ON INDEPENDENCE CONDITIONS IN NONSEPARABLE MODELS: OBSERVABLE AND UNOBSERVABLE INSTRUMENTS

Rosa L. Matzkin*

Department of Economics
University of California - Los Angeles

June 2015

ABSTRACT

This paper develops identification results employing independence conditions among unobservable variables. The independence conditions are used to derive first-stage nonseparable reduced form functions. Once constructed, these reduced form functions are employed to express the derivatives of nonseparable structural functions in terms of the derivatives of the reduced form functions. For models with simultaneity, we obtain the new results by combining the independence assumptions together with parametric specifications and exclusion restrictions. For models with triangularity, we allow all functions to be nonparametric and nonseparable in unobservable random terms. For the latter, we provide several equivalence results and discuss some of the trade-offs between observable and unobservable instruments.

*The support of NSF through grants SES-0241858 and SES-1062090 is gratefully acknowledged. I benefitted from the comments of Esfandiar Maasoumi and a referee, as well as from participants at the 2014 Conference of the Society for Economic Measurement, the Microdata Methods and Practice: A cemmap Celebration Conference, and the 2015 Cowles Summer Conference on Heterogeneous Agents and Microeconometrics. Sections 2 and 5 are based on results previously circulated in the working paper entitled "Unobservable Instruments."

1. Introduction

We consider the use of independence conditions among unobservable variables as a source of identifying restrictions in nonparametric models with endogeneity. We show how to exploit the independence conditions to construct first-stage reduced form functions, which we then use to identify the derivatives of nonparametric, nonseparable functions of interest. We develop the methods for two semiparametric models with simultaneity and for a nonparametric triangular model.

In linear simultaneous equation models with additive unobservable random terms, identification is usually achieved by exclusion restrictions, equality restrictions, or covariance restrictions (see Hausman (1983) and Hsiao (1983) for surveys). In nonparametric models with simultaneity, the most commonly used methods proceed by employing observable variables that are excluded from the equations of interest, correlated with the endogenous variables, and independent of unobservable random terms. This limits the possibilities when such instruments are not observed. The methods that we develop can be considered as possible alternatives in such cases, where instruments that can be used to identify a function of interest are not observed. The model then becomes one with independence restrictions among unobservable variables. We consider triangular models with all equations in the system being nonparametric and nonseparable in unobservable random terms, and simultaneous equation models where the function of interest is nonparametric and nonseparable in an unobservable random term, while the other equations in the system are linear in parameters and additive in unobservable random terms. The specific restrictions that we impose in the simultaneous equations models allow us to derive first-stage reduced form functions without identifying first the structural function of interest. We can then use those identified reduced form functions in the same way as in triangular models, to identify the derivatives of the function of interest.

The approach developed in this paper to identify triangular and simultaneous equations models is an extension of Chesher's (2003) approach. Chesher (2003) showed how to constructively identify the derivatives of a nonparametric and nonseparable function in a triangular model by estimating first stage conditional quantile functions, given exogenous variables, for the endogenous variables in the equation of interest. The derivatives of the function of interest were then obtained as functions of the derivatives of the conditional quantile functions.¹ Instead of using observable instruments, as in Chesher (2003), we use unobservable ones. We extend Chesher's approach to identify functions in simultaneous equations by identifying reduced form functions using Matzkin (2003, 2008, 2010).

For models with simultaneity, other approaches can also be used to exploit independence

¹Previously, using conditional expectations rather than conditional quantiles, Newey, Powell, and Vella (1999) employed a similar approach to identify a separable nonparametric function in a triangular model.

among the unobservable random terms in a system. Two possible such alternative methods are the conditional density method (Matzkin (2007a, 2008, 2010, 2015), Berry and Haile (2011, 2014), Blundell, Kristensen, and Matzkin (2013)) and nonparametric IV (Newey and Powell (2003), Ai and Chen (2003), Chernozhukov and Hansen (2005), Hall and Horowitz (2005), Darolles, Fan, Florens, and Renault (2011), Chen and Pouzo (2012), Chen, Chernozhukov, Lee, and Newey (2014), and several more recent works.) Applying these methods would usually require a different set of assumptions than the ones we use in this paper.

We consider two models with simultaneity. The first model is a two equation system. In this system, the equation of interest has no excluded observable variables, but it includes as one of its arguments an observable variable that can be used as an instrument to identify the second equation. This allows one to identify the unobservable random term of the second equation, which is then used as an instrument to identify the equation of interest. Hausman and Taylor (1983) showed identification of the functions in such system, through covariance restrictions, when the first and second functions were linear in parameters and additive in the random terms. We allow the function of interest to be nonparametric and nonadditive in the unobservable random term, and show how it can be identified using the distribution of observable and identified unobservable variables. As an application, suppose we are interested in estimating the market demand function for a product. We observe income of the consumers but possess no observations on input prices or other variables that could shift the supply equation. The assumption that those unobserved variables are independent of the unobservable in the demand equation would in many cases be satisfied. In such cases, we could use the method in this paper to identify an unobservable variable representing the effect of those unobservable supply shifters, and use them to identify demand using conditional distribution functions.

Our second model is one with three endogenous variables and no observable exogenous variables. We use the mutual independence of the unobservable random terms together with exclusion restrictions in the equations and functional restrictions in two of the equations to identify a nonadditive function of interest. The model is a nonparametric extension of the last example in Hausman and Taylor (1983). In that example, the first endogenous variable is a linear function of a second endogenous variable and a random term, the second endogenous variable is a linear function of a third endogenous variable and another random term, and the third endogenous variable is a linear function of the first endogenous variable and another random term. The model is suitable for analyzing, for example, the actions of members in a group, where each one is influenced by a different member of the group. We keep the linear assumptions on two of the functions but let the third function be nonparametric and nonadditive. We adapt the methods in Matzkin (2010, Section 4) to identify the unobservable random terms of the linear equations, employing expressions for the unknown

parameters in the linear functions in terms of logs of the density of the observable variables. We then use these two unobservable random terms to construct reduced-form functions for the endogenous variables that are used to identify the nonseparable function of interest.

Our results for triangular models do not impose either parametric or separability conditions. We show that conditional independence among the unobservable variables in the system, at a unique value of the conditioning variable, provides identification of the derivatives of the nonparametric, nonseparable function of interest. An alternative, commonly used, identification method when independence conditions among the unobservable variables in the system are not satisfied, is the control function approach (Heckman and Robb (1985)). In nonseparable models, this requires a variable that is independent of the vector of unobservable random variables in the system. This observable instrument is used to identify an unobservable variable, excluded from the equation of interest. The identified unobservable variable is then used as an additional conditioning variable in the identification of the equation of interest. The control function approach has been used in semiparametric and nonparametric models by Ng and Pinkse (1995), Newey, Powell and Vella (1999), Pinkse (2000), Chesher (2003), and Imbens and Newey (2009), among others. Florens, Heckman, Meghir, and Vytlačil (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 (2004) used this approach to develop estimation methods for a semiparametric binary response model. (See Blundell and Powell (2003) for a survey of these and other methods used in nonparametric and semiparametric regression models. See also Rothe (2009) for a more recent method for estimating semiparametric binary response models using control functions.) Although very useful for triangular models when an instrument is observed, the control function approach can be employed for identification of models with simultaneity only in very specific cases. (See Blundell and Matzkin (2014).) More recent work on triangular models with observable instruments include Kasy (2014), who assumes that the relationship between the endogeneous variable and the instrument is monotone, and Torgovitsky (2015) and D'Haultfœuille and P. Février (2015), who allow the observable instrument to be discrete. Also for triangular models, Altonji and Matzkin (2005) provide estimators for the structural function and distribution, and for average derivatives, when the conditioning variable is endogenous, either continuously distributed or discrete.

The structure of the paper is as follows. In the next section we discuss triangular models, constructions of first-stage reduced form functions, and trade-offs between observable and unobservable instruments. In Sections 3 and 4 we present the results for models with simultaneity. In Section 5 we provide equivalence results for identification in triangular

models, and discuss applications to binary response models. Section 6 concludes.

2. Identification using first-stage reduced form functions

The objective of this section is to discuss previous results on estimation of nonseparable models, to show how they can be used to construct first-stage reduced form functions, and to demonstrate in a triangular model how these can be used to identify the derivatives of a function of interest.

