathbf{z}_{1i}\boldsymbol{\gamma}_1 + \xi_i\} \tag{4.6} \\
&= 1\{\beta_0 + \mathbf{z}_{1i}\boldsymbol{\beta}_1 + y_{2i}\boldsymbol{\beta}_2 + u_i > 0\} \tag{4.7}
\end{aligned}
$$

32

where y_{2i} is the log other income variable ($\ln \mu_i$). This other income variable is assumed to be determined by the reduced form

$$
\begin{aligned}
y_{2i} &= E[y_{2i}|\mathbf{z}_i] + v_i \\
&= \mathbf{\Pi}(\mathbf{z}_i) + v_i
\end{aligned}
\tag{4.8}
$$

and $\mathbf{z}_i = [\mathbf{z}_{1i}, \mathbf{z}_{2i}]$.

In the empirical application we have already selected households by cohort, region and demographic structure. Consequently we are able to work with a fairly parsimonious specification in which $\mathbf{z}_{1i}$ simply contains the education level indicator. The excluded variables $\mathbf{z}_{2i}$ contain the log benefit income variable (denoted z_{21i}) described above and the education level of the spouse (z_{22i}).

4.3. Empirical Results

In Table 4.2 we present the empirical results for the joint normal (parametric) simultaneous probit model using the conditional likelihood approach, see (2.8). This consists of a linear reduced form for the log other income variable and a conditional Probit specification for the participation decision. The first column of Table 4.2 presents the parametric reduced form estimates. Given the selection by region, cohort, demographic structure and time period, the reduced form simply contains the education variables and the log exogenous benefit income variable. The reduced form results show a strong role for the benefit income variable in the determination of other income.

The first of the Probit results in Table 4.2 refer to the model without adjustment for the endogeneity of other income. These results show a positive and significant coefficient estimate for the education dummy variable and a small but

significantly negative estimated coefficient on other income. The other income coefficient in Table 4.2 is the coefficient normalized by the education coefficient for comparability with the results from the semiparametric specification to be presented below.

As is evident from the results in the last columns of Table 4.2, the impact of adjusting for endogeneity is quite dramatic. The income coefficient is now considerably larger in magnitude and quite significant. The estimated education coefficient remains positive and significant. The asymptotic t-test for the null of exogeneity (see Blundell and Smith (1986)), strongly rejects the exogeneity of the log other income variable in this parametric binary response formulation of the labour market participation model.

Table 4.2: Results for the Parametric Specification.

Variable	Reduced Form		Probit		Probit	
	y_2 Coeff.	Std. Err.	Pr[Work] Coeff.	Std. Err.	Pr[Work$|v$] Coeff.	Std. Err.
education: z_1	0.0603	0.0224	1.007	.1474	1.4166	.1677
ln(other inc): y_2	-	-	-.3364	.1293	-2.8376	.5124
ln(benefit inc): z_{21}	0.0867	0.0093	-	-	-	-
education(sp): z_{22}	0.0799	0.0219	-	-	-	-
R^2	.0708		.0550		.0885	
F	30.69(3)		67.84 ($\chi^2_{(2)}$)		109.29 ($\chi^2_{(3)}$)	
exog test					5.896 (t)	

We now turn to the estimation results for the semiparametric estimator. Table

4.3 presents the estimation results for the $\boldsymbol{\beta}_0$ coefficients. The bootstrap distribution relate to 500 bootstrap samples of size n ($=1606$); the standard errors for the semiparametric methods are computed from a standardized interquartile range for the bootstrap distribution, and are calculated using the usual asymptotic formulae for the Probit and linear probability estimators.

The education coefficient in the binary response specification is normalized to unity and so the $-\theta_2$ estimates in Table 4.3 correspond to the ratio of estimates of the other income coefficient to the education coefficient. In this application, bandwidths were chosen according to the $1.06\sigma_z n^{-\frac{1}{5}}$ rule (see Silverman(1986)). These may well be too smooth for the estimation of $\boldsymbol{\beta}_0$ and in the the third row we present results which use half this bandwidth $(\frac{1}{2}h)$. This suggests the estimates are relatively robust for this sample over this range of this bandwidth.

For comparison, Table 4.3 presents results for the ratio of coefficients estimated assuming the errors are normally distributed (i.e., Probit, as in Table 4.2), as well as corresponding results from classical least squares and two-stage least squares (i.e., Linear Probability) estimators. The differing estimation methods yield qualitatively-similar conclusions concerning the endogeneity correction.

