Identification of Treatment Effects in a Triangular System of Equations*

Sung Jae Jun,† Joris Pinkse,‡ Haiqing Xu,§ and Neşe Yildiz¶

this version: October 2012

We consider a model in which an outcome depends on two discrete treatment variables, where one treatment is given before the other. We formulate a three–equation triangular system with weak separability conditions. Without assuming assignment is random, we establish the identification of treatment effects using two–step matching. We allow for the treatment variables to be nonbinary and do not appeal to an identification–at–infinity argument.

Key Words: Nonparametric Identification; Discrete Endogenous Regressors; Triangular Models.
JEL Codes: C01, C14, C31.

*This paper is based on research supported by NSF grant SES–0922127. We thank the Human Capital Foundation (http://www.hcfoundation.ru), and especially Andrey P. Vavilov, for their support of the Center for Auctions, Procurements and Competition Policy (CAPCP, http://capcp.psu.edu) at Penn State University. We thank Andrew Chesher, Elie Tamer, Xavier DHaultfoeuille, and (other) participants of the 2010 Cowles foundation workshop and 2012 CIREQ–CEMMAP conference for their helpful comments.
†(Corresponding Author) CAPCP and Department of Economics, Pennsylvania State University, 608 Kern Graduate Building, University Park, PA 16802, sjun@psu.edu
‡CAPCP and Department of Economics, Pennsylvania State University, joris@psu.edu
§Department of Economics, University of Texas at Austin, h.xu@austin.utexas.edu
¶Department of Economics, University of Rochester, nese.yildiz@rochester.edu

1. Introduction

This paper deals with with nonparametric identification in a three–equation nonparametric model with discrete endogenous regressors. We provide conditions under which treatment effects can be (point) identified. Like Jun, Pinkse, and Xu (2011, 2012) we use a *Dynkin system* approach.

The model we study is similar to that of Vytlacil and Yildiz (2007); Jun, Pinkse, and Xu (2012) and others in that we make and exploit a weak separability assumption. However, Vytlacil and Yildiz (2007) specifically excludes the possibility of nonbinary categorical endogenous regressors, imposes extreme support restrictions on the covariates, and only deals with the two–equation case. The nonbinary categorical regressor case is not discussed in (the published version of) Jun, Pinkse, and Xu (2012), which further does not deal with the present, more complicated, three–equation model featuring two discrete endogenous regressors. There are other papers that do have a three–equation model and/or allow for nonbinary regressors (e.g. Lewbel, 2007; Imai and Van Dyk, 2004; Black and Smith, 2004), but the model and/or the object of interest is generally different.

There are many examples in which (a (semi)parametric version of) our structure has been used. We mention only a few. Flores and Flores-Lagunes (2009) studies the effects of smoking on birth weight through the mechanism of gestation time. Dearden, Ferri, and Meghir (2002) analyzes the effects of school type and class size on earnings and educational attainment. Lechner (2001) is a similar application with a simpler dependence structure than ours.

The focus here is on point identification. There are several papers (e.g. Shaikh and Vytlacil, 2011; Chiburis, 2010; Mourifié, 2012) which develop bounds on treatment effects in models that are similar to, but simpler than, the one in this paper using weaker monotonicity assumptions than are imposed here. As shown in Jun, Pinkse, and Xu (2011), the Dynkin system approach can be used to obtain sharp bounds in an environment in which there is only partial identification. We do not pursue this possibility in the current paper.

The Dynkin system approach is a natural scheme that allows one to collect and aggregate information contained in the data in a natural and thorough fashion through a recursion scheme.[1] Each combination of observables implies that the unobservable error terms belong to certain sets. From these sets one can infer additional information through various operations on these sets. In this paper we use a version of the Dynkin system approach, first used in Jun, Pinkse, and

[1] d'Haultfœuille and Février (2011) also uses a recursion scheme for the purpose of identification, but both their method and their model is different from ours.

