

# ENDOGENEITY IN NONPARAMETRIC AND SEMIPARAMETRIC REGRESSION MODELS

---

*Richard Blundell*  
*James L Powell*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
**cemmap** working paper CWP09/01

# Endogeneity in Nonparametric and Semiparametric Regression Models

Richard Blundell\* and James L. Powell†

November 2001‡

## Abstract

This paper considers the nonparametric and semiparametric methods for estimating regression models with continuous endogenous regressors. We list a number of different generalizations of the linear structural equation model, and discuss how two common estimation approaches for linear equations – the “instrumental variables” and “control function” approaches – may be extended to nonparametric generalizations of the linear model and to their semiparametric variants. We consider the identification and estimation of the “Average Structural Function” and argue that this is a parameter of central interest in the analysis of semiparametric and nonparametric models with endogenous regressors. We consider a particular semiparametric model, the binary response model with linear index function and nonparametric error distribution, and describes in detail how estimation of the parameters of interest can be constructed using the “control function” approach. This estimator is applied to estimating the relation of labor force participation to nonlabor income, viewed as an endogenous regressor.

---

\*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.

‡**Acknowledgements:** Prepared for the Econometric Society World Meetings Seattle, August 2000. We are grateful to Manuel Arellano, David Card, Kenneth Chay, Lars Hansen, James Heckman, Hidehiko Ichimura, Arthur Lewbel, Oliver Linton, Costas Meghir, Aviv Nevo, Whitney Newey, Thomas Rothenberg, Adrian Pagan, Joris Pinkse, Paul Ruud, James Stock, Thomas Stoker, Frank Vella, Edward Vytlačil, and seminar participants at Berkeley, CREST, MIT, Northwestern, Princeton, Rutgers, UBC and UCL for helpful comments on this and related work. We would also like to thank Chuan Goh for able research assistance. Blundell acknowledges financial support for this research from the Leverhulme Foundation and the ESRC. The usual disclaimer applies.

## 1. Introduction

The analysis of data with endogenous regressors – that is, observable explanatory variables that are correlated with unobservable error terms – is arguably the main contribution of econometrics to statistical science. While “endogeneity” can arise from a number of different sources, including mismeasured regressors, sample selection, heterogeneous treatment effects, correlated random effects in panel data, etc., the term originally arose in the context of “simultaneity,” where the explanatory variables were, with the dependent variable, determined through a system of equations, so that their correlation with error terms arose from feedback from the dependent to the explanatory variables. Analysis of linear “supply and demand” systems (with normal errors) yielded the familiar rank and order conditions for identification, two- and three-stage estimation methods, and analysis of structural interventions; while these multi-step estimation procedures have been extended to nonlinear parametric models with additive nonnormal errors (e.g., Amemiya 1974, Hansen 1982), extensions to nonparametric and semiparametric models have only recently been considered.

The aim of this paper is to examine the existing literature on estimation of some “non-parametric” models with endogenous explanatory variables, and to compare the different identifying assumptions and estimation approaches for particular models and determine their applicability to others. To maintain a manageable scope for the paper, we restrict attention to nonparametric and semiparametric extensions of the usual simultaneous equations models (with endogenous regressors that are continuously distributed). We consider the identification and estimation of the “Average Structural Function” and argue that this parameter is a central parameter of interest in the analysis of semiparametric and nonparametric models with endogenous regressors. The two leading cases we consider are ad-

ditive nonparametric specifications in which the regression function is unknown, and nonadditive models in which there is some known transformation function that is monotone but not invertible. An important example of the latter, and one which we use as an empirical illustration, is the binary response model with endogenous regressors. We will not explicitly consider the closely-related problems of selectivity, heterogeneous treatment effects, correlated random effects or measurement error (see Heckman, Ichimura, Smith and Todd (1998), Angrist, Imbens and Rubin (1996), and Arellano and Honore (1999) for lucid treatments of these topics). Moreover, we consider only recent work on nonparametric and semiparametric variants of the two-stage least squares (2SLS) estimation procedure; Matzkin (1994) gives a broader survey of identification and estimation of nonlinear models with endogenous variables. Also, for convenience, we restrict attention to randomly-sampled data, though most of our discussion applies in non-i.i.d. contexts, provided the structural equations and stochastic restrictions involve only a finite number of observable random variables.

In the next subsections, a number of different generalizations of the linear structural equation are presented, and the objects of estimation (the “parameters of interest”) are defined and motivated. The sections which follow consider how two common interpretations of the two-stage least squares (2SLS) estimator for linear equations – the “instrumental variables” and “control function” approaches – may or may not be applicable to nonparametric generalizations of the linear model and to their semiparametric variants. The discussion then turns to a particular semiparametric model, the binary response model with linear index function and nonparametric error distribution, and describes in detail how estimation of the parameters of interest can be constructed using the “control function” approach. This estimator is applied to the empirical problem of the relation of labor force participation to nonlabor income studied in Blundell and

Powell (1999). The results point to the importance of the semiparametric approach developed here and the strong drawbacks of the linear probability model and other parametric specifications.

### 1.1. Structural Equations

A natural starting point for investigation of endogeneity is the classical linear structural equation

$$y = \mathbf{x}'\boldsymbol{\beta} + u, \tag{1.1}$$

where  $(y, \mathbf{x}')$  represents a data point of dimension  $(1 \times (k+1))$ ,  $\boldsymbol{\beta}$  is a conformable parameter vector and  $u$  is an unobservable disturbance term. The explanatory variables  $\mathbf{x}$  are assumed to include a subset of continuous *endogenous* variables, meaning that

$$E(\mathbf{x}u) \neq \mathbf{0}. \tag{1.2}$$

This is the standard single linear equation treated in the literature on simultaneous equations, and is considered here as a base case for comparison with other nonlinear setups. To identify and estimate the coefficient vector  $\boldsymbol{\beta}$  in this setup, the model must be completed by imposing restrictions on the unobservable error terms  $u$  that are consistent with (1.2); the traditional approach assumes that some observable vector  $\mathbf{z}$  of “instrumental variables” is available, satisfying the moment condition

$$E(\mathbf{z}u) = \mathbf{0}, \tag{1.3}$$

which leads to the well-known two-stage least squares (2SLS) estimator of  $\boldsymbol{\beta}$  (Theil 1953, Basmann 1959). The algebraic form of the 2SLS estimator can be derived from a number of different estimation principles based on (1.3), or on stronger conditions which imply it. As we will see, some of these estimation approaches, under

suitably strengthened stochastic restrictions, can be extended to the nonparametric and semiparametric generalizations of the linear model that are considered here, but in certain important cases this turns out not to be so.

At an opposite extreme from the standard linear model, the relation between  $y$  and its observable and unobservable determinants  $\mathbf{x}$  and  $u$  could be assumed to be of a general form

$$y = H(\mathbf{x}, u) \tag{1.4}$$

which may represent a single equation of a nonlinear simultaneous equation system, possibly with a limited or qualitative dependent variable. Of course, without further restrictions on  $H$ , this function would be unidentified even under strong conditions on the unobservables (like independence of  $\mathbf{x}$  and  $u$ ), but it is useful to view the various nonparametric and semiparametric models below as special cases of this setup.

One important class of structural functions, treated in more detail below, assumes  $H$  to be additively separable,

$$y = g(\mathbf{x}) + u, \tag{1.5}$$

which would be the “nonparametric regression model” if the expectation of  $u$  given  $\mathbf{x}$  could be assumed to be zero (i.e., if  $\mathbf{x}$  were exogenous). Identification and estimation of  $g$  when a subset of  $\mathbf{x}$  is endogenous and instrumental variables  $\mathbf{z}$  are available is the subject of a number of recent studies, including those by Newey and Powell (1989), Newey, Powell and Vella (1999), Darolles, Florens, and Renault (1999), and Ng and Pinkse (2000).

A non-separable variant of (1.4) would be Matzkin’s (1991) nonparametric version of Han’s (1987) “generalized regression” model,

$$y = t(g(\mathbf{x}), u), \tag{1.6}$$

in which  $h$  is a known function that is monotone, but not invertible, in its first argument (a "single index" if  $g$  is one-dimensional) and  $g$  is an unknown function satisfying appropriate normalization and identification restrictions. A leading special case of this specification is the nonparametric binary choice model (Matzkin 1992), in which  $H$  is an indicator function for positivity of the sum of  $g(\mathbf{x})$  and  $u$ :

$$t(g(\mathbf{x}), u) = 1(g(\mathbf{x}) + u > 0).$$

We group all of the models (1.4)-(1.6) in the *nonparametric* category, where the term refers to the lack of parametric structure to the structural function  $H$  or  $h$  or the "regression" function  $g$ . *Semiparametric* models restrict  $H$  (and possibly the distribution of  $u$ ) to have finite-dimensional parametric components, i.e.,

$$y = h(\mathbf{x}, \boldsymbol{\beta}, u). \tag{1.7}$$

For example, another special case of (1.4) is Han's (1987) original model, where the "single index" function  $g$  is assumed to be linear in  $\mathbf{x}$ ,

$$y = t(\mathbf{x}'\boldsymbol{\beta}, u); \tag{1.8}$$

estimation of this model when  $\mathbf{x}$  is endogenous is considered by, e.g., Lewbel (1998) and Blundell and Powell (1999) (which focuses on the binary response version of this linear index model). Yet another semiparametric special case

$$y = s(\mathbf{x}, \boldsymbol{\beta}, g(\cdot)) + u, \tag{1.9}$$

where  $\boldsymbol{\beta}$  is a finite parameter vector and  $h$  is a known function, has been considered in the recent work by Ai and Chen (2000). While estimation of the coefficient vector  $\boldsymbol{\beta}$  is typically the main objective of semiparametric analysis of such models, estimation of the distribution of  $u$ , or at least certain functionals of it, will also be needed to evaluate the response of the dependent variable  $y$  to possible "exogenous" movements in the explanatory variables.

## 1.2. Parameters of Interest

For a nonparametric model, the “parameters of interest” are actually unknown functions which summarize important characteristics of the structural function  $H$  and the distribution of the errors  $u$ ; these parameters will be identified if they can be extracted from the distributions of the observable random variables. From a random sample of observations on the dependent variable  $y$ , regressors  $\mathbf{x}$ , and instrumental variables  $\mathbf{z}$ , the joint distribution of  $y$ ,  $\mathbf{x}$ ,  $\mathbf{z}$ , is by definition identified, and conditional distributions and moments can be consistently estimated using standard nonparametric methods. In particular, the conditional expectation of functions of  $y$ , given either  $\mathbf{x}$  or  $\mathbf{z}$  or both, can be estimated without imposing additional restrictions (besides, say, smoothness and finite moment restrictions) on the joint distribution of the observable data, and these conditional expectations clearly summarize key features of the structural function and error distribution. However, as is clear from the well-worn supply and demand examples, knowledge only of the conditional distributions of observables is insufficient for analysis of the results of certain types of structural interventions which affect the distribution of the regressors  $\mathbf{x}$  but not the structural error terms  $u$ . Thus, the expectation of  $y$  given the instruments  $\mathbf{z}$ , called the *reduced form for  $y$* , may be of interest if the values of the instrumental variables are control variables for the policymaker, but for interventions which alter the explanatory variables  $\mathbf{x}$  directly, independently of the error terms  $u$ , knowledge only of the joint distribution of the observables will be insufficient. Similarly, potential results of interventions which directly affect some components of a vector-valued structural function  $H$  – e.g., a change in the supply function in a supply-and-demand system due to rationing – clearly could not be analyzed solely from knowledge of the reduced form for  $y$  prior to the intervention, nor could the results of policies intended to change the distribution



of the unobservable component  $u$ .

For analysis of such policies, it would be most useful to know the form of the structural function  $H(\mathbf{x}, u)$  from (1.4), along with the joint distribution of the errors  $u$  and  $\mathbf{x}, \mathbf{z}$ , but these may not be identifiable, at least for models without additive errors. An alternative summary version of the structural function  $H$  which can be more straightforward to estimate is the *average structural function (ASF)*, where the average is taken over the *marginal* distribution of the error terms  $u$

$$G(\mathbf{x}) \equiv \int H(\mathbf{x}, u) dF_u, \quad (1.10)$$

for  $F_u$  the marginal cumulative distribution function of  $u$ . In models with additively separable errors, i.e.,

$$H(\mathbf{x}, u) = g(\mathbf{x}) + u, \quad (1.11)$$

as in (1.5), the average structural function  $G(\mathbf{x})$  reduces to the usual “regression function”  $g(\mathbf{x})$ , which would correspond to  $E[y|\mathbf{x}]$  if the error terms  $u$  had conditional mean zero given  $\mathbf{x}$ . More generally, the average structural function  $G$  would be the counter-factual conditional expectation of  $y$  given  $\mathbf{x}$  if the endogeneity of  $\mathbf{x}$  were absent, i.e., if the regressors  $\mathbf{x}$  could be manipulated independently of the errors, which would be considered invariant to the structural change. For the heterogeneous treatment effect model (see Heckman and Robb (1985) and Imbens and Angrist (1994)), the average structural function is directly related to the average treatment effect recovered from an experimental design – specifically, the average treatment effect for a binary regressor would be  $G(1) - G(0)$ .

In some structural interventions, where the regressors  $\mathbf{x}$  can be manipulated directly, knowledge of the function  $G$ , or its derivatives, would be sufficient to assess the impact of the policy. For example, in the classical supply-and-demand

example, with  $y$  corresponding to quantity demanded and  $\mathbf{x}$  representing price, the ASF would suffice to determine expected demand if the market supply function were replaced with a fixed price (by fiat or by the world market), and if the distribution of  $u$  were assumed to be invariant to this structural change. And, for the additively-separable model (1.11), the ASF embodies the direct effect of the regressors for a particular observation, holding the error terms fixed; in this case the individual-specific, not just average, effects of changes in  $\mathbf{x}$  can be analyzed if  $G$  is known.

However, for interventions which do not directly determine the endogenous regressors, knowledge of the ASF is not enough; additional structural information about the structural function  $H$  and the distribution of  $\mathbf{x}$  (and possibly  $\mathbf{z}$ ) would be required to evaluate the policy effect. For example, to analyze the effects of imposition of a sales tax, which would rescale  $\mathbf{x}$  in the structural function by a fixed amount exceeding unity, the “inverse supply function” relating price  $\mathbf{x}$  to quantity supplied  $y$  and other observable and unobservable covariates would have to be specified to account for the joint determination of price and quantity following the structural intervention. More generally, policies which alter some components of the function  $H$  rather than manipulating the argument  $\mathbf{x}$  require specification, identification, and consistent estimation of all components of  $H$  (including equations for all endogenous regressors) and the distribution of  $u$ . As shown by Roehrig (1988) and Imbens and Newey (2001), nonparametric identification of a fully-specified system of simultaneous equations is possible under strong restrictions on the forms of the structural function – for example, invertibility of  $H(\mathbf{x}, u)$  in the unobservable component  $u$  – but such restrictions may be untenable for limited dependent variable models like the binary response model analyzed below. Thus, the average structural function  $G$  may be the only feasible measure of the direct effect of  $\mathbf{x}$  on  $y$  for limited dependent variable models

and for “limited information” settings in which the structural relations for the endogenous regressors in a single structural equation are incompletely specified.

Of course, the expected value of the dependent variable  $y$  need not be the only summary measure of interest; a complete evaluation of policy could require the entire distribution of  $y$  given an “exogenous”  $\mathbf{x}$ . Since that distribution can be equivalently characterized by the expectation of all measurable functions of  $y$ , a more ambitious objective is calculation of the ASF for any transformation  $\tau(y)$  of  $y$  with finite first moment,

$$G_\tau(\mathbf{x}) \equiv \int \tau(H(\mathbf{x}, u))dF_u.$$

For those sets of stochastic restrictions on  $u$  that require additively-separable form (1.11) for identification and estimation of the ASF, this collection of expectations can be evaluated directly from the marginal distribution of  $u = y - g(\mathbf{x})$ ; for those stochastic restrictions which do require additivity of errors for identification of  $G$ , the collection of functions  $G_\tau$  (and thus the structural distribution of  $y$ ) can be derived by redefinition of the dependent variable to  $\tau(y)$ .

For semiparametric problems of the form (1.7), the finite-dimensional parameter vector  $\boldsymbol{\beta}$  is typically of interest in its own right, since economic hypotheses can impose testable restrictions on the signs or magnitudes of its components. A primary goal of statistical analysis of semiparametric models is to construct consistent estimators of  $\boldsymbol{\beta}$  that converge at the parametric rate (the inverse of the square root of the sample size), with asymptotically normal distributions and consistently-estimable asymptotic covariance matrices. For some sets of restrictions on error distributions, this objective may be feasible even when estimation of the ASF  $G(\mathbf{x})$  is not. For example, for the semiparametric model (1.7), the form of the reduced form for  $y$ ,

$$E[y|\mathbf{z}] = E[h(\mathbf{x}'\boldsymbol{\beta}, u)|\mathbf{z}],$$

can, under appropriate restrictions, be exploited to obtain estimates of the “single index coefficients”  $\beta$  even if  $G$  is not identified, as noted below.

