Identification in Nonparametric
Limited Dependent Variable Models with
Simultaneity and Unobserved Heterogeneity

Rosa L. Matzkin\textsuperscript{1}
Department of Economics
University of California, Los Angeles

First version: January 2010
This version: March 2011

Abstract

We extend the identification results for nonparametric simultaneous equations models in Matzkin (2008) to situations where the observations on the vector of dependent variables might be limited, and where the number of exogenous unobservable variables is larger than the number of dependent variables.

\textsuperscript{1}The support of NSF through grants SES-0833058 and BCS-0852261 and of NIA through grant F014613//AG012846-1251 is gratefully acknowledged. I am also thankful for their insightful comments to Elie Tamer, two referees, Gary Chamberlain, Andrew Chesher, Yong Hyeon Yang, and the participants at the May 2009 Conference on Identification and Decisions, in honor of Charles Manski, where this paper was presented.
1 Introduction

This paper develops new identification results for nonparametric limited dependent variable models. The models can be subject to simultaneity in latent or observable continuous variables, and may depend on a large number of unobservable variables. We provide conditions under which the distribution of the vector of unobservable variables as well as the functions of the unobservable and observable variables are nonparametrically identified.

The results can be applied to a wide range of models. In particular, they can be applied to binary threshold crossing models, ordered dependent variable models, and censored dependent variable models with continuous endogenous explanatory variables and no dummy or other limited explanatory variables. Models with "no structural shifts" considered in Heckman (1978) belong to this type. Examples of empirical situations that can be analyzed using such models include variations of the labor participation decision of one member of the family as a function of other household income, as in Smith and Blundell (1986), Blundell and Smith (1989), and Blundell and Powell (2003b), and the decision model for post-secondary vocational school training in Nelson and Olson (1978).

Our main method proceeds by first transforming the model with limited dependent variables and a large number of unobservable variables into one with ‘observable’ continuous dependent variables and with the same number of unobservable variables as the number of dependent variables. The transformed model will satisfy the conditions of the models analyzed in Matzkin (2008). To determine the identification of this transformed model, we use the identification results for simultaneous equations models developed in Matzkin (2008). We then use the identified elements in the transformed model together with the particular structures of the original model to determine the identification of the elements of the original model. Such elements are the particular functions and distributions generating the distribution of the unobservable variables in the transformed models and the structural equations
that describe the interaction among the latent and observable dependent variables, the observable exogenous variables, and the unobservable exogenous variables.

To transform the nonparametric model with limited dependent variables into another nonparametric model with ‘observable’ continuous dependent variables, we extend some of the original idea in Manski (1975). In this pioneering paper, Manski showed, among other things, that one can identify and consistently estimate the coefficients of linear subutility functions in discrete choice models without specifying parametrically the distribution of the unobservable random subutilities. Two of the key elements in Manski’s (1975) proof of identification were a scale normalization on the vector of coefficients and assumptions on a regressor, requiring it to possess a strictly increasing distribution conditional on the other regressors and a nonzero coefficient. Those same assumptions were used in a large number of distribution-free methods for binary and multinomial models that were later developed. In particular, Cosslett (1983) showed how these same elements deliver as well identification of the distribution of the unobservable random term, when this is independent of the regressors. Matzkin (1992, examples 3 and 5) also used an additive regressor with strictly increasing conditional distribution and nonzero coefficient. Instead of the scale normalization used by Manski (1975) and Cosslett (1983), she specified the coefficient of this regressor to have the value one. Matzkin (1992) showed that when the latent variable in binary threshold crossing or binary choice models is the sum of this regressor plus a nonparametric function of the other regressors, plus the unobservable random term, the nonparametric function and the distribution of the unobservable random term are identified. Matzkin (1992) developed the results assuming that the additive unobservable random term was distributed independently of the vector of all observable explanatory variables. Lewbel (2000) showed that the continuous, large support regressor need only be independent of the unobservable variables, conditional on the other explanatory
variables. This allowed the regressor to be used in the identification of a variety of models with endogeneous regressors. In this paper, we will require statistical independence between the vector of all exogenously determined, observed explanatory variables and the vector of unobservable variables, as in Matzkin (1992), instead of only a conditional independence assumption, as in Lewbel (2000).²

To review the literature related to our models, consider first simultaneous equations models in latent or observable continuous dependent variables, where each structural equation contains only one unobservable exogeneous variable. The parametric identification and estimation of such models can be analyzed using the results in Amemiya (1974, 1977, 1979), Heckman (1974, 1976, 1978), Nelson and Olson (1978), and Blundell and Smith (1989) by restricting attention to the cases where the structural system of simultaneous equations involve only latent or observable dependent variables. See, in particular, Heckman (1974), Nelson and Olson (1978), Heckman (1978, Cases 1, 3, and 5), Amemiya (1979), and Blundell and Smith (1989). (Amemiya (1981), Lee (1981), and Maddala (1985) compare various such models.) Newey (1985) developed semiparametric estimation methods for some of these models. (Blundell and Powell (2003a) provide a lucid survey.) Berry and Haile (2009a, 2009b) and Chiappori and Komunjer (2009) analyzed nonparametric identification in multinomial choice models with simultaneity and unobserved heterogeneity with special emphasis on the model developed by Berry, Levinsohn, and Pakes (1995)). These papers deal with simultaneity by making a completeness assumption, as in Newey and Powell (1989, 2003) and Chernozhukov and Hansen (2005). Berry and Haile (2009b) show identification also using a transformation of variables approach, related to


The main method we develop in this paper to deal with simultaneity in latent and observable variables is an extension of the unpublished Section 5, in Matzkin (2005), which considered a simultaneous equation model in latent variables with discrete observability on the latent variables. The method imposes specified scales on the measurement of the unobserved values of the limited dependent variables. Intuitively, fixing the scale allows us to identify a nonparametric function of the unobserved values of the limited dependent variables. Although this known scale assumption can be relaxed, this would typically be at the expense of adding functional restrictions.

unobserved heterogeneity. (See also Matzkin (2007b).) The explanatory variables in the models mentioned above are either independent or conditionally independent of the unobservable random terms. Identification of the distribution of unobserved heterogeneity in models with simultaneity was studied in the previously mentioned papers by Berry and Haile (2009a, 2009b) and Chiappori and Komunjer (2009).

The method that we develop to deal with unobserved heterogeneity is based on specifying a relationship among nonadditive functions of unobservable random terms, and identifying the nonparametric distributions of these unobservable random terms as well as the functions of which they are arguments of, using the identification results in Matzkin (2003) and Matzkin (2008). Specifically, our techniques allow to decompose the value of an ‘aggregate’ unobservable random vector into functions of multiple other unobservable random terms. One can use these techniques any time the distribution of an unobservable random vector conditional on a vector of observable variables is determined to be identified. In particular, the techniques we develop, to decompose the conditional distribution of a vector of unobservable variables conditional on a vector of observable variables, can be used instead of or together with some of the other techniques that have been mentioned above to deal with unobserved heterogeneity in various models.

The structure of the paper is as follows. In the next section, we describe the model and specify the assumptions. In Section 3, we present our identification result. Examples are provided in Section 4. Section 5 concludes.

2 The Model

We consider a model, in which an unknown function, $V$, represents a relationship between a vector of potentially latent, endogenous variables, $Y^* \in R^G$, a vector of observable exogenous variables, $(X, Z, W) \in R^{K_x + K_z + G}$, and
a vector of unobservable exogenous variables, \((\delta, \eta) \in R^{G + K_\eta}\). The relationship among these vectors is described by the unknown function \(V : R^{G + K_G + K_x + K_\eta + G + K_\eta} \rightarrow R^G\), by

\[
(1) \quad V (Y^*, W, X, Z, \delta, \eta) = 0
\]

The function \(V\) will be assumed to satisfy certain invertibility and weak separability restrictions, specified below. The vector \(\delta\) will represent structural errors, generating dependence across the coordinates of \(Y^*\). The vector \(\eta\) will represent unobserved heterogeneity. The potentially unobservable endogenous vector \(Y^*\) is assumed to be at least partially observed. Specifically, we assume that there exists a known transformation \(T : R^G \rightarrow R^G\) such that for any value \(y^*\) of \(Y^*\), \(T (y^*)\) is observed. We denote \(T (Y^*)\) by \(Y\). Hence, the model is (1) and

\[
(2) \quad Y = T (Y^*)
\]

where \(Y\) is the vector of observable endogenous variables and \((X, Z, W)\) is the vector of observable exogenous variables. Note that \(T(Y^*)\) is not an argument in (1). The simultaneity is in terms of only the latent variables. Particular cases of such models are the ones in Heckman’s (1978) with no structural shifts. In such models the coherency conditions are satisfied because they impose zero-valued coefficients for the dummy variables in the structural equations.

