UNOBSERVABLE INSTRUMENTS

Rosa L. Matzkin*
Department of Economics
Northwestern University

June 2004
(Incomplete Version)

Abstract

We introduce several estimators for nonparametric functions with nonadditive unobservable random terms in models with endogenous explanatory variables and, more generally, in models with simultaneity. The methods use unobservable exogenous variation as instruments. For models with endogeneity, we present an equivalence result that unifies various approaches that have been used to deal with these models, and allows one to extend techniques used for linear models to nonparametric models. We develop nonparametric estimation methods using unobservable exogenous variation in the explanatory variable, and show how, in some models, one can use a functional restriction to generate such required exogenous variation. For the case where the nonparametric function belongs to a system of simultaneous equations, we develop a method that uses an argument of the function to derive an estimator for an unobservable instrument, which is then used to estimate the nonparametric function. We demonstrate how our approach can be used to estimate nonparametric models for panel data with fixed effects, measurement error, or unobservable factors, as well as nonparametric discrete choice models where unobservable variables are not independent of observable characteristics of the alternatives, duration models with endogenous regressors, and models of demand and supply.

---

* The support of NSF is gratefully acknowledged. I have benefitted from my on-going collaborations/discussions with Joseph Altonji, Richard Blundell, James Heckman, Daniel McFadden, Lars Nesheim, and my colleagues at Northwestern. I am also grateful for the comments on previous versions of this paper by Richard Blundell, Joel Horowitz, Lung-Fei Lee, Arthur Lewbel, Costas Meghir, Ariel Pakes, Robert Porter, Christopher Taber, Kenneth Train, participants at the econometrics seminar at Ohio State University (December 2003) and Northwestern University (March 2004), participants at the NSF-NBER-UCLA Conference on Panel Data (April 2004) and at the 6th Invitational Choice Symposium (June 2004), and, specially, James Heckman and Daniel McFadden.

1. INTRODUCTION

One of the main issues in econometrics is the treatment of endogenous regressors. Economic systems, which involve agent's optimization and equilibrium conditions, generate many interrelationships among economic variables. When, as it is typically the case, only very few of the variables in an economic system can be observed, analyzing the effect of one observable variable, X, on another variable, Y, is difficult, because unobservable variables, ε, that also affect Y may be correlated, though the equilibrium or optimization conditions, with X.

Instrumental variables is, by far, the most commonly used method in applied econometrics to deal with endogenous regressors in linear models. The standard approach, which is used in most applications, requires finding an observable variable, Z, which is uncorrelated with the unobservable variable ε and correlated with the variable X. This approach has been extended to estimation of semiparametric and nonparametric models, using methods based on the conditional moment restriction that the expectation of ε given Z is zero. (Newey and Powell (1989, 2003), Darolles, Florens and Renault (2002), Hall and Horowitz (2003), Ai and Chen (2003).) Local identification of models using local independence of Z in nonparametric models with nonadditive disturbances was considered in Chesher (2001, 2002a,b,2003), and its estimation using quantile regression was recently developed by Ma and Koenker (2003). Brown and Matzkin (1996) and Chernozhukov and Hansen (2001) also used observable instruments Z.

In some cases, an instrumental variable is used in a different way. First, the residual of the relation that determines X as a function of Z is estimated, and then, this residual is used as an additional regressor in the equation of interest. This approach is based on Heckman (1976, 1978, 1980), and later works by Heckman and Robb (1985), Blundell and Smith (1986, 1989), and Rivers and Vuong (1988). The use of this approach to develop estimation methods for nonparametric triangular systems of equations can be found in Ng and Pinkse (1995), Newey, Powell and Vella (1999), Pinkse (2000), and Imbens and Newey (2001).[1] Florens, Heckman, Meghir, and Vytlacil (2003) consider identification of average treatment effects in models with continuous endogenous variables using this approach, among others. Ma and Koenker (2003) developed quantile regression estimation methods (originally developed in Koenker and Bassett (1978)). Blundell and Powell (2003b) used this approach to develop estimation methods for a semiparametric binary response model. Altonji and Matzkin (2001) proposed it to estimate an average derivative in a nonseparable model. (See Blundell and Powell (2003a) for a survey of these and other methods used in nonparametric and semiparametric regression models.) [2]

In many situations, being able to observe a variable that qualifies as an instrument is not easy. Typically, discussion about the quality of an instrument becomes central to evaluating the results of an empirical analysis.

In this paper, we develop an alternative approach for estimation of some limited dependent variable, semiparametric, and nonparametric models with endogenous explanatory variables and with simultaneity. The approach does not require observing an instrumental variable, Z. Suppose that the model of interest is

$$(1.1) \qquad Y = m(X, \varepsilon)$$

where ε and X are not independently distributed. An instrumental variable, Z, would be required to satisfy some type of independence with ε. Our approach uses, instead of Z, another variable, $\widetilde{X}$,

[1] While, in some papers, this is called a "control function" approach, the definition of a control function is more general that this and it does not necessarily require the use of an instrument. (See Heckman and Robb (1985).)

[2] When the observations are from panel or group data, an instrument, Z, is not usually needed. For example, Chamberlain (1984) proposed conditioning on past and future observations of the endogenous variables to deal with fixed effects in panel data models. Altonji and Matzkin (2001) used group data in conjunction with exchangeability of the conditional distribution of ε given the vector of endogenous variables for the members of the group, to deal with endogenous regressors.

that is not required to be independent of the unobservable variable ε. The main requirement on $\widetilde{X}$ is that the endogenous variable X be an exogenous perturbation of $\widetilde{X}$. By this we mean that for some function s and some unobservable variable η,

$$(1.2) \qquad X = s\left(\widetilde{X}, \eta\right)$$

where η satisfies some type of independence with ε. In this sense, the approach can be thought of as a dual of the observable instrument approach. We show, in Section 2, that (1.2) is equivalent to the existence of a function, r, and an unobservable, δ, such that $\varepsilon = r\left(\widetilde{X}, \delta\right)$ and δ satisfies a similar type of independence with $\left(X, \widetilde{X}\right)$. When the model (1.1)-(1.2) exhibits simultaneity, as would be the case when $\widetilde{X} = Y$, our approach provides a method for the estimation of (1.1), using one observable instrumental variable for $\widetilde{X}$, which may be included as an argument of m, to estimate both (1.1) and (1.2).

To provide an example, consider the model

$$c_{it} = m\left(I_{it}, r\left(\varepsilon_i, \delta_{it}\right)\right)$$

where c_{it} and I_{it} denote consumption and income, respectively, of individual i at time t, ε_i is an unobserved fixed effect that is correlated with I_{it}, and δ_{it} is an unobserved variable that is independent of $(\varepsilon_i, I_{it'})$ for all t'. The standard practice for estimation of these models assumes that m and r are linear, and estimates m by differencing. In contrast, our method would not impose any particular structure. It would only require that for some period t', some unknown function s, and some unobservable η that is distributed independently of $(\varepsilon_i, \delta_{it})$, conditional on $I_{it'}$

$$I_{it} = s\left(I_{it'}, \eta\right)$$

In other words, our method would require to observe the income of the individual at some other period such that the variation between both incomes is independent of the fixed effect, ε_i, conditional on the other period income. As we show below, this is enough to identify m and the joint distribution of $(I_{it}, r(\varepsilon_i, \delta_{it}))$. Imposing some structure on the function r would allow one to identify the joint distribution of (I_{it}, ε_i).

When ε is interpreted as an unobservable factor and X and $\widetilde{X}$ are interpreted as two different indicators of ε, our approach provides a way of estimating nonparametric versions of parametric models with unobservable factors, such as those recently considered in Heckman and Scheinkman (1987) and Carneiro, Hansen, and Heckman (2003). (See Joreskog (1970), Goldberger (1972), Chamberlain and Griliches (1975), Chamberlain (1977a,b), Heckman (1981), Pudney (1982), and McFadden (1984,1986) for previous work.)

For another example, consider the model in Imbens and Newey (2001)[3], where $Y = m(X, \varepsilon)$ denotes observed lifetime earnings, X denotes observed schooling, and Z is an observable cost shifter. The variable X is chosen optimally by maximizing expected lifetime earnings, given Z and an unobservable noisy signal of ε, which may be some test results, which we will denote by η^*. The optimization leads to a relationship: $X = h(Z, \eta^*)$. It is assumed that Z is independent of (ε, η^*). Suppose that instead of observing the cost shifter, Z, and not observing the noisy signal η^*, we observed the endogenous η^* and did not observe the exogenous Z. Then, as we argue in Section 2, we will still be able to estimate m and the distribution of (ε, X). Moreover, for this, we would not need to require that Z is independent of (ε, η^*); it would suffice to require that the now

[3]See also Card (2001) and Das (2001), among others.

unobservable cost shifter, Z, is independent of ε for only one value of the observable signal η^*.[4] [5]

For an example with simultaneity, consider a demand and supply model, where the quantity demanded, q, depends on price, p, consumer's income, I, and an unobservable demand shock, ε_1. The price demanded by producers depends on q and unobserved costs, represented by ε_2, which is assumed to be independent of (I, ε_1) :

$$\begin{aligned} q &= m_1(p, I, \varepsilon_1) \\ p &= m_2(q, \varepsilon_2) \end{aligned}$$

Income, I, is assumed to be independent of $(\varepsilon_1, \varepsilon_2)$. Since there is no observable instrument, Z, to identify the demand function m_1, a method based on such Z cannot be used. However, we introduce, in Section 3, a method where after using I as an observable instrumental variable to estimate m_2, we use the derived estimator for ε_2 to estimate m_1. In contrast to previously developed instrumental variable methods for nonparametric functions, the method that we introduce to estimate m_1 using ε_2 as an instrument does not require either solving an optimization problem or finding a solution to an integral equation. We show that our easily computable estimator is consistent and asymptotically normally distributed.

In models that are known to satisfy some functional restrictions, the variable $\widetilde{X}$ in (1.2) may just be one of the observable arguments of the function. Consider, for example, a model where one of the arguments, X_2^*, is measured with error, ε, so that $X_2 = X_2^* + \varepsilon$, and

$$Y = m(X_1, X_2 - \varepsilon)$$

We show, in Section 2, that if ε is independent of X_1, given X_2, then we can let X_2 be $\widetilde{X}$, in which case, m and the joint distribution of (X_2, ε) are identified.

A similar result can be applied to a semiparametric transformation model of the type

$$Y_1 = \Lambda\left(Y_2\gamma + X'\beta + \varepsilon\right)$$

where Λ is an unknown, strictly increasing function, the variables Y_1, Y_2 and the vector X are observable, and where ε is an unobservable random variable of unknown distribution, which is not necessarily independent of Y_2. This specification may arise, for example, from a proportional hazard model with unobserved heterogeneity. Suppose, for example, that Y_1 denotes the length of unemployment of an individual, Y_2 denotes years of education, X denotes a vector of other characteristics of the individual, and the unobserved heterogeneity is ability of the individual. Specifying the always needed scale normalization as $\gamma = 1$, and letting $\widetilde{X} = Y_2$, we show that Λ, β, and the joint distribution of (ε, Y_2) can be identified when ε is independent of X given Y_2.

Our results can be applied to estimate binary response as well as other limited dependent variable models. Consider, for example, a model of an individual's choice between two products. Define Y_1 by

$$Y_1 = \left\{ \begin{array}{ll} 1 & \text{if} \quad \beta\, X + \gamma\, Y_2 + \xi \geq \varepsilon \\ 0 & \text{otherwise} \end{array} \right\}$$

where X denotes the difference in observed characteristics of the two products, Y_2 denotes the difference in the store prices of the two products, ξ denotes the difference of unobserved characteristics of the two products, and ε is an unobservable random term that is independent of (X, Y_2, ξ). The interpretation is that $Y_1 = 1$ if the first of the two products is chosen. A standard estimation

[4] In our notation, we would let $\eta = Z$.

[5] Chesher (2002a, 2003) analyzed the local identification of a model similar to the one in Imbens and Newey, using local independence conditions among the observable cost shifter, Z, and the unobservables.

approach[6] proceeds by first finding an observable variable Z that is independent of ξ and correlated with Y_2, estimating first stage residuals, and including these residuals, $\widehat{v}$, as additional regressors in the function of interest. Hence, the model estimated in the second stage becomes

$$Y_1 = \left\{ \begin{array}{ll} 1 & \text{if} \quad \beta\, X + \gamma\, Y_2 + \tau\, \widehat{v} \geq \widetilde{\varepsilon} \\ 0 & \text{otherwise} \end{array} \right\}$$

where, assuming that ε is independent of (X, Y_2, v), $\widetilde{\varepsilon}$ is independent of $(X, Y_2, \widehat{v})$. (See Petrin and Train (2002) for an empirical application where this approach is used and its results are compared with those obtained using the fixed effects approach developed in Berry (1994) and Berry, Levinsohn, and Pakes (2001).) In contrast, the approach that we propose uses a variable, $\widetilde{Y}_2$, which in some cases could be the producer's price. Assume that the unobservable disturbance, η, which may represent salesmen's wages, in the relationship

$$Y_2 = s\left(\widetilde{Y}_2, \eta\right)$$

is independent of $\left(\xi, \widetilde{Y}_2\right)$, and that ε is independent of $\left(X, Y_2, \widetilde{Y}_2\right)$. Then, as we show in the following sections, for some function r and an unobservable δ that is distributed independently of $\left(X, Y_2, \widetilde{Y}_2\right)$, $\xi = r\left(\widetilde{Y}_2, \delta\right)$. Under a linearity restriction on r, one can then estimate the coefficients in the discrete choice model by estimating the model

$$Y_1 = \left\{ \begin{array}{ll} 1 & \text{if} \quad \beta\, X + \gamma\, Y_2 + \phi\, \widetilde{Y}_2 \geq \varepsilon^* \\ 0 & \text{otherwise} \end{array} \right\}$$

where $\varepsilon^* = \varepsilon - \delta$ is independent of $\left(X, Y_2, \widetilde{Y}_2\right)$. When the distribution of ε^* is specified parametrically, this model can be estimated by any usual parametric estimation method.[7] In Section 5, we show how relaxation of the linearity and parametric distributional restrictions can be easily handled by modifying existing semiparametric and nonparametric methods.