Even for a fully nonparametric model, some finite-dimensional summary measures of  $G$  may be of interest. In particular, the “average derivative” of  $G(\mathbf{x})$  with respect to  $\mathbf{x}$  (Stoker, 1986) can be an important measure of marginal effects of an exogenous shift in the regressors. Altonji and Ichimura (2000) consider the estimation of the derivative  $y$  with respect to  $x$  for the case when  $y$  is censored. They are able to derive a consistent estimator for the non-additive case. Unlike the estimation of “single index coefficients”, which are generally only identified up to a scale factor, estimation of the average derivatives of the ASF  $G$  is problematic unless  $G$  itself is identifiable.

## 2. Nonparametric Estimation Under Alternative Stochastic Restrictions

As in the traditional treatment of linear simultaneous equations, we will assume that there exists a  $1 \times m$  vector  $\mathbf{z}$  of “instrumental variables”, typically with  $m \geq k$ . The particular stochastic restrictions on  $\mathbf{z}$ ,  $\mathbf{x}$  and  $u$  will determine what parameters are identified and what estimators are available in each of the model specifications (1.4) - (1.9). Each of the stochastic restrictions is a stronger form of the moment condition (1.3), and each can be used to motivate the familiar 2SLS estimator in the linear model with additive errors (1.1) under the usual rank condition, but their applicability to the nonparametric and semiparametric models varies according to the form of the structural function  $H$ .

## 2.1. “Instrumental Variables” Methods

### 2.1.1. *The linear model*

The “instrumental variables” (IV) version of the standard 2SLS estimator is the sample analogue to the solution of a weaker implication of (1.3), namely,

$$0 = E(P[\mathbf{x}|\mathbf{z}]u) \equiv E(\Pi'\mathbf{z}u) = E(\Pi'\mathbf{z}(y - \mathbf{x}'\beta)), \quad (2.1)$$

where  $P[x|z]$  is the population least-squares projection of  $\mathbf{x}$  on  $\mathbf{z}$ , with

$$\Pi \equiv \{E(\mathbf{z}\mathbf{z}')\}^{-1}E(\mathbf{z}'\mathbf{x}). \quad (2.2)$$

Replacing population expectations with sample averages in (2.1) yields the 2SLS estimator

$$\hat{\boldsymbol{\beta}}_{2SLS} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}, \quad (2.3)$$

with

$$\hat{\mathbf{X}} = \mathbf{Z}\hat{\Pi} \text{ and } \hat{\Pi} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}, \quad (2.4)$$

and where  $\mathbf{X}$ ,  $\mathbf{Z}$  and  $\mathbf{y}$  are the  $N \times k$ ,  $N \times m$  and  $N \times 1$  data arrays corresponding respectively to  $\mathbf{x}$ ,  $\mathbf{z}$  and  $y$  for a sample of size  $N$ . When the linear form of the residual  $u = y - \mathbf{x}'\beta$  is replaced with a nonlinear, parametric version  $u = m(y, \mathbf{x}, \beta)$ , extension of this estimation approach yields the generalized instrumental variable (GIVE) estimator of Sargan (1959), the nonlinear two-stage least squares (NLLS) estimator of Amemiya (1974) and the generalized method of moments (GMM) estimator (Hansen 1982).<sup>1</sup>

---

<sup>1</sup>When  $\beta$  is “overidentified” – that is, the dimension  $m$  of the instruments  $\mathbf{x}$  exceeds the dimension  $k$  of  $\beta$  – asymptotically efficient estimation would be based upon a different implication of (1.3), in which  $\Pi$  is replaced by a different  $(m \times k)$  “selection” matrix, as noted by Hansen (1982).

Another closely-related formulation of 2SLS exploits a different implication of (1.3), namely,

$$0 = P[u|\mathbf{z}] = P[y|\mathbf{z}] - P[\mathbf{x}|\mathbf{z}]'\beta, \quad (2.5)$$

where the population linear projection coefficients of  $u$  and  $y$  on  $\mathbf{z}$  are defined analogously to (2.2). Replacing  $P[y|\mathbf{z}]$  and  $P[\mathbf{x}|\mathbf{z}]$  with their sample counterparts and applying least squares yields Basmann's (1959) version of 2SLS,

$$\widehat{\boldsymbol{\beta}}_{2SLS} = (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\widehat{\mathbf{y}}, \quad (2.6)$$

where now

$$\widehat{\mathbf{y}} = \mathbf{Z}\widehat{\boldsymbol{\pi}} \equiv \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

Although logically distinct from the IV interpretation of 2SLS in (2.3), extension of the estimation approaches in either (2.3) or (2.6) yield the same NLLS and GMM estimators in the nonlinear parametric case, and we refer to generalization of either approach to the nonparametric or semiparametric structural equations as an *instrumental variable method*.

#### *Extensions to additive nonparametric models*

To extend the instrumental variables methods to non-parametric settings, it is natural to strengthen the unconditional moment restriction  $E(\mathbf{z}u) = \mathbf{0}$  to a conditional mean restriction

$$E(u|\mathbf{z}) = 0, \quad (2.7)$$

just as the assumption of  $E[\mathbf{x}u] = \mathbf{0}$  is strengthened to  $E[u|\mathbf{x}] = 0$  for a nonparametric regression model. For the additive structural function (1.11), identification and estimation of  $g(\mathbf{x})$  was considered by Newey and Powell (1989) and Darolles, Florens, and Renault (1999). Substitution of the error term  $u = y - g(\mathbf{x})$  into

condition (2.7) yields a relationship between the reduced form  $E[y|\mathbf{z}]$  and the structural function  $g$  :

$$\begin{aligned} E[y|\mathbf{z}] &= E[g(\mathbf{x})|\mathbf{z}] \\ &= \int g(\mathbf{x})dF_{\mathbf{x}|\mathbf{z}}, \end{aligned} \tag{2.8}$$

where  $F_{\mathbf{x}|\mathbf{z}}$  is the conditional c.d.f. of  $\mathbf{x}$  given  $\mathbf{z}$ . The reduced form for  $y$ ,  $E[y|\mathbf{z}]$ , and the conditional distribution of  $\mathbf{x}$  given  $\mathbf{z}$  are functionals of the joint distribution of the observable variables  $y$ ,  $\mathbf{x}$ , and  $\mathbf{z}$  are identified; identifiability of the structural function  $g$  therefore reduces to the uniqueness of the solution of the integral equation (2.8). And, as noted in the Newey-Powell and Darolles-Florens-Renault manuscripts, this in turn reduces to the question of statistical completeness of the family of conditional distributions  $F_{\mathbf{x}|\mathbf{z}}$  in the “parameter”  $\mathbf{z}$ . (See, for example, Ferguson 1967, Section 3.6, for a definition of completeness and its connection to minimum variance unbiased estimation in parametric problems.) While conditions for completeness of  $F_{\mathbf{x}|\mathbf{z}}$  are known for certain parametric classes of distributions (e.g., exponential families), and generally the “order condition”  $\dim(\mathbf{z}) \geq \dim(\mathbf{x})$  must be satisfied, in a nonparametric estimation setting uniqueness of the solution of (2.8) must be imposed as a primitive assumption. Darolles, Florens, and Renault (2000) give a more thorough discussion of the conditions for existence and uniqueness of the solution of (2.8) and its variants.

In the special case where  $\mathbf{x}$  and  $\mathbf{z}$  have a joint distribution which is discrete with finite support, conditions for identification and consistent estimation of the ASF  $g(\mathbf{x})$  are straightforward to derive. Suppose  $\{\boldsymbol{\xi}_j, j = 1, \dots, J\}$  are the set of possible values for  $\mathbf{x}$  and  $\{\boldsymbol{\zeta}_l, l = 1, \dots, L\}$  are the support points for  $\mathbf{z}$ , and let

$$\begin{aligned} \boldsymbol{\pi} &\equiv \text{vec}(E[y|\mathbf{z} = \boldsymbol{\zeta}_l]) \text{ and} \\ \mathbf{P} &\equiv [P_{jl}] \equiv [\text{Pr}\{\mathbf{x} = \boldsymbol{\xi}_j|\mathbf{z}_l = \boldsymbol{\zeta}_j\}] \end{aligned} \tag{2.9}$$

denote the vector of reduced form values  $E[y|\mathbf{z}]$  and the matrix of conditional probabilities that  $\mathbf{x} = \boldsymbol{\xi}_j$  given that  $\mathbf{z} = \boldsymbol{\zeta}_l$ , respectively; these would clearly be identified, and consistently estimable, from a random sample of observations on  $y$ ,  $\mathbf{x}$ , and  $\mathbf{z}$ . If  $\mathbf{g} \equiv \text{vec}(g(\boldsymbol{\xi}_j))$  denotes the vector of possible values of  $g(\mathbf{x})$ , the question of identifiability of  $\mathbf{g}$  using (2.8) is the question of uniqueness of the solution to the set of linear equations

$$\boldsymbol{\pi} = \mathbf{P}\mathbf{g}, \tag{2.10}$$

so that  $\mathbf{g}$  is identified if and only if  $\text{rank}\{\mathbf{P}\} = J = \dim\{\mathbf{g}\}$ , which requires the order condition  $L = \dim\{\boldsymbol{\pi}\} \geq \dim\{\mathbf{g}\} = J$ . When  $\mathbf{g}$  is identified, it may be consistently estimated (at a parametric rate) by replacing  $\boldsymbol{\pi}$  and  $\mathbf{P}$  by estimators using the empirical c.d.f. of the observed vectors in (2.10), and solving for  $\hat{\mathbf{g}}$  in the “just-identified” case  $J = L$ , or using a minimum chi-square procedure when  $\mathbf{g}$  is “overidentified” ( $J < L$ ).<sup>2</sup> More details for this finite-support case are given by Das (1999).

*The “ill-posed inverse” problem*

Unfortunately, the simple structure of this finite support example does not easily translate to the general case, where  $\mathbf{x}$  and  $\mathbf{z}$  may have continuously-distributed components. Unlike in typical nonparametric estimation problems, where identification results can be easily translated into consistent estimators of the identified functions under smoothness or monotonicity restrictions, identification of  $g$  and consistent estimators of the components  $E[y|\mathbf{z}]$  and  $F_{\mathbf{x}|\mathbf{z}}$  are not, by themselves, sufficient for a solution of a sample analogue of (2.8) to be a consistent estimator of  $g$ . First, it is clear that, unlike the standard nonparametric regression problem, the function  $g(\mathbf{x})$  (and the reduced form and conditional dis-

---

<sup>2</sup>Note that when  $J=K=2$  this is the “treatment effect” in the homogeneous treatment effect model case with additive errors. The heterogeneous treatment effect case is a specific form of the general nonadditive model.



tribution function) must be estimated for all values of  $\mathbf{x}$  in the support of the conditional distribution, and not just at a particular value  $\mathbf{x}_0$  of interest; thus, consistency of  $g$  must be defined in terms of convergence of a suitable measure of distance between the functions  $\hat{g}(\cdot)$  and  $g(\cdot)$  (e.g., the maximum absolute difference over possible  $\mathbf{x}$  or the integrated squared difference) to zero in probability. Moreover, the integral equation (2.8), a generalization of the “Fredholm integral equation of the first kind,” is a notorious example of an “ill-posed inverse” problem: the integral  $T_{\mathbf{z}}(g) \equiv \int g(\mathbf{x})dF_{\mathbf{x}|\mathbf{z}}$ , while continuous in  $g$  for the standard functional distance measures, has inverse which is not continuous in general, even if the inverse is well-defined. That is, even if a unique solution  $\hat{g}$  of the sample version

$$\hat{E}[y|\mathbf{z}] = \int g(\mathbf{x})d\hat{F}_{\mathbf{x}|\mathbf{z}} \quad (2.11)$$

of (2.8) exists, that solution  $\tilde{g} \equiv \hat{T}_{\mathbf{z}}^{-1}(\hat{E}[y|\mathbf{z}])$  is not continuous in the argument  $\hat{E}[y|\mathbf{z}]$ , so consistency of the reduced form estimator (and the estimator of the conditional distribution of  $\mathbf{x}$  given  $\mathbf{z}$ ) does not imply consistency of  $\hat{g}$ . Heuristically, the reduced form  $E[y|\mathbf{z}]$  can be substantially smoother than the structural function  $g(\mathbf{x})$ , so that very different structural functions can yield very similar reduced forms; the ill-posed inverse problem is a functional analogue to the problem of multicollinearity in a classical linear regression model, where large differences in regression coefficients can correspond to small differences in fitted values of the regression function. Such “ill-posed inverse” problems are well-known in applied mathematics and statistics, arising, for example, in the problem of estimation of the density of an unobservable variable  $x$  which measured with error; that is, observations are available only on  $y = x + u$ , where  $u$  is an unobservable error term with known density function (the “deconvolution problem”). O’Sullivan (1986) surveys the statistical literature on ill-posed inverse problems, and describes the

general “regularization” approaches to construction of consistent estimators for such problems. If the joint distribution of  $\mathbf{x}$  and  $\mathbf{z}$  is approximated by a distribution with a finite support (as common for “binning” approaches to nonparametric estimation of conditional distributions), then the ill-posed inverse problem would manifest itself as an extreme sensitivity of the “transition” matrix  $\mathbf{P}$  to choice of  $J$  and  $L$ , and its near-singularity as  $J$  and  $L$  increase to infinity.

*Consistent estimation methods*

Newey and Powell (1989) impose further restrictions on the set of possible structural functions  $g$  to obtain a consistent estimator, exploiting the fact that the inverse of a bounded linear functional such as  $T_{\mathbf{z}}(g)$  will be continuous if the domain of the functional is compact. For this problem, compactness of the set of possible  $g$  functions with respect to, say, the “sup norm” measure of distance between functions, can be ensured by restricting the “Sobolev norm” of all possible  $g$  functions is bounded above by a known constant. This Sobolev norm (denoted here as  $\|g\|_S$ ) is a different but related distance measure on functions which involves a sum of the integrated squared values of  $g(\mathbf{x})$  and a certain number of its derivatives. In effect, the requirement that  $\|g\|_S$  is bounded, which ensures that  $g$  is sufficiently smooth, counteracts the “ill-posed inverse” problem by substantially restricting the possible candidates for the inverse function  $\hat{T}_{\mathbf{z}}^{-1}(\hat{E}[y|\mathbf{z}])$ .

To obtain a computationally -feasible estimation procedure, Newey and Powell assume the structural function  $g$  can be well-approximated by a function which is linear in parameters:

$$g(\mathbf{x}) \cong g_J(\mathbf{x}) \equiv \sum_{j=1}^J \alpha_j \rho_j(\mathbf{x}), \quad (2.12)$$

where the  $\{\rho_j\}$  are known, suitably-chosen “basis functions” (like polynomials or trigonometric functions) which yield an arbitrarily-close approximation to  $g$  as the number of terms  $J$  in the sum is increased. For this approximation, the

corresponding approximation to the reduced form  $E[y|\mathbf{z}] = E[g(\mathbf{x})|\mathbf{z}]$  is

$$E[g_J(\mathbf{x})|\mathbf{z}] = \sum_{j=1}^J \alpha_j E[\rho_j(\mathbf{x})|\mathbf{z}], \quad (2.13)$$

which is itself linear in the same parameters  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)'$ , so that constrained least-squares regression of  $y$  on nonparametric estimates  $\hat{E}[\rho_j(\mathbf{x})|\mathbf{z}]$  of the conditional means of the basis functions  $\rho_j(\mathbf{x})$  can be used to estimate the coefficients of the approximate structural function  $g_J$  under the compactness restriction. Furthermore, the square of the Sobolev norm  $\|g_J\|_S$  of the linear approximating function  $g_J$  can be written as a quadratic form,

$$\|g_J\|_S^2 = \frac{1}{2} \boldsymbol{\alpha}' \mathbf{S}_J \boldsymbol{\alpha}, \quad (2.14)$$

where the matrix  $\mathbf{S}_J$  is a known matrix constructed using integrals involving the basis functions  $\rho_j(\mathbf{x})$  and their derivatives. Minimization of the sum of squared differences between observed values of  $y$  and the estimators  $\{\hat{E}[\rho_j(\mathbf{x})|\mathbf{z}]\}$ , subject to the restriction that the quadratic form in (2.14) is bounded above by a known constant  $B$ , yields an estimator of the coefficient vector  $\boldsymbol{\alpha}$  that is of “penalized least squares” form:

$$\hat{\boldsymbol{\alpha}} = (\hat{\mathbf{R}}' \hat{\mathbf{R}} + \hat{\lambda} \mathbf{S}_J)^{-1} \hat{\mathbf{R}}' \mathbf{y}, \quad (2.15)$$

where  $\hat{\mathbf{R}}$  is the matrix of the “first-stage” estimators  $\{\hat{E}[\rho_j(\mathbf{x})|\mathbf{z}]\}$  for the sample and  $\hat{\lambda}$  is a Lagrange multiplier for the constraint  $\boldsymbol{\alpha}' \mathbf{S}_J \boldsymbol{\alpha} \leq 2B$ . Imposition of the compactness condition thus introduces an adjustment for multicollinearity (through the term  $\hat{\lambda} \mathbf{S}_J$ ) to the otherwise-familiar 2SLS formula, to account for the near-singularity of the fitted values in the first-stage, which is at the heart of the ill-posed inverse problem.