In simple cases, the transformation, \(T\), determines whether the model corresponds to a multinomial model, a Tobit model, or any other limited dependent variables models. To provide some simple examples, suppose that the model contains one latent variable, \(Y_1^*\) and one observable endogenous variable, \(Y_2 = Y_2^*\). Then, the case where the econometrician observes only if
$Y_1^*$ is above or below a threshold corresponds to the transformation

$$T(y_1^*, y_2^*) = (1, y_2^*) \quad \text{if} \quad y_1^* > 0$$

$$= (0, y_2^*) \quad \text{otherwise}$$

The case where the econometrician observes the value of $Y_1^*$ only when this value is above 0 corresponds to the transformation

$$T(y_1^*, y_2^*) = (y_1^*, y_2^*) \quad \text{if} \quad y_1^* > 0$$

$$= (0, y_2^*) \quad \text{otherwise}$$

Instead of specifying a known coefficient for the limited dependent variable and additivity in $\varepsilon_1$ and $\varepsilon_2$, as other models do, our assumptions impose a piecewise linear specification the functional $y_1^*$ the median of $\varepsilon$ is 0.

If the value of $Y_1^*$ is always observed, the transformation is

$$T(y_1^*, y_2^*) = (y_1^*, y_2^*)$$

Some examples with two latent variables are the Tobit type models

$$T(y_1^*, y_2^*) = (y_1^*, y_2^*) \quad \text{if} \quad (y_1^*, y_2^*) > (0, 0)$$

$$= 0 \quad \text{otherwise}$$

$$T(y_1^*, y_2^*) = (y_1^*, 0) \quad \text{if} \quad y_1^* > 0, \ y_2^* \leq 0$$

$$= (0, y_2^*) \quad \text{if} \quad y_1^* \leq 0, \ y_2^* > 0$$

$$= (y_1^*, y_2^*) \quad \text{if} \quad y_1^* > 0, \ y_2^* > 0$$

$$= (0, 0) \quad \text{otherwise}$$
and the multinomial model

\[ T(y_1^*, y_2^*) = (1, 0) \quad \text{if} \quad y_1^* \geq 0 \geq y_2^* \]
\[ = (0, 1) \quad \text{if} \quad y_1^* \leq 0 < y_2^* \]

### 2.1 Assumptions on distributions

We will first specify assumptions on the distributions of the observable and unobservable exogenous variables. Assumption D1 requires independence between the $G$ dimensional vector $\delta$, the large dimensional vector $\eta$, and the vector of observable exogenous $(X, Z, W)$. The vector $\delta$ will represent structural errors, generating dependence across the coordinates of $Y^*$. The coordinates of $\delta$ may enter one or all of the structural functions. The vector $\delta$ will be linked to a subvector, $Z_0$, of $Z$, in relatively unspecified ways. On the other hand, each coordinate of the vector $\eta$, representing an element of unobserved heterogeneity, will enter only one of the structural equations and it will do so linked to a subvector of $Z$ in restricted ways. Assumption D2 requires independence across the coordinates of $\eta$. This assumption is only needed to identify the distribution of the whole vector $\eta$ from the distribution of the marginal distributions of the coordinates of $\eta$. Assumption D3 requires full support for the unobservables $\delta$ and $\eta$ and for the vector of exogenous variables $X$. Note that while the coordinates of the vector $\eta$ will be assumed to be mutually independent, no such requirement is imposed on $\delta$.

Assumption D1: $\delta, \eta, \text{and } (X, Z, W)$ are mutually independent.
Assumption D2: The coordinates of \( \eta \) are independently distributed.

Assumption D3: The distributions of \( \delta, \eta, \) and \( X \) are absolutely continuous and possess everywhere positive densities.

2.2 Assumptions on \( V \)

We next specify our assumptions on the function \( V \). These concern the observable as well as the unobservable exogenous variables. In the specification of the model in (1) and (2), we have denoted the vector of observable explanatory variables, \((X, Z, W) \in R^{K_x+K_z+G}\), in terms of three subvectors, \(X, Z,\) and \(W\), because we wanted to emphasize the different roles that each of these subvectors has in identification. The vector of observable exogenous variables \( W \in R^G \) will have the role of allowing us to identify the conditional distribution, given \((X, Z, W)\), of the vector of latent variables, \(Y^*\), from the conditional distribution, given \((X, Z, W)\), of the vector of observable endogenous variables, \(Y\). The vector of observable exogenous variables \( Z \in R^{K_x} \) will have the role of allowing us to identify the distribution of \((\delta, \eta)\) and nonparametric functions linking elements of \((\delta, \eta)\) to elements of \(Z\). These latter non-parametric functions are \(m_0(z_0, \delta)\), whose values belong to \(R^G\), and \(K_\eta\) real-valued nonparametric functions, \(m_1(z_1, \eta_1), \ldots, m_{K_\eta}(z_{K_\eta}, \eta_{K_\eta})\), each non-additive in the unobservable random terms. The vectors \(Z_0, Z_1, \ldots, Z_{K_\eta}\) are subvectors of \(Z\). The following assumptions impose separability restrictions on \(V\) that will be instrumental into making the roles of \(W\) and \(Z\) effective.

**Assumption B (separability in \((Y^*, W)\))** : There exists a known function \(B : R^{2G} \rightarrow R^G\) such that (i) the function \(V\) depends on \((y^*, w)\) only through
the values of $B(y^*, w)$, (ii) for any value $w$ of $W$, $B(\cdot, w)$ is continuous, 1–1, and onto $R^G$, (iii) for any $b \in R^G$, there exists a value $w$ in the support of $W$ and a value $y$ in the support of $Y$ such that for any $(x, z)$

$$
Pr(B(Y^*, W) \leq b|W = w, X = x, Z = z) = Pr(Y \leq y|W = w, X = x, Z = z)
$$

This assumption turns the distribution of the potentially discrete $Y$ into the distribution of a continuous random vector $B(Y^*, W)$. It is a generalization of the condition used in Matzkin (2005, Section 5) to show identification in models with simultaneity in latent variables. To provide some examples, disregard the existence of $(X, Z)$. Consider the transformation where

$$
T(y_1^*, y_2^*) = (1, y_2^*) \quad \text{if} \quad y_1^* > 0
$$

$$
= (0, y_2^*) \quad \text{otherwise}
$$

Let $B(y_1^*, y_2^*, w_1, w_2) = (y_1^* + w_1, y_2^*)$. Assume that the support of $W_1$ is $R$ and that 0 belongs to the support of $W_2$. For any given $b = (b_1, b_2)$, let $w_1 = b_1$, $w_2 = 0$, $y_1 = 0$, $y_2 = b_2$. Then

$$
Pr(B(Y^*, W) \leq (b_1, b_2)|W_1 = b_1)
$$

$$
= Pr(Y_1^* + W_1 \leq b_1, Y_2^* \leq b_2|W_1 = b_1)
$$

$$
= Pr(Y_1^* \leq 0, Y_2^* \leq b_2|W_1 = b_1)
$$

$$
= Pr(Y_1 \leq 0, Y_2 \leq b_2|W_1 = b_1)
$$

11
Consider next the transformation defined by

\[ T(y_1^*, y_2^*) = (y_1^*, y_2^*) \quad \text{if} \quad y_1^* > 0 \]
\[ = (0, y_2^*) \quad \text{otherwise} \]

Let again \( B(y_1^*, y_2^*, w_1, w_2) = (y_1^* + w_1, y_2^*) \). Assume that the support of \( W_1 \) is the set of all nonpositive numbers and that 0 belongs to the support of \( W_2 \). If \( b_1 > 0 \), let \( w_1 = 0 \) and \( w_2 = 0 \). Then, since conditional on \( W_1 = 0 \), \( B(Y^*, W) = (Y_1^*, Y_2^*) \), the condition is satisfied letting \( y_1 = b_1 \) and \( y_2 = b_2 \). If \( b_1 < 0 \), let \( w_1 = b_1 \), \( w_2 = 0 \), \( y_1 = 0 \), and \( y_2 = b_2 \). Then, as in the previous example,

\[
\Pr (B(Y^*, W) \leq (b_1, b_2) | W_1 = b_1) \\
= \Pr (Y_1^* + W_1 \leq b_1, Y_2^* \leq b_2 | W_1 = b_1) \\
= \Pr (Y_1^* \leq 0, Y_2^* \leq b_2 | W_1 = b_1) \\
= \Pr (Y_1 \leq 0, Y_2 \leq b_2 | W_1 = b_1)
\]

Hence, the condition is again satisfied.