Figure 2 graphs the estimate of the average structural function ASF, $G(\mathbf{x}\boldsymbol{\beta}_0)$, derived from semiparametric estimation including $\widehat{v}$ as described in (2.22). This controls for the endogeneity of log other income. The ASF is plotted over the 5%-95% range of the log other income distribution for the lower education group. Bootstrap 95% confidence bands are presented at the 10, 25, 50, 75 and 90 percentile points of the log income density for the lower education subsample. The

35

Table 4.3: Semiparametric Results and Bootstrap Distribution.

Specification	$-\widehat{\theta}_2$	$\sigma_{\widehat{\theta}_2}$	10%	25%	50%	75%	90%
Semi-P (with $\widehat{v}$)	-2.2590	0.5621	-4.3299	-3.6879	-2.3275	-1.4643	-1.0101
Semi-P (without $\widehat{v}$)	-0.1871	0.0812	-0.2768	-0.2291	-0.1728	-0.1027	-0.0675
Semi-P (with $\widehat{v}$):$\frac{1}{2}h$	-2.1324	0.6751	-4.8976	-3.8087	-2.2932	-1.1193	-0.06753
Probit (with $\widehat{v}$)	-2.8376	.5124	−3.8124	−3.3304	−2.9167	−2.4451	-1.8487
Probit (without $\widehat{v}$)	-0.3364	.1293	−0.4989	−0.4045	−0.3354	−0.2672	−0.1991
Lin.Prob.(with $\widehat{v}$)	-3.1241	.4679	−3.8451	-3.3811	-3.1422	-2.8998	-2.5425
Lin.Prob.(without $\widehat{v}$)	-0.4199	.1486	-0.6898	-0.5643	-0.4012	-0.3132	-0.2412

regression line shows a strong monotonic decline with log other income. This contrasts with the much shallower slope for the estimated ASF when the control for endogeneity is excluded. The latter is the semiparametric binary response model assuming other income to be exogenous for the work decision. In this case, although the ASF remains monotonic and negative, the marginal impact of an exogenous change in other income is much smaller for almost all values of other income. Note that the degree of bias from ignoring the endogeneity of other income is such that the curves cross, this could never happen if the $\mathbf{x}$ and v were distributed independently.
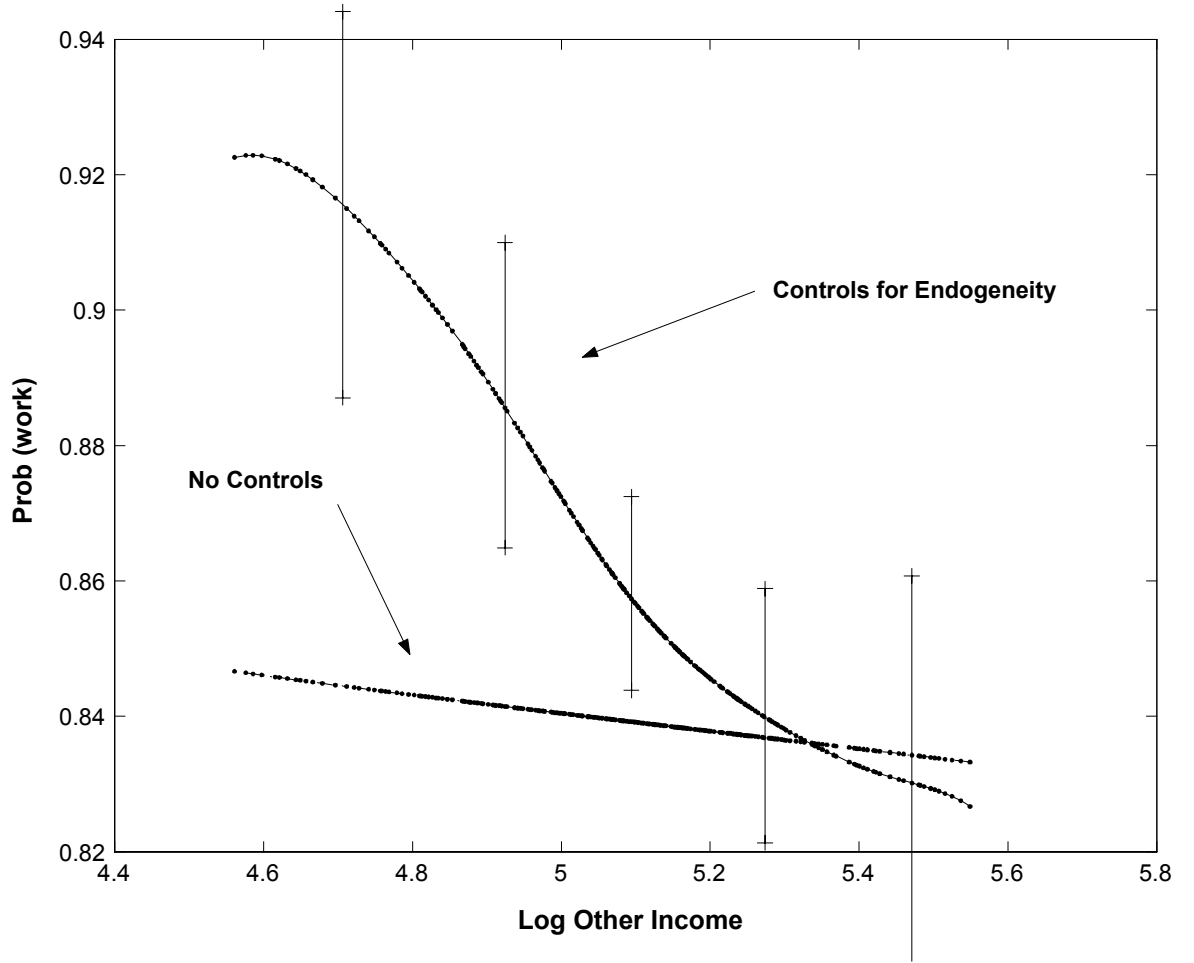
Figure 2: Semiparametric regression with and without controls for endogeneity