Xu (2012), which exploits *matching* in addition to the union and difference operators used in Jun, Pinkse, and Xu (2011). Matching has been used frequently in the past. For instance, Pinkse (2001) used it to avoid support conditions in estimating weakly separable nonparametric regression functions. The way we use matching in this paper is closer to Vytlacil and Yildiz (2007) albeit that our procedure, as already mentioned, can be applied more generally.

Although the fact that the Dynkin system approach requires only weak covariate support restrictions is an attractive feature, this paper will focus on the other extensions since the support restrictions issue was discussed at length in Jun, Pinkse, and Xu (2012), albeit for the two–equation binary endogenous regressor case.

The remainder of the paper is organized as follows. In section 2 we lay out our model and discuss the objects we want to identify and the rationale for our desire to do so. Section 3 provides a rough description of the basic ideas underlying our identification approach. These ideas are formalized and illustrated using more complete examples in sections 4 and 5. Finally, section 6 provides a brief sketch of how the identification methods proposed here could be implemented.

2. Model

Imposing weak separability in multiple places, we consider the model

$$\begin{cases} \boldsymbol{y} = g\{\alpha(\boldsymbol{x}, \boldsymbol{s}, \boldsymbol{d}), \boldsymbol{\epsilon}\}, \\ \boldsymbol{s} = \sum_{j=1}^{\eta_s} \mathbb{1}\{\boldsymbol{v} > m_j(\boldsymbol{w}, \boldsymbol{d})\}, \\ \boldsymbol{d} = \sum_{j=1}^{\eta_d} \mathbb{1}\{\boldsymbol{u} > p_j(\boldsymbol{z})\}, \end{cases} \tag{1}$$

where $\eta_s, \eta_d \geq 1$, and $g, m_1, \ldots, m_{\eta_s}, p_1, \ldots, p_{\eta_d}$ are unknown functions. We impose that $p_0(z) = m_0(w, d) = 0$, $p_j(z) < p_{j+1}(z)$, $m_j(w, d) < m_{j+1}(w, d)$, and $p_{\eta_d+1}(z) = m_{\eta_s+1}(w, d) = 1$. This is without loss of generality in view of assumption B below. Before exploring the model in more detail, we make several model assumptions. Let $\mathscr{U} = (0, 1]$.

Assumption A. *$(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{\epsilon})$ is independent of $(\boldsymbol{w}, \boldsymbol{z}, \boldsymbol{x})$.*

Assumption B. *The distribution of $(\boldsymbol{u}, \boldsymbol{v})$ is absolutely continuous with respect to the Lebesque measure μ with support $\mathscr{S}_{\boldsymbol{uv}} = \mathscr{U}^2$ and $\boldsymbol{u}, \boldsymbol{v}$ have marginal uniform distributions on $\mathscr{U}$.*

Assumption C. *$\mathbb{E}\{g(\alpha, \boldsymbol{\epsilon}) | \boldsymbol{u} = u, \boldsymbol{v} = v\}$ is for all $u, v \in \mathscr{U}$ strictly monotonic in α.*

Both $\boldsymbol{s}$ and $\boldsymbol{d}$ are general ordered response variables, which are allowed to be endogenous. Instead of having one variable with $(1 + \eta_s)(1 + \eta_d)$ support points, we explicitly have two treatment variables here. As we discuss in detail below, we do this to consider various structural

parameters based on counterfactual outcomes. So, the model in (1) is more general than Vytlacil and Yildiz (2007, VY) and Jun, Pinkse, and Xu (2012, JPX).

Assumption A is strong but almost indispensable in the fully nonparametric identification literature. The second half of assumption B constitutes a normalization. The first part is restrictive, but is difficult to avoid. Please note, however, that $\boldsymbol{u}$ and $\boldsymbol{v}$ are allowed to be correlated and that the support of $(\boldsymbol{u}, \boldsymbol{v})$ given $\boldsymbol{\epsilon}$ need not be $\mathscr{U}^2$.