Note that in the Binary Threshold Crossing model, the support of \( W \) is required to be the whole line, while in the Tobit Model the support of \( W \) need only be the set of nonpositive numbers. Intuitively, the large support of \( W \) in the first case is needed to compensate for the fact that the only observed property of \( Y^* \) is whether it is positive or negative. In the Tobit model, the values of \( Y^* \) are observed when they are positive. Hence, there is no need to use values of \( W \) in order to recover the distribution of \( Y^* \) for those values. Lewbel and Matzkin (2009) analyze support requirements on
‘special regressors’ across different Limited Dependent Variable models.

Under additional restrictions, one could apply directly the results in Matzkin (2008) without use of the vector $W$. One such case is where identification of functions in the model can be established using only the region where the distribution of $Y$ given $X$ coincides with the distribution of $Y^*$ given $X$. Along this line, it might be useful contrasting our assumptions with those used in some nonparametric versions of the tobit model when $y_2^*$ is not endogenous. These models have been studied recently by, among others, Lewbel and Linton (2002) and Chen, Dahl, and Kahn (2005). Denote $x = (x_1, x_2)$ where $x_1$ is a scalar. Lewbel and Linton (2002) consider the model

$$y_1^* = m(x_1, x_2) + \varepsilon$$
$$T(y_1^*) = \begin{cases} y_1^* & \text{if } y_1^* > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\varepsilon$ and $x$ are independently distributed. Chen, Dahl, and Kahn (2005) consider the model

$$y_1^* = m(x_1, x_2) + \sigma(x_1, x_2) \varepsilon$$
$$T(y_1^*) = \begin{cases} y_1^* & \text{if } y_1^* > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\varepsilon$ and $x$ are independently distributed, and where two quantiles of $\varepsilon$ are known. In contrast, our model would be specified as

$$y_1^* = x_1 + h(x_2, \varepsilon)$$
$$T(y_1^*) = \begin{cases} y_1^* & \text{if } y_1^* > 0 \\ 0 & \text{otherwise,} \end{cases}$$
where $\varepsilon$ and $(x_1, x_2)$ are independently distributed, $x_1$ is continuously distributed conditional on $x_2$, and $h$ is strictly increasing in $\varepsilon$. Our additivity in $x_1$ and specified coefficient for $x_1$ compensates for the additivity in $\varepsilon$, the separability between $y'_1$ and $x_2$, and the specified coefficient for $y'_1$, imposed in the models of Lewbel and Linton (2002) and Chen, Dahl, and Kahn (2005). One could use the assumptions in those models, together with additional ones, to obtain identification results when simultaneity is present. This would avoid relaying on the observable vector $W$ at the expense of additional structure. Consider for example, the model

$$
\begin{align*}
y_1^* &= \beta y_2^* + m_1(x_1) + \varepsilon_1 \\
y_2^* &= \gamma y_1^* + m_2(x_2) + \varepsilon_2
\end{align*}
$$

$$
T(y_1^*, y_2^*) = (y_1^*, y_2^*) \quad \text{if} \quad y_1^* > 0, \ y_2^* > 0
$$

$$
= (0, 0) \quad \text{otherwise,}
$$

where $x_1$ and $x_2$ be both multivariate continuous random vectors, $\beta$ and $\gamma$ are parameters of unknown values, and $m_1$ and $m_2$ are nonparametric, continuous functions. If the joint distribution of $(y_1^*, y_2^*, x_1, x_2)$ were observed, one could use Matzkin (2008) to easily derive identification results for $\beta$, $\gamma$, $m_1$ and $m_2$. However, even when this is not the case, one can show that, because of the particular structure, under some pointwise conditions on the value of the derivative of the density of $(\varepsilon_1, \varepsilon_2)$, the distribution of $(y_1^*, y_2^*, x_1, x_2)$ only over the region where $(y_1^*, y_2^*) > (0, 0)$ is enough to identify $\beta$, $\gamma$, $m_1$, and $m_2$. This could not be possible, however, in more general structures.

The next assumption assumes the existence of nonparametric functions $m_k (z_k, \eta_k)$ for each coordinate, $k$, of $\eta$. Each of these functions is assumed to satisfy any one of the restrictions that were used in Matzkin (2003) to show identification of nonadditive random functions. In addition, it is assumed
that for each of them, there exists a value, \(z_k\), of the subvector \(Z_k\), at which the value of the function \(m_k\) is known. This is used to ‘shut down’ the effect of each of the functions at the specified values of \(z_k\). Instead of using this trick, one could use, under certain structures, deconvolution techniques as in some of the papers mentioned in the Introduction.

**Assumption M (unknown monotone functions of \(\eta\)):** There exist \(K_\eta\) nonparametric, unknown functions \(m_1(z_1, \eta_1), \ldots, m_{K_\eta}(z_{K_\eta}, \eta_{K_\eta})\), where \(\eta = (\eta_1, \ldots, \eta_{K_\eta})\) and where each \(z_1, \ldots, z_{K_\eta}\) denote the values of subvectors of \(Z = (Z_1, \ldots, Z_{K_\eta})\) satisfying (i) each \(m_k\) is continuously differentiable and strictly increasing in \(\eta_k\) a.e. in \(z_k\) (ii) for each \(k\), there exists a known scalar \(\alpha_k\) and a known value \(z_k\) in the support of \(Z_k\) such that for all values of \(\eta_k\), \(m_k(z_k, \eta_k) = \alpha_k\), (iii) either (iii.a) \(m_k\) is homogenous of degree one in \((z_k, \eta_k)\) and for some value \((z_k, \eta_k)\) in the support of \((Z_k, \eta_k)\) and some scalar \(\gamma_k\), \(m_k(z_k, \eta_k) = \gamma_k\), or (iii.b) for some value \(z_k\) in the support of \(Z_k\) and all values of \(\eta_k\), \(m_k(z_k, \eta_k) = \gamma_k\), or (iii.c) the subvector \(Z_k\) can be partitioned into \(Z_k = (Z^1_k, Z^2_k)\) where \(Z^2_k \in R\) and there exists a function \(n_k\) such that for all \((z^1_k, z^2_k)\), \(m_k(z^1_k, z^2_k, \eta_k) = n_k(z^1_k, z^2_k + \eta_k)\) and for some known value \(\gamma_k\) and known vector value \((z^1_k, \tilde{\tau}), n_k(z^1_k, \tilde{\tau}) = \gamma_k\).

Assumption M(ii) is the one made to be able to ‘shut down’ the effect of \(m_k\) by letting \(z_k\) equal a specified value. Identification of \((m_1, \ldots, m_{K_\eta})\) within a set of functions \((\tilde{m}_1, \ldots, \tilde{m}_{K_\eta})\) would require all functions in the set to satisfy Assumption M in exactly the same way that \((m_1, \ldots, m_{K_\eta})\) satisfies it. In particular, all functions in the set would be required to possess the same values of \(\alpha_k, \gamma_k, z_k\), and \(z_k\) \((k = 1, \ldots, K_\eta)\) as \((m_1, \ldots, m_{K_\eta})\). In some cases, establishing identification within the set will not require knowing the values of \(\tilde{z}_k\), as long as these values are common to all functions in the set. Estimation however, would typically require specifying known values for \(\tilde{z}_k\). An example of a function \(m_k\) satisfying M(ii) is \(m_k(z_k, \eta_k) = \eta_k z_k^2\), for \(z_k \in R\).
In this example $\tilde{z}_k = 0$. When the effect of all $m_k$’s except for one, $m_k^{**}$, is ‘shut down’, we will be able to identify the distribution of the values of $m_k^{**}$ conditional $(\tilde{z}_1, ..., \tilde{z}_{k^{**}-1}, z_{k^{**}}, \tilde{z}_{k^{**}+1}, ..., \tilde{z}_{K_n})$. Since by Assumption D.1, $\eta_{k^*}$ is distributed independently of $Z$, the distribution of $\eta_{k^*}$ will be identified from this distribution of $m_k^{**}$ conditional on $(\tilde{z}_1, ..., \tilde{z}_{k^{**}-1}, z_{k^{**}}, \tilde{z}_{k^{**}+1}, ..., \tilde{z}_{K_n})$. Assumptions M(i) and M(iii) allow then to identify the distribution of $\eta_{k^*}$ from the distribution of $m_k^{**}$ conditional on $Z_k^{**}$. Assumptions M(i) and M(iii) were used in Matzkin (2003) to show identification of functions $m_k(z_k, \eta_k)$, nonadditive in $\eta_k$, when $\eta_k$ was distributed independently of $z_k$ with an everywhere positive density. Let $F_{\mu_k|Z_k=z_k}(\cdot)$ denote the cumulative distribution of $\mu_k = m_k(z_k, \eta_k)$ conditional on $Z_k = z_k$. Matzkin (2003) showed that if the function $m_k$ satisfies $m_k(\bar{z}_k, \eta_k) = \eta_k$ for some specified value $\bar{z}_k$ in the support of $Z_k$, then for all $t, z_k$