In Figure 3 we present corresponding results for the estimates of $G(\mathbf{x}\boldsymbol{\beta}_0)$ using the Probit and Linear Probability models. Bootstrap 95% confidence bands are again presented at the 10, 25, 50, 75 and 90 percentile points of the log income density for the lower education subsample. This application is likely to be a particularly good source on which to carry out this comparison. First, we know from Table 4.3 that the correction for endogeneity induces a large change in the

estimated $\boldsymbol{\beta}_0$ coefficients. Second, the proportion participating in the sample is around 85% which suggests that the choice of probability model should matter as the tail probabilities in the Probit and linear probability models will behave quite differently. The plots show considerable sensitivity of the estimated $G(\mathbf{x}\boldsymbol{\beta}_0)$, after allowing for endogeneity, across these alternative parametric models. Both the Linear Probability and Probit model estimates result in estimated probability curves that are very much steeper than those implied by the semiparametric approach. For example, the Linear Probability model estimates a probability that is more than ten percentage points higher at the 20 percentile point of the log other income distribution.
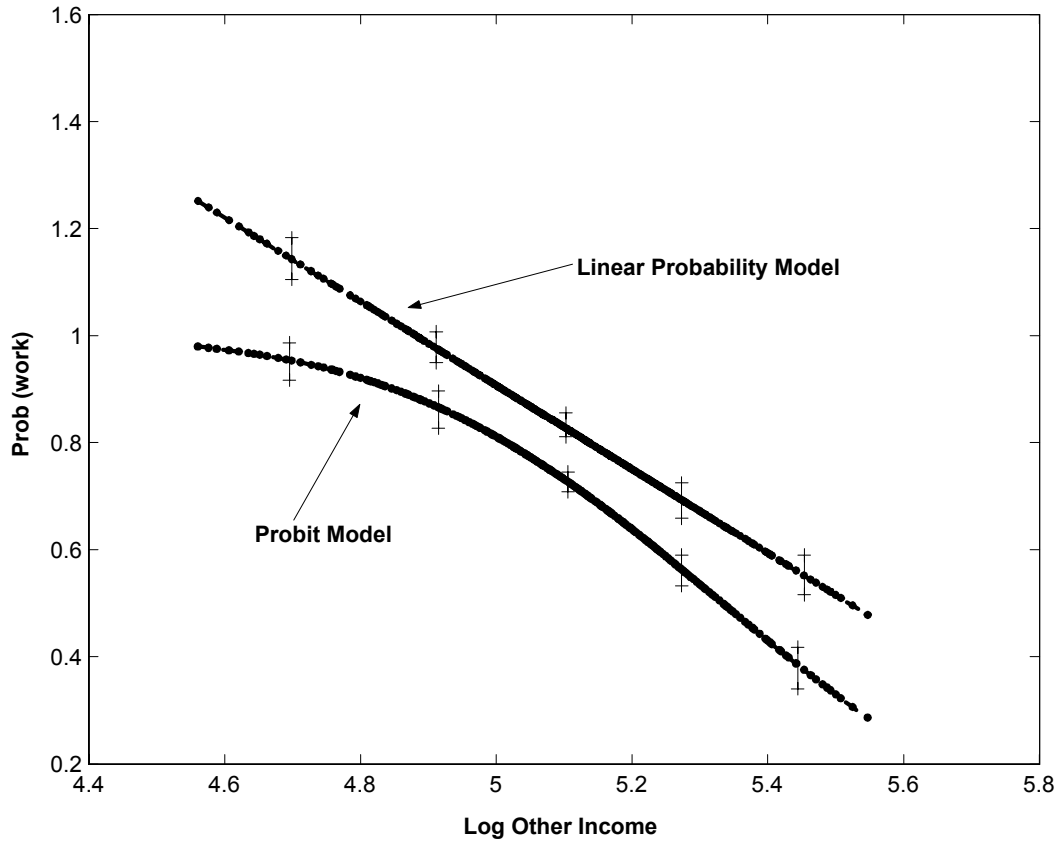
Figure 3: Linear Probability and Probit Results with Endogeneity Controls