Consider first a model where

$$(2.1) \quad Y_1 = m(X, \varepsilon_1)$$

and where m is strictly increasing in ε_1 . If X and ε_2 are distributed independently and if the distribution F_{ε_1} of ε_1 is strictly increasing, then by Matzkin (2003) it follows that for all x, ε_1 at which the conditional distribution of Y_1 given $X = x$ is defined,

$$(2.2) \quad m(x, \varepsilon_1) = F_{Y_1|X=x}^{-1}(F_{\varepsilon_1}(\varepsilon_1))$$

In other words, the function m is identified, up to a monotone transformation of ε_1 , from the distribution of (Y_1, X) . The invertibility of m in ε_1 implies that given (y_1, x) there exists a unique value of ε_1 satisfying (2.1). Assuming differentiability, the partial derivative, $\partial m(x, \varepsilon_1) / \partial x$, of m with respect to x , when the value of ε_1 remains fixed at such value, is given by

$$\frac{\partial m(x, \varepsilon_1)}{\partial x} = \frac{-\frac{\partial F_{Y_1|X=x}(y_1)}{\partial x}}{\frac{\partial F_{Y_1|X=x}(y_1)}{\partial y_1}} = \frac{-\frac{\partial F_{Y_1|X=x}(y_1)}{\partial x}}{f_{Y_1|X=x}(y_1)}$$

This expression can be obtained by taking derivatives with respect to x on both sides of (2.2), and solving for $\partial m(x, \varepsilon_1) / \partial x$. (See Matzkin (1999), or Section 3.3.1 in Matzkin (2007b). Matzkin (1999) introduces an estimator for $\partial m(x, \varepsilon_1) / \partial x$ using kernels, and develops its asymptotic distribution.) Note that this derivative is independent of any normalization on the distribution of ε_1 . It is also independent on whether m is strictly increasing or strictly decreasing in ε (see the proof of Theorem 3.1 or Matzkin (2015).) Imposing a normalization either on the function m or on the distribution of ε_1 , allows one to identify both the function and the distribution of ε_1 . For example, if $F_{\varepsilon_1}(\varepsilon_1) = \varepsilon_1$, and m is strictly increasing in ε_1 , then

$$m(x, \varepsilon_1) = F_{Y_1|X=x}^{-1}(\varepsilon_1)$$

is the conditional quantile function of Y_1 given $X = x$,² and the value of ε_1 is identified by the transformation

$$\varepsilon_1 = F_{Y_1|X=x}(Y_1).$$

Suppose next that the model is

$$(2.3) \quad Y_1 = m(Y_2, \varepsilon_1)$$

where m is strictly increasing in ε_1 and Y_2 and ε_2 are *not* distributed independently. Suppose also that for some observable Y_3 , unobservable ε_2 , and some strictly increasing function $s(y_3, \cdot)$,

$$(2.4) \quad Y_2 = s(Y_3, \varepsilon_2)$$

with ε_2 independent of ε_1 , conditional on $Y_3 = \bar{y}_3$. By the above arguments,

$$F_{\varepsilon_2|Y_3=\bar{y}_3}(\varepsilon_2) = F_{Y_2|Y_3=\bar{y}_3}(Y_2)$$

and

$$s(\bar{y}_3, \varepsilon_2) = F_{Y_2|Y_3=\bar{y}_3}^{-1}(F_{\varepsilon_2|Y_3=\bar{y}_3}(s(\bar{y}_3, \varepsilon_2)))$$

For simplicity, we will denote $F_{\varepsilon_2|Y_3=\bar{y}_3}(\varepsilon_2)$ by ε_2 . Since then, conditional on $Y_3 = \bar{y}_3$, the joint distribution of (Y_1, ε_2) is identified, we can also identify the reduced form function $h^1(\bar{y}_3, \varepsilon_2, \varepsilon_1)$, which assigns a value of y_1 to each $(\varepsilon_2, \varepsilon_1)$, when $Y_3 = \bar{y}_3$:

$$h^1(\bar{y}_3, \varepsilon_2, \varepsilon_1) = F_{Y_1|Y_3=\bar{y}_3, \varepsilon_2}^{-1}(\varepsilon_1)$$

Again in this expression, we have denoted $F_{\varepsilon_1|Y_3=\bar{y}_3}(\varepsilon_1)$ by ε_1 . The functions $s(\bar{y}_3, \cdot)$ and $h^1(\bar{y}_3, \cdot, \cdot)$, which depend only on variables determined outside the system, are first-stage reduced form functions. To show how they can be used to identify the derivatives of the function m , substitute Y_1 by $h^1(\bar{y}_3, \varepsilon_2, \varepsilon_1)$ and Y_2 by $s(\bar{y}_3, \varepsilon_2)$ in (2.3). We then get

$$h^1(\bar{y}_3, \varepsilon_2, \varepsilon_1) = m(s(\bar{y}_3, \varepsilon_2), \varepsilon_1)$$

We will assume that the functions $s(\bar{y}_3, \cdot)$ and $m(\cdot, \cdot)$ are twice continuously differentiable and the density of $(\varepsilon_1, \varepsilon_2)$ conditional on $Y_3 = \bar{y}_3$ is everywhere positive and differentiable. Under such conditions, the derivatives of $h^1(\bar{y}_3, \cdot, \cdot)$ and $s(\bar{y}_3, \cdot)$ exist and are identified. Taking derivatives with respect to ε_2 one gets

$$h_{\varepsilon_2}^1(\bar{y}_3, \varepsilon_2, \varepsilon_1) = m_{y_2}(s(\bar{y}_3, \varepsilon_2), \varepsilon_1) s_{\varepsilon_2}(\bar{y}_3, \varepsilon_2)$$

²See Koenker (2005) for an extensive treatment of conditional quantile functions

where subscripts denote the variable over which the derivative is taken. In particular, given (y_1, y_2) , let ε_1 denote the unique value satisfying

$$y_1 = m(y_2, \varepsilon_1)$$

Then, the partial derivative of m with respect to y_2 evaluated at (y_2, ε_1) is

$$\frac{\partial m(y_2, \varepsilon_1)}{\partial y_2} = \frac{h_{\varepsilon_2}^1 \left(\bar{y}_3, F_{Y_2|Y_3=\bar{y}_3}(y_2), F_{Y_1|Y_3=\bar{y}_3, \varepsilon_2=F_{Y_2|Y_3=\bar{y}_3}(y_2)}(y_1) \right)}{s_{\varepsilon_2}(\bar{y}_3, F_{Y_2|Y_3=\bar{y}_3}(y_2))}$$

Replacing $h_{\varepsilon_2}^1$ and s_{ε_2} by their expressions in terms of the conditional distribution functions used to identify them, one gets

$$\frac{\partial m(y_2, \varepsilon_1)}{\partial y_2} = \frac{-f_{Y_2|Y_3=\bar{y}_3}(y_2)}{f_{Y_1|Y_3=\bar{y}_3, \varepsilon_2=F_{Y_2|Y_3=\bar{y}_3}(y_2)}(y_1)} \frac{\partial F_{Y_1|Y_3=\bar{y}_3, \varepsilon_2=F_{Y_2|Y_3=\bar{y}_3}(y_2)}(y_1)}{\partial \varepsilon_2}$$

We state this result in next proposition:

Proposition 2.1: *Suppose that the model is given by (2.3) and (2.4), where $s(\bar{y}_3, \cdot)$ and $m(\cdot, \cdot)$ are twice continuously differentiable, s is strictly increasing in ε_2 and m is strictly increasing in ε_1 , and where the density of ε_1 and ε_2 conditional on $Y_3 = \bar{y}_3$ is everywhere positive and differentiable. Let $y_1 = m(y_2, \varepsilon_1)$. Then,*

$$\frac{\partial m(y_2, \varepsilon_1)}{\partial y_2} = \frac{-f_{Y_2|Y_3=\bar{y}_3}(y_2)}{f_{Y_1|Y_3=\bar{y}_3, \varepsilon_2=F_{Y_2|Y_3=\bar{y}_3}(y_2)}(y_1)} \frac{\partial F_{Y_1|Y_3=\bar{y}_3, \varepsilon_2=F_{Y_2|Y_3=\bar{y}_3}(y_2)}(y_1)}{\partial \varepsilon_2}$$

where $\varepsilon_2 = F_{Y_2|Y_3=\bar{y}_3}(Y_2)$.