If the functions m and s in $(1.1)-(1.2)$ were linear and η were assumed to be uncorrelated with $\left(\widetilde{X}, \varepsilon\right)$, our identification method would correspond to the analysis of identification using covariance restrictions (Fisher (1966), Wegge (1965), Hausman and Taylor (1983), Hausman, Newey, Taylor (1987)). In contrast, the identification and estimation analysis in Imbens and Newey (2001) would correspond to 2SLS where instead of using a first stage estimate of the dependent variable one used the first stage residual as an argument in the second stage estimation. The identification analysis in Chesher (2001, 2003) corresponds to the order and rank conditions for identification, as in Koopmans, Rubin, and Leipnik (1950).[8]

The methods that we develop in the paper are based on conditional independence type conditions (Dawid (1979)). These conditions are at the heart of the ignorability of treatment assumption made by Rosenbaum and Rubin (1983) in their analysis of treatment effects, and it is used in other matching models. (See Heckman, Ichimura, and Todd (1997,1998) for matching techniques, Heckman and Vytlacil (2000) and Heckman and Navarro-Lozano (2001) for the analysis of these versus other methods, and Vytlacil (2001) for related work.)

[6] The approach is based on Heckman (1976, 1978, 1980), Heckman and Robb (1985), Blundell and Smith (1986, 1989), and Rivers and Vuong (1988).

[7] Note that in this case there is no need to correct the standard errors, since no first stage residuals are used in this estimation.

[8] While most of the analysis in Chesher (2003) deals with identification of triangular systems of equations under a local insensitivity assumption between observable variables and ε, the paper considers also identification of these systems using covariance restrictions, in particular in footnote 16 (page 1423).

The use of functional restrictions to identify systems of equations, which we use in our methods, reduces to imposing identity restrictions in linear and nonlinear models. Such conditions were considered in Malinvaud (1970), Fisher (1961, 1965), Kelejian (1969), and Brown (1985). More recently, Klein and Vella (2003) considered their use in the identification of parametric and semiparametric systems of equations, and Heckman, Matzkin, and Nesheim (2002) used them to estimate nonparametric, nonadditive marginal utilities and product functions in hedonic models.

The outline of the paper is as follows. In the next section, we deal with models with endogenous explanatory variables. We establish an equivalence result between various approaches, and use it to discuss the identification of these models. In the same section, we also define unobservable instruments and show how they can be used to estimate nonparametric models with endogenous explanatory variables. In Section 3, we develop an estimation method for a nonparametric, nonadditive function is a system of two simultaneous equations, when only one observable instrument is available. We show that the estimator is consistent and asymptotically normally distributed. In Section 5, we consider a discrete choice model and a semiparametric transformation model. We demonstrate how well known estimation methods for these models can be easily adapted to use unobservable, instead of observable, instruments. Section 6 presents the conclusions.

2. MODELS WITH ENDOGENOUS REGRESSORS

In this section, we analyze the identification of the function m and the distribution of (X, ε) in the model

$$(2.1) \qquad Y = m(X, \varepsilon)$$

where $Y \in R$ and $X \in R$ are observable variables, $\widetilde{X} \in R$ is another observable variable, ε is unobservable, and m is an unknown function that is strictly increasing in ε.[9] One of the objectives of this section is to establish events, which depend on $\widetilde{X}$, such that conditional on those events, X and ε are independent. We will consider two different specifications that relate X and ε to $\widetilde{X}$, and which guarantee conditional independence between X and ε, given $\widetilde{X}$. The first specification provides a basis under which model (2.1) can be identified using a dual of the approach that uses an observable instrument to derive a control function. Under this specification, for some function, s, and for an unobservable random term, η, which is assumed to satisfy some type of independence condition with ε

$$(2.2) \qquad X = s(\widetilde{X}, \eta)$$

The interpretation is that, in the specification (2.1)-(2.2), the unobservable η acts as an unobservable instrument for X. The second specification provides a basis under which model (2.1) can be transformed into one where the only unobservable random term in the transformed model is one that is distributed independently of $\left(\widetilde{X}, X\right)$. Under this specification, for some function r and a random term, δ, that is distributed independently of $\left(\widetilde{X}, X\right)$

$$(2.3) \qquad \varepsilon = r\left(\widetilde{X}, \delta\right).$$

We will show that, conditional on $\widetilde{X}$, independence between X and ε in (2.1) is equivalent to the existence of a function s and an unobservable η satisfying (2.2), with η independent of ε, and it is also equivalent to the existence of a function r and an unobservable δ satisfying (2.3) with δ independent of η and of X. If either of these conditional independence statements held at

[9] Even though we will concentrate on the case where X is a scalar, the results can be easily extended to the case where X is a vector. Also, for notational simplicity, we will not carry over a vector, W, of variables that may enter as additional arguments in the functions of interest.

only one value $\overline{\overline{x}}$ of $\widetilde{X}$, then one could establish the other statements conditional also on $\widetilde{X} = \overline{\overline{x}}$. Unconditional results, where η is independent of $\left(\widetilde{X}, \varepsilon\right)$ and δ is independent of $\left(\widetilde{X}, X\right)$ can also be established.

The equivalence between (2.2)-(2.3), together with the conditional independence among the various variables, allow one to express model (2.1) in a variety of equivalent formulations, each involving different unobservable random terms:

$$\begin{aligned} Y &= m\left(X, \varepsilon\right) \\ &= m\left(s\left(\widetilde{X}, \eta\right), \varepsilon\right) \\ &= m\left(s\left(\widetilde{X}, \eta\right), r\left(\widetilde{X}, \delta\right)\right) \\ &= m\left(X, r\left(\widetilde{X}, \delta\right)\right) \end{aligned}$$

If we know one of the last three formulations, we can derive the other ones. The first formulation is, of course, the one of interest. The second will in many cases be the easiest to establish. In this specification, η denotes an unobservable instrument, which satisfies some type of independence with ε, and $\widetilde{X}$ is not independent of ε. The third expression highlights the fact that if η and δ are independently distributed, even if this is just for one value of $\widetilde{X}$, then, this provides sufficient exogenous variation between X and ε to identify m. The fourth expression is useful for when one wants to estimate m using standard estimation procedures that require independence between the explanatory variables and the unobservable random terms in the model. It is also useful for the estimation method for systems of equations the we present in Section 4.

When, in the above formulations, η is interpreted as an observable instrument and $\widetilde{X}$ is interpreted as an unobserved residual, our equivalence results can be used to develop nonparametric estimation methods for cases where observable instruments are available.

In Section 2.1, we present our equivalence result for the case where conditional independence is established at only one value of $\widetilde{X}$. Section 2.2 and 2.3 considers the cases where stronger independence conditions can be established. In Section 5, it is demonstrated how these equivalence results can be used to estimate discrete choice models with endogeneity.

2.1. UNOBSERVABLE INSTRUMENTS USING LOCAL CONDITIONAL INDEPENDENCE

Let $\overline{\overline{x}}$ denote a value in the support of $\widetilde{X}$. We will make the following assumptions:

ASSUMPTION 2.1: *The function m in (2.1) is strictly increasing in ε.*

ASSUMPTION 2.2: *The conditional distribution of ε given $\widetilde{X} = \overline{\overline{x}}$ is strictly increasing.*

ASSUMPTION 2.3: *The conditional distribution of X* given $\widetilde{X} = \overline{\overline{x}}$ *is strictly increasing.*

LEMMA 2.1: *Suppose that assumptions 2.1-2.3 are satisfied. Then, the following statements are equivalent:*

(i) ε *is independent of* X, *conditional on* $\widetilde{X} = \overline{\widetilde{x}}$.

(ii) There exists a strictly increasing function $s_{\overline{\widetilde{x}}}(\cdot)$ *and an unobservable random term* η *such that*

$X = s_{\overline{\widetilde{x}}}(\eta)$ *and*
η *is independent of* ε, *conditional on* $\widetilde{X} = \overline{\widetilde{x}}$.

(iii) There exists a strictly increasing function $r_{\overline{\widetilde{x}}}(\cdot)$ *and an unobservable random term* δ *such that*

$\varepsilon = r_{\overline{\widetilde{x}}}(\delta)$,
δ *is independent of* η, *conditional on* $\widetilde{X} = \overline{\widetilde{x}}$, *and*
δ *is independent of* X, *conditional on* $\widetilde{X} = \overline{\widetilde{x}}$.

(iv) There exists a strictly increasing function $r_{\overline{\widetilde{x}}}(\cdot)$ *and an unobservable random term* δ *such that*

$\varepsilon = r_{\overline{\widetilde{x}}}(\delta)$, *and*
δ *is independent of* X, *conditional on* $\widetilde{X} = \overline{\widetilde{x}}$.

An example of a parametric relationship between X, ε, and $\widetilde{X}$ where either of the statements in Lemma 2.1 is satisfied is one where for unknown parameters, γ_{11}, γ_{12}, γ_{21}, γ_{22}, $\widetilde{\sigma}_{12}$, $\widetilde{\sigma}_{11} > 0$, $\widetilde{\sigma}_{22} > 0$, $\alpha_1 > 0$, and $\alpha_2 > 0$, and unobservable (δ_1, δ_2) that is distributed normal with mean $(0,0)$, variances $\sigma_{11} = \left(\alpha_1 \widetilde{X}^2 + \widetilde{\sigma}_{11}\right)$ and $\sigma_{22} = \left(\alpha_2 \widetilde{X}^2 + \widetilde{\sigma}_{22}\right)$, and covariance $\sigma_{12} = \left(\left(\widetilde{X} - 1\right)^2 \widetilde{\sigma}_{12}\right)$,

$$\begin{aligned} X &= \gamma_{11} + \gamma_{12}\widetilde{X} + \delta_1 \\ \varepsilon &= \gamma_{21} + \gamma_{22}\widetilde{X} + \delta_2 \end{aligned}$$

When $\widetilde{X} = 1$, $\sigma_{12} = 0$; hence, δ_1 and δ_2 are independent. The statements in the Lemma are then satisfied for $\overline{\widetilde{x}} = 1$, $\eta = \delta_1$ and $\delta = \delta_2$.

In the next theorem we show that the function m and the distribution of (X, ε) can be identified from the distribution of Y conditional on $\left(X, \widetilde{X} = \overline{\widetilde{x}}\right)$. For this theorem, we will need one more assumption, which will be used to guarantee the existence of $F^{-1}_{Y|X=x,\widetilde{X}=\overline{\widetilde{x}}}$.

ASSUMPTION 2.4: *For each* x, *the conditional distribution of* ε *given* $\left(X, \widetilde{X}\right) = \left(x, \overline{\widetilde{x}}\right)$ *is strictly increasing.*

Then, we can establish the following:

THEOREM 2.1: *Suppose that Assumptions 2.1-2.4 are satisfied. If at least one of the equivalent statements in Lemma 2.1 is satisfied, then for all* x, e

$$(2.4) \quad m(x,e) = F^{-1}_{Y|\widetilde{X}=\overline{\widetilde{x}},X=x}\left(F_{\varepsilon|\widetilde{X}=\overline{\widetilde{x}}}(e)\right)$$

and

$$(2.5) \quad F_{\varepsilon|X=x}(e) = F_{Y|X=x}\left(F^{-1}_{Y|\widetilde{X}=\overline{\widetilde{x}},X=x}\left(F_{\varepsilon|\widetilde{X}=\overline{\widetilde{x}}}(e)\right)\right)$$

Theorem 2.1 establishes the global identification of the function m and the distribution of (X, ε), up to a normalization on the conditional distribution $F_{\varepsilon|\widetilde{X}=\overline{\widetilde{x}}}$. If, for example, we normalized the distribution of ε conditional on $\widetilde{X} = \overline{\widetilde{x}}$ to be $U(0,1)$, then, for all $e \in (0,1)$

$$(2.6) \quad m(x,e) = F^{-1}_{Y|X=x,\widetilde{X}=\overline{\widetilde{x}}}(e)$$

and

$$(2.7) \quad F_{\varepsilon|X=x}(e) = F_{Y|X=x}\left(F^{-1}_{Y|\widetilde{X}=\overline{\widetilde{x}},X=x}(e)\right)$$

To define nonparametric estimators for m and $F_{\varepsilon|X=x}$, one only needs to substitute the conditional distributions of the observable variables, in (2.6) and (2.7), by nonparametric estimators for them. If, for example, the method of kernels were used to derive these estimators, one could use the results in Matzkin (1999, 2003) and in Altonji and Matzkin (2001, 2003) to determine the asymptotic distributions of the estimators for the unknown functions, their derivatives, and the average of their derivatives. If the method of series were used, one could use Newey (1997) and the results in Imbens and Newey (2001) to derive the asymptotic distributions of similar estimators.

2.2 UNOBSERVABLE INSTRUMENTS USING FUNCTIONAL RELATIONSHIPS

In many situations, we have information about the way in which some variables interact. Sometimes, economic theory implies known functional restrictions. For example, in a pure exchange model, income equals price times endowment; in a model of a firm, revenue equals price times output; in a supply equation, the total quantity sold is the sum of the quantities sold to different groups of consumers. In a model where an explanatory variable, X^*, is measured with error, X^* is replaced by $X - \widetilde{\eta}$, where $\widetilde{\eta}$ is an unobservable variable, which depends on X. In semiparametric models, the parametric structure that is used can also be seen as a functional restriction. In this subsection, we show that when we are interested in identifying the function m in the relationship

$$(2.8) \quad Y = m(X, \varepsilon)$$

where ε and X are not independently distributed, we can use such functional restrictions to help us find a variable $\widetilde{X}$, conditional on which X and ε are independent.