Finally, in Figure 4 we present the analogous analysis using the low education sub-sample only, with bootstrap 95% confidence bands at the 10, 25, 50, 75 and 90 percentile points of the log other income density for this lower education sub-sample. For this sample the education dummy is equal to zero for all observations and is therefore excluded. Since x is now simply the log other income variable this analysis is purely nonparametric. As can be seen by comparison to Figure 3 this shows a slightly shallower slope. Similar results to Figure 3 can be found for the linear probability and probit models for this case and are available from the
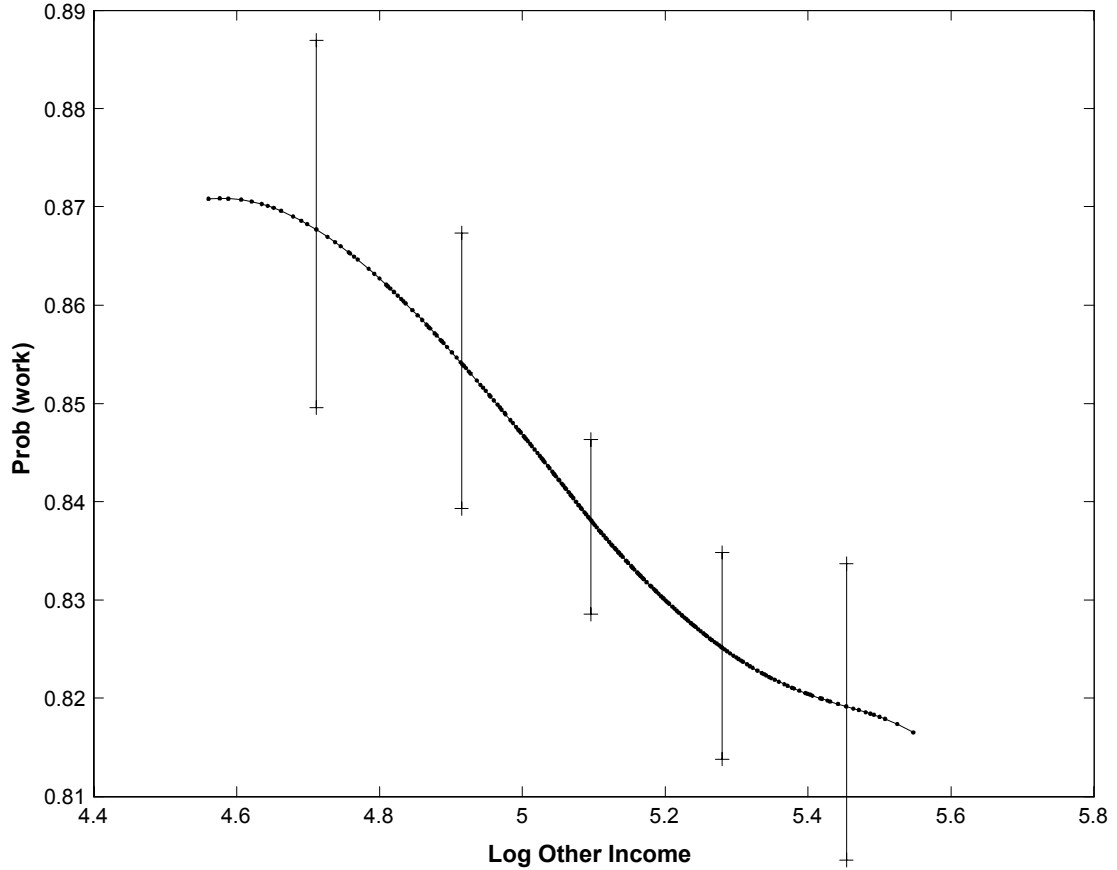
authors on request.