Note that, in contrast to control function methods, in model (2.3)-(2.4), conditioning on only one value of Y_3 is enough to identify the derivatives of the function m for all values of (y_1, y_2) . We next discuss the trade-offs between these two approaches. Consider again the nonseparable model

$$(2.5) \quad Y_1 = m(Y_2, \varepsilon_1)$$

where the unknown function m is differentiable and strictly increasing in ε_1 , the variables Y_1 and Y_2 are observable, and the variable ε_1 is unobservable. Consider two alternative models for Y_2 . In the first model

$$(2.6) \quad Y_2 = v(Z, \varepsilon_2)$$

where Z is an observable variable, ε_2 is an unobservable variable, the unknown function v is strictly increasing in ε_2 , and Z is distributed independently of $(\varepsilon_1, \varepsilon_2)$. In the second model

$$(2.7) \quad Y_2 = s(Y_3, \eta)$$

where η is an unobservable random variable, Y_3 is an observable variable, the unknown function s is strictly increasing in η , and η is distributed independently of (ε_1, Y_3) .

The independence between Z and $(\varepsilon_1, \varepsilon_2)$, in model (2.6), implies that conditional on ε_2 , ε_1 is independent of Z ; while the independence between η and (ε_1, Y_3) , in model (2.7), implies that conditional on Y_3 , ε_1 is independent of η . Substituting model (2.6) into (2.5), we get

$$Y_1 = m(v(Z, \varepsilon_2), \varepsilon_1)$$

This makes it clear that conditional on ε_2 , Y_2 , as a function of Z , and ε_1 provide separate, independent variation in the coordinates of m . Similarly, substituting model (2.7) into (2.5), we get

$$Y_1 = m(s(Y_3, \eta), \varepsilon_1)$$

which makes it clear that conditional on Y_3 , Y_2 , as a function of η , and ε_1 provide separate, independent variation in the coordinates of m .

Hence, both (2.5)–(2.6) and (2.5)–(2.7) are very similar, with η acting as Z and Y_3 acting as ε_2 . The main difference between (2.6) and (2.7) is that in (2.6), v is strictly increasing in the "endogenous" ε_2 , while in (2.7), s is strictly increasing in the "exogenous" η . Specification (2.6) does not require monotonicity conditions on the instrument, while specification (2.7) does. On the other hand, if the specification (2.6) is used, the conditioning variable ε_2 has to be estimated in a first stage. If the specification (2.7) is used, there is no need to estimate the conditioning variable, Y_3 , since it is observed. In particular, specification (2.7) allows to identify the function m everywhere, by weakening the independence assumption between η and (ε_1, Y_3) to the assumption that for some known value $\bar{y}_3$ in the support of Y_3 , ε_1 and η are independent conditional on $Y_3 = \bar{y}_3$.

In Section 5, we provide conditions that are equivalent to specification (2.7) and that in particular applications may be easier to justify. In the next two sections, we extend the use of first-stage reduced form functions to show constructive, recursive identification, in semiparametric models with simultaneity. In Section 3, we consider the two equation model

$$(2.8) \quad \begin{aligned} Y_1 &= m^1(Y_2, X, \varepsilon_1) \\ Y_2 &= \beta Y_1 + \varepsilon_2 \end{aligned}$$

where the function of interest, m^1 , is strictly increasing in ε_1 but otherwise unknown, $Y_1, Y_2, X \in R$ are observable, ε_1 and ε_2 are unobservable, and $\varepsilon_1, \varepsilon_2$ and X are mutually independent. As in the triangular model, the identification approach proceeds by first obtaining first-stage reduced-form functions, $h^1(x, \varepsilon_2, \varepsilon_1)$ and $h^2(x, \varepsilon_2, \varepsilon_1)$, to replace Y_1 and Y_2 in the function of interest:

$$h^1(x, \varepsilon_2, \varepsilon_1) = m^1(h^2(x, \varepsilon_2, \varepsilon_1), x, \varepsilon_1)$$

In the next step, derivatives are taken on both sides, with respect to ε_2 and x . This allows one to obtain expressions for $m_{y_2}^1$ and m_x^1 in terms of the derivatives of the identified functions h^1 and h^2 . While the derivatives of m^1 can be identified with a less restrictive structure on the second equation, by adapting for example the results in Matzkin (2010, 2015), such modification would require more restrictive conditions on the other elements of the model.

In Section 4, we consider a model with no observable instrument but with exclusion restrictions in the endogenous variables. The model is

$$\begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ Y_2 &= \beta Y_3 + \varepsilon_2 \\ Y_3 &= \gamma Y_1 + \varepsilon_3 \end{aligned}$$

where the function of interest is m^1 and where $\varepsilon_1, \varepsilon_2$ and ε_3 are mutually independent. In this model, we use the transformation of variables equation between (Y_1, Y_2, Y_3) and $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ to identify β and γ . This allows us to calculate the distributions of $(Y_1, \varepsilon_2, \varepsilon_3)$ and $(Y_2, \varepsilon_2, \varepsilon_3)$, from which we obtain reduced form functions $h^1(\varepsilon_2, \varepsilon_3, \varepsilon_1)$ and $h^2(\varepsilon_2, \varepsilon_3, \varepsilon_1)$. Substituting into the first equation, we get

$$h^1(\varepsilon_2, \varepsilon_3, \varepsilon_1) = m^1(h^2(\varepsilon_2, \varepsilon_3, \varepsilon_1), \varepsilon_1)$$

Again, taking derivatives on both sides of the equation, we obtain an expression for the derivative of m^1 in terms of derivatives of the constructed functions h^1 and h^2 . As in the previous model, the restrictive parametric assumptions can be relaxed at the cost of additional restrictions on the model. In particular, a similar identification method can be developed, under stronger conditions, when the model is specified, for some nonparametric function m^3 ,

as

$$\begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ Y_2 &= \beta_2 Y_3 + \varepsilon_2 \\ Y_3 &= m^3(Y_1) + \varepsilon_3 \end{aligned}$$

3. Identification with one included exogenous variable

We consider in this section the model

$$\begin{aligned} (3.1) \quad Y_1 &= m^1(Y_2, X, \varepsilon_1) \\ Y_2 &= \beta Y_1 + \varepsilon_2 \end{aligned}$$

where the function of interest, m^1 , is strictly increasing in ε_1 but otherwise unknown, $Y_1, Y_2, X \in R$ are observable, ε_1 and ε_2 are unobservable, and $\varepsilon_1, \varepsilon_2$ and X are mutually independent. We make the following assumptions:

Assumption 3.1: m^1 is a twice continuously differentiable function and it is strictly increasing in ε_1 .

Assumption 3.2: X, ε_1 and ε_2 are mutually independent, their densities are differentiable, and the densities of ε_1 and ε_2 are everywhere positive.

Assumption 3.3: $\beta \neq 0$ and for all $(y_2, x_2, \varepsilon_1)$, $\left(1 - \beta \frac{\partial m^1(y_2, x_2, \varepsilon_1)}{\partial y_2}\right) > 0$.

The differentiability of m^1 is needed to identify the derivatives. The invertibility of m^1 in ε_1 is used to derive a unique relationship between the distribution of ε_1 and that of Y_1 conditional on (Y_2, X) . The statistical independence between X and ε_2 allows to identify ε_2 . Of course, if there were a vector W of observable variables such that the independence condition is satisfied conditional on W , then our results would still hold, conditional on W . The statistical independence between (X, ε_2) and ε_1 allows to construct from the distributions of (Y_1, X, ε_2) and of (Y_2, X, ε_2) reduced form functions for, respectively, Y_1 and Y_2 . These reduced form functions are the nonseparable reduced form functions that are used for the identification of m^1 . Assumption 3.3 together with the monotonicity of m^1 in ε_1 guarantee

that for any value of the exogenous variables $(X, \varepsilon_2, \varepsilon_1)$, there exists a unique value of (Y_1, Y_2) satisfying model (3.1).