Newey and Powell (1989) give conditions under which the resulting estimator

of the structural function  $g$ ,

$$\hat{g}(\mathbf{x}) \equiv \sum_{j=1}^J \hat{\alpha}_j \rho_j(\mathbf{x}), \quad (2.16)$$

is consistent; in addition to the compactness restrictions on the set of possible structural functions, these conditions restrict the form of the basis functions  $\rho_j$  and require that the number of terms  $J$  in the approximating function increase to infinity with the sample size. However, unlike some other nonparametric regression methods based upon series approximations,  $J$  can be arbitrarily large for finite samples, and its value needs not be related to the sample size to ensure convergence of bias and variance to zero, but instead is governed by the tradeoff between numerical precision of the series approximation to  $g$  and computational convenience. The Newey and Powell manuscript does not discuss the rate of convergence or asymptotic distribution of  $\hat{g}$ , nor appropriate choice of the constraint constant  $B$  or, equivalently, the Lagrange multiplier  $\hat{\lambda} = \hat{\lambda}(B)$ , which acts as a “smoothing parameter” in the second-stage estimator.

A conceptually-simple variant of this estimation strategy can be based upon a finite-support approximation to the joint distribution of  $\mathbf{x}$  and  $\mathbf{z}$ . Once the data are “binned” into partitions with representative values  $\{\boldsymbol{\xi}_j\}$  and  $\{\boldsymbol{\zeta}_i\}$ , the linear relation (2.10) between the vector  $\mathbf{g}$  of structural function values and the reduced form vector  $\boldsymbol{\pi}$  and transition matrix  $\mathbf{P}$  will hold approximately (with the approximation improving as the number of bins increases), and the components  $\boldsymbol{\pi}$  and  $\mathbf{P}$  can be estimated using bin averages and frequencies. Though the estimated transition matrix  $\hat{\mathbf{P}}$  may be nearly singular even if the approximating bins are chosen with  $L \gg J$  for  $J$  large, the structural function vector  $\mathbf{g}$  could be estimated by “ridge regression,” i.e.,

$$\hat{\mathbf{g}} = (\hat{\mathbf{P}}' \hat{\mathbf{P}} + \lambda \mathbf{S})^{-1} \hat{\mathbf{P}}' \hat{\boldsymbol{\pi}} \quad (2.17)$$

for some nonsingular matrix  $\mathbf{S}$  and “smoothing parameter”  $\lambda$  which shrinks to zero as the sample size increases. This can be viewed as a “histogram” version of the series estimator proposed by Newey and Powell, which uses bin indicators as the basis functions and kernel regression (with uniform kernels) in the first stage.

Darolles, Florens, and Renault (2000) take a different approach to estimation of the structural function  $g$  in (2.8). They embed the problem of the solution of (2.8) in the mean-squared error minimization problem, defining the structural function  $g$  as

$$g(\cdot) \equiv \arg \min_{\phi(\cdot)} E[\|E[y|\mathbf{z}] - \int \phi(\mathbf{x}) dF_{\mathbf{x}|\mathbf{z}}\|^2], \quad (2.18)$$

and note that the “normal equations” for this functional minimization problem are of the form

$$\begin{aligned} E[E[y|\mathbf{z}|\mathbf{x}]] &\equiv \tau(\mathbf{x}) \\ &= E[E[g(\mathbf{x})|\mathbf{z}|\mathbf{x}]] \\ &\equiv T^*(g)(\mathbf{x}). \end{aligned} \quad (2.19)$$

That is, they transform (2.8) into another integral equation by taking conditional expectations of the reduced form  $E[y|\mathbf{z}]$  given the original explanatory variables  $\mathbf{x}$ ; an advantage of this formulation is that the transformation  $T^*(g) = \tau$  has the same argument ( $\mathbf{x}$ ) as the structural function  $g$ , as is standard for the literature on the solution of Fredholm integral equations. While the ill-posed inverse problem persists, a standard solution method for this formulation is “Tikhonov regularization,” which replaces the integral equation (2.19) with the approximate problem

$$\tau(\mathbf{x}) = T^*(g^\lambda)(\mathbf{x}) + \lambda g^\lambda(\mathbf{x}), \quad (2.20)$$

for  $\lambda$  a small, nonnegative “smoothing” parameter. While (2.20) reduces to (2.19) as  $\lambda \rightarrow 0$ , it is a “Fredholm integral equation of the *second* kind,” which is free

of the ill-posed inverse problem, when  $\lambda$  is nonzero. Again approximating the solution function  $g^\lambda$  as a linear combination of basis functions,

$$g^\lambda(\mathbf{x}) \cong g_J^\lambda(\mathbf{x}) \equiv \sum_{j=1}^J \alpha_j^\lambda \rho_j(\mathbf{x}), \quad (2.21)$$

as in (2.12), a further approximation to the equation (2.19) is

$$E[E[y|\mathbf{z}|\mathbf{x}]] \cong \sum_{j=1}^J \alpha_j^\lambda \{E[E[\rho_j(\mathbf{x})|\mathbf{z}|\mathbf{x}]] + \lambda \rho_j(\mathbf{x})\}. \quad (2.22)$$

This suggests a two-stage strategy for estimation of the  $\alpha^\lambda$  coefficients: in the first stage, obtain nonparametric estimators of the components  $\tau(\mathbf{x}) = E[E[y|\mathbf{z}|\mathbf{x}]]$  and the doubly-averaged basis functions  $\{T^*(\rho_j)(\mathbf{x}) = E[E[\rho_j(\mathbf{x})|\mathbf{z}|\mathbf{x}]]\}$  using standard nonparametric estimation methods, then, in the second stage, regress the fitted  $\hat{\tau}(\mathbf{x})$  on the constructed regressors  $\{\hat{T}^*(\rho_j)(\mathbf{x}) + \lambda \rho_j(\mathbf{x}), j = 1, \dots, J\}$ . The terms  $\lambda \rho_j(\mathbf{x})$  serve to attenuate the severe multicollinearity of the doubly-averaged basis functions in this second-stage regression.

Darolles, Florens, and Renault take the basis functions  $\rho_j$  to be the eigenfunctions of the estimated double-averaging operator  $\hat{T}^*$ , i.e., the solutions to the functional equations  $\hat{T}^*(\rho_j) = \nu_j \rho_j$  for scalar eigenvalues  $\nu_j$ , which simplifies both computation of the estimator and derivation of the asymptotic theory. For this choice of basis function, they derive the rate of convergence and asymptotic normal distribution of the estimator  $\tilde{g}^\lambda(\mathbf{x}) = \sum_{j=1}^J \hat{\alpha}_j^\lambda \rho_j(\mathbf{x})$  under certain regularity conditions. The rate of convergence is comparable to, but slower than, the rate of convergence of standard nonparametric estimators of the reduced form  $E[y|\mathbf{z}]$ , due to the bias introduced by approximating (2.19) by (2.20) for nonzero  $\lambda$ . Their manuscript also proposes an alternative estimator of  $g$  based upon regression of  $\hat{\tau}(\mathbf{x})$  on the a subset of the doubly-averaged basis functions with eigenvalues bounded away from zero, i.e.,  $\{\hat{T}^*(\rho_j)(\mathbf{x}) : |\nu_j| > b_n\}$  for some  $b_n \rightarrow 0$ , and

extends the identification analysis to permit the structural function  $g$  to be additively separable in its endogenous and exogenous components.

### 2.1.2. *Nonadditive models*

Both the Newey-Powell and Darolles-Florens-Renault approaches exploit the additive separability of the error terms  $u$  in the structural function for  $y$ ; for models with nonadditive errors –  $H(\mathbf{x}, u) \neq g(\mathbf{x}) + u$  – the “instrumental variables” assumption imposed in these papers apparently does not suffice to identify the average structural function  $G$  of (1.10). Of course, it is clear that, for a nonadditive model, the conditional mean assumption (2.7) would not suffice to yield identification even for parametric structural functions, but imposition of the still-stronger assumption of independence of  $u$  and  $\mathbf{z}$ , denoted here as

$$u \perp\!\!\!\perp \mathbf{z}, \tag{2.23}$$

(which implies (2.7), and thus (1.3) provided  $u$  has finite expectation which can be normalized to zero) will still not suffice in general for identification of the ASF  $G$ . This is evident from inspection of the reduced form  $E[y|\mathbf{z}]$  in the nonadditive case:

$$\begin{aligned} E[y|\mathbf{z}] &= E[H(\mathbf{x}, u)|\mathbf{z}] && (2.24) \\ &= \int H(\mathbf{x}, u) dF_{u, \mathbf{x}|\mathbf{z}} \\ &= \int \left[ \int H(\mathbf{x}, u) dF_{u|\mathbf{x}, \mathbf{z}} \right] dF_{\mathbf{x}|\mathbf{z}} \\ &\neq \int \left[ \int H(\mathbf{x}, u) dF_u \right] dF_{\mathbf{x}|\mathbf{z}} \\ &= E[G(\mathbf{x})|\mathbf{z}]. \end{aligned}$$

That is, independence of  $u$  and  $\mathbf{z}$  does not imply independence of  $u$  and  $\mathbf{x}$ ,  $\mathbf{z}$ , or even conditional independence of  $u$  and  $\mathbf{z}$  given  $\mathbf{x}$ , etc. In the additive case

$H(\mathbf{x}, u) = g(\mathbf{x}) + u$ , conditional expectations of each component only require the conditional distributions of  $u$  given  $\mathbf{z}$  and of  $\mathbf{x}$  given  $\mathbf{z}$ , and not the joint distribution of  $u, \mathbf{x}$  given  $\mathbf{z}$ , which is not identified under (2.23) without further conditions on the relation of  $\mathbf{x}$  to  $\mathbf{z}$ . And, since (2.24) also holds in general for any function of  $y$ , the restriction (2.23) does not yield restrictions on the conditional distribution of the observable  $y$  given  $\mathbf{z}$  which might be used to identify the ASF  $G$  for general nonseparable structural functions  $H$ .

Of course, failure of the reduced-form relation (2.24) to identify the ASF  $G$  does not directly imply that it could not be identified using some other functionals of the joint distribution of the observables  $y, \mathbf{x}$ , and  $\mathbf{z}$ , and it is difficult to provide a constructive proof of nonidentification of the ASF under the independence restriction (2.23) at this level of generality (i.e., with structural function  $H$  and the joint distribution of  $u, \mathbf{x}$ , and  $\mathbf{z}$  otherwise unspecified). Still, the general nonidentification result can be illustrated by considering a simple (slightly pathological) binary response example in which the ASF is unidentified under (2.23). Suppose  $H$  is binary, with  $y$  generated as

$$y = 1(x+u \geq 0), \tag{2.25}$$

for a scalar regressor  $x$  generated by a multiplicative model

$$x = z \cdot e, \tag{2.26}$$

for some scalar instrumental variable  $z$  with  $\Pr\{z \geq 0\} = 1$ . For this example, the average structural function is

$$G(x) \equiv 1 - F_u(x), \tag{2.27}$$

with  $F_u$  the marginal c.d.f. of  $u$ . Now suppose the errors  $u$  and  $e$  are generated as

$$\begin{aligned} u &\equiv \varepsilon \cdot \text{sgn}(\eta), \\ e &\equiv \eta \cdot \text{sgn}(\varepsilon), \end{aligned} \tag{2.28}$$



with  $\varepsilon$ ,  $\eta$ , and  $z$  being mutually independently distributed and

$$\text{sgn}(\eta) \equiv 1 - 2 \cdot 1(\eta < 0).$$

This model is “pathological” because  $\text{sgn}(u) = \text{sgn}(x)$  by construction, and thus  $y = 1(x \geq 0)$ , whenever  $z \neq 0$ . Still, the independence condition (2.23) is satisfied, the dependent variable  $y$  is non-constant if the support of  $e$  includes positive and negative values, and the endogenous regressor  $x$  has conditional expectation which is a nontrivial function of the instrument  $z$  when  $E[e] = E[\eta] \cdot (1 - 2 \cdot \Pr\{\varepsilon < 0\}) \neq 0$ . Yet the ASF  $G(x)$  is only identified at  $x = 0$ , when zero is in the support of  $z$  (with  $G(0) = E[y|z = 0]$ ), and is not identified elsewhere.

This example demonstrates that, without further restrictions on the form of the structural function  $H$  and/or the conditional distribution of  $u$  given  $\mathbf{x}$  and  $\mathbf{z}$ , the assumption of independence of the structural error  $u$  and the instruments  $\mathbf{z}$  is insufficient to identify the ASF in nonadditive models even when the endogenous regressors  $\mathbf{x}$  are not independent of the instruments  $\mathbf{z}$ . The nonidentification of the ASF here is a consequence of the dependence of the support of the endogenous variable  $x$  (either zero or the positive or negative half-line) on the realized value of the error term  $e$ . In general, if the nature of the endogeneity of  $\mathbf{x}$  is restricted by assuming

$$\mathbf{x} = h(\mathbf{z}, \mathbf{e}) \tag{2.29}$$

for some function  $h$  that is invertible in the error terms  $\mathbf{e}$ , i.e.,

$$\mathbf{e} = k(\mathbf{z}, \mathbf{x}), \tag{2.30}$$

and if the support of  $\mathbf{x}$  given  $\mathbf{e}$  is independent of  $\mathbf{e}$ , Imbens (2000) has shown how the ASF  $G$  is identified under the independence restriction (2.23) and these additional restrictions on the nature of the endogeneity of  $\mathbf{x}$ . Imbens’ identification

argument is based upon the “control function” approach described in more detail below.

As an alternative to imposing such restrictions on the nature of the endogeneity of  $\mathbf{x}$ , additional structure on the form of the structural function  $H$  – such as invertibility of the structural function  $H(\mathbf{x}, u)$  in the error term  $u$  – may yield more scope for identification of the ASF when the stochastic restrictions only involve the conditional distribution of  $u$  given  $\mathbf{z}$ . For example, suppose there is some invertible transformation  $t(y)$  of  $y$  for which the additive form (1.11) holds:

$$t(y) = g(\mathbf{x}) + u, \quad (2.31)$$

where  $u$  satisfies the conditional mean restriction (2.7). If the transformation  $t$  were known, then estimation of  $g$  could proceed using the Newey-Powell or Darolles-Florens-Renault approaches, and the ASF could be estimated by averaging the estimator of  $H(\mathbf{x}, u) = t^{-1}(g(\mathbf{x}) + u)$  over the marginal empirical distribution of  $u = t(y) - g(\mathbf{x})$ . When  $t$  is unknown, the conditional mean restriction (2.7) yields an integral equation

$$\begin{aligned} 0 &= E[u|\mathbf{z}] \\ &= \int t(y)dF_{y|\mathbf{z}} - \int g(\mathbf{x})dF_{\mathbf{x}|\mathbf{z}} \end{aligned} \quad (2.32)$$

which has multiple solutions, e.g.,  $t(y) \equiv g(\mathbf{x}) \equiv k$  for any constant  $k$ . Still, with appropriate normalizations on the unknown functions  $t$  and  $g$ , like  $E[t(y)] = 0$  and  $E[(t(y))^2] = 1$ , it may be possible to extend the estimation approaches for the “ill-posed inverse problem” to joint estimation of  $t$  and  $g$ , though this may require “overidentification,” that is,  $m = \dim(\mathbf{z}) > \dim(\mathbf{x}) = k$ .

For the semiparametric problems (1.7), the parametric components  $\beta$  of the structural function may well be identified and consistently-estimable, at the parametric (root- $N$ ) rate, even if the ASF  $G$  is not identified. Ai and Chen (1999)

propose sieve estimation of semiparametric models of the form (1.9) under the assumption that the instrumental variables  $\mathbf{z}$  are independent of the error terms  $u$ ; while the estimator of the infinite-dimensional nuisance function  $h(\cdot)$  is not generally consistent with respect to the usual distance measures (like integrated square differences), the corresponding estimator  $\hat{\boldsymbol{\beta}}$  of the parametric component  $\boldsymbol{\beta}$  is root-N consistent and asymptotically normal under the regularity conditions they impose.