$$F_{\eta_k}(t) = F_{\mu_k|Z_k=z_k}(t)$$

and

$$m(z_k, t) = F^{-1}_{\mu_k|Z_k=z_k}(F_{\mu_k|Z_k=z_k}(t))$$

If the function $m_k$ satisfies $m_k(\bar{z}_k, \eta_k) = \gamma_k$ and is homogeneous of degree one, then for all $t, z_k$

$$F_{\eta_k}(t) = F_{\mu_k|Z_k=(\bar{z}_k, \eta_k)}(t)$$

and

$$m(z_k, t) = F^{-1}_{\mu_k|Z_k=z_k}(F_{\mu_k|Z_k=(\bar{z}_k, \eta_k)}(t))$$

If the function satisfies $m_k(z_k^1, z_k^2, \eta_k) = n_k(z_k^1, z_k^2 + \eta_k)$ and for some specified $\gamma_k$ and vector value $(\bar{z}_k^1, \bar{t})$, $n_k(\bar{z}_k^1, \bar{t}) = \gamma_k$, then

$$F_{n_k}(t) = F_{\mu_k|Z_k=(\bar{z}_k^1, \bar{t}-t)}(\gamma_k)$$

and

$$m(z_k, t) = F^{-1}_{\mu_k|Z_k=z_k}(F_{\mu_k|Z_k=(\bar{z}_k^1, \bar{t}-t)}(\gamma_k))$$
We note that Assumption M does not require all the $m_k$ functions to satisfy the same restrictions. The assumption only requires that for each $k$, $m_k$ satisfies either (iii.a), or (iii.b), or (iii.c). We also note that even though we have argued that Assumption M allows identification of the distribution of $\eta_k$ for each $k$, this with Assumption D.3 imply that the distribution of the whole vector $\eta$ is identified.

The following assumption will be made to guarantee identification of the distribution of $\delta$ and of nonparametric functions of $\delta$. We will assume that the $G$-dimensional vector $\delta$ is an argument, together with a subvector $Z_0$ of the observable vector $Z$, of an unknown nonparametric function $m_0$, which has range $R^G$. Denote the value of $m_0$ at $(z_0, \delta)$ by $\mu_0 : \mu_0 = m_0(z_0, \delta)$. We want to impose restrictions on $m_0$ guaranteeing that $m_0$ and the distribution of $\delta$ are identified from the distribution of $(\mu_0, z_0)$. Matzkin (2008) provides a method to find such restrictions. To use her result, we need to introduce additional notation. We will let $m_0$ and $f_\delta$ satisfy the assumptions made on the reduced form system of the simultaneous equations models analyzed in Matzkin (2008). For any two functions $m_0(z_0, \delta)$ and $\tilde{m}_0(z_0, \delta)$, both invertible in $\delta$, denote the inverses with respect to $\delta$ by, respectively, $m_0^{-1}(z_0, \mu_0)$ and $\tilde{m}_0^{-1}(z_0, \mu_0)$. Denote the matrix of partial derivatives of $\tilde{m}_0^{-1}(z_0, \mu_0)$ with respect to $\mu_0$ by $\partial \tilde{m}_0^{-1}(z_0, \mu_0) / \partial \mu_0$, and the matrix of partial derivatives of $\tilde{m}_0^{-1}(z_0, \mu_0)$ with respect to $z_0$ by $\partial \tilde{m}_0^{-1}(z_0, \mu_0) / \partial z_0$. Denote the matrices corresponding to $m_0^{-1}(z_0, \mu_0)$ in the same way. The Jacobian determinants of $\tilde{m}_0^{-1}(z_0, \mu_0)$ and $m_0^{-1}(z_0, \mu_0)$ will be denoted respectively by $|\partial \tilde{m}_0^{-1}(z_0, \mu_0) / \partial \mu_0|$ and $|\partial m_0^{-1}(z_0, \mu_0) / \partial \mu_0|$ and be both assumed to be.
positive. We define

$$\Delta_{\mu_0} (z_0, \mu_0) = \frac{\partial}{\partial \mu_0} \log \left| \frac{\partial m_0^{-1}(z_0, \mu_0)}{\partial \mu_0} \right| - \frac{\partial}{\partial \mu_0} \log \left| \frac{\partial \tilde{m}_0^{-1}(z_0, \mu_0)}{\partial \mu_0} \right|$$

and

$$\Delta_{z_0} (z_0, \mu_0) = \frac{\partial}{\partial z_0} \log \left| \frac{\partial m_0^{-1}(z_0, \mu_0)}{\partial \mu_0} \right| - \frac{\partial}{\partial z_0} \log \left| \frac{\partial \tilde{m}_0^{-1}(z_0, \mu_0)}{\partial \mu_0} \right|$$

For any possible density $f_\delta$ of $\delta$ and any value of $(\mu_0, z_0)$, we then define the matrix $B (\mu_0, z_0; m, \tilde{m}, f_\delta)$ by

$$B (\mu_0, z_0; m, \tilde{m}, f_\delta) = \left( \begin{array}{c} \left( \frac{\partial \tilde{m}_0^{-1}(z_0, \mu_0)}{\partial \mu_0} \right)^\prime \Delta_{\mu_0} (z_0, \mu_0) + \left( \frac{\partial m_0^{-1}(z_0, \mu_0)}{\partial \mu_0} \right)^\prime \frac{\partial \log (f_\delta (m_0^{-1}(z_0, \mu_0)))}{\partial \mu_0} \\ \left( \frac{\partial \tilde{m}_0^{-1}(z_0, \mu_0)}{\partial z_0} \right)^\prime \Delta_{z_0} (z_0, \mu_0) + \left( \frac{\partial m_0^{-1}(z_0, \mu_0)}{\partial z_0} \right)^\prime \frac{\partial \log (f_\delta (m_0^{-1}(z_0, \mu_0)))}{\partial \delta} \end{array} \right)$$

The following assumption will be used, together with the results on nonparametric identification of simultaneous equations models in Matzkin (2008), to guarantee that the distribution of $\delta$ and the nonparametric function $m_0$ are identified from the distribution of $(\mu_0, z_0)$.

Assumption I (unknown invertible function of $\delta$): There exist a nonparametric, unknown function, $m_0(z_0, \delta)$, such that (i) for any value of $z_0$, the range of $m_0(z_0, \cdot)$ is $R^G$ (ii) a.e. in $z_0$, $m_0(z_0, \cdot)$ is invertible and $m_0$ and its inverse are continuously differentiable, (iii) for any value of $z_0$, $|\partial m_0(z_0, \delta)/\partial \delta| > 0$, (iv) for a known value $\tilde{z}_0$ of $Z_0$ and a known value $\alpha_0 \in R^G$ of a vector $\alpha \in \tilde{R}^G$, $m_0(\tilde{z}_0, \delta) = \alpha_0$ for all $\delta$, and (v) $m_0$ satisfies
shape restrictions guaranteeing that for any \( \tilde{m}_0 \neq m_0 \) satisfying the same restrictions that \( m_0 \) is assumed to satisfy there exists \((\mu_0, z_0)\) such that for any \( f_\delta \) satisfying the restrictions that \( f_\delta \) is assumed to satisfy, the rank of the matrix \( B(\mu_0, z_0; m, \tilde{m}, f_\delta) \) is \( G + 1 \).

Note that, as in Assumption M, identification of \( \mu_0 \) within a set of functions \( \tilde{m}_0 \) will require all functions \( \tilde{m}_0 \) within the set to satisfy \( I(iv) \) with the same \( z_0 \) and same \( \alpha_0 \) as \( m_0 \).

**Assumption L (known links):** There exist a known continuously differentiable function \( C : R^{G+K_0} \rightarrow R^G \) such that (i) the function \( V \) depends on \((z, \delta, \eta)\) only through the values of \( C\left( m_0(z_0, \delta), m_1(z_1, \eta_1), ..., m_{K_\eta}(z_{K_\eta}, \eta_{K_\eta}) \right) \), (ii) for each \( k \geq 1 \), \( m_k \) is an argument in only one coordinate, \( C_g \), of \( C \), and (iii) for each \( k \geq 1 \) and \( g \), if \( m_k \) is an argument of \( C_g \) then when for all \( j \geq 0 \), all the subvectors \( z_j \) of \( Z \) other than \( z_k \) equal the value \( \tilde{z}_j \) at which \( m_j(\tilde{z}_j, \eta_j) = \alpha_j \), \( C_g \) is strictly increasing in the value of \( m_k \), (v) \( C\left( m_0(z_0, \delta), \alpha_1, ..., \alpha_{K_\eta} \right) \) is invertible in \( m_0 \).