Suppose that the model is

$$(2.9) \quad Y = m(X_2, \widetilde{\varepsilon})$$

where X_2 and $\widetilde{\varepsilon}$ are not necessarily independent, for some known function q

$$(2.10) \quad \widetilde{\varepsilon} = q(X_1, \varepsilon)$$

and where for at least some value $\overline{x}_1$ of X_1

$$(2.11) \qquad F_{\varepsilon|X_1=\overline{x}_1,X_2} = F_{\varepsilon|X_1=\overline{x}_1}$$

Consider, for example, a model where Y is a function of X_2 and a variable, X_1^*, which can only be measured with error. If X_1 denotes the measurement of X_1^*, so that $X_1 = X_1^* + \varepsilon$, then, $Y = m(X_2, X_1 - \varepsilon)$, which satisfies (2.9) and (2.10). Clearly, ε and X_1 are not independently distributed, and, in general, ε will also not be independently distributed of (X_1, X_2). However, if (2.11) holds, we will be able to identify m and the joint distribution of $(X_2, \widetilde{\varepsilon})$. From this and (2.10) we can identify the distribution of ε given X_1. Note that if the function q is only known when $X_1 = \overline{x}_1$, the function m and the distribution of $(X_2, \widetilde{\varepsilon})$ will still be identified.

To analyze the identification of this model, let $\overline{x}_1$ denote a value of X_1, and make the following assumptions:

ASSUMPTION 2.1*: *The function m in (2.8) is strictly increasing in ε.*

ASSUMPTION 2.2*: *The conditional distribution of ε given $X_1 = \overline{x}_1$ is strictly increasing.*

ASSUMPTION 2.3*: *The conditional distribution of X_2 given $X_1 = \overline{x}_1$ is strictly increasing.*

ASSUMPTION 2.4*: $F_{\varepsilon|X_1=\overline{x}_1} = F_{\varepsilon|X_1=\overline{x}_1,X_2}$

Consider model (2.9). Let $\widetilde{X} = X_1$. Then, by (2.10) and (2.11), using Lemma 2.1, we can establish the existence of a strictly function $s_{\overline{x}_1}(\cdot)$ and an unobservable η that is distributed independently of ε, conditional on $X_1 = \overline{\widetilde{x}}$, such that

$$(2.12) \qquad X_2 = s_{\overline{x}_1}(\eta)$$

Using then Theorem 2.1, with $X = X_2$ and $\widetilde{X} = X_1$, it follows that for all x, e

$$(2.13) \qquad m(x,e) = F^{-1}_{Y|X_1=\overline{x}_1,X_2=x}\left(F_{\widetilde{\varepsilon}|X_1=\overline{x}_1}(e^*)\right)$$

and

$$(2.14) \qquad F_{\widetilde{\varepsilon}|X=x}(e) = F_{Y|X=x}\left(F^{-1}_{Y|X_1=\overline{x}_1,X_2=x}\left(F_{\widetilde{\varepsilon}|X_1=\overline{x}_1}(e^*)\right)\right)$$

where e^* is any value of $\widetilde{\varepsilon}$ such that $q(\overline{x}_1, e^*) = e$. Hence, the model described by (2.9)-(2.11) is identified up to a normalization on the conditional distribution of $\widetilde{\varepsilon}$ given $X_1 = \overline{x}_1$.

2.3. UNOBSERVABLE INSTRUMENTS USING GLOBAL CONDITIONAL INDEPENDENCE

In some situations, we may be able to determine that an unobservable instrument is independent of the unobservable, ε, conditional on all values, $\widetilde{x}$, of $\widetilde{X}$, instead of on only one value of $\widetilde{X}$, as considered in the previous subsection. In cases where this condition is satisfied, we will be able to identify nonparametrically not only m and the distribution of (X, ε), as in Subsection 2.1, but also the nonparametric random relationship between ε and $\widetilde{X}$. The added identification result is useful, in particular, when one is interested in estimating m directly using the specification

$$Y = m\left(X, r\left(\widetilde{X}, \delta\right)\right)$$

where δ is independent of $\left(X, \widetilde{X}\right)$. Any nonlinear, semiparametric, or nonparametric method that can be used when the unobservable in the model is independent of the observable explanatory variables can be used to estimate m and r using this specification. In Section 4, we demonstrate this use of the function r in the estimation of a discrete choice model where the unobservable random term is not independent of the observable characteristics.

Consider the following assumptions:

ASSUMPTION 2.2': *For all values* $\widetilde{x}$ of $\widetilde{X}$, t*he conditional distribution of* ε *given* $\widetilde{X} = \widetilde{x}$ *is strictly increasing.*

ASSUMPTION 2.3': *For all values* $\widetilde{x}$ of $\widetilde{X}$, t*he conditional distribution of* X *given* $\widetilde{X} = \widetilde{x}$ *is strictly increasing.*

ASSUMPTION 2.4': *For all values* $(x, \widetilde{x})$ of $\left(X, \widetilde{X}\right)$, t*he conditional distribution of* ε *given* $\left(X, \widetilde{X}\right) = (x, \widetilde{x})$ *is strictly increasing.*

In Lemma 2.2 of Appendix A, we show that if Assumptions 2.1 and 2.2'-2.4' are satisfied, and if

$$X = s\left(\widetilde{X}, \eta\right)$$

where s is strictly increasing in the unobservable η, and η is distributed independently of ε, given $\widetilde{X}$, then there exists a function r and an unobservable δ such that r is strictly increasing in δ,

$$\varepsilon = r\left(\widetilde{X}, \delta\right) \tag{2.19}$$

and δ is independent of $\left(X, \widetilde{X}\right)$. Lemma 2.2 also shows that the inverse is true, and that either of the above statements is equivalent to ε being independent of X, conditional on $\widetilde{X}$. In the parametric example described after the statement of Lemma 2.1, these will be satisfied as long as $\sigma_{12} = 0$. Suppose that, in addition to Assumptions 2.2'-2.4', the following assumption is also satisfied:

ASSUMPTION 2.5: *For all values* $\widetilde{x}$ of $\widetilde{X}$, *the conditional distribution* η *given* $\widetilde{X} = \widetilde{x}$ *is strictly increasing.*

We can then establish the following:

THEOREM 2.2: *Suppose that Assumptions 2.1, 2.2'-2.4', and 2.5 are satisfied. Suppose also that (2.19) is satisfied for* δ independent of $\left(X, \widetilde{X}\right)$ and r strictly increasing in δ. *Then,* for any $\widetilde{x}, t, x$

$$r\left(\widetilde{x}, t\right) = F^{-1}_{\varepsilon|\widetilde{X}=\widetilde{\widetilde{x}}}\left(F_{Y|\widetilde{X}=\widetilde{\widetilde{x}}, X=x}\left(F^{-1}_{Y|\widetilde{X}=\widetilde{x}, X=x}\left(F_\delta\left(t\right)\right)\right)\right) \tag{2.20}$$

Theorem 2.2 states that the function r is identified from the joint distribution of $\left(Y, \widetilde{X}, X\right)$ up to a normalization on the distribution of δ and a normalization on the conditional distribution of ε, given $\widetilde{X} = \overline{\widetilde{x}}$. If we normalized these distributions to be, for example, $U(0,1)$, then, we could estimate the function r using the relationship:

$$(2.21) \qquad r\left(\widetilde{x}, t\right) = F_{Y|\widetilde{X}=\overline{\widetilde{x}}, X=x}\left(F^{-1}_{Y|\widetilde{X}=\widetilde{x}, X=x}\left(t\right)\right)$$

Moreover, since the choice of x in (2.21) is arbitrary, we could use a weighted average. Let $w(x)$ denote a weight function on x. Then,

$$(2.22) \qquad r\left(\widetilde{x}, t\right) = \int \left[F_{Y|\widetilde{X}=\overline{\widetilde{x}}, X=x}\left(F^{-1}_{Y|\widetilde{X}=\widetilde{x}, X=x}\left(t\right)\right)\right] \; w\left(x\right) \; dx$$

To provide an example of the usefulness of this result, consider the measurement error model discussed in the previous section, where $Y = m\left(X_2, X_1 - \varepsilon\right)$ and X_1 is a measurement of X_1^*. Assume that the distribution of $\widetilde{\varepsilon} = X_1 - \varepsilon$, conditional on X_1, is independent of X_2. Using (2.20) together with a normalization on the distribution of δ, one can immediately derive the conditional distribution of $\widetilde{\varepsilon} = r\left(X_1, \delta\right)$, given X_1. Since $\varepsilon = X_1 - \widetilde{\varepsilon}$, the distribution of ε given X_1 is also easily derived.

2.4 UNOBSERVABLE INSTRUMENTS USING GLOBAL INDEPENDENCE CONDITIONS

One of the benefits of the estimation methods that we developed in the previous sections is that none of them required *estimating* the unobservable instrument. It sufficed to establish the *existence* of the unobservable instrument. In some cases, however, we might need to estimate the unobservable instrument, such as, for example, when we need to use it for estimating a system of equations, as discussed in Section 4. Suppose that we were able to establish that for an observable $\widetilde{X}$ and for some function s, which is strictly increasing in η

$$(2.23) \qquad X = s\left(\widetilde{X}, \eta\right)$$

where η is independent of $\left(\widetilde{X}, \varepsilon\right)$. (For example, the parametric specification presented after Lemma 2.1 satisfies this conditions when $\sigma_{12} = 0$ and $\alpha_1 = \alpha_2 = 0$.) Then, using Matzkin (1999), it would follow that for all $\widetilde{x}, t$

$$s\left(\widetilde{x}, t\right) = F^{-1}_{X|\widetilde{X}=\widetilde{x}}\left(F_\eta\left(t\right)\right)$$

If, for example, we used the normalization that η is $U(0,1)$[10], this would imply that we could estimate η from the joint distribution of $\left(\widetilde{X}, X\right)$ using the relationship $\eta = F_{X|\widetilde{X}=\widetilde{x}}\left(x\right)$.

Lemma 2.3 in Appendix A extends the equivalence results from the previous sections to the case where (2.23) can be established for η independent of $\left(\widetilde{X}, \varepsilon\right)$.

[10] Imbens and Newey (2001) used, for example, such a normalization for the estimation of a nonparametric, non-additive function.

3. MODELS WITH SIMULTANEITY

In some cases, (2.23) with η independent of $\left(\widetilde{X},\varepsilon\right)$, or even the weaker conditional independence conditions considered in the previous sections might be difficult to establish. Consider, for example, a demand and supply model, where the quantity demanded, q, depends on price, p, consumers' income, I, and an unobservable demand shock, ε_1, and where the price demanded by producers depends on q and unobserved marginal costs, ε_2, which are independent of I and ε_1 :

$$\begin{aligned} q &= m_1\left(p, I, \varepsilon_1\right) \\ p &= m_2\left(q, \varepsilon_2\right) \end{aligned}$$

In this case, we cannot assume that the unobservable variable in m_2 is independent of ε_1 conditional on q. For this case, we next introduce a method to estimate m_1 and the distribution of ε_1. Our method applies to the case where m_1 is a nonparametric function that is strictly increasing in ε_1, I is an instrument for the estimation of m_2, and ε_1 is distributed independently of (I,ε_2). We provide conditions under which our easily computable estimator for the function m_1 is consistent and asymptotically normally distributed.

Our method requires a first step estimator, $\widehat{\varepsilon}_2$, for ε_2. Such an estimator can be obtained using I as an instrumental variable to estimate the function m_2. For this, one may specify a parametric structure for m_2, or use a method based on recently developed semiparametric and nonparametric instrumental variable methods, such as the ones developed by Newey and Powell (1989, 2003), Brown and Matzkin (1996), Chernozhukov and Hansen (2001,2003), Darolles, Florens and Renault (2002), Brown and Wegkamp (2002), Ai and Chen (2003), or Hall and Horowitz (2003).

Given an estimator, $\widehat{m}_2$, for the function m_2, and assuming that m_2 is either strictly increasing or strictly decreasing in ε_2, we can define, for each i, $\widehat{\varepsilon}_2^i$ by

$$\widehat{\varepsilon}_2^i = \widehat{m}_2^{-1}\left(q^i, p^i\right)$$

Define the function $\widetilde{p}$ by

$$p = \widetilde{p}\left(I, \varepsilon_2, \varepsilon_1\right)$$

Assume that $\widetilde{p}$ is strictly increasing in ε_1. (In the demand and supply example above, this will be satisfied under the standard assumptions that $\partial m_2/\partial q > 0$, $\partial m_1/\partial p < 0$, and $\partial m_1/\partial \varepsilon_1 > 0$.) Assume also that ε_1 has an everywhere positive density. Then, since ε_1 is distributed independently of (I,ε_2), it follows that

$$\widetilde{p}\left(I, \varepsilon_2, e_1\right) = F_{p|I,\varepsilon_2}^{-1}\left(F_{\varepsilon_1}(e_1)\right)$$

(See Matzkin (1999)). Hence, the function $\widetilde{p}$ is identified, up to a normalization, if ε_2 is observed. If, for example, $F_{\varepsilon_1}(t) = t$, an estimator for $\widetilde{p}$ can be obtained as

$$\widehat{\widetilde{p}}\left(I, \varepsilon_2, e\right) = \widehat{F}_{p|I,\widehat{\varepsilon}_2}^{-1}(e)$$

where $\widehat{F}_{p|I,\widehat{\varepsilon}_2}(e)$ is a kernel estimator for the conditional distribution of p given $(I,\widehat{\varepsilon}_2)$:

$$\widehat{F}_{p|I,\widehat{\varepsilon}_2}(e) = \frac{\sum_{i=1}^{N} \widetilde{k}\left(\frac{p-p^i}{\sigma}\right) K\left(\frac{I-I^i}{\sigma}, \frac{\widehat{\varepsilon}_2-\widehat{\varepsilon}_2^i}{\sigma}\right)}{\sum_{i=1}^{N} K\left(\frac{I-I^i}{\sigma}, \frac{\widehat{\varepsilon}_2-\widehat{\varepsilon}_2^i}{\sigma}\right)}$$

with $\widetilde{k}$ denoting the integral of a univariate kernel function, K a kernel function, and σ a bandwidth. In a similar way, one can define the function $\widetilde{q}$ by

$$q = \widetilde{q}\left(I, \varepsilon_2, \varepsilon_1\right)$$

and, if $\widetilde{q}$ is strictly increasing in ε_1, we can obtain an estimator for it using the relationship

$$\widetilde{q}(I, \varepsilon_2, e_1) = F^{-1}_{q|I,\varepsilon_2}\left(F_{\varepsilon_1}(e_1)\right)$$

(In the demand and supply example, $\widetilde{q}$ is strictly increasing in ε_1 if $\partial m_2/\partial q > 0$, $\partial m_1/\partial p < 0$, and $\partial m_1/\partial \varepsilon_1 > 0$.).