Figure 4: Nonparametric regression with controls for endogeneity

These results point to the attractiveness of the approach developed in this paper. For this data set we have found relatively small reductions in precision from adopting the semiparametric control function approach while finding quite different estimated responses from those estimated using the parametric Probit or Linear Probability models. In the next section we consider the implementation

40

of the proposed estimator to an alternative representation of the simultaneous binary choice framework.

5. An Alternative Specification: Fixed Costs of Work and The Coherency Model

One interpretation of the endogenous linear index binary response model described above is as the 'triangular form' of some underlying joint decision problem in terms of latent endogenous variables. Partitioning $\mathbf{z}$ as before

$$\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2), \tag{5.1}$$

we can express the model as

$$y_{1i} = 1\{y_{1i}^* > 0\}, \tag{5.2}$$

$$y_{1i}^* = \mathbf{z}_{1i}\boldsymbol{\beta}_1 + \mathbf{y}_{2i}\boldsymbol{\beta}_2 + u_i \tag{5.3}$$

and

$$\mathbf{y}_{2i} = \mathbf{z}_{2i}\Xi_1 + y_{1i}^*\boldsymbol{\gamma}_2 + \boldsymbol{\varepsilon}_i. \tag{5.4}$$

Substitution of (5.3) in (5.4) delivers the first-stage regression model

$$\mathbf{y}_{2i} = \mathbf{z}_i\Pi + \mathbf{v}_i \tag{5.5}$$

for some coefficient matrix Π. This 'triangular' structure has $\mathbf{y}_2$ first being determined by $\mathbf{z}$ and the error terms $\mathbf{v}$, while y_1 is then determined by y_2, $\mathbf{z}$, and the structural error u.

In some economic applications, however, joint decision making may be in terms of the *observed* outcomes rather than latent outcomes implicit in (5.3) and (5.4).

41

In this alternative specification (5.4) is replaced with a model incorporating feedback between the *observed* dependent variable y_1 and $\mathbf{y}_2$

$$\mathbf{y}_{2i} = \mathbf{z}_{2i}\Xi_1 + y_{1i}\boldsymbol{\alpha}_2 + \boldsymbol{\varepsilon}_i \tag{5.6}$$

that is, the realization $y_1 = 1$ results in a discrete shift $y_{1i}\boldsymbol{\alpha}_2$. Due to the non-linearity in the binary response rule (5.2), there is no explicit reduced form for this system. Indeed, Heckman (1978), in his analysis of simultaneous models with dummy endogenous variables, shows that (5.2), (5.3) and (5.6) is only a statistically 'coherent' system, i.e. one that processes a unique (if not explicit) reduced form, when $\boldsymbol{\alpha}_2 = \mathbf{0}$, removing the direct feedback.

To provide a fully simultaneous system in terms of observed outcomes, and one that is also statistically coherent, Heckman (1978) suggests incorporating a structural jump in the equation for y_{1i}^*,

$$y_{1i}^* = y_{1i}\alpha_1 + \mathbf{z}_{1i}\boldsymbol{\beta}_1 + \mathbf{y}_{2i}\boldsymbol{\beta}_2 + u_i, \tag{5.7}$$

with the added restriction

$$\alpha_1 + \boldsymbol{\alpha}_2'\boldsymbol{\beta}_2 = 0 \tag{5.8}$$

This Heckman (1978) labels the Principal Assumption.[9] Thus for a consistent probability model with general distributions for the unobservables and exogenous covariates we require the coherency condition (5.8). Below we show that the semi-parametric control function approach developed in this paper extends naturally

---

[9] To derive the condition (5.8) notice that from (5.2), (5.6) and (5.7) we can write

$$y_{1i}^* = 1\{y_{1i}^* > 0\}(\alpha_1 + \alpha_2\beta_2) + \mathbf{z}_{1i}\boldsymbol{\beta}_1 + \mathbf{z}_{2i}\gamma_1\beta_2 + u_i + \boldsymbol{\varepsilon}_i\beta_2, \tag{5.9}$$

or

$$y_{1i}^* \lessgtr 0 \Leftrightarrow 1\{y_{1i}^* > 0\}(\alpha_1 + \alpha_2\beta_2) + \mathbf{z}_{1i}\boldsymbol{\beta}_1 + \mathbf{z}_{2i}\gamma_1\beta_2 + u_i + \boldsymbol{\varepsilon}_i\beta_2 \lessgtr 0. \tag{5.10}$$

to this framework. First we relate this coherency specification to a fixed costs model of participation.