Lewbel (1998, 1999a) considers the single-index generalized regression model (1.8), constructing consistent estimators of the index coefficients  $\boldsymbol{\beta}$  under the assumption that one of the components of the explanatory variables  $\mathbf{x}$ , say  $x_1$ , is continuously distributed and independent of the structural error  $u$  (and is thus a component of the set of instruments  $\mathbf{z}$  satisfying (2.23) above). Provided there exist an exogenous variable  $x_1$  which satisfy these conditions, Lewbel’s approach permits a weaker stochastic restriction than independence of  $\mathbf{z}$  (including the “special regressor”  $\mathbf{x}_1$ ) and  $u$  – namely, that  $u$  need only be independent of  $x_1$  conditionally on the other components of  $\mathbf{x}$  and of the instrument vector  $\mathbf{z}$ . The conditional mean restriction  $E[u|\mathbf{z}] = 0$  can also be weakened to the moment restriction  $E[\mathbf{z}u] = 0$  in this setup. The conditional independence restriction is similar to the restrictions imposed for the “control function” methods described below. Nevertheless, even if the coefficient vector  $\boldsymbol{\beta}$  were known *a priori*, the endogeneity of the remaining components of  $\mathbf{x}$ , and thus of the index  $\mathbf{x}'\boldsymbol{\beta}$ , would yield the same difficulties in identification of the ASF  $G$  as in (2.24) above.

### 2.1.3. “Fitted Value” Methods

When the conditional mean (2.7) or independence (2.23) restrictions of the instrumental variables approach does not suffice to identify the ASF  $G$  in a non-parametric model, the researcher can either abandon the average structural func-

tion concept and focus on alternative summary measures that are identified, or impose stronger restrictions on the structural function or error distributions to achieve identification of the ASF. Though imposing additional restrictions on the structural function  $H$  (such as additivity of the error terms  $u$ ) can clearly help achieve identifiability of  $G$ , such restrictions may be implausible when the range of  $y$  is restricted (e.g., when  $y$  is binary), and it is more customary to strengthen the restrictions on the conditional error distribution  $u$  given the instruments  $\mathbf{z}$  to identify the parameters of interest.

One alternative set of restrictions and estimation procedures are suggested by Theil's (1953) version of the 2SLS estimator for simultaneous equations. Defining the "first stage residuals"  $\mathbf{v}$  as the difference between the regressors  $\mathbf{x}$  and their linear projections onto  $\mathbf{z}$ ,

$$\mathbf{v} \equiv \mathbf{x} - P[\mathbf{x}|\mathbf{z}] \equiv \mathbf{x} - \Pi'\mathbf{z}, \quad (2.33)$$

where  $\Pi$  is defined in (2.2), the condition (1.3), when combined with the definition of  $\mathbf{v}$  and the linear structural function (1.1), yields the restriction

$$\begin{aligned} 0 &= E[P[\mathbf{x}|\mathbf{z}](u + \mathbf{v}'\boldsymbol{\beta})] \\ &= E[(\Pi'\mathbf{z})(y - (\mathbf{z}'\Pi)\boldsymbol{\beta})], \end{aligned} \quad (2.34)$$

so that the structural coefficients  $\boldsymbol{\beta}$  are the least squares regression coefficients of the regression of the dependent variable  $y$  on the "fitted values"  $\Pi'\mathbf{z}$  of the regressors  $\mathbf{x}$ . The sample analogue of the population regression coefficients of  $y$  on  $\Pi'\mathbf{z}$  is Theil's version of 2SLS,

$$\widehat{\boldsymbol{\beta}}_{2SLS} = (\widehat{\mathbf{X}}'\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}'\mathbf{y}, \quad (2.35)$$

where  $\widehat{\mathbf{X}} = \mathbf{Z}\widehat{\Pi}$  is defined as in (2.4) above. The motivation for this form of 2SLS is the replacement of the endogenous regressors  $\mathbf{x}$  with that part of  $\mathbf{x}$  (its linear

projection on  $\mathbf{z}$ ) that is uncorrelated with the error  $u$  in the linear structural equation.

In a nonparametric setting, it is natural to define the first-stage residuals  $\mathbf{v}$  as deviations from conditional expectations, rather than linear projections:

$$\begin{aligned}\mathbf{v} &\equiv \mathbf{x} - E[\mathbf{x}|\mathbf{z}] & (2.36) \\ &\equiv \mathbf{x} - \Pi(\mathbf{z}).\end{aligned}$$

By construction  $E[\mathbf{v}|\mathbf{z}] = \mathbf{0}$ , and, as for the “instrumental variables” approaches, the moment condition (1.3) would be replaced by the stronger conditional mean restriction (2.7), or the still-stronger assumption of independence of the errors and the instruments,

$$(u, \mathbf{v}) \perp\!\!\!\perp \mathbf{z}. \quad (2.37)$$

A nonparametric generalization of Theil’s version of 2SLS would estimate  $E[\mathbf{x}|\mathbf{z}] = \Pi(\mathbf{z})$  by a suitable nonparametric method in the first stage, and then substitute the fitted values  $\tilde{\Pi}(\mathbf{z})$  into the structural function in a second-stage estimation procedure. As noted by Amemiya (1974), though, substitution of fitted values into nonlinear structural functions generally yields inconsistent estimates of the structural parameters, even in parametric problems; estimation methods which use substitution of fitted values into the structural function rely heavily on linearity of the regression function, so that the model can be written in terms of a composite error  $u + \mathbf{v}'\boldsymbol{\beta}$  with similar stochastic properties to the structural error  $u$ . For the general (nonadditive) structural function  $H$  of (1.4), substitution of the reduced form into the structural function yields  $y = H(\Pi(\mathbf{z}) + \mathbf{v}, u)$ , and, analogously to (2.24), the reduced form for  $y$  bears no obvious relation to the average structural function  $G$  under condition (2.37). Even when the structural function  $H$  is additive,  $H(\mathbf{x}, u) = g(\mathbf{x}) + u$ , the reduced form for  $y$  can be written

as

$$E[y|\mathbf{z}] = E[\mathbf{y}|\Pi(\mathbf{z})] = \int g(\Pi(\mathbf{z}) + \mathbf{v})dF_{\mathbf{v}},$$

so insertion of the first-stage equation for  $\mathbf{x}$  into the structural function yields an ill-posed inverse relation between the reduced form for  $y$  and the ASF  $g$ . Thus the fitted value approach inherits similar limitations to the instrumental variables approach, though it may simplify the resulting integral equation, which depends upon  $\Pi(\mathbf{z})$  rather than  $\mathbf{z}$  and involves the marginal distribution of  $\mathbf{v}$  rather than the conditional distribution of  $\mathbf{x}$  given  $\mathbf{z}$ .

Of course, for structural equations where the conditional expectations  $E[\mathbf{x}|\mathbf{z}] = \Pi(\mathbf{z})$  are the “right” explanatory variables, the fitted value estimation method, using nonparametric estimates of  $\Pi(\mathbf{z})$ , is an obvious way to proceed<sup>3</sup>. And, as was true for the instrumental variables approach, for semiparametric problems consistent estimation of the parametric component  $\beta$  may be feasible even when the average structural function  $G$  is not identified. For example, for the generalized regression model  $y = h(\mathbf{x}'\beta, u)$  of (1.8), the reduced form for  $y$  given  $\mathbf{z}$  is of single-index form when the errors  $u, \mathbf{v}$  are independent of the instrument vector  $\mathbf{z}$ :  $E[y|\mathbf{z}] \equiv G^*(\Pi(\mathbf{z})'\beta)$  for some function  $G^*$ , so that a nonparametric estimator of the first-stage regression function  $\Pi(\mathbf{z})$  can be combined with a standard estimation method for single index regression models<sup>4</sup>.

---

<sup>3</sup>Some asymptotic results for such estimators were given by Ahn and Manski (1993) and Ahn (1995), which showed, e.g., that the rate of convergence of the restricted reduced form estimator  $\hat{E}[y|\hat{\Pi}(\mathbf{z})]$  is the smaller of the rate of convergence of  $\hat{\Pi}$  to  $\Pi$  and of  $\hat{E}[y|\Pi(\mathbf{z})]$  to  $E[y|\Pi(\mathbf{z})]$  for  $\Pi$  known.

<sup>4</sup>See, e.g., Ichimura (1993) and others described in Horowitz (1993) and Powell (1994).

## 2.2. “Control Function” Methods

### 2.2.1. *The linear model*

While insertion of the fitted values from the first-stage nonparametric regression of  $\mathbf{x}$  on  $\mathbf{z}$  is not generally helpful in identification and estimation of the ASF  $G$ , alternative assumptions and procedures involving the use of the residuals  $\mathbf{v}$  from this first-stage regression to “control for endogeneity of the regressors”  $\mathbf{x}$  do yield identification of the ASF even for the general, nonadditive structural function (1.4). This *control function* approach has its antecedent in another algebraically-equivalent interpretation of the 2SLS estimator  $\hat{\beta}_{2SLS}$  as the coefficients on  $\mathbf{x}$  in a least-squares regression of  $y$  on  $\mathbf{x}$  and the residuals  $\hat{\mathbf{v}}$  from a linear regression of  $\mathbf{x}$  on  $\mathbf{v}$  :

$$\begin{pmatrix} \hat{\beta}_{2SLS} \\ \hat{\rho}_{2SLS} \end{pmatrix} = (\widehat{\mathbf{W}}' \widehat{\mathbf{W}})^{-1} \widehat{\mathbf{W}}' \mathbf{y}, \quad (2.38)$$

where

$$\widehat{\mathbf{W}} = [\mathbf{X} \ \widehat{\mathbf{V}}] \text{ and } \widehat{\mathbf{V}} = \mathbf{X} - \widehat{\mathbf{X}} = \mathbf{X} - \mathbf{Z}\widehat{\Pi},$$

and where  $\hat{\rho}_{2SLS}$  are the coefficients on the first-stage residuals  $\widehat{\mathbf{V}}$ .<sup>5</sup> This construction exploits another consequence of the moment condition  $E(\mathbf{z}u) = \mathbf{0}$ , that

$$\begin{aligned} P[u|\mathbf{x}, \mathbf{z}] &= P[u|\Pi'\mathbf{z} + \mathbf{v}, \mathbf{z}] & (2.39) \\ &= P[u|\mathbf{v}, \mathbf{z}] \\ &= P[u|\mathbf{v}] \\ &\equiv \mathbf{v}'\boldsymbol{\rho} \end{aligned}$$

---

<sup>5</sup>It has been difficult to locate a definitive early reference to the control function version of 2SLS. Dhrymes (1970, equation 4.3.57) shows that the 2SLS coefficients can be obtained by a least-squares regression of  $\mathbf{y}$  on  $\widehat{\mathbf{X}}$  and  $\widehat{\mathbf{V}}$ , while Telser (1964) shows how the seemingly-unrelated regressions model can be estimated by using residuals from other equations as regressors in a particular equation of interest. Heckman (1978) references this paper in his comprehensive discussion of estimating simultaneous models with discrete endogenous variables.

for some coefficient vector  $\boldsymbol{\rho}$ ; the third equality follows from the orthogonality of both error terms  $u$  and  $\mathbf{v}$  with  $\mathbf{z}$ . Thus, this particular linear combination of the first-stage errors  $\mathbf{v}$  is a function which controls for the endogeneity of the regressors  $\mathbf{x}$ . It also follows from this formulation that

$$P[u|\mathbf{x}, \mathbf{z}] = P[u|\mathbf{x}, \mathbf{v}], \quad (2.40)$$

which can be used as the basis for a test of overidentification if  $\dim(\mathbf{z}) > \dim(\mathbf{v}) = \dim(\mathbf{x})$ .

This approach treats endogeneity as an omitted variable problem, where the inclusion of estimates of the first-stage errors  $\mathbf{v}$  (the part of the regressors  $\mathbf{x}$  that is correlated with  $\mathbf{z}$ ) as a covariate corrects the inconsistency of least-squares regression of  $y$  on  $\mathbf{x}$ , in the same way that the Heckman (1979) two-step estimator corrects for selectivity bias through introduction of an appropriately-estimated regressor derived from a parametric form for the error distribution. The control function approach to correct for endogeneity has been extended to nonlinear parametric models by Blundell and Smith (1986,1989), who show how introduction of first-stage residuals into single-equation probit or Tobit procedures yields consistent estimators of the underlying regression coefficients when some of the regressors are endogenous.

### **2.2.2. Extensions to additive nonparametric models**

Application of the control function approach to nonparametric and semiparametric settings requires strengthening of the linear projection restrictions (2.39) and (2.40) to conditional mean restrictions

$$\begin{aligned} E(u|\mathbf{x}, \mathbf{z}) &= E(u|\mathbf{x}, \mathbf{v}) \\ &= E(u|\mathbf{v}), \end{aligned} \quad (2.41)$$



or, for nonadditive models, the stronger conditional independence assumptions

$$\begin{aligned} u|\mathbf{x}, \mathbf{z} &\sim u|\mathbf{x}, \mathbf{v} \\ &\sim u|\mathbf{v}. \end{aligned} \tag{2.42}$$

While the “control variates”<sup>6</sup>  $\mathbf{v}$  are typically taken to be deviations of  $\mathbf{x}$  from its conditional mean  $E[\mathbf{x}|\mathbf{z}]$ , as in (2.36), this is not required; more generally,  $\mathbf{v}$  can be any function of the observable random vectors

$$\mathbf{v} = \boldsymbol{\nu}(y, \mathbf{x}, \mathbf{z}) \tag{2.43}$$

that is identified and consistently estimable, provided (2.41) or (2.42) holds for this formulation (and  $\mathbf{v}$  is not  $\mathbf{x}$  a nontrivial function of  $\mathbf{x}$ ). This permits the control function approach to be applied to some systems of nonlinear simultaneous equations for which the appropriate reduced form is difficult or impossible to derive, such as the “coherent simultaneous binary response models” considered below. In comparison to the identifying assumptions (2.7) or (2.23) for the instrumental variables approaches, the corresponding assumptions (2.41) or (2.42) for the control function approach are no more nor less general. Both sets of assumptions are implied by the independence restriction (2.37), which may be plausible for certain applications, and which permits a choice between the two estimation approaches.

Estimation of the ASF  $g$  in the additive model (1.11) under the conditional mean exclusion restriction (2.41) was considered by Newey, Powell, and Vella (1999) and by Ng and Pinkse (1995) and Pinkse (2000). When the errors are additive, substitution of  $u = y - g(\mathbf{x})$  into (2.41) yields a “generalized additive

---

<sup>6</sup>This use of the term “control variate” is logically distinct, but similar in spirit, to its use in the literature on Monte Carlo methods – see, e.g., Hammersley and Handscomb (1964), section 5.5.

regression” form for  $y$  :

$$\begin{aligned} E[y|\mathbf{x}, \mathbf{v}] &= E[(g(\mathbf{x}) + u)|\mathbf{x}, \mathbf{v}] \\ &= g(\mathbf{x}) + \boldsymbol{\eta}(\mathbf{v}) \end{aligned} \tag{2.44}$$

for some “control function”  $\eta$ . With a suitable normalization (e.g.,  $E[\eta(\mathbf{v})] = 0$ ) the ASF  $g$  can be estimated using standard additive nonparametric regression methods applied to the regression of  $y$  on  $\mathbf{x}$  and the first-stage residuals  $\hat{\mathbf{v}}$ . Both Newey-Powell-Vella and Ng-Pinkse propose estimation of (2.44) using a series approximation for the functions  $g$  and  $\eta$  :

$$g(\mathbf{x}) + \boldsymbol{\eta}(\mathbf{v}) \cong \sum_{j=1}^J \alpha_j \rho_j(\mathbf{x}) + \sum_{l=1}^L \gamma_l \psi_l(\mathbf{v}), \tag{2.45}$$

where  $\{\rho_j\}$  and  $\{\psi_l\}$  are appropriate basis functions, and where the number of terms  $J$  and  $L$  for each approximating series increase to infinity as the sample size increases. The second-stage using this series approximation is a least-squares regression of  $y$  on the basis functions  $\{\rho_j(\mathbf{x})\}$  and  $\{\psi_l(\hat{\mathbf{v}})\}$ , and the estimator of  $g$  is given by (2.16), assuming  $\rho_1(\mathbf{x}) \equiv 1$  (which enforces the normalization  $E[\eta(\mathbf{v})] = 0$ ). The cited manuscripts give regularity conditions for consistency of the estimator  $\hat{g}$ , and derive its rate of convergence, which is the same as the rate for a direct nonparametric regression of  $y$  on the regressors  $\mathbf{x}$ ; Newey, Powell, and Vella also give conditions under which the estimator  $\hat{g}(\mathbf{x})$  is asymptotically normal.