As an example, suppose that m^1 is the demand function for a product, where Y_1 denotes quantity demanded, Y_2 denotes price, X denotes observable income, and ε_1 denotes unobservable taste for the product. Suppose that the supply function is linear in price and additive in an unobservable productivity variable, ε_2 . In this example, the assumption that unobservable taste and unobservable productivity are independently distributed can be justified. The assumption that income is independent of unobservable taste would usually not be satisfied. However, these two variables might be independently distributed conditional on an observable variable, such as income in a previous period that was subject to different income taxes. Specifically, if for some function r and unobservable δ ,

$$X = r(W, \delta)$$

where W is income in the previous period and δ is mutually independent of ε_1 and ε_2 , then, conditional on W , the variables X , ε_1 , and ε_2 would be independently distributed. Our identification results will then hold conditional on W . Another specification of the model that might not be satisfied in a particular application is the linearity assumption on the supply equation. This specification can be relaxed to a nonlinear or nonparametric one, as long as ε_2 can be identified under such more flexible specifications using X as an instrument. In general, this would require imposing additional restrictions on the whole system of equations.

The next theorem provides explicit equations for the partial derivatives of the function m^1 with respect to y_2 and with respect to x when ε_1 is the (unique) value that satisfies $y_1 = m^1(y_2, x, \varepsilon_1)$ and when ε_2 is obtained as the residual of the second equation. In the proof of the theorem, the independence between X and ε_2 is used to identify β using standard instrumental variable methods. This is then used to identify the value of ε_2 for any given value (y_1, y_2) of (Y_1, Y_2) . Given this, the distributions of (Y_1, X, ε_2) and of (Y_2, X, ε_2) can also be identified. We use these two distributions to construct first-stage reduced form functions, and then use the derivatives of these reduced form functions to identify the derivatives of m^1 .

Theorem 3.1: *Let (y_1, y_2, x) be such that (3.1) is satisfied for a value of $(\varepsilon_1, \varepsilon_2)$ and the density f_X of X is strictly positive at x . Let ε_1 denote the unique value satisfying*

$y_1 = m^1(y_2, x, \varepsilon_1)$. Then, if assumptions 3.1-3.3 are satisfied

$$\frac{\partial m^1(y_2, x, \varepsilon_1)}{\partial y_2} = \frac{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial y_2} \right]}{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]} \frac{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial \varepsilon_2} \right]}{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial \varepsilon_2} \right]}$$

and

$$\begin{aligned} \frac{\partial m^1(y_2, x, \varepsilon_1)}{\partial x} = & - \frac{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial x} \right]}{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]} \\ & + \frac{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial x} \right]}{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]} \frac{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial \varepsilon_2} \right]}{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial \varepsilon_2} \right]} \end{aligned}$$

Proof: Denote by r^1 the inverse of m^1 with respect to ε_1 . Then, $r^1(y_1, y_2, x)$ is such that $y_1 = m^1(y_2, x, r^1(y_1, y_2, x))$. Let $r^2(y_1, y_2) = y_2 - \beta y_1$. Then, $y_2 = \beta y_1 + r^2(y_1, y_2)$. Consider the transformation, conditional on $X = x$,

$$\begin{aligned} \varepsilon_1 &= r^1(y_1, y_2, x) \\ \varepsilon_2 &= r^2(y_1, y_2) \end{aligned}$$

The Jacobian of this transformation is

$$\begin{bmatrix} r_{y_1}^1 & r_{y_2}^1 \\ -\beta & 1 \end{bmatrix}$$

The monotonicity of m in ε_1 implies that $r_{y_1}^1 > 0$, and Assumption 3.3 implies that the determinant of this Jacobian matrix is positive. It then follows by Gale and Nikaido (1965) that there exists a reduced form system for (3.1), of the form

$$\begin{aligned} y_1 &= h^1(x, \varepsilon_1, \varepsilon_2) \\ y_2 &= h^2(x, \varepsilon_1, \varepsilon_2) \end{aligned}$$

which for each $(\varepsilon_1, \varepsilon_2)$ satisfies

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \begin{bmatrix} r^1(h^1(x, \varepsilon_1, \varepsilon_2), h^2(x, \varepsilon_1, \varepsilon_2), x) \\ h^2(x, \varepsilon_1, \varepsilon_2) - \beta h^1(x, \varepsilon_1, \varepsilon_2) \end{bmatrix}$$

By Assumption 3.1 and the Implicit Function Theorem, h^1 and h^2 are differentiable. Taking derivatives of both equations with respect to ε_1 and ε_2 gives

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} r_{y_1}^1 & r_{y_2}^1 \\ -\beta & 1 \end{bmatrix} \begin{bmatrix} h_{\varepsilon_1}^1 & h_{\varepsilon_2}^1 \\ h_{\varepsilon_1}^2 & h_{\varepsilon_2}^2 \end{bmatrix}$$

Solving for $(h_{\varepsilon_1}^1, h_{\varepsilon_1}^2)$, we get

$$\begin{bmatrix} h_{\varepsilon_1}^1 \\ h_{\varepsilon_1}^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{[r_{y_1}^1 + \beta r_{y_2}^1]} \\ \frac{\beta}{[r_{y_1}^1 + \beta r_{y_2}^1]} \end{bmatrix}$$

Assumption 3.3 and the monotonicity of m with respect to ε_1 imply that $[r_{y_1}^1 + \beta r_{y_2}^1] > 0$. Hence, h^1 is strictly increasing in ε_1 and h^2 is strictly increasing in ε_1 if β is positive and strictly decreasing in ε_1 if β is negative.

Since $\varepsilon_2 = Y_2 - \beta Y_1$ and the independence between X and ε_2 imply that β is identified, the distributions of (Y_1, X, ε_2) and of (Y_2, X, ε_2) are identified. By Assumptions 3.1 and 3.2, the densities of (Y_1, X, ε_2) and of (Y_2, X, ε_2) are differentiable.

Since ε_1 is, by Assumption 3.2, independent of (X, ε_2) , it follows by Matzkin (2003) that

$$F_{\varepsilon_1}(\varepsilon_1) = F_{Y_1|X=x, \varepsilon_2}(h^1(x, \varepsilon_1, \varepsilon_2))$$

and

$$\begin{aligned} F_{\varepsilon_1}(\varepsilon_1) &= F_{Y_2|X=x, \varepsilon_2}(h^2(x, \varepsilon_1, \varepsilon_2)) & \text{if } \beta > 0 \\ F_{\varepsilon_1}(\varepsilon_1) &= 1 - F_{Y_2|X=x, \varepsilon_2}(h^2(x, \varepsilon_1, \varepsilon_2)) & \text{if } \beta < 0 \end{aligned}$$

Taking derivatives with respect to x and ε_2 , and solving for the derivatives of h^1 and h^2 , we get

$$h_{\varepsilon_2}^1 = - \left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]^{-1} \left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial \varepsilon_2} \right]$$

$$\begin{aligned}
h_{\varepsilon_2}^2 &= - \left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial y_2} \right]^{-1} \left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial \varepsilon_2} \right] \\
h_x^1 &= - \left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]^{-1} \left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial x} \right] \\
h_x^2 &= - \left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial y_2} \right]^{-1} \left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial x} \right]
\end{aligned}$$

Hence, these derivatives are identified.