Assumption L specifies the structure generating the heterogeneity in the model. Define the function \( E : R^{K_z+G+K_\eta} \rightarrow R^G \) by

\[
E(z, \delta, \eta) = C\left( m_0(z_0, \delta), m_1(z_1, \eta_1), ..., m_{K_\eta}(z_{K_\eta}, \eta_{K_\eta}) \right).
\]

Let \( \varepsilon \) denote the value of \( E(z, \delta, \eta) \). In our identification theorem, we will show that under the assumptions of our model, the joint distribution of \((\varepsilon, Z)\) is identified. Then, we will show that from the joint distribution of \((\varepsilon, Z)\) we will be able to identify the distributions of \( \delta \) and \( \eta \) and the functions \( m_0, m_1, ..., m_{K_\eta} \). This latter step will use Assumptions M, I, and L together with D1-D3. To provide an intuition for the way in which this result will be
obtained, note that when \( z_k = \tilde{z}_k \) for \( k = 1, \ldots, K_\eta \),

\[
E (z, \delta, \eta) = \varepsilon = C \left( m_0 (z_0, \delta), \alpha_1, \ldots, \alpha_{K_\eta} \right)
\]

Since the function \( C \) is known and invertible in \( m_0 \), from the distribution of \( \varepsilon \) given \( Z = (z_0, \tilde{z}_1, \ldots, \tilde{z}_{K_\eta}) \) we will be able to calculate the distribution of \( \mu_0 = m_0 (z_0, \delta) \) given \( Z \). Let \( \mu_0 = C_\alpha^{-1} (\varepsilon) \) denote the inverse of \( C \) with respect to \( \mu_0 \) when \( Z = (z_0, \tilde{z}_1, \ldots, \tilde{z}_{K_\eta}) \). We then have that the distribution of \( \mu_0 = m_0 (z_0, \delta) \) conditional on \( Z = (z_0, \tilde{z}_1, \ldots, \tilde{z}_{K_\eta}) \) is identified. Using then the results in Matzkin (2008), we can show that under Assumption I, \( m_0 \) and the distribution of \( \delta \) are identified. Next, letting \( Z = (\tilde{z}_0, z_1, \tilde{z}_2, \ldots, \tilde{z}_{K_\eta}) \), we get that

\[
E (z, \delta, \eta) = \varepsilon = C \left( \alpha_0, m_1 (z_1, \eta_1), \alpha_2, \ldots, \alpha_{K_\eta} \right)
\]

Since the distribution of \( \varepsilon \) given \( Z = (\tilde{z}_0, z_1, \tilde{z}_2, \ldots, \tilde{z}_{K_\eta}) \) is identified, and the known function \( C \) is strictly increasing in \( m_1 \), we can calculate the distribution of \( \mu_1 = m_1 (z_1, \eta_1) \) given \( Z = (\tilde{z}_0, z_1, \tilde{z}_2, \ldots, \tilde{z}_{K_\eta}) \). Since by Assumption M, \( m_1 \) satisfies one of the assumptions in Matzkin (2003), \( m_1 \) and the distribution of \( \eta_1 \) are identified. From this it follows that the distribution of \( \eta \) is identified, since by assumption, its coordinates are independent.

We next make an assumption on \( V \) that allows us to transform the model with limited dependent variables and a large number of unobservables into a model with observable continuous dependent variables and the same number of unobservable random terms as the number of endogenous variables.

**Assumption S (separability of \( V \)):** For some function \( V^S : \mathbb{R}^{G + K_\eta + G} \to \mathbb{R}^G \) and all values \((y^*, w, x, z, \delta, \eta)\) of \((Y^*, W, X, Z, \delta, \eta)\) in the support of this vector
\[ V (y^*, w, x, z, \delta, \eta) = 0 \quad \iff \quad V^S (B (y^*, w), x, E (z, \delta, \eta)) = 0 \]

The next assumption relates to conditions on the function \( V^S \) satisfying

\[ (3) \quad V^S (b, x, \varepsilon) = 0 \]

It requires that for any values \((x, \varepsilon)\) there exists a unique value of \(b\) satisfying (3), and for any values \((x, b)\), there exists a unique value of \(\varepsilon\) satisfying (3). Moreover, it also requires that the functions mapping \(x, \varepsilon\) into \(b\) and \(x, b\) into \(\varepsilon\) be twice continuously differentiable, onto \(R^G\), and have positive Jacobian determinants, given \(x\). These assumptions were made in Matzkin (2008) to show identification of nonparametric simultaneous equations models where the number of unobservable variables equal the number of observable endogenous variables, and where all variables are continuous and with everywhere positive densities. In many simultaneous equation models, the uniqueness of \(b\) may be too strong. It implies uniqueness of equilibrium, which in most cases is known to be satisfied under very strong assumptions. Our separability assumptions on \(V\) have the effect of transforming the model with limited dependent variables and large number of unobservables into a model of the type analyzed in Matzkin (2008), where there were no latent endogenous variables and the number of unobservable exogenous variables was the same as that of the observable endogenous variables.

**Assumption U (uniqueness of \(b\) and \(\varepsilon\)):** There exist unique twice continuously differentiable functions \(h : R^{K_x+G} \to R^G\) and \(r : R^{G+K_x} \to R^G\) such that \(h\) and \(r\) are onto \(R^G\), for each \((x, \varepsilon) \in R^{K_x+G}\)

\[ V^S (h(x, \varepsilon), x, \varepsilon) = 0 \]
and for each \((b, x) \in R^{G+K_e}\)

\[ V^S(b, x, r(b, x)) = 0 \]

In addition, the Jacobian determinants, \(\left| \partial r(b, x) / \partial b \right|\) and \(\left| \partial h(x, \varepsilon) / \partial \varepsilon \right|\), are assumed to be positive for all \((b, x, \varepsilon)\).

3 Identification

The assumptions imply that model (1) can be expressed as

\[ \varepsilon = r(b, x) \]

where

\[ \varepsilon = E(z, \delta, \eta) = C\left( m_0(z_0, \delta), m_1(z_1, \eta_1), \ldots, m_{K_\eta}(z_{K_\eta}, \eta_{K_\eta}) \right) \]

and

\[ b = B(y^*, w) \]

This is the structural equations model, involving in each of the \(G\) equations interactions among the endogenous variables in the vector \(y^*\). Since the functions \(C\) and \(B\) are known, this structural model will be identified if the functions \(m_0, m_1, \ldots, m_{K_\eta}\), and \(r\), and the distribution of \((\delta, \eta)\) are identified.

The reduced form system of the transformed model is

\[ b = h(x, \varepsilon) \]

for a function \(h\), which is the inverse of \(r\) with respect to \(b\). If \(r\) is identified, \(h\) is also identified.

The analysis of identification will proceed by applying Matzkin (2008) to
the model $\varepsilon = r(b, x)$ to show that under our assumptions the function $r$ is identified. The identification of $r$ will then be used to show identification of $m_0, m_1, ..., m_{K_q}$ and the distributions of $\delta$ and $\eta$. Instead of imposing enough assumptions, as we do, to guarantee the identification of $r$, one could consider imposing weaker assumptions to achieve identification of only a functional, $\mu(r)$, of $r$. Given an alternative function, $\tilde{r}$, Matzkin (2008) can be used to determine whether $\mu(\tilde{r})$ is observationally equivalent to $\mu(r)$.

Some parts of our identification analysis will be constructive, while those that invoke Matzkin (2008) will be less so. Constructive identification is appealing because it immediately suggests an estimation method. While the identification results in Matzkin (2008) are not constructive, a recent paper (Matzkin (2010)) develops an easily computable estimator for models that are shown to be identified using Matzkin (2008). The estimation method is based on the same transformation of variables approach from which the results in Matzkin (2008) were developed. A control function approach (e.g., Blundell and Powell (2003b), Chesher (2003), Imbens and Newey (2003, 2009) could also be used to estimate some simultaneous equation models. (See Blundell and Matzkin (2010).)

Our identification theorem gives conditions on the set of functions $\tilde{r}$ and on distributions $f_{\tilde{\varepsilon} | Z = z}$, satisfying all the assumptions that were made on the function $r$ and all the assumptions that were made on the distribution, $f_{\varepsilon | Z = z}$ of $\varepsilon$ given $Z = z$. When these conditions are satisfied, and all the above assumptions are also satisfied, the theorem states that the distribution of $(\delta, \eta)$, the functions $m_k$, and the function $r$ are identified. To state the conditions of the theorem we will first define a set to which $(r, f_{\varepsilon, Z})$ is assumed to belong, and then define a matrix that depends on any pair of functions and any distribution within that set.