Suppose that the fixed cost of work is given by α_2. In determining participation, the fixed cost α_2 will have to be subtracted from other income (or consumption) for those who choose to work. In this case the model for other income (y_{2i}) will depend on the discrete employment decision (y_{1i}), *not* the latent variable (y_{1i}^*). So that for those employment other income is defined net of fixed costs

$$\widetilde{y}_{2i} \equiv y_{2i} - y_{1i}\alpha_2. \tag{5.11}$$

The coherency restrictions (5.8) imply that (5.7) can be rewritten

$$y_{1i}^* = (y_{2i} - y_{1i}\alpha_2)\boldsymbol{\beta}_2 + \mathbf{z}_{1i}\boldsymbol{\beta}_1 + u_i. \tag{5.12}$$

The adjustment to y_{2i} which guarantees statistical coherency is therefore *identical* to the correction to other income in the fixed cost model of labour market participation.

Defining $\widetilde{y}_{2i} \equiv y_{2i} - y_{1i}\alpha_2$, the coherency condition implies that the model can be rewritten

$$y_{1i} = 1\{\mathbf{z}_{1i}\boldsymbol{\beta}_1 + \widetilde{y}_{2i}\beta_2 + u_i > 0\} \tag{5.13}$$

and

$$\widetilde{y}_{2i} = \mathbf{z}_{2i}\boldsymbol{\gamma}_2 + \varepsilon_i. \tag{5.14}$$

If α_2 were known then the equations (5.14) and (5.13) are analogous to (5.2), (5.3) and (2.5). The semiparametric estimator using the control function approach would simply apply the estimation approach described in this paper to the conditional model.[10] Following the previous discussion, assumptions (2.11) and

---

[10]Blundell and Smith (1994) develop this estimator for the simultaneous parametric normal Probit and Tobit models.

(2.12) would be replaced by the modified conditional independence restrictions

$$u|\mathbf{z}_1, y_2, \mathbf{z}_2 \quad \sim \quad u|\mathbf{z}_1, \widetilde{y}_2, \varepsilon \tag{5.15}$$

$$\sim \quad u|\varepsilon. \tag{5.16}$$

The conditional expectation of the binary variable y_1 given the regressors $\mathbf{z}_1, \widetilde{y}_2$ and errors ε would then take the form

$$\begin{aligned} E[y_1|\mathbf{z}_1, \widetilde{y}_{2i}, \varepsilon] &= \Pr[-u \leq \mathbf{z}_{1i}\boldsymbol{\beta}_1 + \widetilde{y}_{2i}\beta_2 | \mathbf{z}_1, \widetilde{y}_{2i}, \varepsilon] \\ &\equiv F(\mathbf{z}_{1i}\boldsymbol{\beta}_1 + \widetilde{y}_{2i}\beta_2, \varepsilon). \end{aligned}$$

Finally, note that although α_2 is unknown, given sufficient exclusion restrictions on $\mathbf{z}_{2i}$, a root-n consistent estimator for α_2 can be recovered from (linear) 2SLS estimation of (5.6). More generally, if the linear form $\mathbf{z}_{2i}\boldsymbol{\gamma}_1$ of the regression function for y_2 is replaced by a nonparametric form $\gamma(\mathbf{z}_{2i})$ for some unknown (smooth) function γ, then a $\sqrt{n}$-consistent estimator of α_2 in the resulting partially linear specification for y_{2i} could be based on the estimation approach proposed by Robinson (1988), using nonparametric estimators of instruments $(\mathbf{z}_{1i} - E[\mathbf{z}_{1i}|\mathbf{z}_{2i}])$ in an IV regression of y_{2i} on y_{1i}.

5.1. The Estimates of the Coherency Model

The first column of Table 5.1 presents the estimates of the parameters of the structural equation for y_2 (5.6). These are recovered from instrumental variables estimation using the education of the husband as an excluded variable. The 'fixed cost of work' parameter seems reasonable for the income variable, whose mean is around £165 per week. The two sets of Probit results differ according to whether or not they control for ε. Notice that having removed the direct simultaneity of y_1

on y_2 through the adjustment $\widetilde{y}_2$, there is much less evidence of endogeneity bias. Indeed the coefficients on the adjusted other income variable in the two columns are quite similar (these are normalized relative to the education coefficient). If anything, after adjusting for fixed costs, controlling for endogeneity leads to a downward correction to the income coefficient.

Table 5.1: Results for the Coherency Specification.