Let y_1, y_2 , and x be as given in the statement of the theorem, and let ε_1 denote the unique value such that

$$y_1 = m^1(y_2, x, \varepsilon_1)$$

Substituting $h^1(x, \varepsilon_1, \varepsilon_2)$ for y_1 and $h^2(x, \varepsilon_1, \varepsilon_2)$ for y_2 , we get

$$h^1(x, \varepsilon_1, \varepsilon_2) = m^1(h^2(x, \varepsilon_1, \varepsilon_2), x, \varepsilon_1)$$

Taking derivatives on both sides with respect to x and ε_2 , we get

$$h_x^1 = m_{y_2}^1 h_x^2 + m_x^1$$

$$h_{\varepsilon_2}^1 = m_{y_2}^1 h_{\varepsilon_2}^2$$

Then, solving for $m_{y_2}^1$ and m_x^1 in the above equations, it follows that

$$\frac{\partial m^1(y_2, x, \varepsilon_1)}{\partial y_2} = \frac{h_{\varepsilon_2}^1}{h_{\varepsilon_2}^2}$$

and

$$\frac{\partial m^1(y_2, x, \varepsilon_1)}{\partial x} = h_x^1 - \frac{h_{\varepsilon_2}^1}{h_{\varepsilon_2}^2} h_x^2$$

Substituting the expressions for the derivatives of the functions h^1 and h^2 in terms of the expressions of these functions in terms of the distribution of the observed or previous identified variables, we get

$$\frac{\partial m^1(y_2, x, \varepsilon_1)}{\partial y_2} = \frac{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial y_2} \right]}{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]} \frac{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial \varepsilon_2} \right]}{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial \varepsilon_2} \right]}$$

and

$$\begin{aligned} \frac{\partial m^1(y_2, x, \varepsilon_1)}{\partial x} &= - \frac{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial x} \right]}{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]} \\ &+ \frac{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial x} \right]}{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial y_1} \right]} \frac{\left[\frac{\partial F_{Y_1|X=x, \varepsilon_2=e_2}(y_1)}{\partial \varepsilon_2} \right]}{\left[\frac{\partial F_{Y_2|X=x, \varepsilon_2=e_2}(y_2)}{\partial \varepsilon_2} \right]} \end{aligned}$$

This completes the proof. //

4. Identification with no observable exogenous variables

Instead of imposing a functional restriction that allows to identify ε_2 using an instrument, one can identify ε_2 employing an additional equation and exclusion restrictions in the endogenous variables. We consider here the following model, where again m^1 is the function of interest, assumed to be strictly increasing in ε_1 :

$$\begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ (4.1) \quad Y_2 &= \beta Y_3 + \varepsilon_2 \\ Y_3 &= \gamma Y_1 + \varepsilon_3 \end{aligned}$$

We will make the following assumptions:

Assumption 4.1: m^1 is a twice continuously differentiable function and it is strictly increasing in ε_1 .

Assumption 4.2: ε_1 , ε_2 and ε_3 are mutually independent and their densities are continuously differentiable of order 3 and everywhere positive.

Assumption 4.3: For all (y_2, x, ε_1) , $\left(1 - \beta \gamma \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2}\right) > 0$ and $(\beta \gamma) < 0$.

Assumption 4.4: For all (y_1, y_2, y_3) in an open neighborhood, $\partial^3 \log f_{\varepsilon_2}(y_2 - \beta y_3) / \partial \varepsilon_2^3 \neq 0$ and $\partial^3 \log f_{\varepsilon_3}(y_3 - \gamma y_1) / \partial \varepsilon_3^3 \neq 0$.

The differentiability of the function m^1 and of the densities is needed to be able to establish existence of and to identify derivatives of functions and distributions. The monotonicity of m^1 in ε_1 is needed to guarantee the invertibility of m^1 in ε_1 and the existence of reduced form functions for the system of simultaneous equations. The monotonicity of m^1 together with Assumption 4.3 are imposed to guarantee the existence of unique values of (Y_1, Y_2, Y_3) for any given values of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$. Assumption 4.2 allows to transform the log of the joint density of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ into the sum of three functions, each corresponding to the log of a different marginal densities. This together with the exclusion restrictions in the structural functions guarantee that the value of the log of each marginal density does not depend on the value of all the endogenous variables. Assumptions 4.4 guarantee that higher order derivatives of the logs of the marginal densities of ε_2 and of ε_3 do not vanish everywhere. An example of a density satisfying these assumptions is $f_\varepsilon(\varepsilon) = e^{-e^{-\varepsilon}} e^{-\varepsilon}$. Assumptions 4.4 and 4.5 are testable. They correspond to analogous properties of the derivatives of the log of the density of (Y_1, Y_2, Y_3) . For identification of the derivatives of m^1 at a particular value (y_1, y_2) , they do not need to hold for all values of (y_1, y_2, y_2) . If they are satisfied for all (y_1, y_2, y_3) , it is possible to allow the third function to be a nonparametric, of the form

$$y_3 = m^3(y_1) + \varepsilon_3$$

where for some specified value $\bar{y}_1$ of Y_1 , and some known constant α , $m^3(\bar{y}_1) = \alpha$.

As an example, consider a situation where Y_1, Y_2 and Y_3 represent survey responses of a typical individual, which can be affected by the survey design of the particular questions. (See Matzkin (2007c).) Suppose Y_1 is the individual's report of the time he or she spends doing household duties, Y_2 is the individual's report of the time his or her partner spends doing household duties, and Y_3 is the individual's report of the time his or her partner spends at work. In such situation the linearity between Y_2 and Y_3 might be justified. If the assumption of mutually independent "response errors" were not satisfied, one could consider conditioning the distribution of (Y_1, Y_2, Y_3) on an observable vector of characteristics of the survey and of the individual, such that conditioning on those, the response errors would be independent. Our identification analysis would then be performed conditionally on such observable vector.

Theorem 4.1 establishes identification of the derivatives of m^1 . In the proof of the theorem, a relationship between the density of the observable variables at (y_1, y_2, y_3) and the density of the unobservable variables $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ is established, employing a transformation of variables equation as in Matzkin (2008). The linearity of the second and third structural

equations, the exclusion restrictions in the three structural functions, and the property that the joint density of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ is the multiplication of the marginal densities of ε_1 , ε_2 , and ε_3 are then used to obtain expressions for the unknown parameters β and γ . These expressions are obtained by adapting the procedure in Matzkin (2010, Section 4), where by taking derivatives of the log of the density of the observable variables and of the corresponding expression in terms of the unknown density and functions, one derives expressions for the unknown structural elements in terms of the known elements. In the model considered in this section, these lead to expressions that relate high order derivatives of the log of the density of the observable variables to high order derivatives of the log of the marginal densities of ε_1 , ε_2 , and ε_3 , and to the unknown parameters β and γ . The identification of β and of γ allows to identify the values of ε_2 and ε_3 from any value of (y_1, y_2, y_3) . These are then used to obtain a reduced form functions for Y_1 , which determines the value of Y_1 for any value of $(\varepsilon_2, \varepsilon_3, \varepsilon_1)$, and a reduced form function for Y_2 , which determines the value of Y_2 for any value of $(\varepsilon_2, \varepsilon_3, \varepsilon_1)$. The existence of such functions is guaranteed by Assumptions 4.1-4.3. Their identification from the distribution of, respectively, $(Y_1, \varepsilon_2, \varepsilon_3)$ and $(Y_2, \varepsilon_2, \varepsilon_3)$ follows by the monotonicity of the reduced form functions in ε_1 and the independence between ε_1 and $(\varepsilon_2, \varepsilon_3)$. The independence between ε_1 and $(\varepsilon_2, \varepsilon_3)$ is guaranteed by Assumption 4.2. The monotonicity of the reduced form functions in ε_1 follows from Assumptions 4.1 and 4.3. Once the reduced form functions are shown to be identified, identification of the derivatives of the function m^1 follows as in the previous result, by substituting Y_1 and Y_2 by their corresponding functions in terms of the exogenous variables, and taking derivatives on both sides of the equation $Y_1 = m^1(Y_2, \varepsilon_1)$. Although not critical for the result, the linearity of the second and third equations allows us to decrease the dimension of the arguments of the reduced functions by one. We provide the detailed proof after the statement of the theorem.

For any (y_1, y_2, y_3) and any $i, j, k \in \{1, 2, 3\}$, denote $g_{y_i, y_j, y_k}(y_1, y_2, y_3)$ by

$$g_{y_i, y_j, y_k}(y_1, y_2, y_3) = \frac{\partial^3 \log f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3)}{\partial y_i \partial y_j \partial y_k}$$

Theorem 4.1: *Suppose that Assumptions 4.1-4.4 are satisfied. Then, for any (y_1, y_2, y_3) satisfying Assumption A.4,*

$$\beta = - \frac{g_{y_3, y_3, y_2}(y_1, y_2, y_3)}{g_{y_3, y_2, y_2}(y_1, y_2, y_3)}$$

and

$$\gamma = - \frac{g_{y_3, y_1, y_1}(y_1, y_2, y_3)}{g_{y_3, y_3, y_1}(y_1, y_2, y_3)}$$

Moreover, let (y_1, y_2) be given and let ε_1 denote the unique value satisfying $y_1 = m^1(y_2, \varepsilon_1)$.