Our method makes use of the relationship

$$\begin{aligned} q &= \widetilde{q}(I, \varepsilon_2, \varepsilon_1) \\ &= m_1(p, I, \varepsilon_1) \\ &= m_1(\widetilde{p}(I, \varepsilon_2, \varepsilon_1), I, \varepsilon_1) \end{aligned}$$

from which it follows that we can use the "reduced form" estimators $\widehat{\widetilde{p}}$ and $\widehat{\widetilde{q}}$, to define an estimator for m_1 by

$$\widehat{m}_1(p, I, e_1) = \widehat{\widetilde{q}}(I, \widehat{\varepsilon}_2^*, e_1)$$

where $\widehat{\varepsilon}_2^*$ is the value of $\widehat{\varepsilon}_2$ that satisfies the equation

$$p = \widehat{\widetilde{p}}(I, \widehat{\varepsilon}_2^*, e_1)$$

When kernel estimators as described above are used, we can establish the consistency and asymptotic normality of our estimator for m_1 under the following assumptions:

ASSUMPTION 3.1: *The observations* $\left\{p^i, q^i, I^i\right\}_{i=1}^N$ *are i.i.d. across i.*

ASSUMPTION 3.2: *The density* $f(p, q, I, \varepsilon_2)$ *has compact support and is continuously differentiable up to order* $s' \geq 2$.

ASSUMPTION 3.3: *The kernel function* K *is differentiable of order* $\widetilde{s}$, *the derivatives of* K *of order* $\widetilde{s}$ *are Lispchitz,* K *vanishes outside a compact set, integrates to* 1, *and is of order* s'', *where* $\widetilde{s} + s'' \leq s'$, $\widetilde{s} \geq 2$ *and* $s'' \geq 2$.

ASSUMPTION 3.4: *As* $N \to \infty$, $\sigma_N \to 0$, $\ln(N)/\left(N\sigma_N^5\right) \to 0$, $\sqrt{N}\sigma_N \to \infty$, $\sqrt{N}\sigma_N^{1+s''} \to 0$, *and* $\sqrt{N\sigma_N^2}\left(\sqrt{\left(\ln(N)/(N\sigma_N^6\right)} + \sigma_N^{s''}\right)^2 \to 0$.

ASSUMPTION 3.5: $f(I, \varepsilon_2) > 0$, $f(p, I, \varepsilon_2) > 0$, *and*
$\left[\varepsilon_1\left(\partial f(I, \varepsilon_2)/\partial \varepsilon_2\right) - \int_{-\infty}^{p}\left(\partial f(p', I, \varepsilon_2)/\partial \varepsilon_2\right)\ dp'\right] \neq 0$.

ASSUMPTION 3.6: *The estimator for* $\widehat{\varepsilon}_2$ *satisfies* $\sqrt{N\sigma_N^2}\ \sup|\widehat{\varepsilon}_2 - \varepsilon_2| \to 0$.

The proof of the following theorem is presented in Appendix A:

THEOREM 3.1: *If assumptions 3.1-3.6 are satisfied, then*

$$\widehat{m}_1(p, I, e_1) \to m_1(p, I, e_1) \text{ in probability}$$

and

$$\sqrt{N\sigma^2}\left(\widehat{m}_1(p,I,e_1)-m_1(p,I,e_1)\right)\to N(0,V)$$

where $V=\widetilde{V}\int K(I,e_2)^2 dIde_2$ *and*

$$\widetilde{V} = \frac{\left[\int\left[e_1 - 1\left[q'\le m_1(p,I,e_1)\right]\right]^2 f\left(q',I,\varepsilon_2^*\right)\, dq'\right]}{\left[f\left(m_1(p,I,e_1),I,\varepsilon_2^*\right)\right]^2}$$

$$+\frac{\left[\int\left[e_1 - 1\left[p'\le p\right]\right]^2 f\left(p',I,\varepsilon_2^*\right)\, dp'\right]}{\left[f\left(m_1(p,I,e_1),I,\varepsilon_2^*\right)\right]^2}.$$

$$\frac{\left[e_1\frac{\partial f\left(I,\varepsilon_2^{*\prime}\right)}{\partial\varepsilon_2}-\int^{m_1(p,I,e_1)}\frac{\partial f\left(q',I,\varepsilon_2^{*\prime}\right)}{\partial\varepsilon_2}dq'\right]^2}{\left[e_1\frac{\partial f\left(I,\varepsilon_2^{*}\right)}{\partial\varepsilon_2}-\int^{p}\frac{\partial f\left(p',I,\varepsilon_2^{*}\right)}{\partial\varepsilon_2}dp'\right]^2}$$

5. EXAMPLES

5.1. A DISCRETE CHOICE MODEL

The methods developed in the previous section can be used to identify and estimate discrete choice models that possess an unobservable variable that is correlated with an observable explanatory variable.

Consider, for example, the model

$$Y_1=\left\{\begin{array}{ll}1 \quad if & X_0+v(X_1,Y_2,\xi)\ge\varepsilon\\ 0 & otherwise\end{array}\right\}$$

where Y_1, Y_2, and X are observable and ξ and ε are unobservable. The identification of this model has been established in Briesch, Chintagunta and Matzkin (1997), under the assumptions that ξ is independent of (X,Y_2,ε), ε is independent of (X,Y_2,ξ), and X_0 is independent of $(X_1,Y_2,\xi,\varepsilon)$. Their identification arguments can be easily modified to allow ξ to depend on Y_2. Specifically, let

$$W=v(X_1,Y_2,\xi)$$

and assume that for some unknown function s, observable $\widetilde{Y}_2$, and unobservable η

$$Y_2=s\left(\widetilde{Y}_2,\eta\right)$$

From the distribution of Y_1 given $\left(X_0,X_1,Y_2,\widetilde{Y}_2\right)$, one can follow the arguments in Briesch, Chintagunta, and Matzkin (1997) to identify the distribution of W given $\left(X_1,Y_2,\widetilde{Y}_2\right)$. Then, using the methods developed in the previous sections, one can identify v and the distribution of (ξ, Y_2), under conditional independence assumptions between X and ξ.

Consider a more parameterized version of the above model, of the type

$$(5.1) \qquad Y_1 = \left\{ \begin{array}{ll} 1 \quad if & \beta\, X + \gamma\, Y_2 + \xi \geq \varepsilon \\ 0 & \qquad otherwise \end{array} \right\}$$

where ξ is independent of (X, ε) but not of Y_2, and ε is independent of (X, Y_1, Y_2). As mentioned in the introduction, a particular application of this model that has received a lot of attention is one where $Y_1 = 1$ if one of two differentiated products is chosen and $Y_1 = 0$ otherwise, X denotes the difference in observed characteristics of the two products, Y_2 denotes the difference in the store prices of the two products, and ξ denotes the difference of unobserved characteristics of the two products. Another important application for such a model is where ξ denotes an individual effect, that is correlated with Y_2, which might be the case, for example, if Y_2 denoted income.

To estimate model (5.1) using our unobservable instruments, suppose that there exists an observable variable, $\widetilde{Y}_2$, such that at least one of the conditions in the statements of either Lemma 2.2 or Lemma 2.3 is satisfied with $\widetilde{X} = \widetilde{Y}_2$ and $X = Y_2$. Then, we can transform the above model into

$$(5.2) \qquad Y_1 = \left\{ \begin{array}{ll} 1 \quad if & \beta\, X + \gamma\, Y_2 + r\left(\widetilde{Y}_2, \delta\right) \geq \varepsilon \\ 0 & \qquad otherwise \end{array} \right\}$$

where r is a function that is strictly increasing in its last argument and δ is distributed independently of $\left(X, Y_2, \widetilde{Y}_2\right)$. In the choice of product example, $\widetilde{Y}_2$ might be the producer's price. Suppose, for example, that the stores are perfectly competitive and each store's technology is constant returns to scale. Then, the store price would be determined by the price paid by the store to the producer and other variables that, in most cases, will be independent of the products' unobservable characteristics, such as salesmen's wages. In the case where ξ denotes an individual effect and Y_2 denotes current income, $\widetilde{Y}_2$ could, in some situations, be some other period's income.

If we are willing to assume that the function r is linear in $\widetilde{Y}_2$ and additive in δ, then we can estimate consistently the coefficients in this discrete choice model by estimating the model

$$Y_1 = \left\{ \begin{array}{ll} 1 \quad \text{if} & \beta\, X + \gamma\, Y_2 + \phi\, \widetilde{Y}_2 \geq \varepsilon^* \\ \quad 0 \quad \text{otherwise} & \end{array} \right\}$$

where $\varepsilon^* \equiv \varepsilon - \delta$ is independent of $\left(X, Y_2, \widetilde{Y}_2\right)$. When the distribution of ε^* is specified parametrically, this model can be estimated by any of the usual parametric methods. Note that since $\widetilde{Y}_2$ is an observed quantity, rather than an estimated one, there is no need to adjust the standard errors, as would be required when first stage residuals were used instead on $\widetilde{Y}_2$.

If, instead, we do not want to restrict either the function r or the distribution of ε or δ to belong to parametric families, then, as we next show, we can estimate the model using a slight modification of the approach in Blundell and Powell (2003b). For this, we would first need to specify some normalizations. One such possible set of normalizations is that (i) δ is $U(0,1)$, (ii) the distribution of ξ at some value $\overline{\widetilde{y}}_2$ of $\widetilde{Y}_2$ is $U(0,1)$, and (iii) one of the coefficients of X equals 1. It is easy to show that (i) and (ii) imply that for all δ,

$$r\left(\overline{\widetilde{y}}_2, \delta\right) = \delta$$

Let

$$\widetilde{\eta} = \varepsilon - r\left(\widetilde{Y}_2, \delta\right)$$

Then, the model

$$Y_1 = \left\{ \begin{array}{lc} 1 \quad if & \beta\ X + \gamma\ Y_2 + \xi \geq \varepsilon \\ 0 & otherwise \end{array} \right\}$$

becomes

$$Y_1 = \left\{ \begin{array}{lc} 1 \quad if & \beta\ X + \gamma\ Y_2 \geq \widetilde{\eta} \\ 0 & otherwise \end{array} \right\}$$

where $\widetilde{\eta}$ is an unobservable random term whose distribution depends on $\widetilde{Y}_2$. Since ε and δ are, by assumption, independent of (X, Y_2), $\widetilde{\eta}$ is independent of (X, Y_2) conditional on $\widetilde{Y}_2$. It then follows that the conditional expectation of Y_1 given $\left(X, Y_2, \widetilde{Y}_2\right)$ possesses the multiple index form given by

$$E\left[Y_1 = 1 | X, Y_2, \widetilde{Y}_2\right] = G\left(\beta\ X + \gamma\ Y_2, \widetilde{Y}_2\right)$$

where G is an unknown function. This conditional expectation possesses the same form as that considered in Blundell and Powell (2003b), except for the fact that, in Blundell and Powell (2003b), the residual, $\widehat{v}$, from a first stage regression, takes up the place of the observable variable $\widetilde{Y}_2$. In other words, the expression that is used to estimate the coefficients in Blundell and Powell (2003b) is

$$E\left[Y_1 = 1 | X, Y_2, \widehat{v}\right] = G\left(\beta\ X + \gamma\ Y_2, \widehat{v}\right)$$

Blundell and Powell (2003b) describe several estimation methods for the coefficients of the index and for the function G, in this model. Those same estimation methods can be applied when $\widehat{v}$ in their estimators is substituted with $\widetilde{Y}_2$. Once an estimator, $\widehat{G}$, of G, and estimators, $\widehat{\beta}$ and $\widehat{\gamma}$, of β and γ, are obtained, one can estimate the distributions of ε and the function r. To see this, note that, for any $\widetilde{y}_2$ and t,

$$G\left(t, \widetilde{y}_2\right) = \Pr\left(\widetilde{\eta} \leq t | \widetilde{Y}_2 = \widetilde{y}_2\right)$$

Since, by (5.3) and (5.4), when $Y_2 = \overline{\overline{y}}_2$, $\widetilde{\eta} = \varepsilon - \delta$, and since by assumption, $\varepsilon - \delta$ is independent of $\widetilde{Y}_2$,

$$(5.5) \qquad G\left(t, \overline{\overline{y}}_2\right) = \Pr\left(\varepsilon - \delta \leq t\right)$$

Since, δ is $U(0,1)$, one can use (5.5) and $\widehat{G}\left(t, \overline{\overline{y}}_2\right)$ to estimate the distribution of ε, by deconvolution. Next, using the estimators for the distribution of ε and for the distribution of $\widetilde{\eta}$ conditional on $\widetilde{Y}_2$, one can estimate the distribution of $(\varepsilon - \widetilde{\eta})$, conditional on $\widetilde{Y}_2$. From this, one can estimate the function $r\left(\widetilde{y}_2, \delta\right)$, using the expression

$$r\left(\widetilde{y}_2, t\right) = F^{-1}_{\varepsilon - \widetilde{\eta} | \widetilde{Y}_2 = \widetilde{y}_2}(t),$$

which follows because

$$\begin{aligned} t &= \Pr\left(\delta \leq t\right) \\ &= \Pr\left(\delta \leq t | \widetilde{Y}_2 = \widetilde{y}_2\right) \\ &= \Pr\left(r\left(\widetilde{Y}_2, \delta\right) \leq r\left(\widetilde{y}_2, t\right) | \widetilde{Y}_2 = \widetilde{y}_2\right) \\ &= \Pr\left(\varepsilon - \widetilde{\eta} \leq r\left(\widetilde{y}_2, t\right) | \widetilde{Y}_2 = \widetilde{y}_2\right). \end{aligned}$$