We will let $(\Gamma \times F)$ denote the set of pairs $(\tilde{r}, f_{\tilde{\varepsilon}, Z})$ such that $\tilde{r}$ satisfies all the assumptions made about $r$ and $f_{\tilde{\varepsilon}, Z}$ is such that for all $z$ in the support of $Z$, $f_{\varepsilon, Z}$ satisfies all the assumptions that $f_{\varepsilon, Z}$ is assumed to satisfy. In
particular, both \( r \) and \( \tilde{r} \) possess the same values of \( \alpha_k, \gamma_k, \tilde{z}_k, \) and \( \tilde{z}_k \) that satisfy assumptions M and I. We will also let \( (\Gamma_0 \times F_0) \) denote the set of pairs \( (\tilde{m}_0, \tilde{f}_\delta) \) such that \( \tilde{m}_0 \) satisfies all the assumptions made about \( m_0 \) and \( \tilde{f}_\delta \) satisfies all the assumptions made about \( f_\delta \). Similarly to the notation we introduced before Assumption I, for any function \( \tilde{r} \) in \( \Gamma \), we will denote the matrix of partial derivatives of \( \tilde{r} \) with respect to \( \beta \) by \( \tilde{\mathbf{\nabla}}_{\tilde{r}}(\beta, x) \) and the matrix of partial derivatives of \( \tilde{r} \) with respect to \( x \) by \( \tilde{\mathbf{\nabla}}_{x}(\tilde{r}(b, x) / \partial b) \). The Jacobian determinant of \( \tilde{r} \) will be denoted by \( |\tilde{\nabla} \tilde{r}(b, x) / \partial b| \). Note that by assumption this determinant is positive. The value of a conditional density \( \phi_{\| | \phi = \zeta} \) at \( \tilde{r}(\beta, x) \) will be denoted by \( \phi_{\| | \phi = \zeta}(\tilde{r}(\beta, x)) \). For any two functions \( \tilde{r} \) and \( \tau \), we will denote by \( \Delta_b \) the difference in the rate of change of their Jacobian determinants with respect to \( b \) and by \( \Delta_x \) the difference in the rate of change of their Jacobian determinants with respect to \( x \). That is

\[
\Delta_b = \frac{\partial}{\partial b} \log \left| \frac{\partial \tau(b, x)}{\partial b} \right| - \frac{\partial}{\partial b} \log \left| \frac{\partial \tilde{r}(b, x)}{\partial b} \right|
\]

\[
\Delta_x = \frac{\partial}{\partial x} \log \left| \frac{\partial \tau(b, x)}{\partial b} \right| - \frac{\partial}{\partial x} \log \left| \frac{\partial \tilde{r}(b, x)}{\partial b} \right|
\]

For any two possible functions, \( \tilde{r} \) and \( \tau \) in \( \Gamma \) and any possible density \( f_{\phi, z} \) in \( F \), we define the matrix \( A(b, x; \tilde{r}, \tau, f_{\phi, z}) \) on \( (b, x, z) \in R^{G+K_x+K_z} \) by

\[
A(b, x; z; \tilde{r}, \tau, f_{\phi, z}) = \begin{pmatrix}
\left( \frac{\partial \tilde{r}(b, x)}{\partial b} \right) & \Delta_b + \left( \frac{\partial \tau(b, x)}{\partial b} \right) \frac{\partial \log(f_{\phi, z}(\tau(b, x)))}{\partial \phi} \\
\left( \frac{\partial \tilde{r}(b, x)}{\partial x} \right) & \Delta_x + \left( \frac{\partial \tau(b, x)}{\partial x} \right) \frac{\partial \log(f_{\phi, z}(\tau(b, x)))}{\partial \phi}
\end{pmatrix}
\]
**Theorem:** Consider the model described by (1)-(2) and satisfying Assumptions B,M,I,L,S,U,D1,D2, and D3. Suppose that for any \( \tilde{\tau} \) and \( \tau \) in \( \Gamma \) such that \( \tilde{\tau} \neq \tau \) and any density \( f_{\tau,Z} \) in \( F \) there exists \((b,x,z)\) such that the rank of the matrix \( A(b,x,z;\tilde{\tau},\tau,f_{\tau,Z}) \) is larger than \( G \). Then, \( r \) is identified in \( \Gamma \), \( m_0 \) is identified in \( \Gamma_0 \), the distribution of \( \delta \) is identified in \( F_0 \), and the functions \( m_1, \ldots, m_{K_\pi} \), and the distribution of \( \eta \) and therefore of \((\delta,\eta)\) are identified in their respective sets.

To prove the theorem, we will first show that we can identify the distribution of \( b = B(Y^*,W) \) conditional on \((X,Z)\) and that the distribution of \( \varepsilon = E(Z,\delta,\eta) \) conditional on \( Z \) is independent of \( X \). We then show that, conditional on \( Z \), the model can be transformed into one of the form

\[
\varepsilon = r(b,x)
\]

where all the assumptions in the basic simultaneous equations model considered in Matzkin (2008) are satisfied. The rank condition then implies by Matzkin (2008) that the function \( r(b,x) \) is identified. This implies that its inverse with respect to \( b, h(x,\varepsilon) \), is also identified.

To identify the distribution of \( \varepsilon \), given any \( z \) in the support of \( Z \), we note that for any \( e \in R^G \)

\[
f_{\varepsilon|Z=z}(e) = f_{b|X=x,Z=z}(h(x,e)) \left| \frac{\partial h(x,e)}{\partial e} \right|
\]

Finally, by choosing the appropriate values, \( z^{(k)} \), for \( Z \), we can identify the distribution of \( \delta \), the distribution of any coordinate \( \eta_k \) of \( \eta \), and the distribution of any \( m_k(z_k,\eta_k) \). From the latter, we can identify the function \( m_k \). The detailed proof follows.

In the proof, instead of showing identification of the functions \( m_k \) when
$m_k$ satisfies any of the cases in M(iii), we will show identification only for the case where $m_k(\mathcal{Z}_k, \eta_k) = \eta_k$. The proof for the case where the function $m_k$ satisfies M(iii.a) or M(iii.c) is very similar. The only difference is that we would be showing identification of $m_k$ and the distribution of $\eta_k$ using a different functional on the distribution of observable variables.

**Proof of the Theorem:** Let the distribution of $Y$ given $(X, Z, W)$ be given. By Assumption B, for any $b \in \mathbb{R}^G$, there exists a value $w$ in the support of $W$ and $y$ in the support of $Y$ such that for any $(x, z)$

$$\Pr(B(Y^*, W) \leq b|W = w, X = x, Z = z) = \Pr(Y \leq y|W = w, X = x, Z = z)$$

By Assumption D1, $(\delta, \eta)$ is distributed independently of $(X, Z, W)$. This implies that conditional on $(X, Z)$, $(\delta, \eta)$ is independent of $W$. By Assumption U, $B(Y^*, W) = r(X, E(Z, \delta, \eta))$. Hence, by Assumption U, conditional on $(X, Z)$, $B(Y^*, W)$ is a function of $(\delta, \eta)$. Since $(\delta, \eta)$ is independent of $W$ conditional on $(X, Z)$, this implies that $B(Y^*, W)$ is independent of $W$, conditional on $(X, Z)$. Hence, for any $b$,

$$\Pr(B(Y^*, W) \leq b|W = w, X = x, Z = z) = \Pr(B(Y^*, W) \leq b|X = x, Z = z)$$

Since

$$\Pr(B(Y^*, W) \leq b|W = w, X = x, Z = z) = \Pr(Y \leq y|W = w, X = x, Z = z)$$

we have that the distribution of $b$ conditional on $(X, Z)$ can be recovered from the distribution of $Y$ conditional on $(W, X, Z)$ by

$$\Pr(B(Y^*, W) \leq b|X = x, Z = z) = \Pr(Y \leq y|W = w, X = x, Z = z)$$
where \( y \) and \( w \) are such that \( y = T(y^*) \) for \( y^* \) such that \( B(y^*, w) = b \).

The independence between \((\delta, \eta)\) and \((X, Z, W)\) which follows by Assumption D1 also implies that \((\delta, \eta)\) is independent of \((X, Z)\). This implies that conditional on \( Z \), \((\delta, \eta)\) is independent of \( X \). Hence, conditional on \( Z \), \( \varepsilon = E(Z, \delta, \eta) \) is independent of \( X \). Consider the model

\[
\varepsilon = r(b, x)
\]

where the probability density \( f_{b, X, Z} \) and the marginal pdf of \((X, Z)\) are also known. Under our assumptions, a pair \((\tilde{r}, f_{\tilde{r}, Z})\) in \((\Gamma \times F)\) is observationally equivalent to another pair \((\tilde{r}, f_{\tilde{r}, Z})\) in \((\Gamma \times F)\) iff for all \( z \) in the support of \( Z \) and all \((b, x)\)

\[
f_{z|Z=z}(\tilde{r}(b, x)) \left| \frac{\partial \tilde{r}(b, x)}{\partial b} \right| = f_{\pi|Z=z}(\pi(b, x)) \left| \frac{\partial \pi(b, x)}{\partial b} \right|
\]

Following Matzkin (2008), one can show that under our assumptions, this equality is equivalent to requiring that for all \((b, x)\) the rank of the matrix \( A(b, x, z; \tilde{r}, \pi, f_{\pi, Z}) \) is \( G \). Note that the rank of this matrix is at least \( G \) and can’t be strictly larger than \( G + 1 \). Hence, since we have that whenever \( \tilde{r} \neq \pi \), for any \( f_{\pi, Z} \) in \( F \), there exists \( z, b, x \) such that the rank of \( A(b, x, z; \tilde{r}, \pi, f_{\pi, Z}) \) is \( G + 1 \) then \( r \) is identified. Since \( r \) is identified, its inverse, \( h \), conditional on \( x \), is also identified.