Let $\eta = \beta\varepsilon_3 + \varepsilon_2$, where $\varepsilon_3 = Y_3 - \gamma Y_2$ and $\varepsilon_2 = Y_2 - \beta Y_3$. Then, the partial derivative of m^1 with respect to y_2 evaluated at (y_2, ε_1) is identified and it is given by

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = \frac{\frac{\partial F_{Y_1|\eta}(y_1)}{\partial \eta}}{f_{Y_1|\eta}(y_1)} \frac{f_{Y_2|\eta}(y_2)}{\frac{\partial F_{Y_2|\eta}(y_2)}{\partial \eta}}.$$

Proof: Denote by $r^1(y_1, y_2)$ the value of ε_1 satisfying $y_1 = m^1(y_2, r^1(y_1, y_2))$. Following Matzkin (2008), we analyze identification employing the transformation of variables equation that determines the density of (Y_1, Y_2, Y_3) from the density of $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ and the structural functions and parameters, m^1 , β , and γ . The equation is given by

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = f_{\varepsilon_1, \varepsilon_2, \varepsilon_3}(r^1(y_1, y_2), y_2 - \beta y_3, y_3 - \gamma y_1) \left| \begin{bmatrix} r_{y_1}^1 & r_{y_2}^1 & 0 \\ 0 & 1 & -\beta \\ -\gamma & 0 & 1 \end{bmatrix} \right|$$

The determinant equals

$$r_{y_1}^1 + \beta \gamma r_{y_2}^1 = r_{y_1}^1 \left[1 - \beta \gamma \frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} \right]$$

This determinant is positive because the monotonicity of m^1 with respect to ε_1 implies that $r_{y_1}^1 > 0$, and by Assumption 4.3, the term in brackets is positive. Note that this determinant is constant over y_3 . The mutual independence of $\varepsilon_1, \varepsilon_2$ and ε_3 implies that

$$f_{\varepsilon_1, \varepsilon_2, \varepsilon_3}(r^1(y_1, y_2), y_2 - \beta y_3, y_3 - \gamma y_1) = f_{\varepsilon_1}(r^1(y_1, y_2)) f_{\varepsilon_2}(y_2 - \beta y_3) f_{\varepsilon_3}(y_3 - \gamma y_1)$$

where f_{ε_i} denotes the marginal density of ε_i ($i = 1, 2, 3$). The equation above then becomes

$$f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3) = f_{\varepsilon_1}(r^1(y_1, y_2)) f_{\varepsilon_2}(y_2 - \beta y_3) f_{\varepsilon_3}(y_3 - \gamma y_1) [r_{y_1}^1 + \beta \gamma r_{y_2}^1]$$

Following the procedure in Matzkin (2010, Section 4), we take logs on both sides and differentiate with respect to the observable variables to obtain expressions that would allow us to solve for the structural unknown elements in the model. Taking logs and differentiating with respect to y_3 gives

$$(4.1) \quad \frac{\partial \log f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3)}{\partial y_3} = -\beta \frac{\partial \log f_{\varepsilon_2}(y_2 - \beta y_3)}{\partial \varepsilon_2} + \frac{\partial \log f_{\varepsilon_3}(y_3 - \gamma y_1)}{\partial \varepsilon_3}$$

Differentiating (4.1) with respect to y_3 gives

$$(4.2) \quad \frac{\partial^2 \log f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3)}{\partial y_3^2} = \beta^2 \frac{\partial^2 \log f_{\varepsilon_2}(y_2 - \beta y_3)}{\partial \varepsilon_2^2} + \frac{\partial^2 \log f_{\varepsilon_3}(y_3 - \gamma y_1)}{\partial \varepsilon_3^2}$$

Differentiating (4.2) with respect to y_1 gives

$$(4.3) \quad \frac{\partial^3 \log f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3)}{\partial y_3^2 \partial y_1} = -\gamma \frac{\partial^3 \log f_{\varepsilon_3}(y_3 - \gamma y_1)}{\partial \varepsilon_3^3}$$

and differentiating (4.2) with respect to y_2 gives

$$(4.4) \quad \frac{\partial^3 \log f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3)}{\partial y_3^2 \partial y_2} = \beta^2 \frac{\partial^3 \log f_{\varepsilon_2}(y_2 - \beta y_3)}{\partial \varepsilon_2^3}$$

Differentiating (4.1) twice with respect to y_2 gives

$$(4.5) \quad \frac{\partial^3 \log f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3)}{\partial y_3 \partial y_2^2} = -\beta \frac{\partial^3 \log f_{\varepsilon_2}(y_2 - \beta y_3)}{\partial \varepsilon_2^3}$$

and differentiating (4.1) twice with respect to y_1 gives

$$(4.6) \quad \frac{\partial^2 \log f_{Y_1, Y_2, Y_3}(y_1, y_2, y_3)}{\partial y_3 \partial y_1^2} = \gamma^2 \frac{\partial^3 \log f_{\varepsilon_3}(y_3 - \gamma y_1)}{\partial \varepsilon_3^3}$$

Dividing (4.6) by (4.3) and (4.4) by (4.5), it follows that for (y_1, y_2, y_3) satisfying Assumption 4.4

$$\frac{g_{y_3, y_1, y_1}(y_1, y_2, y_3)}{g_{y_3, y_3, y_1}(y_1, y_2, y_3)} = -\gamma$$

and

$$\frac{g_{y_3, y_3, y_2}(y_1, y_2, y_3)}{g_{y_3, y_2, y_2}(y_1, y_2, y_3)} = -\beta$$

Hence, β and γ are identified.

Let $\eta = \beta \varepsilon_3 + \varepsilon_2$, where $\varepsilon_3 = Y_3 - \gamma Y_2$ and $\varepsilon_2 = Y_2 - \beta Y_3$. Notice that the system (4.1) can be expressed in terms of (ε_1, η) as:

$$\begin{aligned} Y_1 &= m^1(Y_2, \varepsilon_1) \\ Y_2 &= \beta \gamma Y_1 + \eta \end{aligned}$$

Since $\beta\gamma$ is identified and η and ε_1 are independently distributed, we can follow the same steps as in the model in Section 3. In particular, the derivatives with respect to η of the

reduced form functions

$$h^1(\eta, \varepsilon_1) \quad \text{and} \quad h^2(\eta, \varepsilon_1)$$

can be identified by

$$h_\eta^1 = - \frac{\frac{\partial F_{Y_1|\eta}(y_1)}{\partial \eta}}{\frac{\partial F_{Y_1|\eta}(y_1)}{\partial y_1}}$$

and

$$h_\eta^2 = - \frac{\frac{\partial F_{Y_2|\eta}(y_2)}{\partial \eta}}{\frac{\partial F_{Y_2|\eta}(y_2)}{\partial y_2}}$$

when the derivatives exist.

Let y_1, y_2 be given, and let ε_1 denote the unique value satisfying

$$y_1 = m^1(y_2, \varepsilon_1)$$

Substituting y_1 and y_2 with the reduced form functions, we get

$$h^1(\eta, \varepsilon_1) = m^1(h^2(\eta, \varepsilon_1), \varepsilon_1)$$

Differentiating with respect to η , it follows that

$$h_\eta^1 = m_{y_2}^1 h_\eta^2$$

Hence

$$m_{y_2}^1 = \frac{h_\eta^1}{h_\eta^2}$$

Substituting h_η^1 and h_η^2 by their expressions in terms of conditional distribution functions, it follows that

$$\frac{\partial m^1(y_2, \varepsilon_1)}{\partial y_2} = \frac{\frac{\partial F_{Y_1|\eta}(y_1)}{\partial \eta}}{f_{Y_1|\eta}(y_1)} \frac{f_{Y_2|\eta}(y_2)}{\frac{\partial F_{Y_2|\eta}(y_2)}{\partial \eta}}.$$

This completes the proof. //

5. Triangular models and discrete choice

In this section, we provide conditions that are equivalent to independence among unobservable random terms and that in some empirical situations may be easier to justify. In particular, we provide insight into how in nonseparable models with triangularity, indepen-

dence across unobservable variables can be used in nonseparable models to generate proxies. We concentrate on the problem of identification of the function m and the distribution of (Y_2, ε_1) in the model

$$(5.1) \quad Y_1 = m(Y_2, \varepsilon_1)$$

where $Y_1, Y_2 \in R$ are observable variables, ε_1 is unobservable, and m is an unknown function that is strictly increasing in ε_1 .³ Given an observable variable Y_3 , we show the equivalence between the existence of a function s and an unobservable random term η distributed independently of (Y_3, ε_1) such that

$$(5.2) \quad Y_2 = s(Y_3, \eta)$$

and the existence of a function r and a random term δ distributed independently of (Y_2, Y_3) such that

$$(5.3) \quad \varepsilon_1 = r(Y_3, \delta).$$

The equivalence is important because a particular empirical situation may be consistent with one expression while needed estimators can be derived using the alternative expression. Also, in some situations one may need to estimate either η or δ to be used as an explanatory variable in another model. Being able to determine its properties would then be important as well.