5.2. A TRANSFORMATION MODEL

Consider a semiparametric transformation model, of the form

$$(5.6) \qquad Y = \Lambda\left(X_1 + X_2\beta + \varepsilon\right)$$

where Λ is strictly increasing. Horowitz (1996) considered the semiparametric identification and estimation of this model, when Λ and the distribution of ε are unknown, and where ε is distributed independent of (X_1, X_2). For identification, Horowitz (1996) required that the coefficient of a variable, X_1, be specified to be some number, and that the value of the function Λ be known at one point. Matzkin (2003) considered a nonadditive version of (5.6), of the type (5.10) below, also assuming independence between ε and X. Suppose, instead, that ε is not independent of X_1. An example of such a model is where $Y = \Lambda\left(X^* + X_2\beta + \nu\right)$, X^* is unobservable, $X_1 = X^* + \kappa$ is observable, ν is independent of (X^*, X, X_2), and κ is unobservable. In such model,

$$(5.7) \qquad Y = \Lambda\left(X_1 + X_2\beta + \varepsilon\right),$$

where $\varepsilon = \nu - \kappa$ is not independent of X_1. This specification fits the model considered in Section 2.2, where

$$Y = m\left(X_2, q(X_1, \varepsilon)\right)$$

for $q(X_1, \varepsilon) = X_1 + \varepsilon$. Hence, using the results in that section, we get that under Assumptions 2.1*-2.4*, for any values x_2, x_1, e, of X_1, X_2, ε

$$(5.8) \qquad m\left(x_2, x_1 + e\right) = F^{-1}_{Y|X_1=\overline{x}_1, X_2=x_2}\left(x_1 + e - \overline{x}_1\right)$$

and

$$(5.9) \qquad F_{\varepsilon|X_1=x_1}(e) = \int F_{Y|X_1=x_1, X_2=x_2}\left(F^{-1}_{Y|X_1=\overline{x}_1, X_2=x_2}\left(x_1 + e - \overline{x}_1\right)\right) \; w(x_2)\; dx_2$$

Expressions (5.8) and (5.9) yield two main implications. First, they show that model (5.7) is identified without requiring that ε be distributed independently of (X_1, X_2). The only additional normalization that we have imposed on (5.7) in order to obtain (5.8) and (5.9) is a normalization on the conditional distribution of ε at a particular value, $\overline{x}_1$ of X_1. Second, for identification, the additive structure is not needed. Equations (5.8) and (5.9) would be the same if we specified instead of model (5.7) that

$$(5.10) \qquad Y = \widetilde{\Lambda}\left(X_2, X_1 + \varepsilon\right)$$

Imposing an additive structure would allow one to obtain more efficient estimators, if the added structure is a correct description of the relationship among the variables. But, for identification, requiring that the model satisfies (5.10) for some unknown function $\widetilde{\Lambda}$ that is strictly increasing in its last coordinate is all that is it needed. A third implication of this analysis is that the identification and estimation of this model does not require that either X_1 or X_2 be continuously distributed.

In model (5.6), we could have also specified that $\beta = 1$, so that

$$\begin{aligned} (5.11) \qquad Y &= \Lambda\left(X_1 + X_2\beta + \varepsilon\right) \\ &= \widetilde{\Lambda}\left(X_1 + X_2, \varepsilon\right) \end{aligned}$$

or, if X_2=(X_{21}, X_{22}) where $X_{22} \in R$,

$$Y = \widetilde{\Delta}\left(X_1 + X_{21}, X_{22}, \varepsilon\right)$$

The results in Section 2.2 show that, in such model, for any e and x_1

$$(5.12) \qquad F_{\varepsilon|X_1=x_1}(e) = \int \left[F_{Y|X_1=x_1,X_2=x_2}\left(F^{-1}_{Y|X_1=\overline{x}_1,X_2=x_1+x_2-\overline{x}_1}(e)\right)\right] \; w(x_2)\; dx_2$$

and for any t_1, t_2

$$\widetilde{\Lambda}(t_1,t_2) = \int \left[F^{-1}_{Y|X_1=x_1,X_2=t_1-x_1}\left(F_{\varepsilon|X_1=x_1}(e)\right)\right] \; w(x_1)\; dx_1$$

where $w(x_2)$ and $w(x_1)$ are any weight functions on x_2 and x_1, respectively, and where $F_{\varepsilon|X_1=x_1}(e)$ is given by (5.12).

6. CONCLUSIONS

We have developed techniques that use unobserved exogenous variations as instruments in the estimation of models with endogenous explanatory variables and in simultaneous equations. When the unobserved variable in the model of interest and the exogenous unobserved variation are either conditionally independent or conditionally independent at only one point, we have shown that the model is identified and can be easily estimated. This does not require estimating the unobserved exogenous variation. When the model is part of a system of simultaneous equations, we have shown how it can still be estimated using an unobservable instrument. The method that we have developed for this case requires first estimating this unobservable instrument in a first step. In either case, our easily computable estimators for these single equation models have been shown to be consistent and asymptotically normally distributed.

We have described how some of the methods developed in this paper can be used to estimate discrete choice models and transformation models with endogenous explanatory variables, by modifying existent methods for the estimation of those models.

APPENDIX A

In this Appendix, we first present the proof of Lemma 2.1. Then, we present the extensions of Lemma 2.1 to cases where independence between an unobservable instrument, η and the unobservable random term ε in the relationship of interest can be established conditional on all values of $\widetilde{X}$ (Lemma 2.2) and, even stronger, the extension to the case where η is jointly independent of $\left(\widetilde{X},\varepsilon\right)$ (Lemma 2.3).

PROOF OF LEMMA 2.1: To show that (i) implies (ii), define $s_{\overline{\widetilde{x}}}(\eta) = F^{-1}_{X|\widetilde{X}=\overline{\widetilde{x}}}(\eta)$. Then, since the distribution of X conditional on $\widetilde{X}=\overline{\widetilde{x}}$ is strictly increasing, η is distributed $U(0,1)$ and $s_{\overline{\widetilde{x}}}(\eta)$ is strictly increasing in η. (This same argument is used in Lemma 2 of McFadden and Train (2000) to show that an observable random variable can always be written as a function of another observable variable random variable and a $U(0,1)$ unobservable random term.) Since, conditional on $\widetilde{X}=\overline{\widetilde{x}}$, $\eta = F^{-1}_{X|\widetilde{X}=\overline{\widetilde{x}}}(X)$ is a function of X and since, conditional on $\widetilde{X}=\overline{\widetilde{x}}$, ε is independent of X, it follows that η is independent of ε, conditional on $\widetilde{X}=\overline{\widetilde{x}}$. This shows (ii).

To show that (ii) implies (iii), we use an argument similar to that used to show that (i) implies (ii). Define $r_{\overline{\widetilde{x}}}(\delta) = F^{-1}_{\varepsilon|\widetilde{X}=\overline{\widetilde{x}}}(\delta)$. Then, $r_{\overline{\widetilde{x}}}(\delta)$ is strictly increasing and the unobservable δ is $U(0,1)$. Since, conditional on $\widetilde{X}=\overline{\widetilde{x}}$, $\delta = r^{-1}_{\overline{\widetilde{x}}}(\varepsilon) = F_{\varepsilon|\widetilde{X}=\overline{\widetilde{x}}}(\varepsilon)$ is a function of ε, and since, by (ii), η is independent of ε, conditional on $\widetilde{X}=\overline{\overline{x}}$, it follows that δ is independent of η, conditional on $\widetilde{X}=\overline{\overline{x}}$. This shows (iii).

That (iii) implies (iv) is trivial. To show that (iv) implies (i), we note that, conditional on $\widetilde{X}=\overline{\overline{x}}$, ε is a function of δ and, conditional on $\widetilde{X}=\overline{\overline{x}}$, δ and X are independent.

LEMMA 2.2: *Suppose that assumptions 2.1 and 2.2'-2.3' are satisfied. Then, the following statements are equivalent:*

(i) ε *is independent of* X, *conditional on* $\widetilde{X}$.

(ii) There exists a strictly increasing function $s\left(\widetilde{X},\cdot\right)$ *and an unobservable random term* η *such that*

$X=s\left(\widetilde{X},\eta\right)$ *and*

η *is independent of* ε, *conditional on* $\widetilde{X}$.

(iii) There exists a strictly increasing function $r\left(\widetilde{X},\cdot\right)$ *and an unobservable random term* δ *such that*

$\varepsilon=r\left(\widetilde{X},\delta\right)$,

δ *is independent of* η, *conditional on* $\widetilde{X}$, *and*

δ *is independent of* $\left(\widetilde{X},X\right)$.

(iv) There exists a strictly increasing function $r\left(\widetilde{X},\cdot\right)$ *and an unobservable random term* δ *such that*

$\varepsilon=r\left(\widetilde{X},\delta\right)$, *and*

δ *is independent of* $\left(\widetilde{X},X\right)$.

PROOF OF LEMMA 2.2: This proof follows the same arguments as the proof of Lemma 2.1, except that instead of conditioning on $\widetilde{X}=\overline{\overline{x}}$, we condition on $\widetilde{X}=\widetilde{x}$, for arbitrary $\widetilde{x}$.

LEMMA 2.3: *Suppose that assumptions 2.1 and 2.2'-2.3' are satisfied. Then, the following statements are equivalent:*

(i) ε *is independent of* X, *conditional on* $\widetilde{X}$.

(ii) There exists a strictly increasing function $s\left(\widetilde{X},\cdot\right)$ *and an unobservable random term* η *such that*

$X=s\left(\widetilde{X},\eta\right)$ *and*

η *is independent of* $\left(\widetilde{X},\varepsilon\right)$.

(iii) There exists a strictly increasing function $r\left(\widetilde{X},\cdot\right)$ *and an unobservable random term* δ *such that*

$$\varepsilon = r\left(\widetilde{X},\delta\right),$$

$$\delta \text{ is independent of } \left(\widetilde{X},X\right).$$

PROOF OF LEMMA 2.3: We first show that (ii) implies (i), following a dual argument of Imbens and Newey (2001)'s result: Since η is independent of $\left(\varepsilon,\widetilde{X}\right)$, η is independent of ε, conditional on $\widetilde{X}$. Since $X = s\left(\widetilde{X},\eta\right)$, X is a function of η, conditional on $\widetilde{X}$. It follows that, conditional on $\widetilde{X}$, X is independent of ε. This shows (i). Next, to show that (i) implies (iii), define, for any value of $\left(X,\widetilde{X}\right)$ and any value of ε, $\widetilde{\delta} = F_{\varepsilon|X,\widetilde{X}}(\varepsilon)$ and $\delta = F_{\varepsilon|X}(\varepsilon)$. Then, both $\widetilde{\delta}$ and δ are $U(0,1)$; $\widetilde{\delta}$ is independent of $\left(X,\widetilde{X}\right)$ and δ is independent of $\widetilde{X}$. Define $\widetilde{r}\left(X,\widetilde{X},\delta\right) = F^{-1}_{\varepsilon|X,\widetilde{X}}(\delta)$ and $r\left(\widetilde{X},\delta\right)$ $= F^{-1}_{\varepsilon|\widetilde{X}}(\delta)$. Since, by (i), $F_{\varepsilon|X,\widetilde{X}} = F_{\varepsilon|\widetilde{X}}$, it follows that for all δ, $\widetilde{r}\left(X,\widetilde{X},\delta\right) = r\left(\widetilde{X},\delta\right)$. Hence, $\varepsilon = r\left(\widetilde{X},\delta\right)$ where δ is independent of $\left(X,\widetilde{X}\right)$. This shows that (i) implies (iii). To show that (iii) implies (ii), define $s\left(\widetilde{x},\eta\right) = F^{-1}_{X|\widetilde{X}=\widetilde{x}}\left(\eta\right)$. Then, s is strictly increasing in η, and $\eta = s^{-1}\left(\widetilde{x},X\right)$ is $U(0,1)$, for all $\widetilde{x}$. Since, by (iii), δ is independent of $\left(\widetilde{X},X\right)$, it follows that η is independent of δ. To show that η is independent of $\left(\widetilde{X},\varepsilon\right)$, we note that the distribution of η conditional on $\left(\widetilde{X},\varepsilon\right)$ equals the distribution of η conditional on $\left(\widetilde{X},\delta\right)$, since $\varepsilon = r\left(\widetilde{X},\delta\right)$, where r is strictly increasing in δ.Since conditional on $\widetilde{X}$, η is a function of only X, and since δ is independent of $\left(X,\widetilde{X}\right)$, it follows that the distribution of η conditional on $\left(\widetilde{X},\delta\right)$ equals the distribution of η conditional on $\widetilde{X}$. Since η is independent of $\widetilde{X}$, this implies that the distribution of η conditional on $\left(\widetilde{X},\varepsilon\right)$ equals the marginal distribution of η, and it completes the proof.

APPENDIX B: PROOF OF THEOREMS

In this Appendix, we present the proofs of the theorems in the main body of the paper.