From \( h \) and \( f_{b|X, Z} \) we can identify \( f_{\varepsilon|Z} \) since for any \( z \) and any \( e \in R^G \)

\[
f_{\varepsilon|Z=z}(e) = f_{b|X=x, Z=z}(h(x, e)) \left| \frac{\partial h(x, e)}{\partial e} \right|
\]

Next, we use \( f_{\varepsilon|Z=z}(e) \) to identify the functions \( m_k \) for \( k = 0, 1, ..., K_\eta \) and the distribution of \((\delta, \eta)\). Under our assumptions, by appropriately choosing the value of \( Z \), we can ‘shut down’ the effect of all \( m'_k's \) \((k = 1, ..., K_\eta)\) on \( f_{\varepsilon|Z} \). In particular, when \( Z = (z_0, \tilde{z}_1, ..., \tilde{z}_{K_\eta}) \), \( \varepsilon = C(m_0(z_0, \delta), \alpha_1, ..., \alpha_{K_\eta}) \). By
Assumptions I and L, \( C_0 (\cdot) \equiv C (\cdot, \alpha_1, ..., \alpha_{K_n}) \) is 1-1, continuously differentiable, known, and has support \( R^G \). Then, \( \mu_0 = m_0 (z_0, \delta) = C_0^{-1} (\varepsilon) \). It follows that the density of \( \mu_0 \) conditional on \( Z = (z_0, \tilde{z}_1, ..., \tilde{z}_{K_n}) \) is identified, since

\[
f_{\mu_0|Z=(z_0,\tilde{z}_1,\ldots,\tilde{z}_{K_n})}(\mu_0) = f_{\varepsilon|Z=(z_0,\tilde{z}_1,\ldots,\tilde{z}_{K_n})}(C_0(\mu_0)) \left| \frac{\partial C_0(\mu_0)}{\partial \mu_0} \right|
\]

By Matzkin (2008), the rank condition in Assumption I implies that \( m_0 \) and \( f_\delta \) are identified.

To show identification of \( m_1 \) and of the distribution of \( \eta_1 \), we employ a similar argument. First we show identification of the joint distribution of \( (\mu_1, Z) \) and, next, we use this distribution to identify the function \( m_1 (z_1, \eta_1) \) and the distribution of \( \eta_1 \). In this case, we use Matzkin (2003). For this, we let \( Z = (\tilde{z}_0, z_1, \tilde{z}_2, ..., \tilde{z}_{K_n}) \). Assume without loss of generality that \( \mu_1 = m_1 (z_1, \eta_1) \) is an argument of the first coordinate of \( C \). Denote the value of this first coordinate of \( C \) when \( Z = (\tilde{z}_0, z_1, \tilde{z}_2, ..., \tilde{z}_{K_n}) \) by the real-valued function \( C_1 (\mu_1) \). By Assumption L and M, \( C_1 (\cdot) \) is strictly increasing and known. The distribution of \( \mu_1 \) conditional on \( Z = (\tilde{z}_0, z_1, \tilde{z}_2, ..., \tilde{z}_{K_n}) \) is identified by

\[
F_{\mu_1|Z=(\tilde{z}_0,z_1,\tilde{z}_2,\ldots,\tilde{z}_{K_n})}(\mu_1) = F_{\varepsilon|Z=(\tilde{z}_0,z_1,\tilde{z}_2,\ldots,\tilde{z}_{K_n})}(C_1(\mu_1))
\]

We will assume, for simplicity, that \( m_1 \) satisfies M(iii.b). Following the arguments in Matzkin (2003), when \( z_1 = \tau_1 \), we have that for any value \( t_1 \),

\[
F_{\mu_1|Z=(\tilde{z}_0,\tau_1,\tilde{z}_2,\ldots,\tilde{z}_{K_n})}(t_1) = F_{\eta_1|Z=(\tilde{z}_0,\tau_1,\tilde{z}_2,\ldots,\tilde{z}_{K_n})}(t_1) = F_{\tau_1}(t_1).
\]

Hence, \( F_{\eta_1} \) is identified. The function \( m_1 (z_1, \eta_1) \) can then be identified using
the expression

\[ m_1 (z_1, \eta_1) = F_{\mu_1 | Z=(z_0, z_1, \ldots, z_K)}^{-1} \left( F_{\eta_1} (\eta_1) \right). \]

Similar arguments can be used for all \( k \geq 1 \). Hence, \( m_k \) and \( F_{\eta_k} \) are identified. The identification of the distribution of \( \eta = (\eta_1, \ldots, \eta_K) \) then follows by the independence across \( k \). Similarly the identification of the distribution of \( (\delta, \eta) \) follows by the independence between \( \delta \) and \( \eta \).

We have then shown that from the distribution of \( \varepsilon \) given \( Z \), we can identify the functions \( m_0 \) and \( m_1, \ldots, m_K \), as well as the distribution of \( (\eta, \delta) \). This together with the identification of \( r \) shown above concludes the proof.

4 Examples

We next provide some models motivated by empirical examples that have been analyzed in previous studies, and describe variations of them where our results can be applied. We first consider a version of the empirical example studied in Nelson and Olson (1978). In this version of the model, each individual simultaneously allocates his time in each period to vocational school, college, and work, taking prior accumulated time as given, and considering the choice consequences for current wages. The model has four endogenous variables. The variable \( v_t^* \) denotes the allocation of time during period \( t \) to vocational school training, while \( V_t \) denotes observed accumulated vocational school training by the beginning of the period. The variable \( s_t^* \) denotes allocation of time to college study during period \( t \) measured in years, while \( S_t \) denotes accumulated formal schooling by the beginning of the period. The variable \( h_t \) denotes work experience during period \( t \), while \( H_t \) denotes ac-
cumulated work experience by the beginning of the period. The variable $w_t$ denotes the average hourly rate in period $t$. The model has additional conditioning variables, which we will denote by a vector $Q_t$. The variable $v_t^*$ is observed only when it is positive and only an ordered indicator for $s_t^*$ is observed. The equations of the model are

$$
\begin{align*}
v_t^* &= \alpha_0 + \alpha_1 V_t + \alpha_2 S_t + \alpha_3 H_t + \alpha_4 w_t + \alpha_5' Q_t + u_{vt} \\
s_t^* &= \beta_0 + \beta_1 V_t + \beta_2 S_t + \beta_3 H_t + \beta_4 w_t + \beta'_5 Q_t + u_{st} \\
h_t &= \gamma_0 + \gamma_1 V_t + \gamma_2 S_t + \gamma_3 H_t + \gamma_4 w_t + \gamma'_5 Q_t + u_{ht} \\
\ln(w_t) &= \delta_0 + \delta_1 V_t + \delta_2 S_t + \delta_3 H_t + \delta_4 v_t^* + \delta_5 s_t^* + \delta_6 h_t + \delta'_7 Q_t + u_{wt}
\end{align*}
$$

with

$$
v_t = \begin{cases} 
v_t^* & \text{if } v_t^* > 0 \\
0 & \text{otherwise} \end{cases}
$$

and

$$
s_t = \begin{cases} 
3 & \text{if } s_t^* > 1 \\
2 & \text{if } 0 < s_t^* < 1 \\
1 & \text{otherwise} \end{cases}
$$

After accommodating support conditions, this model can satisfy our assumption B when $\alpha_1 = -1, \beta_2 = -1, \delta_1 = \delta_4, \text{ and } \delta_2 = \delta_5$. This would be reasonable if wages were functions of $v_t^* + V_t$ and $s_t^* + S_t$ and the individual would make his decisions about time allocation accordingly. Assume that $(V_t, S_t, H_t, Q_t)$ is distributed independently of $(u_{vt}, u_{st}, u_{ht}, u_{wt})$. Then, again after accommodating if necessary support conditions, one can use our results to relax the linearity in the model.