### 2.2.3. *Nonadditive models*

Unlike the instrumental variables approach, a stronger “independence” version (2.42) of the conditional mean exclusion restrictions (2.41) for the control function approach does lead to a consistent estimator of the ASF  $G$  when the structural function  $H$  is nonadditive, as in (1.4). Blundell and Powell (1999) point

out how averaging the conditional mean of  $y$  given  $\mathbf{x}$  and  $\mathbf{v}$  over the marginal distribution of the first-stage errors  $\mathbf{v}$  gives the ASF  $G$  for the nonadditive model. Since

$$\begin{aligned}
E[y|\mathbf{x}, \mathbf{v}] &= E[H(\mathbf{x}, u)|\mathbf{x}, \mathbf{v}] & (2.46) \\
&= \int H(\mathbf{x}, u) dF_{u|\mathbf{x}, \mathbf{v}} \\
&= \int H(\mathbf{x}, u) dF_{u|\mathbf{v}} \\
&\equiv H^*(\mathbf{x}, \mathbf{v})
\end{aligned}$$

under the strong exclusion restriction (2.42), it follows that the “generalized control function”  $H^*$  can be integrated over the marginal distribution of the (observable) reduced form errors to obtain the ASF:

$$\begin{aligned}
\int H^*(\mathbf{x}, \mathbf{v}) dF_{\mathbf{v}} &= \int \left[ \int H(\mathbf{x}, u) dF_{u|\mathbf{v}} \right] dF_{\mathbf{v}} & (2.47) \\
&= \int H(\mathbf{x}, u) dF_u \\
&\equiv G(\mathbf{x}).
\end{aligned}$$

In a sense, the control function exclusion restriction (2.42) permits replacement of the unidentified structural errors  $u$  with the identified control variable  $\mathbf{v}$  through iterated expectations, so that averaging the structural function  $H$  over the marginal distribution of the structural errors  $u$  is equivalent to averaging the (identified) intermediate regression function  $H^*$  over the marginal distribution of  $\mathbf{v}$ . The intermediate structural function  $H^*$  is a nonadditive generalization of the previous control function  $\eta(\mathbf{v}) = E[u|\mathbf{v}]$  for additive models.

For the binary response example described in (2.25) through (2.28) above, the first-stage residuals are of the form

$$\begin{aligned}
v &= z \cdot (e - E[e]) & (2.48) \\
&= z \cdot [\eta \cdot \text{sgn}(\varepsilon) - E[\eta] \cdot (\Pr\{\varepsilon \geq 0\} - \Pr\{\varepsilon < 0\})],
\end{aligned}$$

and the conditional exclusion restriction (2.42) only holds if the structural error  $u$  is degenerate, i.e.,  $u = 0$  with probability one. In this case,  $x$  is exogenous, and the intermediate structural function reduces to

$$\begin{aligned} H^*(x, v) &= E[1(x \geq 0)|x, v] \\ &= 1(x \geq 0), \end{aligned} \tag{2.49}$$

which trivially integrates to the true ASF  $G(x) = 1(x \geq 0)$ . Thus, imposition of the additional restriction (2.42) serves to identify the average structural function here. An alternative “control variate” to the first-stage errors  $v$  would be

$$v^* \equiv \text{sgn}(x) = \text{sgn}(u), \tag{2.50}$$

which satisfies (2.42) when the structural error  $u$  is nondegenerate. Since  $v^*$  is functionally related to  $x$ , the intermediate structural function  $H^*(x, v^*) = E[y|x, v^*]$  is not identified using this control variate.

Translating the theoretical formulation (2.47) to a sampling context leads to a “partial mean” (Newey 1994b) or “marginal integration” (Linton and Nielson 1995, Tjostheim and Auestad 1996) estimator for the ASF  $G$  under the conditional independence restrictions (2.42) of the control function approach. That is, after obtaining a first-stage estimator  $\hat{\mathbf{v}}$  of the control variate  $\mathbf{v}$  (which would be the residual from a nonparametric regression of  $\mathbf{x}$  on  $\mathbf{z}$  when  $\mathbf{v}$  is defined by (2.36)), an estimator  $\hat{H}^*$  of the function  $H^*$  in (2.46) can be obtained by nonparametric regression of  $y$  on  $\mathbf{x}$  and  $\hat{\mathbf{v}}$ . A final estimation step would average  $H^*$  over the observed values of  $\hat{\mathbf{v}}$ ,

$$\hat{G}(\mathbf{x}) = \int \hat{E}(y|\mathbf{x}, \mathbf{v}) d\hat{F}_{\hat{\mathbf{v}}} \equiv \int \hat{H}^*(\mathbf{x}, \mathbf{v}) d\hat{F}_{\hat{\mathbf{v}}}, \tag{2.51}$$

where  $\hat{F}_{\hat{\mathbf{v}}}$  is the empirical c.d.f. of the residuals  $\hat{\mathbf{v}}$ . Alternatively, if  $\mathbf{v}$  were assumed to be continuously distributed with density  $f_{\mathbf{v}}$ , the ASF  $G$  could be estimated by

integrating  $\hat{H}^*$  over a nonparametric estimator  $\hat{f}_{\mathbf{v}}$  of  $f_{\mathbf{v}}$ ,

$$\tilde{G}(\mathbf{x}) = \int \hat{E}(y|\mathbf{x}, \mathbf{v}) \hat{f}_{\mathbf{v}} d\mathbf{v}. \quad (2.52)$$

Either the partial mean (2.51) or marginal integration (2.52) are alternatives to the series estimators based upon (2.45) for the additive structural function (1.11), but this latter approach is not applicable to general structural functions, since the intermediate regression function  $H^*$  need not be additive in its  $\mathbf{x}$  and  $\mathbf{v}$  components.

#### 2.2.4. *Support Restrictions*

The identification requirements for the ASF  $G$  are simpler to interpret for the control function approach than the corresponding conditions for identification using the instrumental variables approaches, since they are conditions for identification of the nonparametric regression function  $H^*$ , which is based upon observable random vectors. For example, in order for the ASF  $G(\mathbf{x})$  to be identified from the partial-mean formulation (2.47) for a particular value  $\mathbf{x}_0$  of  $\mathbf{x}$ , the support of the conditional distribution of  $\mathbf{v}$  given  $\mathbf{x} = \mathbf{x}_0$  must be the same as the support of the marginal distribution of  $\mathbf{v}$  – otherwise, the regression function  $H^*(\mathbf{x}_0, \mathbf{v})$  will not be well-defined for all  $\mathbf{v}$ , nor will the integral of  $H^*$  over the marginal distribution of  $\mathbf{v}$ . For those components of  $\mathbf{x}$  that are exogenous – that is, those components of  $\mathbf{x}$  which are also components of the instrument vector  $\mathbf{z}$  – the corresponding components of  $\mathbf{v}$  are identically zero, both conditionally on  $\mathbf{x}$  and marginally, so this support requirement is automatically satisfied. However, for the endogenous components of  $\mathbf{x}$ , the fact that  $\mathbf{x}$  and  $\mathbf{v}$  are functionally related through the first-stage relation  $\mathbf{v} = \mathbf{x} - \Pi(\mathbf{z})$  (or the more general form (2.43)) means that the support condition generally requires that  $\mathbf{v}$ , and thus  $\mathbf{x}$ , must be continuously distributed, with unbounded support (conditionally on the

instruments  $\mathbf{z}$ ) if its marginal distribution is nondegenerate. Similar reasoning, imposing the requirement that  $H^*(\mathbf{x}, \mathbf{v})$  be well-defined on the support of  $\mathbf{x}$  for all possible  $\mathbf{v}$ , and noting that

$$E[y|\mathbf{x}, \mathbf{v}] = E[y|\Pi(\mathbf{z}), \mathbf{v}], \quad (2.53)$$

implies that the first-stage regression function  $\Pi(\mathbf{z}) = E[\mathbf{x}|\mathbf{z}]$  must also be continuously distributed, with full-dimensional support, for the nondegenerate components of  $\mathbf{v} = \mathbf{x} - E[\mathbf{x}|\mathbf{z}]$ .

The requirement that the endogenous components of  $\mathbf{x}$  be continuously distributed is the most important limitation of the applicability of the control function approach to estimation of nonparametric and semiparametric models with endogenous regressors. When the structural equations for the endogenous regressors have limited or qualitative dependent variables, the lack of an invertible representation (2.43) of the underlying error terms for such models generally makes it impossible to construct a control variate  $\mathbf{v}$  for which the conditional independence restrictions (2.41) or (2.42) are plausible. The requirement of an additive (or invertible) first-stage relation for the regressors  $\mathbf{x}$  in the control function approach is comparable to the requirement of an additive (or invertible) structural function  $H$  for the identification of the ASF  $G$  using the instrumental variables approach.

On the other hand, when the errors are invertible and the support of  $\mathbf{x}$  does not depend upon them, Imbens (2000) has shown how a particular control variate – the conditional cumulative distribution of  $\mathbf{x}$  given  $\mathbf{z}$ , evaluated at the observed values of those random variables – can be used to identify the ASF  $G$  under the independence restriction (2.23), since it satisfies the conditional independence restriction (2.42). Also, when it is applicable, the control function approach to estimation with endogenous regressors is compatible with other estimation strategies that use “control function” methods to adjust for different sources of specification

bias, like selection bias (e.g., Heckman 1978, Heckman and Robb 1985, Ahn and Powell 1993, Honoré and Powell 1997, Das, Newey, and Vella 1998, and Vytlačil 1999) or correlated random effects in panel data models (Altonji and Matzkin 1997).

### 3. Binary Response Linear Index Models

#### 3.1. Model Specification and Estimation Approach

While the control function approach adopted by Blundell and Powell (1999) applies to fully nonparametric problems (as described above), their discussion focuses attention on estimation of the parametric and nonparametric components of a particular semiparametric single index model (1.8), the binary response model with linear index,

$$y = 1\{\mathbf{x}'\boldsymbol{\beta} + u > 0\}, \quad (3.1)$$

where the conditional independence assumption (2.42) is assumed to hold for  $\mathbf{v} \equiv \mathbf{x} - E[\mathbf{x}|\mathbf{z}]$ . For this linear binary response model, the ASF  $G$  is the marginal c.d.f. of  $-u$  evaluated at the linear index  $\mathbf{x}'\boldsymbol{\beta}$ ,

$$G(\mathbf{x}) = F_{-u}(\mathbf{x}'\boldsymbol{\beta}), \quad (3.2)$$

which is interpreted as the counterfactual conditional probability that  $y = 1$  given  $\mathbf{x}$  if  $\mathbf{x}$  were exogenous, i.e., if the conditional distribution of  $u$  given  $\mathbf{x}$  were assumed to be identical to its true marginal distribution. Under the conditional independence restriction (2.42), the intermediate regression function  $H^*$  is the conditional c.d.f. of  $-u$  given  $\mathbf{v}$ , again evaluated at  $\mathbf{x}'\boldsymbol{\beta}$  :

$$H^*(\mathbf{x}, \mathbf{v}) = F_{-u|\mathbf{v}}(\mathbf{x}'\boldsymbol{\beta}|\mathbf{v}). \quad (3.3)$$

Given a random sample of observations on  $y$ ,  $\mathbf{x}$ , and  $\mathbf{z}$  from the model (3.1) under the exclusion restriction (2.42), the estimation approach proposed by Blundell and Powell for the parameters of interest in this model follows three main steps. The first step uses nonparametric regression methods – specifically, the Nadaraya-Watson kernel regression estimator – to estimate the error term  $\mathbf{v}$  in the reduced form, as well as the unrestricted conditional mean of  $y$  given  $\mathbf{x}$  and  $\mathbf{v}$ ,

$$E[y|\mathbf{x}, \mathbf{v}] \equiv H^*(\mathbf{w}) \quad (3.4)$$

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

$$\mathbf{w} = (\mathbf{x}', \mathbf{v}')'. \quad (3.5)$$

This step can be viewed as an intermediate “structural estimation” step, which imposes the first exclusion restriction of (2.42) but not the second. The remaining estimation steps use semiparametric “pairwise differencing” or “matching” methods to obtain an estimator of the index coefficients  $\boldsymbol{\beta}$ , followed by partial-mean estimation of the ASF  $G$ .

### 3.1.1. *The semiparametric estimator of the index coefficients*

After using kernel regression methods to obtain an estimator  $\hat{H}^*(\mathbf{x}, \hat{\mathbf{v}}) = \hat{E}[y|\mathbf{x}, \hat{\mathbf{v}}]$  of the intermediate regression function  $H^*$ , Blundell and Powell use a semiparametric estimation method to extract an estimator of  $\boldsymbol{\beta}$  from the relation

$$\begin{aligned} H^*(\mathbf{w}) &= E[y|\mathbf{x}, \mathbf{v}] \\ &= E[y|\mathbf{x}'\boldsymbol{\beta}, \mathbf{v}] \\ &\equiv \Gamma(\mathbf{x}'\boldsymbol{\beta}, \mathbf{v}), \end{aligned} \quad (3.6)$$

which is a consequence of the single-index form of the binary response model (3.1). While a number of standard methods for estimation of the coefficients



of the single-index regression model  $E[y|\mathbf{x}] = \Gamma(\mathbf{x}'\boldsymbol{\beta})$  could be extended to the multi-index model<sup>7</sup> (3.6), the particular estimator of  $\boldsymbol{\beta}$  adopted by Blundell and Powell (1999) is an adaptation of a method proposed by Ahn, Ichimura, and Powell (1996), which imposes additional regularity conditions of both continuity and monotonicity of  $H^*(\lambda, \mathbf{v}) = E[y|\mathbf{x}'\boldsymbol{\beta} = \lambda, \mathbf{v}]$  in its first argument. These conditions follow from the assumption that  $u$  is continuously distributed, with support on the entire real line, conditional on  $\mathbf{v}$ , because of (3.3).

Since the “structural” index model is related to the conditional mean of  $y$  given  $\mathbf{w}' = (\mathbf{x}', \mathbf{v}')$  by the relation

$$H^*(\mathbf{w}) = \Gamma(\mathbf{x}'\beta_0, \mathbf{v}), \quad (3.7)$$

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

$$\mathbf{x}'\beta_0 - \psi(g(\mathbf{w}), \mathbf{v}) = 0, \text{ (w.p.1)}, \quad (3.8)$$

where

$$\psi(\cdot, \mathbf{v}) \equiv \Gamma^{-1}(\cdot, \mathbf{v}) \quad (3.9)$$

i.e.,  $\Gamma(\psi(g, \mathbf{v}), \mathbf{v}) \equiv g$ . So, if two observations (with subscripts  $i$  and  $j$ ) have identical conditional means (i.e.,  $H^*(\mathbf{w}_i) = H^*(\mathbf{w}_j)$ ) and identical reduced form error terms ( $\mathbf{v}_i = \mathbf{v}_j$ ), it follows from the assumed invertibility of  $\Gamma$  that their indices  $\mathbf{x}_i\beta_0$  and  $\mathbf{x}_j\beta_0$  are also identical:

$$\begin{aligned} (\mathbf{x}_i - \mathbf{x}_j)\beta_0 &= \boldsymbol{\psi}(g(\mathbf{w}_i), \mathbf{v}_i) - \boldsymbol{\psi}(g(\mathbf{w}_j), \mathbf{v}_j) = \mathbf{0} \text{ if} \\ g(\mathbf{w}_i) &= g(\mathbf{w}_j) \quad \text{and} \quad \mathbf{v}_i = \mathbf{v}_j. \end{aligned} \quad (3.10)$$

For any nonnegative function  $\omega(\mathbf{w}_i, \mathbf{w}_j)$  of the conditioning variables  $\mathbf{w}_i$  and

---

<sup>7</sup>See Horowitz (1993) and Powell (1994).

$\mathbf{w}_j$ , it follows that

$$\begin{aligned} 0 &= E[\omega(\mathbf{w}_i, \mathbf{w}_j) \cdot ((\mathbf{x}_i - \mathbf{x}_j)\beta_0)^2 \mid g(\mathbf{w}_i) = g(\mathbf{w}_j), \mathbf{v}_i = \mathbf{v}_j] \\ &\equiv \beta_0' \Sigma_w \beta_0, \end{aligned} \quad (3.11)$$

where

$$\Sigma_w \equiv E[\omega(\mathbf{w}_i, \mathbf{w}_j) \cdot (\mathbf{x}_i - \mathbf{x}_j)'(\mathbf{x}_i - \mathbf{x}_j) \mid g(\mathbf{w}_i) = g(\mathbf{w}_j), \mathbf{v}_i = \mathbf{v}_j]. \quad (3.12)$$

That is, the nonnegative-definite matrix  $\Sigma_w$  is singular, and, under the identifying assumption that  $\Sigma_w$  has rank  $k - 1 = \dim(\mathbf{w})$  – which requires that any nontrivial linear combination  $(\mathbf{x}_i - \mathbf{x}_j)a$  of the difference in regressors has nonzero variance when  $a \neq 0$ ,  $a \neq \beta_0$ ,  $\mathbf{v}_i = \mathbf{v}_j$ , and  $(\mathbf{x}_i - \mathbf{x}_j)\beta_0 = 0$  – the unknown parameter vector  $\beta_0$  is the eigenvector (with an appropriate normalization) corresponding to the unique zero eigenvalue of  $\Sigma_w$ .