PROOF OF THEOREM 2.1: Using statement (i) in Lemma 2.1 and the strict monotonicity of m in its last coordinate, it follows that

$$\begin{aligned}
F_{\varepsilon|\widetilde{X}=\overline{\widetilde{x}}}(e) &= \Pr\left(\varepsilon \leq e|\widetilde{X}=\overline{\widetilde{x}}\right) \\
&= \Pr\left(\varepsilon \leq e|\widetilde{X}=\overline{\widetilde{x}}, X=x\right) \\
&= \Pr\left(m\left(X,\varepsilon\right) \leq m\left(x,e\right)|\widetilde{X}=\overline{\widetilde{x}}, X=x\right) \\
&= \Pr\left(Y \leq m\left(x,e\right)|\widetilde{X}=\overline{\widetilde{x}}, X=x\right) \\
&= F_{Y|\widetilde{X}=\overline{\widetilde{x}},X=x}\left(m\left(x,e\right)\right)
\end{aligned}$$

Since the distribution of ε given $\left(X = x, \widetilde{X} = \widetilde{\overline{x}}\right)$ is strictly increasing, $F^{-1}_{Y|\widetilde{X}=\widetilde{\overline{x}},X=x}$ exists. Hence,

$$m(x,e) = F^{-1}_{Y|\widetilde{X}=\widetilde{\overline{x}},X=x}\left(F_{\varepsilon|\widetilde{X}=\widetilde{\overline{x}}}(e)\right)$$

By the strict monotonicity of m with respect to its last coordinate

$$\begin{aligned} F_{\varepsilon|X=x}(e) &= \Pr(\varepsilon \leq e|X=x) \\ &= \Pr(m(X,\varepsilon) \leq m(x,e)|X=x) \\ &= F_{Y|X=x}(m(x,e))) \end{aligned}$$

Hence,

$$\begin{aligned} F_{\varepsilon|X=x}(e) &= F_{Y|X=x}(m(x,e))) \\ &= F_{Y|X=x}\left(F^{-1}_{Y|\widetilde{X}=\widetilde{\overline{x}},X=x}\left(F_{\varepsilon|\widetilde{X}=\widetilde{\overline{x}}}(e)\right)\right) \end{aligned}$$

This complete the proof.

PROOF OF THEOREM 2.2: Since δ is independent of $\left(\widetilde{X}, X\right)$, it follows by the strict monotonicity of r and m that for any t and any values $(\widetilde{x}, x)$

$$\begin{aligned} F_\delta(t) &= \Pr(\delta \leq t) \\ &= \Pr\left(\delta \leq t|\widetilde{X} = \widetilde{x}, X = x\right) \\ &= \Pr\left(\delta \leq t|\widetilde{X} = \widetilde{x}, X = x\right) \\ &= \Pr\left(r\left(\widetilde{X},\delta\right) \leq r(\widetilde{x},t)|\widetilde{X} = \widetilde{x}, X = x\right) \\ &= \Pr\left(m\left(X, r\left(\widetilde{X},\delta\right)\right) \leq m(x, r(\widetilde{x},t))|\widetilde{X} = \widetilde{x}, X = x\right) \\ &= \Pr\left(Y \leq m(x, r(\widetilde{x},t))|\widetilde{X} = \widetilde{x}, X = x\right) \\ &= F_{Y|\widetilde{X}=\widetilde{x},X=x}(m(x, r(\widetilde{x},t))) \end{aligned}$$

By Theorem 2.1,

$$m(x, r(\widetilde{x},t)) = F^{-1}_{Y|\widetilde{X}=\widetilde{\overline{x}},X=x}\left(F_{\varepsilon|\widetilde{X}=\widetilde{\overline{x}}}(r(\widetilde{x},t))\right)$$

Combining these results, it follows that

$$F_\delta(t) = F_{Y|\widetilde{X}=\widetilde{x},X=x}\left(F^{-1}_{Y|\widetilde{X}=\widetilde{\overline{x}},X=x}\left(F_{\varepsilon|\widetilde{X}=\widetilde{\overline{x}}}(r(\widetilde{x},t))\right)\right)$$

The result in Theorem 2.2 follows from this last expression.

PROOF OF THEOREM 3.1: We first show that when ε_2 is observed, the asymptotic results regarding $\widehat{m}_1$ hold. Let F denote the joint distribution of (p,q,I,ε_2). Let $1[\cdot]=1$ if $[\cdot]$ is true, with $1[\cdot]=0$ otherwise. For any distribution function G of (p,q,I,ε_2), let g denote the marginal and joint densities, depending on the argument, and define $\widetilde{G_P}(p,I,\varepsilon_2)$ and $\widetilde{G_Q}(q,I,\varepsilon_2)$ by $\widetilde{G_P}(p,I,\varepsilon_2)=\int^p g(p',I,\varepsilon_2)\ dp'=\int 1[p'\leq p]\ g(p',I,\varepsilon_2)\ dp'$ and $\widetilde{G_Q}(q,I,\varepsilon_2)=\int^q g(q',I,\varepsilon_2)\ dq'=\int 1[q'\leq q]\ g(q',I,\varepsilon_2)\ dq'$. Let $\underline{\mathrm{C}}$ denote a compact set in R^4 that strictly includes the support of (p,q,I,ε_2). Let D denote the set of all distribution functions G of (p,q,I,ε_2) such that $g(p,q,I,\varepsilon_2)$ vanishes outside $\underline{\mathrm{C}}$, $g(p,I,\varepsilon_2)$, $g(q,I,\varepsilon_2)$ and $g(I,\varepsilon_2)$ are twice differentiable with respect to ε_2, $g(p,I,\varepsilon_2)$ is differentiable with respect to p, and $g(q,I,\varepsilon_2)$ is differentiable with respect to q. Let D_p denote the set of all functions $\widetilde{G_P}$ that are derived from some G in D. Let D_Q denote the set of all functions $\widetilde{G_Q}$ that are derived from some G in D. Let $\|G\|$ denote the sum of the sup norms of $g(p,q,I,\varepsilon_2)$, $g(I,\varepsilon_2)$, $\partial g(p,I,\varepsilon_2)/\partial\varepsilon_2$, and $\partial g(I,\varepsilon_2)/\partial\varepsilon_2$. We will use $\int^t$ to denote $\int_{-\infty}^t$.

Let t_1, t_2, and e_1 be given numbers. Define the functional $\kappa(G)$ implicitly by

$$(1)\qquad G_{p|I=t_2,\varepsilon_2=\kappa(G)}(t_1)=e_1$$

where $G_{p|I=t_2,\varepsilon_2=\kappa(G)}(t_1)$ denote the value of the conditional cdf of p given $I=t_2$, and $\varepsilon_2=\kappa(G)$. Define the functional $\Phi(G)$ by

$$(2)\qquad \Phi(G)=G^{-1}_{q|I=t_2,\varepsilon_2=\kappa(G)}(e_1)$$

Then, when G is the kernel estimator $\widehat{F}$ of the joint distribution of (p,q,I,ε_2), $\kappa(\widehat{F})$ satisfies $t_1=\widehat{\widetilde{p}}\left(t_2,\kappa(\widehat{F}),e_1\right)$ and $\Phi\left(\widehat{F}\right)=\widehat{m}_1(t_1,t_2,e_1)$. In other words, $\kappa(\widehat{F})$ is the value of ε_2 for which the estimated value of $\widetilde{p}$ equals t_1, when $I=t_2$ and $\varepsilon_1=e_1$; and $\Phi\left(\widehat{F}\right)$ is the value of the estimated value of m_1 when $p=t_1$, $I=t_2$, and $\varepsilon_1=e_1$. When G is the true distribution, F, $\widehat{\widetilde{p}}=\widetilde{p}$ and $\Phi(F)=m_1(t_1,t_2,e_1)$.

To derive the asymptotic properties of $\widehat{m}_1$, we will use a Delta method (see Newey (1994) and Ait-Sahalia (1994)). By (1), letting $G=F$ and $H\in$D be such that $\|H\|$ is sufficiently small, it follows that

$$(3)\qquad \int^{t_1} f(p',t_2,\kappa(F))\ dp'=e_1\ f(t_2,\kappa(F))$$

and

$$(4)\qquad \int^{t_1} f(p',t_2,\kappa(F+H))\ dp'+\int^{t_1} h(p',t_2,\kappa(F+H))\ dp'$$

$$=e_1\ \left(f(t_2,\kappa(F+H))+h(t_2,\kappa(F+H))\right)$$

By Assumption 3.2, the definition of D, and Taylor Theorem,

$$(5)\qquad \int^{t_1} f(p',t_2,\kappa(F+H))\ dp'$$

$$=\int^{t_1} f(p',t_2,\kappa(F))\ dp'+[\kappa(F+H)-\kappa(F)]\int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'+R_1,$$

$$(6)\qquad f(t_2,\kappa(F+H))$$

$$= f\left(t_2, \kappa(F)\right) + \tfrac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2}\left[\kappa(F+H)-\kappa(F)\right] + R_2,$$

(7) $$\int^{t_1} h(p', t_2, \kappa(F+H))\; dp'$$

$$= \int^{t_1} h(p', t_2, \kappa(F))\; dp' + \left[\kappa(F+H)-\kappa(F)\right]\int^{t_1} \tfrac{\partial h(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' + R_3,$$

(8) $$h\left(t_2, \kappa(F+H)\right)$$

$$= h\left(t_2, \kappa(F)\right) + \tfrac{\partial h(t_2,e_h)}{\partial \varepsilon_2}\left[\kappa(F+H)-\kappa(F)\right]$$

where for some a_1, a_2, a_3, and a_4, $|R_1| \leq a_1 \left\|\kappa(F+H)-\kappa(F)\right\|^2$, $|R_2| \leq a_2 \left\|\kappa(F+H)-\kappa(F)\right\|^2$, $|R_3| \leq a_3 \left\|\kappa(F+H)-\kappa(F)\right\|^2$, and $|R_4| \leq a_4 \left\|\kappa(F+H)-\kappa(F)\right\|^2$. Substituting (5)-(8) into (4) and using (3) it follows that

$$\int^{t_1} h(p', t_2, \kappa(F))\; dp' - e_1 h\left(t_2, \kappa(F)\right) + R_5$$

$$= \left[\kappa(F+H)-\kappa(F)\right]\left\{e_1\left(\tfrac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} + \tfrac{\partial h(t_2,\kappa(F))}{\partial \varepsilon_2}\right)\right\}$$

$$-\left[\kappa(F+H)-\kappa(F)\right]\left\{\int^{t_1} \tfrac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' + \int^{t_1} \tfrac{\partial h(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp'\right\}$$

where for some a_5, $|R_5| \leq a_5 \left\|\left[\kappa(F+H)-\kappa(F)\right]\right\|^2$ .Define

(9) $$D\kappa = \left[\int^{t_1} h(p', t_2, \kappa(F))\; dp' - e_1 h\left(t_2, \kappa(F)\right)\right]\cdot\left[e_1 \tfrac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \tfrac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp'\right]^{-1}$$

and

$$\begin{aligned}
R\kappa = &-\left[\int^{t_1} h(p', t_2, \kappa(F))\; dp' - e_1 h\left(t_2, \kappa(F)\right)\right]\cdot\left[\int^{t_1} \tfrac{\partial h(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' - e_1 \tfrac{\partial h(t_2,\kappa(F))}{\partial \varepsilon_2}\right]\cdot \\
&\cdot\left[e_1\left(\tfrac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} + \tfrac{\partial h(t_2,\kappa(F))}{\partial \varepsilon_2}\right) - \int^{t_1} \tfrac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2}\; dp' - \int^{t_1} \tfrac{\partial h(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp'\right]^{-1}\cdot \\
&\cdot\left[\int^{t_1} \tfrac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2}\; dp' - e_1 \tfrac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2}\right]^{-1} \\
&+R_5\cdot\left[\int^{t_1} \tfrac{\partial h(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' - e_1 \tfrac{\partial h(t_2,\kappa(F))}{\partial \varepsilon_2}\right]\cdot \\
&\cdot\left[\int^{t_1} \tfrac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2}\; dp' + \int^{t_1} \tfrac{\partial h(p',t_2,\kappa(F))}{\partial \varepsilon_2}\; dp' - e_1\left(\tfrac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} + \tfrac{\partial h(t_2,\kappa(F))}{\partial \varepsilon_2}\right)\right]^{-1}\cdot \\
&\cdot\left[\int^{t_1} \tfrac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2}\; dp' - e_1 \tfrac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2}\right]^{-1}
\end{aligned}$$

Then,

(10) $$\kappa(F+H)-\kappa(F) = D\kappa + R\kappa,$$

and, by Assumption 3.5 and the restriction that $\|H\|$ is small enough, for some a_6 and a_7,

(12) $$|R\kappa| \leq a_6 \|H\|^2 \text{ and } |D\kappa| \leq a_7 \|H\|.$$

Next, by (2),

(13) $\int^{\Phi(F)} f\left(q', t_2, \kappa(F)\right)\ dq' = e_1\ f\left(t_2, \kappa(F)\right),$

and

(14) $\int^{\Phi(F+H)} f\left(q', t_2, \kappa(F+H)\right)\ dq' + \int^{\Phi(F+H)} h\left(q', t_2, \kappa(F+H)\right)\ dq'$

$= e_1\ \left[f\left(t_2, \kappa(F+H)\right) + h\left(t_2, \kappa(F+H)\right)\right]$

By Taylor Theorem

(15) $\int^{\Phi(F+H)} f\left(q', t_2, \kappa(F+H)\right)\ dq'$

$= \int^{\Phi(F)} f\left(q', t_2, \kappa(F+H)\right)\ dq' + f\left(\Phi(F), t_2, \kappa(F+H)\right)\ \left[\Phi(F+H) - \Phi(F)\right] + R_8$

(16) $\int^{\Phi(F+H)} h\left(q', t_2, \kappa(F+H)\right)\ dq'$

$= \int^{\Phi(F)} h\left(q', t_2, \kappa(F+H)\right)\ dq'$

$+h\left(\Phi(F), t_2, \kappa(F+H)\right)\ \left[\Phi(F+H) - \Phi(F)\right] + R_9$

(17) $\int^{\Phi(F)} f\left(q', t_2, \kappa(F+H)\right)\ dq'$

$= \int^{\Phi(F)} f\left(q', t_2, \kappa(F)\right)\ dq'$

$+\left[\kappa(F+H) - \kappa(F)\right] \int^{\Phi(F)} \frac{\partial f(q', t_2, \kappa(F))}{\partial \varepsilon_2} dq' + R_{10}$

and

(18) $\int^{\Phi(F)} h\left(q', t_2, \kappa(F+H)\right)\ dq'$

$= \int^{\Phi(F)} h\left(q', t_2, \kappa(F)\right)\ dq'$

$+\left[\kappa(F+H) - \kappa(F)\right] \int^{\Phi(F)} \frac{\partial h(q', t_2, \kappa(F))}{\partial \varepsilon_2} dq' + R_{11}$

where for some a_8, a_9, a_{10}, and a_{11}, $|R_j| \le a_j \left\|H\right\|^2$ $(j = 8, 9, 10, 11)$. From (13) – (18),