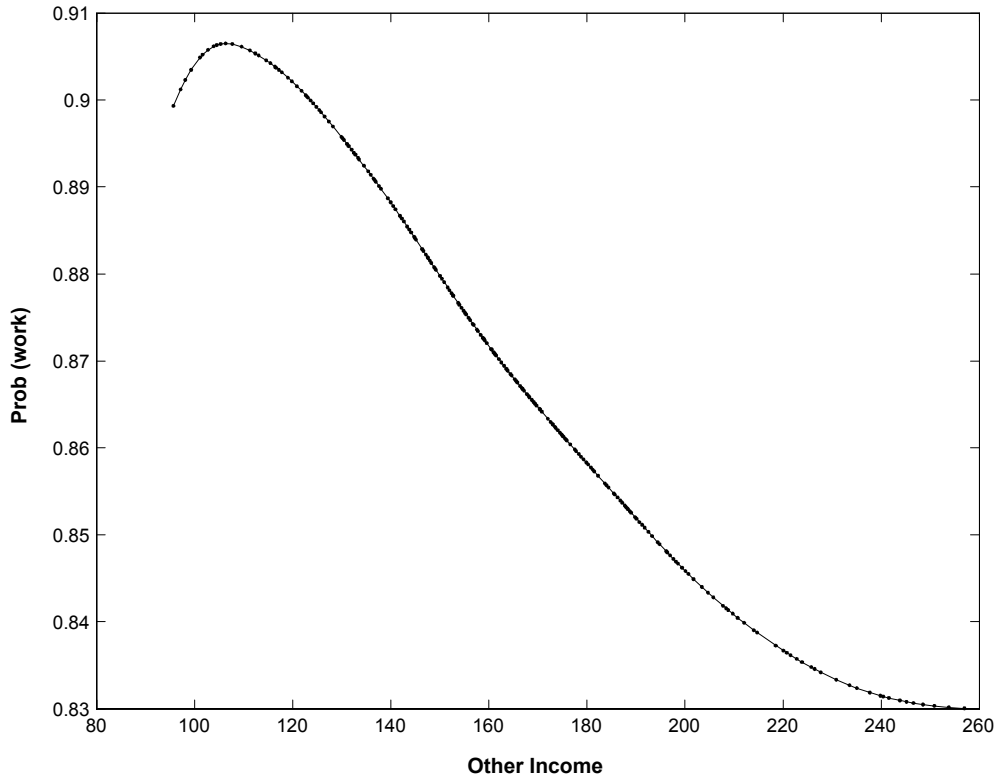| Variable	y_2 Coeff.	Std. Err.	Probit Pr[Work] Coeff.	Std. Err.	Probit Pr[Work$|\varepsilon$] Coeff.	Std. Err.
work: y_1	58.034	8.732	-		-	
education: z_1	-		1.6357	.2989	1.6553	.3012
adjusted income: $\widetilde{y}_2$	-	-	-0.7371	.0643	-.5568	.1433
benefit inc: z_{21}	0.4692	0.1453	-	-	-	-
education(sp): z_{22}	0.1604	0.0421	-	-	-	-
$\sigma_{u\varepsilon} = 0$ (t test)	-		-		2.556	

Figure 5: Semiparametric Estimation of the Coherency Model

Table 5.2: Semiparametric Results for the Coherency Specification.

Variable	Semi-P Pr[Work] *Coeff.*	Std. *Err.*	Semi-P Pr[Work$|\varepsilon$] *Coeff.*	Std. *Err.*
adjusted income: $\widetilde{y}_2$	-1.009	.0689	-0.82256	.2592

The comparable results for the semiparametric specification are presented in Table 5.2. In these we have used the linear structural model estimates for the y_2 equation exactly as in Table 5.1. These show a very similar pattern with only

46

a small difference in the other income coefficient between the specification that control for ε and the one that does not. Again the $\widetilde{y}_2$ adjustment seems to capture much of the endogeneity between work and income in this coherency specification.

In Figure 5 we present the semiparametric estimate of the probability of work across the whole low education sample. To evaluate this probability following the ASF formulation, used in the triangular specification, we have calculated $\widetilde{y}_2$ as if each individual pays the fixed cost.

6. Summary and Conclusions

This paper has proposed and implemented a new semiparametric method for estimating binary response models with continuous endogenous regressors. The method introduces residuals from the reduced form as covariates in the binary response model to control for endogeneity. We considered a specific semiparametric "matching" estimator of the index coefficients which exploits both continuity and monotonicity implicit in the binary response model formulation. We have also shown how the partial mean estimator from the nonparametric regression literature can be used to directly estimate the average structural function. The control function estimation approach, for this semiparametric model, is also shown to be easily adapted to the case where the model specification is not triangular and certain coherency conditions are required to be satisfied.