Given the preliminary nonparametric estimators  $\hat{\mathbf{v}}_i$  and  $\hat{g}(\tilde{\mathbf{w}}_i)$  of  $\mathbf{v}_i$  and  $g(\mathbf{w}_i)$  defined above, and assuming smoothness (continuity and differentiability) of the inverse function  $\psi(\cdot)$  in (3.9), a consistent estimator of  $\Sigma_w$  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  $|\hat{g}(\tilde{\mathbf{w}}_i) - \hat{g}(\tilde{\mathbf{w}}_j)|$  and  $|\mathbf{v}_i - \mathbf{v}_j|$  increase. The details of this semiparametric estimation procedure for  $\beta$  are developed in Blundell and Powell (1999), who demonstrate consistency of the resulting estimator  $\hat{\beta}$  and characterize the form of its asymptotic (normal) distribution.

### 3.1.2. *The partial-mean estimator of the ASF*

Once the consistent estimator  $\hat{\beta}$  of  $\beta$  is obtained, the remaining parameter of interest for this model is  $G(\mathbf{x}'\beta)$ , the marginal probability that  $y_{1i} = 1$  given

an exogenous  $\mathbf{x}$ . The conditional cumulative distribution function  $F_{-u|\mathbf{v}}(\mathbf{x}'\boldsymbol{\beta}|\mathbf{v}) \equiv \Gamma(\mathbf{x}'\boldsymbol{\beta}, \mathbf{v})$  is first estimated using a kernel regression estimator  $\hat{E}[y|\mathbf{x}'\hat{\boldsymbol{\beta}}, \hat{\mathbf{v}}]$ ; the Blundell-Powell approach then estimates  $G(\bar{\lambda})$  from the sample average of  $\hat{\Gamma}(\bar{\lambda}, \hat{\mathbf{v}}_i)$ ,

$$\hat{G}(\bar{\lambda}) = \sum_{i=1}^n \hat{\Gamma}(\bar{\lambda}, \hat{\mathbf{v}}_i) \tau_i \quad (3.13)$$

where  $r_i$  is some “trimming” term which downweights observations for which  $\Gamma$  is imprecisely estimated. Consistency of this approach requires adapting the arguments in Newey (1994b) and Linton and Nielson (1995) for the case of the averaging over the estimated residual  $\hat{\mathbf{v}}_i$ .

#### 4. Coherency and Alternative Simultaneous Representations

One interpretation of the linear index binary response model described in section 3 is as the “triangular form” of some underlying joint decision problem. For simplicity, suppose the explanatory variables  $\mathbf{x}$  can be partitioned as

$$\mathbf{x}' = (\mathbf{z}'_1, y_2), \quad (4.1)$$

where  $y_2$  is a single continuously distributed endogenous regressor; also, suppose the instrument vector  $\mathbf{z}$  is also partitioned into subvectors corresponding to the “included” and “excluded” components of  $\mathbf{x}$ ,

$$\mathbf{z}' = (\mathbf{z}'_1, \mathbf{z}'_2). \quad (4.2)$$

Then, for a random sample  $\{y_i, \mathbf{x}_i, \mathbf{z}_i\}_{i=1}^n$  of observations on  $y \equiv y_1$ ,  $\mathbf{x}$ , and  $\mathbf{z}$ , we can express the model (3.1) as

$$y_{1i} = 1\{y_{1i}^* > 0\}, \quad (4.3)$$

for a latent dependent variable  $y_{1i}^*$  of the form

$$y_{1i}^* = \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + y_{2i}\beta_2 + u_i. \quad (4.4)$$

If the “simultaneity” between  $y_{2i}$  and  $y_{1i}$  can be written in terms of a structural equation for  $y_{2i}$  in terms of the *latent* variable  $y_{1i}^*$ , i.e.,

$$y_{2i} = \mathbf{z}'_{2i}\boldsymbol{\gamma}_1 + y_{1i}^*\boldsymbol{\gamma}_2 + \varepsilon_i, \quad (4.5)$$

for some error term  $\varepsilon_i$ , then substitution of (4.4) in (4.5) delivers the first-stage regression model

$$y_{2i} = \mathbf{z}'_i\Pi + \mathbf{v}_i \quad (4.6)$$

for some coefficient matrix  $\Pi$ . This “triangular” structure has  $y_2$  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 (4.4) and (4.5). For example, consider the joint determination of savings (or consumption) and labor market participation. Let  $y_1$  denote the discrete work decision and let  $y_2$  denote other income including savings. Suppose that work involves a fixed cost  $\alpha_1$ . In this case the structural relationship for other income ( $y_{2i}$ ) will depend on the discrete employment decision ( $y_{1i}$ ), *not* the latent variable ( $y_{1i}^*$ ). It will not be possible therefore to solve explicitly the reduced form for  $y_2$ . Note that, for theoretical consistency, the fixed cost  $\alpha_2$  will also have to be subtracted from the income (or consumption) variable in the participation equation for those in work. So that for those employment other income is defined net of fixed costs

$$\tilde{y}_{2i} \equiv y_{2i} - \alpha_2 y_{1i}. \quad (4.7)$$

We may therefore wish to replace (4.5) with a model incorporating feedback between the *observed* dependent variable  $y_1$  and  $y_2$

$$y_{2i} = \mathbf{z}'_{2i}\gamma_1 + y_{1i}\alpha_2 + \varepsilon_i; \quad (4.8)$$

that is, the realization  $y_1 = 1$  results in a discrete shift  $y_{1i}\alpha_2$  in other income. Due to the nonlinearity in the binary response rule (4.3) there is no explicit reduced form for this system. Indeed, Heckman (1978), in his extensive analysis of simultaneous models with dummy endogenous variables, shows that (4.3), (4.4) and (4.8) is only a statistically “coherent” system, i.e. one that processes a unique (if not explicit) reduced form, when  $\gamma_2 = 0$ .<sup>8</sup>

To provide a fully simultaneous system in terms of observed outcomes, and one that is also coherent, Heckman (1978) further shows that there must be a structural jump in the equation for  $y_{1i}^*$

$$y_{1i}^* = y_{1i}\alpha_1 + \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + y_{2i}\beta_2 + u_i, \quad (4.9)$$

with the added restriction

$$\alpha_1 + \alpha_2\beta_2 = 0 \quad (4.10)$$

This Heckman (1978) labels the Principle Assumption. To derive this condition notice that from (4.3), (4.8) and (4.9) we can write

$$y_{1i}^* = 1\{y_{1i}^* > 0\}(\alpha_1 + \alpha_2\beta_2) + \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \mathbf{z}'_{2i}\gamma_1\beta_2 + u_i + \varepsilon_i\beta_2, \quad (4.11)$$

or

$$y_{1i}^* \leq 0 \Leftrightarrow 1\{y_{1i}^* > 0\}(\alpha_1 + \alpha_2\beta_2) + \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \mathbf{z}'_{2i}\gamma_1\beta_2 + u_i + \varepsilon_i\beta_2 \leq 0. \quad (4.12)$$

---

<sup>8</sup>See Gourieroux, Laffont and Monfort (1980) for further discussion of coherency conditions and Lewbel (1999b) for a recent statement of this result.

Thus for a consistent probability model with general distributions for the unobservables and exogenous covariates we require the coherency condition (4.10).

Substituting for  $\alpha_1$  from (4.10) into (4.9) we have

$$y_{1i}^* = (y_{2i} - y_{1i}\alpha_2)\beta_2 + \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + u_i. \quad (4.13)$$

Note that this adjustment to  $y_{2i}$  which guarantees statistical coherency, is *identical* to the condition for theoretical consistency in the fixed cost model in which fixed cost  $\alpha_2$  is removed from other income for those who participate, as in (4.7) above.

Blundell and Smith (1994) derive a control function like estimator for this set up under joint normality assumptions.<sup>9</sup> But the semiparametric approach developed in the previous section naturally extends to this case. Noting that  $\tilde{y}_{2i} \equiv y_{2i} - \alpha_2 y_{1i}$ , the coherency condition implies that the model can be rewritten

$$y_{1i} = 1\{\mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \tilde{y}_{2i}\beta_2 + u_i > 0\} \quad (4.14)$$

and

$$\tilde{y}_{2i} = \mathbf{z}'_{2i}\boldsymbol{\gamma}_2 + \varepsilon_i. \quad (4.15)$$

This specification could be easily be generalized to allow for a more complex relationship in more complicated models of nonseparable decision making.<sup>10</sup>

---

<sup>9</sup>Blundell and Smith (1986) also develop an exogeneity test based on this estimator and consider results for the equivalent Tobit model. Rivers and Vuong (1988) provide a comprehensive treatment of limited information estimators for this class of parametric limited dependent variable models. They label the Blundell-Smith estimator Two-Stage Conditional Maximum Likelihood (2SCML) and consider alternative LIML estimators. The efficiency and small sample properties of the 2SCML estimator are also considered. These are further refined in Blundell and Smith (1989). See also the important earlier related work of Amemiya (1978) and Lee (1981, 1993) which builds on the Heckman (1978) estimator.

<sup>10</sup>Note that to test this alternative specification against the triangular specification (4.3), (4.4) and (4.5) one may estimate

$$y_{2i} = \mathbf{z}_{2i}\boldsymbol{\gamma}_1 + y_{1i}\alpha_2 + \hat{y}_{2i}\delta_2 + w_i$$

by instrumental variables using  $\mathbf{z}_i$  as instruments, and then test the null hypothesis  $\delta_2 = 0$ , where  $\hat{y}_{2i}$  is the prediction of  $y_{2i}$  under reduced form specification (4.6).

If  $\alpha_2$  were known then the equations (4.15) and (4.14) are analogous to (4.3), (4.4) and (4.6). Consequently, a semiparametric estimator using the control function approach would simply apply the estimation approach described in this paper to the conditional model. Following the previous discussion, assumption (2.42) would be replaced by the modified conditional independence restrictions

$$\begin{aligned} u|\mathbf{z}_1, y_2, \mathbf{z}_2 &\sim u|\mathbf{z}_1, \tilde{y}_2, \varepsilon \\ &\sim u|\varepsilon. \end{aligned} \tag{4.16}$$

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

$$\begin{aligned} E[y_1|\mathbf{z}_1, \tilde{y}_{2i}, \varepsilon] &= \Pr[-u \leq \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \tilde{y}_{2i}\beta_2|\mathbf{z}_1, \tilde{y}_{2i}, \varepsilon] \\ &= F_{-u|\varepsilon}(\mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \tilde{y}_{2i}\beta_2|\varepsilon) \\ &\equiv \Gamma(\mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \tilde{y}_{2i}\beta_2, \varepsilon). \end{aligned} \tag{4.17}$$

Finally, note that although  $\alpha_2$  is unknown, given sufficient exclusion restrictions on  $\mathbf{z}_{2i}$ , a root- $n$  consistent estimator for  $\alpha_2$  can be recovered from (linear) 2SLS estimation of (4.8). More generally, if the linear form  $\mathbf{z}'_{2i}\gamma_1$  of the regression function for  $y_2$  is replaced by a nonparametric form  $\gamma(\mathbf{z}_{2i})$  for some unknown (smooth) function  $\gamma$ , then a  $\sqrt{n}$ -consistent estimator of  $\alpha_2$  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. An Application

The empirical application presented here is taken from the Blundell and Powell (1999) study. In that paper we considered the participation in work 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. Largely as a consequence of the low participation rate this group is subject to much policy discussion. We model the participation decision ( $y_1$ ) in terms of a simple structural 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 benefit rules, the demographic structure of the family, the geographic location and housing costs. 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 it will be a determinant of the reduced form for participation and other income, for the structural model below, it should not enter the participation decision conditional on the inclusion of other income variable.



## 5.1. The Data

The sample consists of married couples 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<sup>11</sup> is used to construct a simulated budget constraint for each individual family given information about age, location, 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<sup>12</sup>, housing benefits, child

---

<sup>11</sup>The IFS tax and benefit simulation model TAXBEN ([www.ifs.org.uk](http://www.ifs.org.uk)), designed for the British Family Expenditure Survey data used in this paper. For an extensive discussion of the use of this data source in the study of male participation, see Blundell, Reed and Stoker (1999).

<sup>12</sup>Unemployment benefit included an earnings-related supplement in 1979, but this was abol-

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 5.1: Descriptive Statistics.

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

**number of observations:** 1606

After making the sample selections described above, our sample contains 1606 observations. A brief summary of the data is provided in Table 5.1.<sup>13</sup> 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 5.1.

---

ished in 1980.

<sup>13</sup>See Blundell and Powell (1999) for further details.

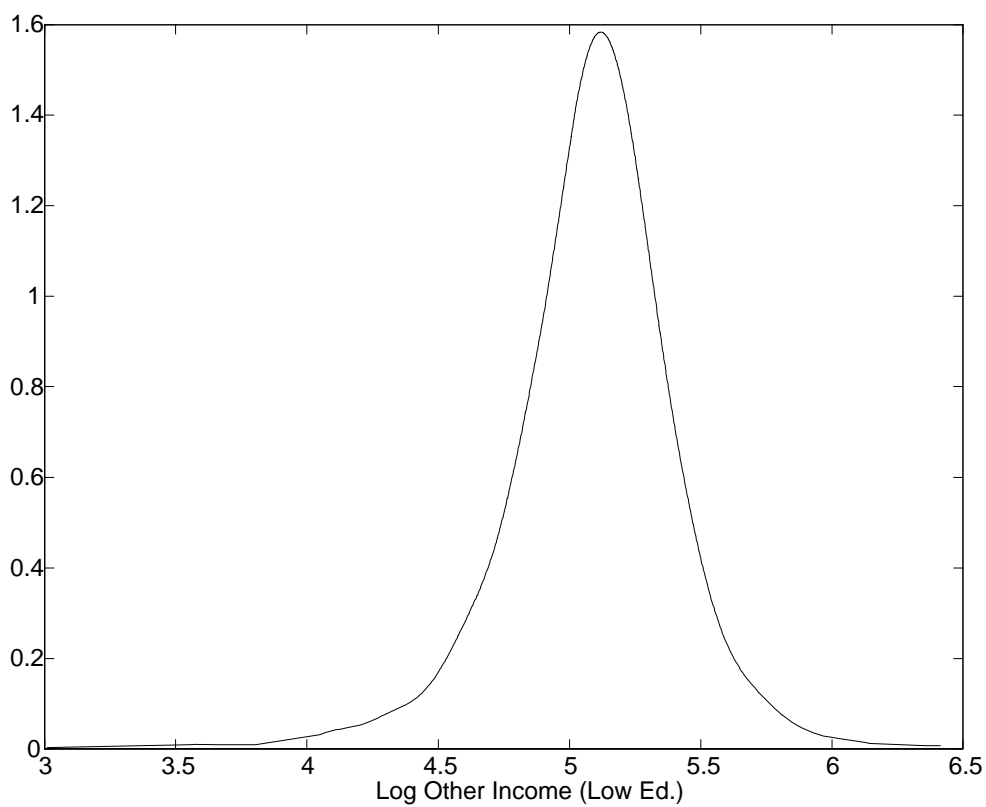


Figure 5.1: Density of Log Other Income for Low Education Sub-Sample

## 5.2. A Model of Participation in Work and Other Family Income

To motivate the 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

$$h_i^* = \delta_0 + \delta_1' \mathbf{z}_{1i} + \delta_2 \ln w_i + \delta_3 \ln \mu_i + \zeta_i \quad (5.1)$$

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  $\zeta_i$  is some unobservable heterogeneity. As  $\ln w_i$  is unobserved for nonparticipants we replace it in (5.1) by the wage equation

$$\ln w_i = \theta_0 + \theta_1' \mathbf{z}_{1i} + \omega_i \quad (5.2)$$

where  $\mathbf{z}_{1i}$  now contains the education level for individual  $i$  as well as other determinants of the market wage. Labor supply (5.1) becomes

$$h_i^* = \phi_0 + \phi_1' \mathbf{z}_{1i} + \phi_2 \ln \mu_i + \nu_i. \quad (5.3)$$

Participation in work occurs according to the binary indicator

$$y_{1i} = 1\{h_i^* > h_i^0\} \quad (5.4)$$

where

$$h_i^0 = \gamma_0 + \gamma_1' \mathbf{z}_{1i} + \xi_i \quad (5.5)$$

is some measure of reservation hours.

Combining these equations, participation is now described by

$$y_{1i} = 1\{\phi_0 + \phi_1' \mathbf{z}_{1i} + \phi_2 \ln \mu_i + \nu_i > \gamma_0 + \gamma_1' \mathbf{z}_{1i} + \xi_i\} \quad (5.6)$$

$$= 1\{\beta_0 + \beta_1' \mathbf{z}_{1i} + \beta_2' y_{2i} + u_i > 0\} \quad (5.7)$$

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

$$\begin{aligned} y_{2i} &= E[y_{2i}|\mathbf{z}_i] + v_i \\ &= \Pi(\mathbf{z}_i) + v_i \end{aligned} \tag{5.8}$$

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  $z_{21i}$ ) and the education level of the spouse ( $z_{22i}$ ).