Another example is a variation on a labor participation decision, which depends on other household income (see, e.g. Blundell and Powell (2003b)). Suppose that the decision is on whether to temporarily increase or decrease work hours, $Y_t^*$, over a fixed exogenously given quantity of hours, $W$, and
suppose that other household income is a function of the total quantity of hours, $Y^*_1 + W$. Then, the model could be

$$Y^*_1 = s^1 (Y_2, X_1 + \varepsilon_1) - W$$
$$Y_2 = s^2 (Y^*_1 + W, X_2 + \varepsilon_2)$$

$$Y_1 = 1 \text{ if } Y^*_1 > 0$$
$$Y_1 = 0 \text{ otherwise}$$

Incorporating unobserved heterogeneity would be possible by letting, for example, $\varepsilon_1 = \mu^1_0 + \mu_1$ and $\varepsilon_2 = \mu^2_0 + \mu_2$, where

$$\mu^1_0 = m^1_0 (Z^1_0 + \delta_1, Z^2_0 + \delta_2)$$
$$\mu^2_0 = m^2_0 (Z^1_0 + \delta_1, Z^2_0 + \delta_2)$$

$$\mu_1 = m^1_1 (Z_1, \eta_1)$$
$$\mu_2 = m^2_2 (Z_2, \eta_2)$$

The model would then be

$$Y^*_1 = s^1 (Y_2, \mu^1_0 + \mu_1 + X_1) - W$$
$$Y_2 = s^2 (Y^*_1 + W, \mu^2_0 + \mu_2 + X_2)$$

$$Y^1 = 1 \text{ if } Y^*_1 > 0$$
$$Y^1 = 0 \text{ otherwise}$$

We can analyze identification of this model making the assumptions we made in the previous sections, and applying the results in those sections.
Note that the unobservables $\delta_1$ and $\delta_2$ enter both equations while each, $\eta_1$ and $\eta_2$, enters into only one of the equations. The observable variables are $(Y_1, Y_2, X, Z, W)$ where $Y_1 \in \{0, 1\}$, $Y_2 \in R$, $X = (X_1, X_2)$, and $Z = (Z_0^1, Z_0^2, Z_1, Z_2)$. In this model, the function $B$ is given by

\[
B(Y^*, W) = \begin{pmatrix} B_1(Y^*, W) \\ B_2(Y^*, W) \end{pmatrix} = \begin{pmatrix} Y_1^* + W \\ Y_2^* \end{pmatrix} = \begin{pmatrix} Y_1^* + W \\ Y_2 \end{pmatrix}
\]

As we have shown in Section 2, this function satisfies Assumption B.

The vector of unobservable random terms, $\varepsilon$, is

\[
\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} = \begin{pmatrix} E_1(Z, \delta, \eta) \\ E_2(Z, \delta, \eta) \end{pmatrix} = \begin{pmatrix} \mu_0^1 + \mu_1 \\ \mu_0^2 + \mu_2 \end{pmatrix} = \begin{pmatrix} m_0^1 (Z_0^1 + \delta_1, Z_0^2 + \delta_2) + m_1 (Z_1, \eta_1) \\ m_0^2 (Z_0^1 + \delta_1, Z_0^2 + \delta_2) + m_2 (Z_2, \eta_2) \end{pmatrix}
\]

Let $(b_1, b_2) = B(y_1^*, y_2^*, w)$. The model obtained by transforming the model with continuous latent endogenous variables into a model with con-
Continuous ‘observable’ variables is

\[ b_1 = s^1 (b_2, \varepsilon_1 + x_1) \]
\[ b_2 = s^2 (b_1, \varepsilon_2 + x_2) \]

Under conditions specified in the previous sections, the joint distribution of \((b, X, Z)\) is identified. Assume that \(s^1\) and \(s^2\) are strictly increasing with respect to their respective last coordinates. Denote the inverse of \(s^1\) with respect to its last coordinate by \(r^1\), and the inverse of \(s^2\) with respect to its last coordinate by \(r^2\). We then get that

\[ \varepsilon_1 = r^1 (b_2, b_1) - x_1 \]
\[ \varepsilon_2 = r^2 (b_1, b_2) - x_2 \]

Under our assumptions, one can use the distribution of \((b_1, b_2)\) conditional on \((X, Z)\) to identify the function \(r = (r^1, r^2)\) and, for each value of \(Z = (z_0^1, z_0^2, z_1, z_2)\), the distribution of \(\varepsilon\) given \(Z = (z_0^1, z_0^2, z_1, z_2)\).

The next step identifies a particular structure generating the distribution of \(\varepsilon\) given \(Z = (z_0^1, z_0^2, z_1, z_2)\). In the example, the particular structure is

\[ \varepsilon_1 = m_0^1 (z_0^1 + \delta_1, z_0^2 + \delta_2) + m_1 (z_1, \eta_1) \]
\[ \varepsilon_2 = m_0^2 (z_0^1 + \delta_1, z_0^2 + \delta_2) + m_2 (z_2, \eta_2) \]

Assume that \(\alpha_0 = (0, 0)\), \(\alpha_1 = 0\), and \(\alpha_2 = 0\). When \(z_1 = \tilde{z}_1\) and \(z_2 = \tilde{z}_2\) this heterogeneity model becomes

\[^3\text{Identification of this model when \(\varepsilon\) is distributed independently of \(Z\) is, in fact, already well known using a variety of normalizations, assumptions, and arguments (Section 5 in Matzkin (2005), Example 4 in Matzkin (2007a), Section 4.2 in Matzkin (2008), Section 6 in Berry and Haile (2009b), and Section 6 in Matzkin (2010).) Estimation for this model has been developed in Matzkin (2010).} \]
\[ \varepsilon_1 = m_1^0 (z_0^1 + \delta_1, z_0^2 + \delta_2) \]
\[ \varepsilon_2 = m_2^0 (z_0^1 + \delta_1, z_0^2 + \delta_2) \]

Under our assumptions, \((m_1^0, m_2^0)\) is invertible. Let \((q^1, q^2)\) denote its inverse. We then have the model

\[ \delta_1 = q^1 (\varepsilon_1, \varepsilon_2) - z_0^1 \]
\[ \delta_2 = q^2 (\varepsilon_1, \varepsilon_2) - z_0^1 \]

By assumption, \((\delta_1, \delta_2)\) is independent of \(Z\). The steps under which the distribution of \((\varepsilon_1, \varepsilon_2)\) conditional on \(Z\) can be identified have been already described above. Hence, we can use this distribution to identify \((q^1, q^2)\) and the distribution of \((\delta_1, \delta_2)\). Since \((m_1^0, m_2^0)\) is the inverse of \((q^1, q^2)\), it is identified once \((q^1, q^2)\) is.

When \(z_0^1 = \tilde{z}_0^1\) and \(z_0^2 = \tilde{z}_0^2\), the heterogeneity model becomes

\[ \varepsilon_1 = m_1 (z_1, \eta_1) \]
\[ \varepsilon_2 = m_2 (z_2, \eta_2) \]

The function \(m_1\) and the distribution of \(\eta_1\) are identified from the distribution of \(\varepsilon_1\) given \(Z = (\tilde{z}_0^1, \tilde{z}_0^2, z_1, \tilde{z}_2)\), and the function \(m_2\) and the distribution of \(\eta_2\) are identified from the distribution of \(\varepsilon_2\) given \(Z = (\tilde{z}_0^1, \tilde{z}_0^2, \tilde{z}_1, z_2)\).

We have then described the steps under which the functions \(m_1^0, m_2^0, m_1, m_2\) and the distribution of \((\delta_1, \delta_2, \eta_1, \eta_2)\) are identified. The identification of \((r^1, r^2)\) imply the identification of \((s^1, s^2)\). Hence, all the elements in the original structural model can be identified under our assumptions.
5 Conclusions

We have presented identification results for limited dependent variable models with simultaneity in, latent or observable, continuous endogenous variables and with unobserved heterogeneity. The identification proceeds by first transforming the model with latent endogenous variables, $Y^*$, into one with continuous endogenous variables, $b$, and with a vector of unobservable variables of the same dimension, $\varepsilon$. We have provided conditions under which the joint distribution of the vector of continuous endogenous variables, $b$, and of observable exogenous variables, $(X, Z)$ is identified, and conditional on $Z$, $\varepsilon$ is independent of $X$. This have been used together with the identification result for simultaneous equation models in Matzkin (2008) to show that the distribution of $\varepsilon$ given $Z$ is identified and the mapping from $(X, \varepsilon)$ to $b$ is also identified. The distribution of $\varepsilon$ given $Z$ has then used to identify nonadditive functions of unobservable random terms. This and the mapping from $(X, \varepsilon)$ to $b$ has then been used to identify the original structural model.
6 References


CHEN, S., G.B. DAHL, and S. KAHN (2005): "Nonparametric Iden-


NEWEY, W. and J. POWELL (2003) “Instrumental Variables Estima-