The proposed estimator was used to investigate the importance of correcting for the endogeneity of other income in a labor market participation model for a sample of married British men. The results show a strong effect of correcting for endogeneity in this example and indicate that adjusting for endogeneity using the

47

standard parametric models, the Probit and linear probability models, can give a highly misleading picture of the impact on participation of an exogenous change in other income.

References

[1] Ahn, H. (1995), "Non-parametric Two Stage Estimation of Conditional Choice Probabilities in a Binary Choice Model under Uncertainty," *Journal of Econometrics*, 67, 337-378.

[2] Ahn, H. and J.L. Powell (1993), "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3-29.

[3] Ahn, H., Ichimura, H. and J.L. Powell (1996), "Simple Estimators for Monotone Index Models," manuscript, Department of Economics, U.C. Berkeley.

[4] Amemiya, T. (1978), "The Estimation of a Simultaneous Equation Generalised Probit Model," *Econometrica*, 46, 1193-1205.

[5] Blundell, R.W. and J.L. Powell (2003), "Endogeneity in Nonparametric and Semiparametric Regression Models," in Dewatripont, M., L.P. Hansen, and S.J. Turnovsky, eds., *Advances in Economics and Econometrics*: *Theory and Applications, Eighth World Congress*, Vol. II, Cambridge: Cambridge University Press..

[6] Blundell, R.W., H. Reed and T. Stoker, (2003), "Interpreting Aggregate Wage Growth: The Role of Labour Market Participation", *American Economic Review,* Vol. 93, No. 4, September, 1114-1131.

[7] Blundell, R.W. and R.J. Smith (1986), "An Exogeneity Test for a Simultaneous Tobit Model", *Econometrica*, 54, 679-685.

[8] Blundell, R.W. and R.J.Smith (1989), "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models", *Review of Economic Studies*, 56, 37-58.

[9] Blundell, R.W. and Smith, R.J. (1994), "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables", *Journal of Econometrics*, 64, 355-373.

[10] Das, M., W.K. Newey, and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70(1), 33-58.

[11] Härdle W. (1990), *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.

[12] Härdle, W. and T. Stoker (1989), "Investigating Smooth Multiple Regression by the Method of Average Derivatives," *Journal of the American Statistical Association*, 84, 986-995.

[13] Heckman, J.J. (1978),"Dummy Endogenous Variable in a Simultaneous Equations System," *Econometrica*, 46, 931-959.

[14] Honoré, B.E. and A. Lewbel (2002), "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica*, 70, 2053-2063.

[15] Hristache, M., A. Juditsky, and V. Spokoiny (2001), "Direct Estimation of the Index Coefficients in a Single Index Model," *Annals of Statistics*, 29, 595-623.

[16] Ichimura, H. (1993), "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71-120.

[17] Klein, R.W. and R.S. Spady (1993), "An Efficient Semiparametric Estimator of the Binary Response Model," *Econometrica*, 61, 387-422.

[18] Lewbel, A. (1998), "Semiparametric Latent Variable Model Estimation With Endogenous or Mismeasured Regressors," *Econometrica*, 66, 105-121.

[19] Lewbel, A. (2000), "Semiparametric Qualitative Response Model Estimation With Instrumental Variables and Unknown Heteroscedasticity," *Journal of Econometrics*, 97, 145-177.

[20] Linton, O. and J.P. Nielson, (1995), "A Kernel Method of Estimating Nonparametric Structured Regression based on a Marginal Distribution," *Biometrika*, 82, 93-100.

[21] Manski, C.F. (1975), "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3, 205-228.

[22] Manski, C.F. (1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 205-228.

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

[24] Newey, W.K., Powell, J.L. and Vella, F. (1999). "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica*, 67, 565-603.

[25] Powell, J., J. Stock and T. Stoker (1989), "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403-1430.

[26] Robinson, P.M. (1988), "Root n-Consistent Semiparametric Regression", *Econometrica*, 56, 931-954.

[27] Silverman, B.W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

[28] Stock, J.H. (1989), "Nonparametric Policy Analysis," *Journal of the American Statistical Association*, 84, 567-575.

[29] Tjosthiem, D. and R.H.Auestad, (1996),"Nonparametric Identification of Nonlinear Time Series: Projections," *Journal of the American Statistical Association,* 89, 1398-1409.