### 5.3. Empirical Results

In Table 5.2 we present the empirical results for the joint normal - simultaneous Probit - model. This consists of a linear reduced form for the log other income variable and a conditional Probit specification for the participation decision. 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 results show a strong role for the benefit income variable in the determination of other income.

The first column of Probit results 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 5.2 is the coefficient normalized by the education coefficient for comparability with the results from the semiparametric specification to be presented below. The impact of adjusting for endogeneity is quite dramatic. The income coeffi-

cient is now considerably larger and quite significant. The estimated education coefficient remains positive and significant.

Table 5.2: Results for the Simultaneous Probit Specification.

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

Table 5.3 presents the semiparametric estimation results for the linear index coefficients. Bandwidths were chosen according to the  $1.06\sigma_z n^{-\frac{1}{5}}$  rule (see Silverman(1986)).<sup>14</sup> The education coefficient in the binary response specification is normalized to unity and so the  $\beta_1$  estimates in Table 5.3 correspond to the ratio of the other income to the education coefficients. We present the standard Probit results in Table 5.2 for comparison (the mean of the bootstrap estimates for the education coefficient was .989 and for the income coefficient it was -.327). The bootstrap figures relate to 500 bootstrap samples of size  $n$  (=1606 ); the standard errors for the semiparametric methods are computed from a standardized

<sup>14</sup>Blundell and Powell (1999) provide results for a similar model specification and also present sensitivity results for this bandwidth choice. In particular, sensitivity to the choice of a smaller bandwidth is investigated and found not to change the overall results.

interquartile range for the bootstrap distribution, and are calculated using the usual asymptotic formulae for the Probit and linear probability estimators.

Table 5.3: Semiparametric Results, Parametric Results and Bootstrap Distributions.

<i>Specification</i>	$\widehat{\beta}_2$	$\sigma_{\widehat{\beta}_2}$	10%	25%	50%	75%	90%
<b>Semi-P</b> (with $\widehat{v}$ )	-2.2590	0.5621	-4.3299	-3.6879	-2.3275	-1.4643	-1.0101
<b>Semi-P</b> (without $\widehat{v}$ )	-0.1871	0.0812	-0.2768	-0.2291	-0.1728	-0.1027	-0.0675
<b>Probit</b> (with $\widehat{v}$ )	-2.8376	.5124	-3.8124	-3.3304	-2.9167	-2.4451	-1.8487
<b>Probit</b> (without $\widehat{v}$ )	-0.3364	.1293	-0.4989	-0.4045	-0.3354	-0.2672	-0.1991
<b>Lin.Prob.</b> (with $\widehat{v}$ )	-3.1241	.4679	-3.8451	-3.3811	-3.1422	-2.8998	-2.5425
<b>Lin.Prob.</b> (without $\widehat{v}$ )	-0.4199	.1486	-0.6898	-0.5643	-0.4012	-0.3132	-0.2412

Figure 5.2 graphs the estimate of the Average Structural Function ASF,  $G(\mathbf{x}'\boldsymbol{\beta})$ , derived from the semiparametric estimation with and without controls for the endogeneity of log other income. These plots cover the 5%-95% range of the log other income distribution for the lower education group.

In Figure 5.3 we compare these semiparametric results with the results of estimating  $G(\mathbf{x}'\boldsymbol{\beta})$  using the Probit and linear probability models. This data set is likely to be a particularly good source on which to carry out this evaluation. First, we know that the correction for endogeneity induces a large change in the  $\beta$  coefficients. Second, the proportion participating in the sample is around 85%

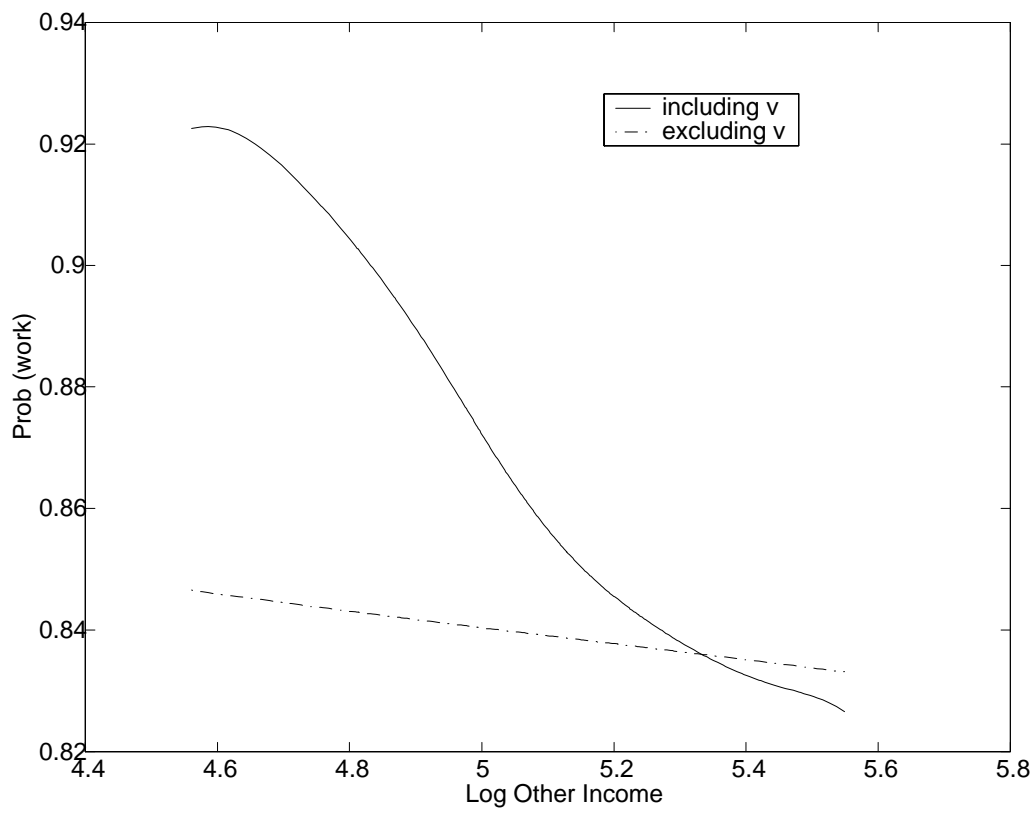


Figure 5.2: Semiparametric Model with and without Controls for Endogeneity



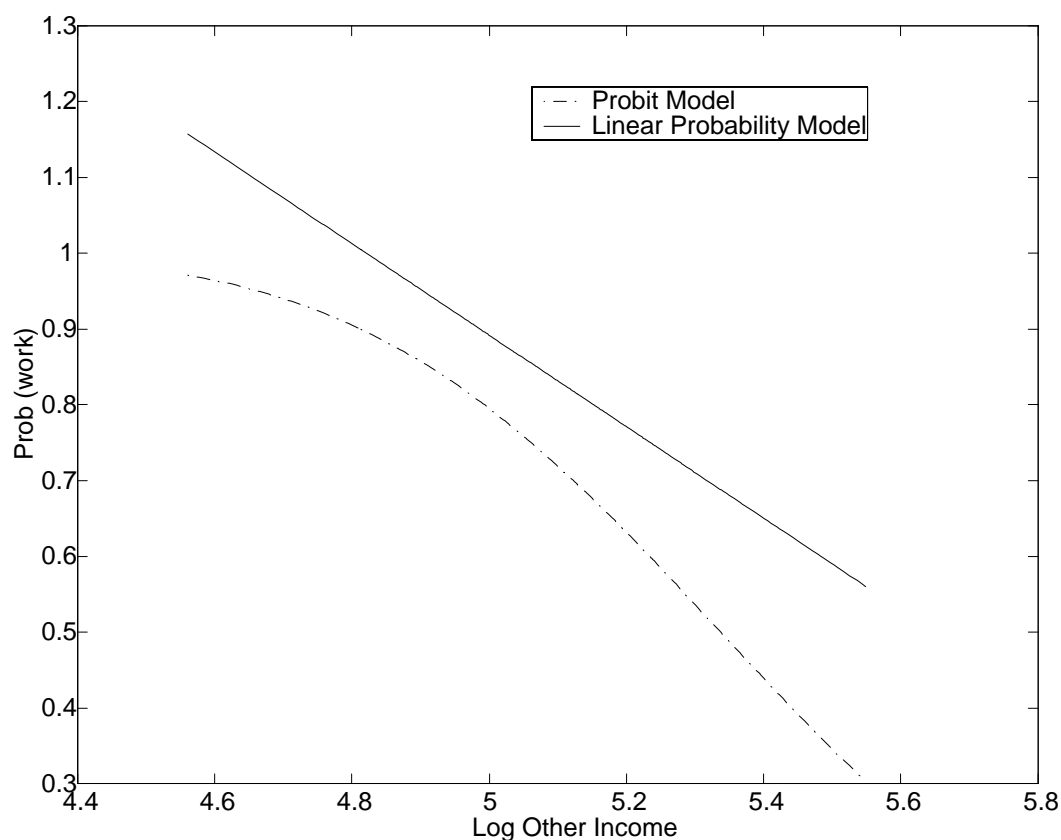


Figure 5.3: Linear Probability and Probit Models with Controls for Endogeneity

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. They show considerable sensitivity of the estimated  $G(\mathbf{x}'\boldsymbol{\beta})$ , after allowing for endogeneity, across these alternative parametric models. Both the linear probability model and the Probit model estimates result in estimated probabilities that are very different from those implied by the semiparametric approach. Figure 5.3 shows this most dramatically. 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.

This example points to the attractiveness of the semiparametric approach de-

veloped 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 in comparison to those from the parametric counterparts.<sup>15</sup> The results show a strong effect of correcting for endogeneity and indicate that adjusting for endogeneity using the 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. This was highlighted in Figure 5.3 where it was shown that the bias correction for endogeneity in the Linear Probability Model was sufficient to produce predicted probabilities larger than unity over a large range of the income distribution. The Probit model did not fair much better. The semiparametric approach showed a strong downward bias in the estimated income responses when endogeneity of other income was ignored. The corrected semiparametric estimates appear plausible and, although there were no shape restrictions imposed, the estimated average structural function was monotonically declining in other income over the large range of the income distribution.

#### 5.4. The Coherency Model

In Blundell and Powell (1999) we also use this application to assess the alternative ‘coherency’ model of Section 4 in which participation itself directly enters the equation determining other income. We interpret this as a fixed cost model in which other income is dependent on whether or not fixed costs are paid, which in turn depends on participation. In this case no explicit reduced form for  $y_2$ , other

---

<sup>15</sup>In Blundell and Powell (2000) we present the analogous analysis using the low education sub-sample only. 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*. The results show a slightly shallower slope.

income, exists. The model for other income may be written

$$y_{2i} = \gamma_0 + y_{1i}\alpha_2 + \mathbf{z}'_{2i}\gamma_1 + \varepsilon_i. \quad (5.9)$$

where we are assuming that  $y_2$  relates to the *level* of other income. Participation is now described by

$$y_{1i} = 1\{\beta_0 + y_{1i}\alpha_1 + \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + y_{2i}\beta_2 + u_i > 0\}$$

with the added coherency restriction

$$\alpha_1 + \beta_2\alpha_2 = 0. \quad (5.10)$$

Together these imply

$$y_{1i} = 1\{\beta_0 + \tilde{y}_{2i}\beta_2 + \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + u_i > 0\} \quad (5.11)$$

with

$$\tilde{y}_{2i} = (y_{2i} - y_{1i}\alpha_2).$$

where we note that this fixed cost adjustment to other income  $y_{2i}$  guarantees statistical coherency.

From the discussion in section 4 we note that the conditional expectation of the binary response variable  $y_{1i}$ , given the regressors  $\mathbf{z}_1$ ,  $\tilde{y}_{2i}$  and errors  $\varepsilon$ , may be expressed as

$$\begin{aligned} E[y_1 | \mathbf{z}_1, \tilde{y}_{2i}, \varepsilon] &= \Pr[-u \leq \mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \tilde{y}_{2i}\beta_2 | \mathbf{z}_1, \tilde{y}_{2i}, \varepsilon] \\ &= F(\mathbf{z}'_{1i}\boldsymbol{\beta}_1 + \tilde{y}_{2i}\beta_2, \varepsilon). \end{aligned} \quad (5.12)$$

Provided  $\tilde{y}_{2i}$  and  $\varepsilon_i$  can be measured, estimation follows the same procedure as in the triangular case.

The first column of Table 5.4 presents the estimates of the parameters of the structural equation for  $y_2$  (5.9) in this coherency specification. These are recovered from instrumental variables estimation using the education of the husband as an excluded variable. The estimated “fixed cost of work” parameter seems reasonable, recall that the income variable has a mean of approximately £165 per week. The two sets of Probit results differ according to whether or not they control for  $\varepsilon$ . Notice that having removed the direct simultaneity of  $y_1$  on  $y_2$  through the adjustment  $\tilde{y}_2$ , 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  $\varepsilon$  leads to a downward correction to the income coefficient.

Table 5.4: Results for the Coherency Specification.

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

The comparable results for the semiparametric specification are presented in Table 5.5. In these the linear structural model estimates for the  $y_2$  equation are used exactly as in Table 5.4. They show a very similar pattern with only a small

Table 5.5: Semiparametric Results for the Coherency Specification.

<i>Variable</i>	Semi-P		Semi-P	
	Pr[Work] <i>Coeff.</i>	Std. <i>Err.</i>	Pr[Work  $\varepsilon$ ] <i>Coeff.</i>	Std. <i>Err.</i>
<b>adjusted income: <math>\tilde{y}_2</math></b>	-1.009	.0689	-0.82256	.2592

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

## 6. Summary and Conclusions

This paper has considered nonparametric and semiparametric methods for estimating regression models of the form  $y = H(\mathbf{x}, u)$  where the  $\mathbf{x}$  contain continuous endogenous regressors and where  $u$  represent unobserved and heterogeneity. It has been assumed that there exist a set of instrumental variables  $\mathbf{z}$  with  $\dim(\mathbf{z}) \geq \dim(\mathbf{x})$ . This general specification was shown to cover a number of non-linear models of interest in econometrics. The leading cases we considered were additive nonparametric specifications  $y = g(\mathbf{x}) + u$ , in which  $g(\mathbf{x})$  is unknown, and nonadditive models  $y = H(g(\mathbf{x}), u)$ , in which  $g(\mathbf{x})$  is unknown but  $H$  is a known function that is monotone but not invertible. An important example of the latter, and one which we used as an empirical illustration, is the binary response model with endogenous regressors. We have focused on identification and estimation in these leading nonparametric regression models and have defined the parameter of interest to be the *average structural function (ASF)*,  $G(\mathbf{x}) \equiv \int H(\mathbf{x}, u) dF_u$ , where the average is taken over the *marginal* distribution of the error terms  $u$  and where  $F_u$  denotes the marginal cumulative distribution function of  $u$ .

In each of these leading cases, and their semiparametric variants, we have considered how three common estimation approaches for linear equations – the “instrumental variables,” “fitted value,” and “control function” approaches – may or may not be applicable. In the case where  $H$  and  $g$  are linear, *iid* distributed errors the covariance restriction  $E(\mathbf{z}u) = 0$  and the rank condition are sufficient to guarantee identification and generate consistent and analytically identical estimators from each of these approaches. In the non-linear models considered here this is no longer the case.

In additive nonparametric specifications we have considered restrictions on the model specification which are sufficient to identify  $g(\mathbf{x})$ , the ASF in this case. The relationship between the reduced form  $E[y|\mathbf{z}]$  and the structural function  $g$  is given by  $E[y|\mathbf{z}] = \int g(\mathbf{x})dF_{\mathbf{x}|\mathbf{z}}$  where  $F_{\mathbf{x}|\mathbf{z}}$  is the conditional c.d.f. of  $\mathbf{x}$  given  $\mathbf{z}$ . Unlike in typical nonparametric estimation problems, identification of  $g$  faces an ill-posed inverse problem and consistent estimators of the components  $E[y|\mathbf{z}]$  and  $F_{\mathbf{x}|\mathbf{z}}$  are not, by themselves, sufficient for consistent instrumental variable estimation of  $g$ . We have reviewed and assessed a number of approaches to instrumental variable estimation that have been proposed in the literature to overcome this problem. For the nonadditive case instrumental variable estimation faces more severe difficulties. Without some further specific structure on  $H$ , such as invertibility, estimation by instrumental variables does not look hopeful. For our leading case in this nonadditive setting, the binary response model,  $H$  is not invertible.