To provide an example, suppose that Y_1 is the demand of an individual for a particular product, Y_2 is the price of the product at a store, and ε_1 includes unobservable quality of the product. Then, ε_1 and Y_2 may not be independently distributed. Suppose that Y_3 is the cost of the product to the store, and η denotes unobservable costs of the store, such as rent and salaries. The assumption that η is independent of (Y_3, ε_1) may then not be unreasonable. (Note that we can condition on additional observable variables if needed.) The equivalence result would imply then that the cost of the product to the store can be employed as a proxy for the quality of the product, and can then be used to eliminate the endogeneity of Y_2 . The variable δ could then represent the unobservable taste of the consumer for the product. Estimation of δ can be used to predict the demand of the same consumer under either price or quality changes.

The equivalence between (5.2)-(5.3) together with conditional independence properties implied by independence allow us to express model (5.1) in the following equivalent formulations, each involving different unobservable random terms:

³Even though we will concentrate on the case where Y_2 is a scalar, the results can be easily extended to the case where Y_2 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.

$$\begin{aligned}
Y_1 &= m(Y_2, \varepsilon_1) \\
&= m(s(Y_3, \eta), \varepsilon_1) \\
&= m(s(Y_3, \eta), r(Y_3, \delta)) \\
&= m(Y_2, r(Y_3, \delta))
\end{aligned}$$

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 variable independent of ε_1 , while Y_3 is not independent of ε_1 . The third expression highlights the fact that if η and δ are independently distributed, even if this is just for one value of Y_3 , then, this provides sufficient exogenous variation between Y_2 and ε_1 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. We present the observational equivalence result under the following assumptions.

Assumption 5.1: *The function m in (5.1) is strictly increasing in ε_1 .*

Assumption 5.2: *For all values y_3 of Y_3 , the conditional distribution of ε_1 given $Y_3 = y_3$ is strictly increasing.*

Assumption 5.3: *For all values y_3 of Y_3 , the conditional distribution of Y_2 given $Y_3 = y_3$ is strictly increasing.*

Proposition 5.1: *Suppose that assumptions 5.1-5.3 are satisfied. Then, the following statements are equivalent:*

- (i) ε_1 is independent of Y_2 , conditional on Y_3 .
- (ii) There exists a strictly increasing function $s(Y_3, \cdot)$ and an unobservable random term η such that

$Y_2 = s(Y_3, \eta)$ and
 η is independent of (Y_3, ε_1) .

(iii) There exists a strictly increasing function $r(Y_3, \cdot)$ and an unobservable random term δ such that

$\varepsilon_1 = r(Y_3, \delta)$,
 δ is independent of (Y_3, Y_2) .

PROOF: We first show that (ii) implies (i): Since η is independent of (ε_1, Y_3) , η is independent of ε_1 , conditional on Y_3 . Since $Y_2 = s(Y_3, \eta)$, Y_2 is a function of η , conditional on Y_3 . It follows that, conditional on Y_3 , Y_2 is independent of ε_1 . This shows (i). Next, to show that (i) implies (iii), define, for any value of (Y_2, Y_3) and any value of ε_1 , $\tilde{\delta} = F_{\varepsilon_1|Y_2, Y_3}^{-1}(\varepsilon_1)$ and $\delta = F_{\varepsilon_1|Y_2}^{-1}(\varepsilon_1)$. Then, both $\tilde{\delta}$ and δ are $U(0, 1)$; $\tilde{\delta}$ is independent of (Y_2, Y_3) and δ is independent of Y_3 . Define $\tilde{r}(Y_2, Y_3, \delta) = F_{\varepsilon_1|Y_2, Y_3}^{-1}(\delta)$ and $r(Y_3, \delta) = F_{\varepsilon_1|Y_3}^{-1}(\delta)$. Since, by (i), $F_{\varepsilon_1|Y_2, Y_3} = F_{\varepsilon_1|Y_3}$, it follows that for all δ , $\tilde{r}(Y_2, Y_3, \delta) = r(Y_3, \delta)$. Hence, $\varepsilon_1 = r(Y_3, \delta)$ where δ is independent of (Y_2, Y_3) . This shows that (i) implies (iii). To show that (iii) implies (ii), define $s(Y_3, \eta) = F_{Y_2|Y_3=Y_3}^{-1}(\eta)$. Then, s is strictly increasing in η , and $\eta = s^{-1}(Y_3, Y_2)$ is $U(0, 1)$, for all Y_3 . Since, by (iii), δ is independent of (Y_3, Y_2) , it follows that η is independent of δ . To show that η is independent of (Y_3, ε_1) , we note that the distribution of η conditional on (Y_3, ε_1) equals the distribution of η conditional on (Y_3, δ) , since $\varepsilon_1 = r(Y_3, \delta)$, where r is strictly increasing in δ . Since conditional on Y_3 , η is a function of only Y_2 , and since δ is independent of (Y_2, Y_3) , it follows that the distribution of η conditional on (Y_3, δ) equals the distribution of η conditional on Y_3 . Since η is independent of Y_3 , this implies that the distribution of η conditional on (Y_3, ε_1) equals the marginal distribution of η , and it completes the proof.

5.1. Application to a discrete choice model

The results presented above can be used to identify and estimate a binary discrete choice model that possess an unobservable variable that is correlated with an observable explanatory variable. Consider, for example, the model

$$(5.4) \quad Y_1 = \begin{cases} 1 & \text{if } X_0 + v(X_1, Y_2, \xi) \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where Y_1 , Y_2 , and X are observable, ξ and ε are unobservable, v is an unknown function, and the distribution of (ξ, ε) is unknown. This model is a generalization of the model in Matzkin

(1992), where it was specified that

$$(5.5) \quad Y_1 = \begin{cases} 1 & \text{if } X_0 + v(X_1) \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

Matzkin (1992) assumed that the unobservable ε is independent of (X_0, X_1) , X_0 is continuously distributed conditional on X_1 , and for some known α and a known value $\bar{x}_1$ of X_1 , $v(\bar{x}_1) = \alpha$. Lewbel (2000) considered a similar model where v is a linear function of X_1 . Lewbel (2000) assumed that ε is independent of X_0 conditional on X_1 , X_0 is continuously distributed with unbounded support, and the conditional expectation of ε given X_1 is 0. The model (5.4) is also a generalization of the model considered in Briesch, Chintagunta, and Matzkin (2010), where it was specified that

$$(5.6) \quad Y_1 = \begin{cases} 1 & \text{if } X_0 + v(X_1, \xi) \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

Briesch, Chintagunta, and Matzkin (2010) assumed that ξ, ε , and (X_0, X_1) are mutually independent, and for some known value α and for a known value $\bar{x}_1$ of X_1 , $v(\bar{x}_1, \xi) = \alpha$ for all ξ .

We can extend the identification arguments in Briesch, Chintagunta, and Matzkin (2010) to allow ξ to depend on Y_2 . Specifically, let

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

and assume that for some unknown function s , observable Y_3 , and unobservable η

$$Y_2 = s(Y_3, \eta)$$

From the distribution of Y_1 given (X_0, X_1, Y_2, Y_3) , one can follow the arguments in Briesch, Chintagunta, and Matzkin (2010), which use Teicher (1961), to identify the distribution of γ given (X_1, Y_2, Y_3) . Then, Proposition 5.1 can be used to identify v and the distribution of (ξ, Y_2) , under the assumption that η is independent of ξ conditional on Y_3 .