$\left[\kappa(F+H) - \kappa(F)\right] \int^{\Phi(F)} \frac{\partial f(q', t_2, \kappa(F))}{\partial \varepsilon_2} dq'$

$+f\left(\Phi(F), t_2, \kappa(F+H)\right)\ \left[\Phi(F+H) - \Phi(F)\right]$

$+\int^{\Phi(F)} h\left(q', t_2, \kappa(F+H)\right)\ dq'$

$+h\left(\Phi(F), t_2, \kappa(F+H)\right)\ \left[\Phi(F+H) - \Phi(F)\right]$

$= e_1\ \left[f\left(t_2, \kappa(F+H)\right) + h\left(t_2, \kappa(F+H)\right)\right]$

$+e_1\ f\left(t_2, \kappa(F)\right) + R_{12}$

where for some a_{12}, $|R_{12}| \leq a_{12} \|H\|^2$. Using Taylor Theorem again, the above becomes

$$[\kappa(F+H)-\kappa(F)] \int^{\Phi(F)} \frac{\partial f(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq'$$

$$+ \left[f\left(\Phi(F), t_2, \kappa(F)\right) + \frac{\partial f(\Phi(F),t_2,\kappa(F))}{\partial \varepsilon_2} \left[\kappa(F+H)-\kappa(F)\right] \right] \; [\Phi(F+H)-\Phi(F)]$$

$$+ \int^{\Phi(F)} h\left(q', t_2, \kappa(F)\right) \; dq' + [\kappa(F+H)-\kappa(F)] \int^{\Phi(F)} \frac{\partial h(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq'$$

$$+ \left[h\left(\Phi(F), t_2, \kappa(F)\right) + \frac{\partial h(\Phi(F),t_2,\kappa(F))}{\partial \varepsilon_2} \left[\kappa(F+H)-\kappa(F)\right] \right] [\Phi(F+H)-\Phi(F)]$$

$$= e_1 \; \left[\frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} \left[\kappa(F+H)-\kappa(F)\right] + h\left(t_2, \kappa(F)\right) + \frac{\partial h(t_2,\kappa(F))}{\partial \varepsilon_2} \left[\kappa(F+H)-\kappa(F)\right] \right]$$

$$+ \; e_1 \; h\left(t_2, \kappa(F)\right) + R_{13}$$

where for some a_{13}, $|R_{13}| \leq a_{13} \|H\|^2$, or

$$[\Phi(F+H)-\Phi(F)] \cdot \left[f\left(\Phi(F), t_2, \kappa(F)\right) + h\left(\Phi(F), t_2, \kappa(F)\right) \right]$$

$$+ [\Phi(F+H)-\Phi(F)] \cdot [\kappa(F+H)-\kappa(F)] \cdot \left[\frac{\partial f(\Phi(F),t_2,\kappa(F))}{\partial \varepsilon_2} + \frac{\partial h(\Phi(F),t_2,\kappa(F))}{\partial \varepsilon_2} \right]$$

$$= e_1 \; \left[h\left(t_2, \kappa(F)\right) \right]$$

$$- \int^{\Phi(F)} h\left(q', t_2, \kappa(F)\right) \; dq'$$

$$- [\kappa(F+H)-\kappa(F)] \cdot \left[\int^{\Phi(F)} \frac{\partial f(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq' + \int^{\Phi(F)} \frac{\partial h(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq' \right]$$

$$+ [\kappa(F+H)-\kappa(F)] \cdot e_1 \cdot \left[\frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} + \frac{\partial h(t_2,\kappa(F))}{\partial \varepsilon_2} \right]$$

$$+R_{13}$$

Using (9) – (12), it follows that

$$[\Phi(F+H)-\Phi(F)] \cdot \left[f\left(\Phi(F), t_2, \kappa(F)\right) + h\left(\Phi(F), t_2, \kappa(F)\right) \right]$$

$$+ [\Phi(F+H)-\Phi(F)] \cdot \left[\int^{t_1} h(p', t_2, \kappa(F)) \; dp' - e_1 h\left(t_2, \kappa(F)\right) \right]$$

$$\cdot \left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' \right]^{-1} \cdot \left[\frac{\partial f(\Phi(F),t_2,\kappa(F))}{\partial \varepsilon_2} + \frac{\partial h(\Phi(F),t_2,\kappa(F))}{\partial \varepsilon_2} \right]$$

$$= e_1 \; h\left(t_2, \kappa(F)\right)$$

$$- \int^{\Phi(F)} h\left(q', t_2, \kappa(F)\right) \; dq'$$

$$- \left[\int^{t_1} h(p', t_2, \kappa(F)) \; dp' - e_1 h\left(t_2, \kappa(F)\right) \right] \cdot \left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' \right]^{-1}$$

$$\cdot \left[\int^{\Phi(F)} \frac{\partial f(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq' + \int^{\Phi(F)} \frac{\partial h(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq' \right]$$

$$+\left[\int^{t_1} h(p',t_2,\kappa(F))\ dp' - e_1 h\left(t_2,\kappa(F)\right)\right]\cdot\left[e_1\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2} - \int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'\right]^{-1}$$

$$\cdot e_1\cdot\left[\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2}+\frac{\partial h(t_2,\kappa(F))}{\partial\varepsilon_2}\right]$$

$$+R_{14}$$

where for some a_{14}, $|R_{14}|\le a_{14}\left\|H\right\|^2.$ Hence,

$$[\Phi(F+H)-\Phi(F)]\cdot[f\left(\Phi(F),t_2,\kappa(F)\right)+h\left(\Phi(F),t_2,\kappa(F)\right)]\left[e_1\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2} - \int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'\right]$$

$$+[\Phi(F+H)-\Phi(F)]\cdot\frac{\partial f(\Phi(F),t_2,\kappa(F))}{\partial\varepsilon_2}\left[e_1 h\left(t_2,\kappa(F)\right)-\int^{t_1} h(p',t_2,\kappa(F))\ dp'\right]$$

$$= e_1\ h\left(t_2,\kappa(F)\right)\left[e_1\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2} - \int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'\right]$$

$$-\int^{\Phi(F)} h\left(q',t_2,\kappa(F)\right)\ dq'\ \left[e_1\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2} - \int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'\right]$$

$$-\int^{\Phi(F)}\frac{\partial f(q',t_2,\kappa(F))}{\partial\varepsilon_2}dq'\ \left[\int^{t_1} h(p',t_2,\kappa(F))\ dp' - e_1 h\left(t_2,\kappa(F)\right)\right]$$

$$+e_1\ \frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2}\ \left[\int^{t_1} h(p',t_2,\kappa(F))\ dp' - e_1 h\left(t_2,\kappa(F)\right)\right]$$

$$+R_{15}$$

It then follows that

$$[\Phi(F+H)-\Phi(F)]=$$

$$e_1\ \left[f\left(t_2,\kappa(F)\right)+h\left(t_2,\kappa(F)\right)\right]\left[e_1\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2} - \int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'\right]\ A^{-1}$$

$$-\int^{\Phi(F)} h\left(q',t_2,\kappa(F)\right)\ dq'\ \left[e_1\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2} - \int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'\right]\ A^{-1}$$

$$-\int^{\Phi(F)}\frac{\partial f(q',t_2,\kappa(F))}{\partial\varepsilon_2}dq'\ \left[\int^{t_1} h(p',t_2,\kappa(F))\ dp' - e_1 h\left(t_2,\kappa(F)\right)\right]\ A^{-1}$$

$$+e_1\ \frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2}\ \left[\int^{t_1} h(p',t_2,\kappa(F))\ dp' - e_1 h\left(t_2,\kappa(F)\right)\right]\ A^{-1}$$

$$+R_{15}$$

where

$$A=[f\left(\Phi(F),t_2,\kappa(F)\right)+h\left(\Phi(F),t_2,\kappa(F)\right)]\left[e_1\frac{\partial f(t_2,\kappa(F))}{\partial\varepsilon_2} - \int^{t_1}\frac{\partial f(p',t_2,\kappa(F))}{\partial\varepsilon_2}dp'\right]$$

$$+\frac{\partial f(\Phi(F),t_2,\kappa(F))}{\partial\varepsilon_2}\left[\int^{t_1} h(p',t_2,\kappa(F))\ dp' - e_1 h\left(t_2,\kappa(F)\right)\right]$$

Let

$$D\Phi = e_1\ h\left(t_2,\kappa(F)\right)\ C(f)\ A(f)^{-1}$$

$$-\ C(f)\ A(f)^{-1}\ \int^{\Phi(F)} h\left(q',t_2,\kappa(F)\right)\ dq'$$

$$- D(h)\ A(f)^{-1}\ \int^{\Phi(F)} \frac{\partial f(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq'$$

$$+e_1\ \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2}\ D(h)\ A(f)^{-1}$$

and

$$R\Phi = [\Phi(F+H) - \Phi(F)] - D\Phi.$$

where

$$A(f) = f(\Phi(F), t_2, \kappa(F)) \left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' \right]$$

$$C(f) = \left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' \right]$$

$$D(h) = \left[\int^{t_1} h(p', t_2, \kappa(F))\ dp' - e_1 h(t_2, \kappa(F)) \right]$$

Then,

$$D\Phi = e_1\ [h(t_2, \kappa(F))]\ [f(\Phi(F), t_2, \kappa(F))]^{-1}$$

$$-\ [f(\Phi(F), t_2, \kappa(F))]^{-1} \int^{\Phi(F)} h(q', t_2, \kappa(F))\ dq'$$

$$- D(h) A(f)^{-1}\ \int^{\Phi(F)} \frac{\partial f(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq'$$

$$+e_1\ \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2}\ D(h)\ A(f)^{-1}$$

$$= \frac{1}{f(\Phi(F),t_2,\kappa(F))} \left[e_1\ h(t_2, \kappa(F)) - \int^{\Phi(F)} h(q', t_2, \kappa(F))\ dq' \right]$$

$$- \frac{\left[e_1\ \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{\Phi(F)} \frac{\partial f(q',t_2,\kappa(F))}{\partial \varepsilon_2} dq' \right]}{f(\Phi(F),t_2,\kappa(F)) \left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp' \right]}\ \left[e_1 h(t_2, \kappa(F)) - \int^{t_1} h(p', t_2, \kappa(F))\ dp' \right]$$

By Assumptions 3.2, 3.5, and the assumption that $\|H\|$ is small enough, there exists $a_{15} > 0$ such that

$$|D\Phi| \leq a_{15} \|H\| \quad \text{and} \quad |R\Phi| \leq a_{15} \|H\|^2 .$$

By Assumptions 3.1-3.4, Lemma B.3 in Newey (1994), and the Lemma in Appendix B of Matzkin (2003), it follows that

$\Phi\left(\widehat{F}\right) \to \Phi(F)$ in probability and

$$\sqrt{N\sigma^2}\left(\Phi\left(\widehat{F}\right) - \Phi(F)\right) \to N(0, V)$$

where $V = \widetilde{V}\ \int K(I, \varepsilon_2)^2\ dI\ d\varepsilon_2$ and

$$\widetilde{V} = \frac{1}{[f(\Phi(F),t_2,\kappa(F))]^2} \int [e_1\ - 1[q' \leq \Phi(F)]]^2\ f(q', t_2, \kappa(F))\ dq'$$

$$
+\frac{\left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{\Phi(F)} \frac{\partial f\left(q',t_2,\kappa(F)\right)}{\partial \varepsilon_2} dq'\right]^2}{[f(\Phi(F),t_2,\kappa(F))]^2 \left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp'\right]^2} \int \left[e_1 - 1\left[p' \leq t\right]\right]^2 f(p',t_2,\kappa(F))\, dp'
$$

$$
= \frac{\left[\int [e_1 - 1[q' \leq \Phi(F)]]^2 \; f(q',t_2,\kappa(F)) \; dq'\right]}{[f(\Phi(F),t_2,\kappa(F))]^2} \frac{\left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp'\right]^2}{\left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp'\right]^2}
$$

$$
+ \frac{\left[\int \; [e_1 - 1[p' \leq t_1]]^2 f(p',t_2,\kappa(F)) \; dp'\right]}{[f(\Phi(F),t_2,\kappa(F))]^2} \frac{\left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{\Phi(F)} \frac{\partial f\left(q',t_2,\kappa(F)\right)}{\partial \varepsilon_2} dq'\right]^2}{\left[e_1 \frac{\partial f(t_2,\kappa(F))}{\partial \varepsilon_2} - \int^{t_1} \frac{\partial f(p',t_2,\kappa(F))}{\partial \varepsilon_2} dp'\right]^2}
$$

To show that the effect of using estimated values of ε_2 instead of actual values of ε_2, in the estimation of m_1, is asymptotically negligible, let $H = H_1 + H_2$, where $H_1 = \widehat{F}_{q,p,I,\varepsilon_2} - F_{q,p,I,\varepsilon_2}$ and $H_2 = \widehat{F}_{q,p,I,\widehat{\varepsilon}_2} - \widehat{F}_{q,p,I,\varepsilon_2}$. Using standard properties of kernel estimators, it is easy to show that under our assumptions, $\sqrt{N\sigma^2} \left\| \widehat{F}_{q,p,I,\widehat{\varepsilon}_2} - \widehat{F}_{q,p,I,\varepsilon_2} \right\| \to 0$ in probability. Hence, the asymptotic properties when $H = H_1$ are the same as those when $H = H_1 + H_2$. This completes the proof.

PROOF OF THEOREM 4.1: Since δ is required to be distributed independently of $\left(X, \widetilde{X}_1, \eta\right)$, the statement of Theorem 5.1 follows directly from the results in Roehrig (1988), by letting $\left(X, \widetilde{X}_1, \eta\right)$ be the vector X, Y_1 be the vector Y, and δ be the vector ε, in Roehrig's formulation.