Apart from some very specific cases, we have argued that the fitted value approach is not well suited to estimation of parameters of interest in these nonlinear models. However, the control function approach, has been shown to provide an attractive solution. This approach treats endogeneity as an omitted variable problem, where the inclusion of estimates of the first-stage errors  $\mathbf{v}$  as a covariate corrects the inconsistency in  $E(y|\mathbf{x})$ . It has been shown to extend naturally

under the conditional independence assumption that the distribution of  $u$  given  $\mathbf{x}$  and  $\mathbf{z}$  is the same as the conditional distribution of  $u$  given  $\mathbf{v}$ . This exclusion restriction permits replacement of the unidentified structural errors  $u$  with the identified control function  $\mathbf{v}$  through iterated expectations, so that averaging the structural function  $H$  over the marginal distribution of the structural errors  $u$  is equivalent to averaging the (identified) intermediate regression function of  $y$  on  $\mathbf{x}$  and  $\mathbf{v}$  over the marginal distribution of  $\mathbf{v}$ . We have derived a general approach to identification and estimation of the ASF  $G(\mathbf{x})$  by this control function approach and have highlighted the importance of support restrictions on the distribution of the endogenous components of  $\mathbf{x}$  and  $\mathbf{z}$ .

We have then considered the particular case of the linear index binary response model. In this semiparametric model we have described in detail how estimation of the parameters of interest can be constructed using the control function approach. 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 estimator, for this semiparametric model, can be easily adapted to the case where the model specification is not triangular and certain coherency conditions are required to be satisfied.

Finally, we have studied the response of labor force participation to nonlabor income, viewed as an endogenous regressor, using these techniques. The procedures we have developed appear to work well and suggest that the usual distributional assumptions underlying Probit and Linear Probability specifications could be highly misleading in binary response models with endogenous regressors. The application found relatively small reductions in precision from adopting the semiparametric approach. The semiparametric approach showed a strong down-

ward bias in the estimated income responses when endogeneity of other income was ignored. The corrected semiparametric estimates appeared plausible and, although there were no shape restrictions imposed, the estimated average structural function was monotonically declining in other income over the large range of the income distribution.

## References

- [1] Ai, C and X. Chen (2000) "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions", mimeo, LSE, July.
- [2] 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.
- [3] Ahn, H. and C.F. Manski (1993), "Distribution Theory for the Analysis of Binary Choice Under Uncertainty with Nonparametric Estimation of Expectations," *Journal of Econometrics*, 56, 291-321.
- [4] Ahn, H. and J.L. Powell (1993), "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3-29.
- [5] Ahn, H., Ichimura, H. and J.L. Powell (1996), "Simple Estimators for Monotone Index Models," manuscript, Department of Economics, U.C. Berkeley.
- [6] Altonji, J. and H. Ichimura (2000), "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables," manuscript, University College London, June (revision of March 1996 paper).
- [7] Altonji, J. and R.L. Matzkin (1997) "Panel Data Estimators for Nonseparable Models with Endogenous Regressors", manuscript, Northwestern University.
- [8] Amemiya, T. (1974), "The Nonlinear Two-Stage Least-Squares Estimator," *Journal of Econometrics*, 2, 105-110.
- [9] Amemiya, T. (1978), "The Estimation of a Simultaneous Equation Generalised Probit Model", *Econometrica*, 46, 1193-1205.



- [10] Andrews, D. (1994): “Asymptotics for Semi-parametric Econometric Models via Stochastic Equicontinuity”, *Econometrica*, 62, 43-72.
- [11] Angrist, J. (1999), “Estimation of Limited-Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice”, mimeo, MIT, October.
- [12] Angrist, J., G. Imbens and D.R. Rubin (1996), “Identification of Causal Effects Using Instrumental Variables”, *Journal of the American Statistical Association*, 91, 444-455.
- [13] Arellano, M. and B. Honoré (1999) “Panel Data Models: Some Recent Developments” prepared for J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, Volume 5, Elsevier-North Holland.
- [14] Basmann, R.L. (1959), “A Generalised Classical Method of Linear Estimation of Coefficients in a Structural Equation”, *Econometrica*, 25, 77-83.
- [15] Blundell, R., M. Browning and I. Crawford (2000): “Nonparametric Engel Curves and Revealed Preference”, UCL Working Paper.
- [16] Blundell, R.W. and A. Duncan, (1998), “Kernel Regression in Empirical Microeconomics”, *Journal of Human Resources: Special Issue on Empirical Microeconometrics*, Winter.
- [17] Blundell, R.W. and J.L.Powell (1999), “Endogeneity in Semiparametric Binary Response Models”, mimeo, Berkeley, July, reproduced as CeMMA Working Paper CWP05/01 <http://cemmap.ifs.org.uk/docs/cwp0501.pdf>
- [18] Blundell, R.W. and R.J. Smith (1986), “An Exogeneity test for a Simultaneous Tobit Model”, *Econometrica*, 54, 679-685.
- [19] 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.
- [20] Blundell, R.W. and R.J. Smith (1993), “Simultaneous Microeconomic Models with Censored or Qualitative Dependent Variables”, in *Handbook of Statistics*, eds. G.S. Maddala, C.R.Rao and H.D.Vinod, Vol. 11, Amsterdam: North Holland.

- [21] 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.
- [22] Blundell, R.W., H. Reed and T. Stoker, (1999), “Interpreting Movements in Aggregate Wages: the Role of Participation”, IFS Working Paper 99/13, May.
- [23] Chamberlain, G. (1987): “Asymptotic Efficiency in Estimation with Conditional Moment Restrictions,” *Journal of Econometrics*, 34, 305-334.
- [24] Chamberlain, G. (1992): “Efficiency Bounds for Semiparametric Regression,” *Econometrica*, 60, 567-596.
- [25] Chen, X. and X. Shen (1998): “Sieve Extremum Estimates for Weakly Dependent Data,” *Econometrica*, 66, 289-314.
- [26] Chen, X. and H. White (1998): “Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators”, *IEEE Tran. Information Theory*, 45, 682-691.
- [27] Chen, X., L.P. Hansen, and J. Scheinkman (1997): “Shape-preserving Estimation of Diffusions”, manuscript, University of Chicago, Dept. of Economics.
- [28] Dagenais, M.G. (1999), “Inconsistency of a proposed nonlinear instrumental variables estimator for probit and logit models with endogenous regressors”, *Economics Letters*, 63, 19-21.
- [29] Das, M. (1999), “Instrumental Variables Estimation of Models with Discrete Endogenous Regressors,” manuscript, Columbia University, Dept. of Economics.
- [30] Das, M., W.K. Newey and F. Vella (1998): “Non-parametric Estimation of the Sample Selection Model”, manuscript, MIT Dept. of Economics.
- [31] Darolles, S., J-P Florens and E. Renault, (2000), “Nonparametric Instrumental Regression”, mimeo, GREMAQ, University of Toulouse, April.
- [32] Dhrymes, P.J. (1970), *Econometrics: Statistical Foundations and Applications*, New York: Springer-Verlag.

- [33] Durbin, J. (1954), "Errors in Variables," *Review of the International Statistical Institute*, 22, 23-32.
- [34] Fenton, V. and A.R. Gallant (1996): "Convergence Rate of SNP Density Estimators," *Econometrica* 64, 719-727.
- [35] Ferguson, T.S. (1967): *Mathematical Statistics: A Decision Theoretic Approach*, New York, Krieger Publishing Company.
- [36] Gourieroux, C. J-J. Laffont and A. Montfort (1980), "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes", *Econometrica*, 48, 675-695.
- [37] Hammersley, J.M and D.C. Handscomb (1964), *Monte Carlo Methods*, London: Chapman-Hall.
- [38] Han, A K.(1987), "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35,.303-316.
- [39] Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.
- [40] Härdle W. (1990) *Applied Nonparametric Regression*, Cambridge: Cambridge University Press.
- [41] Härdle W. and Linton, O. (1995), "Nonparametric Regression analysis" in Engle and McFadden, *Handbook of Econometrics* Vol IV, North-Holland.
- [42] Heckman, J.J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models", *Annals of Economic and Social Measurement*, 5, 475-492.
- [43] Heckman, J.J. (1978), "Dummy Endogenous Variable in a Simultaneous Equations System", *Econometrica*, 46, 931-959.
- [44] Heckman, J.J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-162.
- [45] Heckman, J.J. (1997), "Instrumental Variables", *Journal of Human Resources*, 32, 441-462.

- [46] Heckman, J.J and B.E. Honoré (1990), “The Empirical Content of the Roy Model”, *Econometrica*, 58, 1121-1149.
- [47] Heckman, J.J., H. Ichimura, J. Smith and P. Todd (1998), “Characterizing Selection Bias Using Experimental Data”, *Econometrica*, 66, 1017-1018.
- [48] Heckman, J.J. and Robb (1985), “Alternative Methods for Evaluating the Impact of Interventions”, in Heckman and Singer (eds), *Longitudinal Analysis of Labor Market Data*, Econometric Society Monograph 10, Cambridge: Cambridge University Press.
- [49] Heckman, J.J. and E. Vytlacil (1999), “Local Instrumental Variables”, forthcoming in *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, C. Hsiao, K. Morimune and J. Powell (editors), Cambridge University Press.
- [50] Heckman, J.J. and E. Vytlacil (2000), “The relationship between Treatment Parameters with a Latent Variable Framework”, *Economic Letters*, 66, 33-39.
- [51] Honoré, B.E. and J.L. Powell (1994), “Pairwise Difference Estimators of Linear, Censored and Truncated Regression Models”, *Journal of Econometrics*, 64, 241-278.
- [52] Honoré, B.E. and J.L. Powell (1997), “Pairwise Difference Estimators for Nonlinear Models,” manuscript, Princeton University.
- [53] Horowitz, J.(1993) “.” in *Handbook of Statistics*, eds. G.S. Maddala, C.R.Rao and H.D.Vinod, Vol. 11, Amsterdam: North Holland.
- [54] Horowitz, J. (1996): “Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable,” *Econometrica* 64:1, 103-137.
- [55] Horowitz, J. (1998): *Semiparametric Methods in Econometrics*, Springer-Verlag.
- [56] Imbens, G.W. (2000), “Nonparametric Identification of Triangular Simultaneous Equation Models Without Additivity,” manuscript, University of California at Los Angeles, August.

- [57] Imbens, G.W. and J. Angrist, (1994), “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, 62, 467-476.
- [58] Ichimura, H. (1993), “Local Quantile Regression Estimation of Binary Response Models with Conditional Heteroscedasticity”, manuscript, University of Minnesota, December.
- [59] Ichimura, H. (1993): “Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models,” *Journal of Econometrics* 58, 71-120.
- [60] Ichimura, H. and L.F. Lee (1991): “Semiparametric Least Squares Estimation of Multiple Models: Single Equation Estimation,” in *Nonparametric and Semiparametric Models in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen, Cambridge University Press.
- [61] Klein, R.W. and R.H. Spady (1993), “An Efficient Semiparametric Estimator for Discrete Choice Models”, *Econometrica*, 61, 387-421.
- [62] Lee, L-F, (1981), “Simultaneous Equation Models with Discrete and Censored Dependent Variables”, in: C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Economic Applications*, (MIT Press Cambridge MA).
- [63] Lee, L-F, (1993) “Simultaneous Equation Models with Discrete and Censored Dependent Variables”, in *Handbook of Statistics*, eds. G.S. Maddala, C.R.Rao and H.D.Vinod, Vol. 11, Amsterdam: North Holland.
- [64] Lewbel, Arthur (1998), “Semiparametric Latent Variable Estimation with Endogenous or Mismeasured Regressors”, *Econometrica*, 66, 105-121.
- [65] Lewbel, Arthur (1999a), “Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables”, forthcoming *Journal of Econometrics*.
- [66] Lewbel, Arthur, (1999b), “Coherent Specification of Simultaneous Systems Containing a Binary Choice Equation”, mimeo, Boston College, July.
- [67] Linton, O. and J.P. Nielson, (1995), “A Kernel Method of Estimating Nonparametric Structured Regression based on a Marginal Distribution”, *Biometrika*, 82, 93-100.

- [68] Maddala, G.S. (1983), *Limited -Dependent and Qualitative Variables in Econometrics*, Cambridge: Cambridge University Press.
- [69] Matzkin, R.L. (1991), "A Nonparametric Maximum Rank Correlation Estimator," in *Nonparametric and Semiparametric Methods in Economics and Statistics*, edited by W.A. Barnett, J.L. Powell, and G. E. Tauchen, Cambridge: Cambridge University Press.
- [70] Matzkin, R.L. (1992), "Nonparametric and Distribution-Free Estimation of the Binary Choice and Threshold-Crossing Models," *Econometrica*, 60, 239-270.
- [71] Matzkin, R.L. (1994), "Restrictions of Economic Theory in Nonparametric Methods", in *Handbook of Econometrics: Vol IV*, edited by R.F. Engle and D.L. McFadden, Elsevier Science, North Holland.
- [72] Nelson, F. and L. Olsen (1978), "Specification and estimation in a simultaneous equation model with limited dependent variables", *International Economic Review*, 19, 695-705.
- [73] Newey, W.K. (1985), "Semiparametric Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Annals de l'INSEE*, 59/60, 219-237.
- [74] Newey, W.K. (1987), "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables", *Journal of Econometrics*, 32, 127-237.
- [75] Newey, W.K. (1990): "Efficient Instrumental Variables Estimation of Non-linear Models," *Econometrica*, 58, 809-837.
- [76] Newey, W.K. (1993): "Efficient Estimation of Models with Conditional Moment Restrictions, in *Handbook of Statistics*, Vol. 11, G.S. Maddala, C.R. Rao, and H.D. Vinod, eds., Amsterdam: North-Holland.
- [77] Newey, W.K. (1994a): "The Asymptotic Variance of Semiparametric Estimators", *Econometrica* 62, 1349-1382.
- [78] Newey, W.K. (1994b), "Kernel Estimation of Partial Means and a General Variance Estimator", *Econometric Theory*, 10, 233-253.

- [79] Newey, W.K. (1997): “Convergence Rates and Asymptotic Normality for Series Estimators”, *Journal of Econometrics* 79, 147-168.
- [80] Newey, W.K. and J.L Powell (1998): “Nonparametric Instrumental Variables Estimation”, MIT Working Paper.
- [81] Newey, W.K., J.L. Powell, and F. Vella (1999): “Nonparametric Estimation of Triangular Simultaneous Equations Models”, *Econometrica* 67, 565-603.
- [82] Ng, S. and J. Pinkse, (1995), “Nonparametric Two Step Estimation of Unknown Regression Functions when the Regressors and the Regressor Error are not Independent”, manuscript, University of Montreal.
- [83] O’Sullivan, F., (1986), “Ill-Posed Inverse Problems (with Discussion),” *Statistical Science*, 4, 503-527.
- [84] Pakes, A. and S. Olley (1995): “A Limit Theorem for A Smooth Class of Semiparametric Estimators”, *Journal of Econometrics* 65, 295-332.
- [85] Pagan, A.R. (1986), “Two Stage and Related Estimators and Their Applications”, *Review of Economic Studies*, 53, 513-538.
- [86] Pinkse, Joris (2000), ”Nonparametric Two-Step Regression Estimation When Regressors and Error are Dependent”, *Canadian Journal of Statistics* 28-2, pp. 289-300.
- [87] Powell, J.L. (1998), “Semiparametric Estimation of Censored Selection Models,” forthcoming in *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, edited by C. Hsiao, K.Morimune and J.L. Powell, Cambridge University Press.
- [88] Powell, J.L. (1994), “Estimation of Semiparametric Models”, in *Handbook of Econometrics: Vol IV*, edited by R.F. Engle and D.L. McFadden, Elsevier Science, North Holland.
- [89] Powell, J., J. Stock, and T. Stoker (1989): “Semiparametric Estimation of Index Coefficients”, *Econometrica* 57, 1403-1430.
- [90] Rivers, D. and Q.H.Vuong, (1988), “Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models”, *Journal of Econometrics*, 39, 347-366.

- [91] Robinson, P. (1987): "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of unknown Form," *Econometrica*, 55, 875-892.
- [92] Robinson, P. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica* 56, 931-954.
- [93] Robinson, P. (1991): "Best Nonlinear Three-Stage Least Squares Estimation of Certain Econometric Models," *Econometrica* 59, 755-786.
- [94] Roehrig, C.S. (1988), "Conditions for Identification in Nonparametric and Parametric Models", *Econometrica*, 55, 875-891.
- [95] Sargan J.D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables", *Econometrica*, 26, 393-415.
- [96] Shen, X. (1997): "On Methods of Sieves and Penalization," *The Annals of Statistics* 25(6), 2555-2591.
- [97] Silverman, B. (1986): *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- [98] Stoker, T.M. (1986), "Consistent Estimation of Scaled Coefficients", *Econometrica*, 54, 1461-1481.
- [99] Telser, L.G. (1964), "Iterative Estimation of a Set of Linear Regression Equations," *Journal of the American Statistical Association*, 59, 845-862.
- [100] Theil, H. (1953), "Repeated Least Squares Applied to Complete Equation Systems", The Hague: Central Planning Bureau.
- [101] Tjostheim, D. and R.H.Auestad, (1996), "Nonparametric Identification of Nonlinear Time Series: Projections", *Journal of the American Statistical Association*, 89, 1398-1409.
- [102] Vytlačil, E. (1999), "Semiparametric Identification and the Average Treatment Effect in Non-Separable Models", mimeo, University of Chicago.