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

$$(5.7) \quad Y_1 = \begin{cases} 1 & \text{if } \beta X + \gamma Y_2 + \xi \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where ξ is independent of (X, ε) but not of Y_2 , and ε is independent of (X, Y_2) . A particular application of this model is 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 application for the 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 identify model (5.7), suppose that there exists an observable variable, Y_3 , such that at least one of the conditions in Proposition 5.1 is satisfied. Then, we can transform the above model into

$$(5.8) \quad Y_1 = \begin{cases} 1 & \text{if } \beta X + \gamma Y_2 + r(Y_3, \delta) \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

where r is a function that is strictly increasing in its last argument and δ is distributed independently of (X, Y_2, Y_3) . In the choice of product example, Y_3 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, Y_3 could, in some situations, be some other period's income.

If we are willing to assume that the function r is linear in Y_3 and additive in δ , then we can estimate consistently the coefficients in this discrete choice model by estimating the model

$$Y_1 = \begin{cases} 1 & \text{if } \beta X + \gamma Y_2 + \phi Y_3 \geq \varepsilon^* \\ 0 & \text{otherwise} \end{cases}$$

where $\varepsilon^* \equiv \varepsilon - \delta$ is independent of (X, Y_2, Y_3) . When the distribution of ε^* is specified parametrically, this model can be estimated by any of the usual parametric methods. Note that since Y_3 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 Y_3 .

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, one can estimate the model using a slight modification of the approach in Blundell and Powell (2004). 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 $\bar{y}_3$ of Y_3 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 ,

$$r(\bar{y}_2, \delta) = \delta$$

Let

$$\tilde{\eta} = \varepsilon - r(Y_3, \delta)$$

Then, the model

$$Y_1 = \begin{cases} 1 & \text{if } \beta X + \gamma Y_2 + \xi \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

becomes

$$Y_1 = \begin{cases} 1 & \text{if } \beta X + \gamma Y_2 \geq \tilde{\eta} \\ 0 & \text{otherwise} \end{cases}$$

where $\tilde{\eta}$ is an unobservable random term whose distribution depends on Y_3 . Since ε and δ are, by assumption, independent of (X, Y_2) , $\tilde{\eta}$ is independent of (X, Y_2) conditional on Y_3 . It then follows that the conditional expectation of Y_1 given (X, Y_2, Y_3) possesses the multiple index form given by

$$E[Y_1 = 1 | X, Y_2, Y_3] = G(\beta X + \gamma Y_2, Y_3)$$

where G is an unknown function. This conditional expectation possesses the same form as that considered in Blundell and Powell (2004), except for the fact that, in Blundell and Powell (2004), the residual, $\hat{v}$, from a first stage regression, takes up the place of the observable variable Y_3 . Hence, similar methods can be used. Another alternative follows by modifying the method developed by Rothe (2009).

6. Conclusions

We have developed methods that exploit independence conditions among unobservable variables to construct reduced form functions in triangular and simultaneous equations models. We have demonstrated how these reduced form functions can be used to identify a function of interest in the system. The functions of interest were allowed to be nonparametric and nonseparable in an unobservable random term. We have shown how these reduced form functions can be used to identify nonparametric and nonseparable functions in simultaneous equation models, as well as in triangular models. In triangular models, we have discussed the trade-offs between observable and unobservable instruments. The former give rise to control function approaches. We have also provided several equivalence results that can be used in particular applications and discussed one such application to a binary discrete choice model.

References

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

ALTONJI, J.G. AND R.L. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 3, 1053-1102.

BERRY, S.T. and P.A. HAILE (2009) "Identification in Differentiated Products Markets Using Market Level Data," working paper, Yale University.

BERRY, S.T. and P.A. HAILE (2011) "Identification in a Class of Nonparametric Simultaneous Equations Model," mimeo, Yale University.

BLUNDELL, R., D. KRISTENSEN, and R.L. MATZKIN (2013): "Consumer Demand with Multidimensional Nonseparable Unobserved Heterogeneity," mimeo, UCL.

BLUNDELL, R. and R.L. MATZKIN (2014): "Control Functions in Nonseparable Simultaneous Equations Models," *Quantitative Economics*, Vol. 5, No. 2.

BLUNDELL, R. AND J.L. POWELL (2003): "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.

_____ (2004) "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, Vol. 71, No. 3, 655-679.

BRIESCH, R., P. CHINTAGUNTA, AND R.L. MATZKIN (2010): "Nonparametric Discrete Choice Models with Unobserved Heterogeneity," *Journal of Business and Economic Statistics*, Vol. 28, No. 2, 291-307.

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

CHERNOZHUKOV, V., X. CHEN, S. LEE and W. NEWEY (2014): "Local Identification of Nonparametric and Semiparametric Models," *Econometrica*, Vol. 82, No. 2, 785-809.

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

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

DAROLLES, S., J.P. FLORENS, Y FAN, and E. RENAULT (2011), "Nonparametric Instrumental Regression," *Econometrica*, Vol. 79, No. 5, 1541-1565.

D'HAULTFOEUILLE, X. and P. FEVRIER (2015), "Identification of Nonseparable Triangular Models With Discrete Instruments," *Econometrica*, Vol. 83, 1199-1210.

FLORENS, J.P., J.J. HECKMAN, C. MEGHIR, and E. VYTLACIL (2008), "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, Vol. 76, No. 5, 1191-1206.

GALE, D. and H. NIKAIDO (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Math Annalen* 159, 81-93.

HALL, P. and J.L. HOROWITZ (2003), "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, Vol. 33, No. 6, 2904-2929..

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.

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*, Econometric Society Monograph 10, Cambridge: Cambridge University Press.

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. NEWHEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, Vol. 77, No. 5.

KASY, M (2014), "Instrumental Variables with Unrestricted Heterogeneity and Continuous Treatment", Review of Economic Studies.

KOENKER, R. (2005): *Quantile Regression*, Econometric Society Monographs, Cambridge University Press.

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

LEWBEL, A. (2000): "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity or Instrumental Variables," *Journal of Econometrics*, Vol. 97, No. 1, 145-177.

MA, L. and R. KOENKER (2006): "Quantile Regression Methods for Recursive Structural Equation Models," *Journal of Econometrics*, Vol. 134, No. 2, 471-506.

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

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

MATZKIN, R.L. (2007a): "Heterogeneous Choice," in *Advances in Economics and Econometrics, Theory, and Applications, Ninth World Congress of the Econometric Society*, edited by R. Blundell, W. Newey, and T. Persson, Cambridge University Press.

MATZKIN, R.L. (2007b) "Nonparametric Identification," Chapter 73 in *Handbook of Econometrics*, Vol. 6b, edited by J.J. Heckman and E.E. Leamer, Elsevier B.V., 5307-5368.

MATZKIN, R.L. (2007c): "Nonparametric Survey Response Errors," *International Economic Review*, Vol. 48, No. 4, 1411-1427.

MATZKIN, R.L. (2008) "Identification in Nonparametric Simultaneous Equations," *Econometrica*, Vol. 76, No. 5, 945-978.

MATZKIN, R.L. (2010) "Estimation of Nonparametric Models with Simultaneity," mimeo, UCLA.

MATZKIN, R.L. (2012) "Identification in Nonparametric Limited Dependent Variable Models with Simultaneity and Unobserved Heterogeneity," *Journal of Econometrics*, Vol. 166, No. 1, 106-115.

MATZKIN, R.L. (2015) "Estimation of Nonparametric Models with Simultaneity," *Econometrica*, 83, 1-66.

NEWHEY, W.K. and J.L. POWELL (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1557-1569..

NEWHEY, 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.

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

ROTHER, C. (2009): "Semiparametric Estimation of Binary Response Models with Endogenous Regressors," *Journal of Econometrics*, No. 153, 51-64.

TEICHER, H. (1961) "Identifiability of Mixtures," *Annals of Mathematical Statistics*.

TORGOVITSKY, A. (2015), "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185-1197.