APPENDIX D: USING NORMALIZATIONS ON THE FUNCTIONS

In Section 2, we obtained expressions for the unknown functions and distributions in the models, by using normalizations on the marginal and conditional distributions of random variables. We could have, instead, normalizations on the unknown functions. In this Appendix, we present the expressions that can be obtained using this alternative set of normalization, for the model considered in Section 2.3.

Let $\overline{x}$ and $\overline{\widetilde{x}}$ be fixed values of X and $\widetilde{X}$, respectively. Normalize the functions m, s, and r by requiring that they satisfy, for all ε, η and δ

$$(D.1) \qquad m\left(\overline{x}, \varepsilon\right) = \varepsilon$$

$$(D.2) \qquad s\left(\overline{\widetilde{x}}, \eta\right) = \eta$$

and

$$(D.3) \qquad r\left(\overline{\widetilde{x}}, \delta\right) = \delta$$

These normalizations are satisfied by, for example, linear functions.

THEOREM D.1: *Suppose that Assumptions 2.1, 2.2'-2.4', and 2.5 are satisfied. Suppose also that at least one of the statements in Lemma 2.2 are satisfied. Then, under normalizations (D.1)-(D.3), for any t*

$$(D.4) \qquad \Pr\left(\delta \leq t\right) = F_{Y|X=\overline{x},\widetilde{X}=\overline{\widetilde{x}}}\left(t\right)$$

for any t', t

$$(D.5) \quad m\left(t', t\right) = F^{-1}_{Y|X=t', \widetilde{X}=\widetilde{\overline{x}}}\left(F_{Y|X=\overline{x}, \widetilde{X}=\widetilde{\overline{x}}}\left(t\right)\right)$$

for any $\widetilde{x}, t$

$$(D.6) \quad r\left(\widetilde{x}, t\right) = F^{-1}_{Y|X=\overline{x}, \widetilde{X}=\widetilde{x}}\left(F_{Y|X=\overline{x}, \widetilde{X}=\widetilde{\overline{x}}}\left(t\right)\right)$$

for any x, e

$$(D.7) \quad F_{\varepsilon|X=x}\left(e\right) = F_{Y|X=x}\left(F^{-1}_{Y|X=x, \widetilde{X}=\widetilde{\overline{x}}}\left(F_{Y|X=\overline{x}, \widetilde{X}=\widetilde{\overline{x}}}\left(e\right)\right)\right)$$

and for any $\widetilde{x}, e$

$$(D.8) \quad F_{\varepsilon|\widetilde{X}=\widetilde{x}}(e) = F_{Y|X=x, \widetilde{X}=\widetilde{x}}\left(F^{-1}_{Y|X=x, \widetilde{X}=\widetilde{\overline{x}}}\left(F_{Y|X=\overline{x}, \widetilde{X}=\widetilde{\overline{x}}}\left(e\right)\right)\right)$$

REFERENCES

AI, C. and X. CHEN (2003), "Efficient Estimation of Models with Conditional Moments Restrictions Containing Unknown Functions," *Econometrica*, 71.

ALTONJI, J.G. AND H. ICHIMURA (2001): "Estimating Derivatives in Nonseparable Models with Limited Dependent Variables," mimeo.

ALTONJI, J.G. AND R.L. MATZKIN (2001): "Panel Data Estimators for Nonseparable Models with Endogenous Regressors," mimeo, Northwestern University.

BERRY, S. (1994) "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25.

BERRY, S., J. LEVINSOHN, and A. PAKES (2001) "Automobile Prices in Market Equilibrium," *Econometrica*, 63.

BLUNDELL, R. AND J.L. POWELL (2003a): "Endogeneity in Nonparametric and Semiparametric Regression Models," prepared for the World Congress of the Econometric Society, Seattle, 2000. Published in *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, Vol, II*, ed. by M. Dewatripont, L.P. Hansen, and S.J. Turnovsky, Cambridge: Cambridge University Press.

____________ (2003b) "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, forthcoming.

BLUNDELL, R. and R. SMITH (1986), "An Exogeneity Test for a Simultaneous Equation Tobit Model with and Application to Labor Supply," *Econometrica*, 54.

BLUNDELL, R. and R. SMITH (1989), "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies,* 56, 37-58.

BRIESCH, R., P. CHINTAGUNTA, AND R.L. MATZKIN (1997): "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," mimeo, Northwestern University.

BROWN, B. W. (1983): "The Identification Problem in Systems Nonlinear in the Variables," *Econometrica*, 51, 175-196.

______________ (1985): "The Identification Problem in Simultaneous Equation Models with Identities," *International Economic Review*, 26, 1, 45-66.

BROWN, D.J. AND R.L. MATZKIN (1996): "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," mimeo, first version, Northwestern University.

BROWN, D.J. AND R.L. MATZKIN (2004): "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," mimeo, Northwestern University.

BROWN, D.J. AND M. WEGKAMP (2002): "Weighted Minimum Mean-Square Distance from Independence Estimation," *Econometrica,* 70 (5), 2035-2051.

CARD, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, Vol. 69, 1127-1160.

CARNEIRO, P., K. HANSEN, and J. HECKMAN (2003), "Educational Attainment and Labor Market Outcomes: Estimating Distributions of the Returns to Educational Interventions," *International Economic Review*.

CHAMBERLAIN, G. (1977): "An Instrumental Variable Interpretation of Identification in Variance Components and MIMIC Models," in Paul Taubman, ed. *Kinometrics: Determinants of Socio-economic Success Within and Between Families,* Amsterdam: North Holland.

CHAMBERLAIN, G. (1984): "Panel Data," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. Intriligator. Amsterdam, North Holland, 1247-1318.

CHAMBERLAIN, G. AND Z. GRILICHES (1975): "Unboservables with a Variance-Component Structure: Ability, Schooling, and the Economic Success of Brothers," *International Economic Review,*, v16, n2, 422-449.

CHERNOZHUKOV, V. AND C. HANSEN (2001): "An IV Model of Quantile Treatment Effects," mimeo, MIT.

CHESHER, A. (2001): "Quantile Driven Identification of Structural Derivatives," cemmap working paper # CWP08/01.

____________. (2002a): "Local Identification in Nonseparable Models," cemmap working paper # CWP05/02.

____________ (2002b): "Instrumental Values," cemmap working paper # CWP17/02.

____________ (2003): "Identification in Nonseparable Models," *Econometrica*, Vol. 71, No. 5.

DAROLLES, S., J.P. FLORENS, and E. RENAULT (2002), "Nonparametric Instrumental Regression," mimeo, IDEI, Toulouse.

DAS, M. (2001): "Monotone Comparative Statics and the Estimation of Behavioral Parameters," mimeo, Department of Economics, Columbia University.

DAWID, A.P. (1979): "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B (Methodological),* Vol. 41, No. 1, 1-31.

FLORENS, J.P., J.J. HECKMAN, C. MEGHIR, and E. VYTLACIL (2003), "Instrumental Variables, Local Instrumental Variables and Control Functions," mimeo, UCL.

GOLDBERGER, A.S. (1972): "Structural Equation Methods in the Social Science," *Econometrica,* Vol. 40, No. 6, 979-1001.

HALL, P. and J.L. HOROWITZ (2003), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," mimeo, Northwestern University.

HAUSMAN, J.A. (1983): "Specification and Estimation of Simultaneous Equation Models," in *Handbook of Econometrics*, Vol. I, edited by Z. Griliches and M.D. Intriligator, North Holland.

HAUSMAN, J.A. and W.E. TAYLOR (1983): "Identification in Linear Simultaneous Equations Models with Covariance Restrictions: An Instrumental Variables Interpretation," *Econometrica*, Vol. 51, No. 5, 1527-1550.

HAUSMAN, J.A., W. NEWEY, and W.E. TAYLOR (1987): "Efficient estimation and Identification of Simultaneous Equations Models with Covariance Restrictions," *Econometrica*, Vol. 55, 849-874.

HECKMAN, J. (1976); "Simultaneous Equations Models with Continuous and Discrete Endogenous Variables and Structural Shifts," in *Studies in Nonlinear Estimation*, edited by S. Goldfeld and R. Quandt; Cambridge, Mass.: Ballinger.

HECKMAN, J (1978): "Dummy Endogenous Variables in a Simultaneous Equations System," *Econometrica*, 46, 931-61.

HECKMAN, J (1979): "Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153-61..

HECKMAN, J. (1980); "Addendum to Sample Selection Bias as a Specification Error," in *Evaluation Studies*, vol. 5, edited by E. Stromsdorfer and G. Farkas. San Francisco: Sage.

HECKMAN, J (1981): "Statistical Models for Discrete Panel Data," in C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications,"* M.I.T. Press: 1981.

HECKMAN, J., H. ICHIMURA, and P. TODD (1997): "Matching as an Econometric Evaluation Estimator: Evidence From Evaluating a Job Training Programme," *The Review of Economic Studies,* 64, No. 4, 605-654.

HECKMAN, J., H. ICHIMURA, and P. TODD (1998): "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies,* 65, No. 2, 261-294.

HECKMAN, J., R.L. MATZKIN and L. NESHEIM (2002): “Nonparametric Estimation of Nonadditive Hedonic Models,” mimeo, Northwestern University.

HECKMAN, J. and S. NAVARRO-LOZANO (2001): "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models," mimeo, University of Chicago.

HECKMAN, J., and R. ROBB (1985), "Alternative Methods for Evaluating the Impacts of Interventions," in J.J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data,* Economeric Society Monograph 10, Cambridge: Cambridge University Press.

HECKMAN, J., and J. SCHEINKMAN (1987): "The Importance of Bundling in a Gorman-Lancaster Model of Earnings," *The Review of Economic Studies,* 54, No. 2, 243-255.

HECKMAN, J.J. AND E.J. VYTLACIL (1999): “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects,” *Proceedings of the National Academy of Science*, Vol 96.

HECKMAN, J.J. AND E.J. VYTLACIL (2000): “Structural Equations, Treatment Effects and Econometric Policy Evaluation," forthcoming in *Econometrica*.

____________ (2001): “Local Instrumental Variables,” in *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, ed. by C. Hsiao, K. Morimune, and J. Powell, Cambridge: Cambridge University Press.

HOROWITZ, J.L. (1996): “Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable,” *Econometrica*, 64, 103-137.

HOROWITZ, J.L.and B. LEE (2002): “Semiparametric Estimation of a Panel Data Proportional Hazards Model with Fixed Effects," CEMMAP working paper.

HSIAO, C. (1983): "Identification," in *Handbook of Econometrics*, Vol. I, edited by Z. Griliches and M.D. Intriligator, North Holland.

IMBENS, G.W. AND W.K. NEWEY (2001): “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” mimeo, M.I.T.

JORESKOG, K.G. (1970), "A General Method for Analysis of Covariance Structures," *Biometrika,* Vol. 57, No. 2, 239-251.

KLEIN, R. and F. VELLA (2003), "Identification and Estimation of the Triangular Simultaneous Equations Model in the Absence of Exclusion Restrictions Through the Presence of Heteroskedasticity," mimeo, European University Institute.

KOENKER, R. and G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33-50.

LEWBEL, A. (1997): "Constructing Instruments for Regressions With Measurement Error When No Additional Data are Available, with An Application to Patents and R&D," *Econometrica*, Vol. 65, No. 5, 1201-1213.

LEWBEL, A. (2003): "Identification of Endogenous Heteroskedastic Models," mimeo, Boston College.

MA, L. and R. KOENKER (2003): "Quantile Regression Methods for Recursive Structural Equation Models," mimeo, University of Illinois, Urbana-Champaign.

MANSKI, C.F. (1983) "Closest Empirical Distribution Estimation," *Econometrica*, 51(2), 305–320.

MATZKIN, R.L. (1999): "Nonparametric Estimation of Nonadditive Random Functions," mimeo, first version, Northwestern University.

MATZKIN, R.L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, Vol. 71, No. 5, 1339-1375.

McFADDEN, D. (1984) "Econometric Analysis of Qualitative Response Models," in *Handbook of Econometrics,* Vol. II, Z. Griliches and M. Intriligator, eds. Amsterdam, North-Holland,

McFADDEN, D. (1986) "The Choice Theory Approach to Market Research," *Marketing Science,* Vol. 5, No. 4, Special Issue on Consumer Choice Models, 275-297.

McFADDEN, D. and K. TRAIN (2000) "Mixed MNL Models for Discrete Response" mimeo, University of California at Berkeley.

NEWEY, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics,* 79, 147-168.

NEWEY, W.K. and J.L. POWELL (1989): "Instrumental Variables Estimation for Nonparametric Models," mimeo, Princeton University.

__________________ (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1557-1569..

NEWEY, W.K., J.L. POWELL, and F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models", *Econometrica* 67, 565-603.

NG, S. and J. PINKSE (1995): "Nonparametric Two Step Estimation of Unknown Regression Functions when the Regressors and the Regressor Error are not Independent," mimeo, University of Montreal.

OLLEY, G.S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, Vol. 64, 6, 1263-1297.

PETRIN, A. and K. TRAIN (2002): "Omitted Product Attributes in Discrete Choice Models," mimeo, University of California, Berkeley.

PINKSE, J. (2000): "Nonparametric Two-Step Regression Estimation when Regressors and Errors are Dependent," *Canadian Journal of Statistics,*, 28-2, 289-300.

PUDNEY, S.E. (1982): "Estimating Latent Variable Systems When Specification is Uncertain: Generalized Component Analysis and the Eliminant Method," *Journal of the American Statistical Association*, 77, 883-889.

RIVERS, D. and Q.H. VUONG (1988): "Limited Information Estimators and Exogneity Tests for Simultaneous Probit Models," *Journal of Econometrics,* 39, pp. 347-366.

ROEHRIG, C.S. (1988): “Conditions for Identification in Nonparametric and Parametric Models”, *Econometrica*, 56, 433-447.

ROSENBAUM, P.R. and D.B. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika,* 70, 41-55.

VYTLACIL, E. and N. YILDIZ (2004): “Semiparametric Identification of the Average Treatment Effect in Nonseparable Model,” mimeo, Stanford University.

VYTLACIL, E. (2003): “Dummy Endogenous Variables in Weakly Separable Models,” mimeo, Stanford University.