Monotonicity is a common assumption in the nonparametric identification literature, but unlike e.g. Chernozhukov and Hansen (2005); Chesher (2003); Imbens and Newey (2009), assumption C does not require monotonicity in the error term of the structural function itself but it requires monotonicity of the expectation.[2] For the use of the Dynkin system idea to identify a structural function under a stronger form of monotonicity, see Jun, Pinkse, and Xu (2011).

The parameters of interest will be average treatment effects, where $\boldsymbol{s}$ and $\boldsymbol{d}$ are treatment variables. Let $\boldsymbol{y}_{sd} = g\{\alpha(\boldsymbol{x}, s, d), \boldsymbol{\epsilon}\}$. Thus, $\boldsymbol{y}_{sd} = \boldsymbol{y}$ if $(\boldsymbol{s}, \boldsymbol{d}) = (s, d)$, but if $(\boldsymbol{s}, \boldsymbol{d}) \neq (s, d)$ then $\boldsymbol{y}_{sd}$ is the value $\boldsymbol{y}$ would have taken if the same individual had $\boldsymbol{s} = s, \boldsymbol{d} = d$. So $\boldsymbol{y}_{sd}$ is a typical counterfactual outcome variable but with two indices instead of the usual one. The focus in this paper will be on the identification of

$$\psi(x^*, s^*, d^*) = \mathbb{E}(\boldsymbol{y}_{s^*d^*} | \boldsymbol{x} = x^*) = \mathbb{E}g\{\alpha(x^*, s^*, d^*), \boldsymbol{\epsilon}\},$$

where x^*, s^*, d^* is chosen by the researcher. For instance, $\psi(x^*, s^*, d^*)$ could be the expected wage of a male worker ($\boldsymbol{x} = x^*$) if he had both a college degree ($\boldsymbol{d} = s^*$) and received on the job training ($\boldsymbol{s} = s^*$).

The function ψ can be used to obtain several causal effects of interest. One such object are the (conditional) integrated average treatment effects of $\boldsymbol{d}$ and $\boldsymbol{s}$, i.e.

$$\mathcal{T}_{\boldsymbol{d}}(d^*, \tilde{d}^* | x^*) = \mathbb{E}(\boldsymbol{y}_{\boldsymbol{s}d^*} - \boldsymbol{y}_{\boldsymbol{s}\tilde{d}^*} | \boldsymbol{x} = x^*) = \mathbb{E}\{\psi(x^*, \boldsymbol{s}, d^*) - \psi(x^*, \boldsymbol{s}, \tilde{d}^*) | \boldsymbol{x} = x^*\}, \tag{2}$$

$$\mathcal{T}_{\boldsymbol{s}}(s^*, \tilde{s}^* | x^*) = \mathbb{E}(\boldsymbol{y}_{s^*\boldsymbol{d}} - \boldsymbol{y}_{\tilde{s}^*\boldsymbol{d}} | \boldsymbol{x} = x^*) = \mathbb{E}\{\psi(x^*, s^*, \boldsymbol{d}) - \psi(x^*, \tilde{s}^*, \boldsymbol{d}) | \boldsymbol{x} = x^*\}. \tag{3}$$

The 'marginal' distributions of $\boldsymbol{s}$ and $\boldsymbol{d}$ (of the subpopulation with $\boldsymbol{x} = x^*$) are used to capture the causal effects of $\boldsymbol{d}$ and $\boldsymbol{s}$, respectively. The exogenous variable $\boldsymbol{x}$ can be integrated out afterwards to obtain causal effects for the entire population. Please see Heckman (2005); Heckman and Vytlacil (2007) among many others for more discussion on integration and causal parameters.

Thus, the identification of the function ψ is our main objective. Our approach, like VY, is based on the idea of matching. However, VY cannot be used to deal with the extra complexity of

[2] Under additive separability of the error term, both types of monotonicity are satisfied.

the current model, especially the presence of nonbinary endogenous regressors, so we employ a version of JPX here.

Finally, since both $\boldsymbol{s}$ and $\boldsymbol{d}$ take finitely many values, it is possible to combine them into a single discrete variable $\boldsymbol{r}$ taking $(\eta_s+1)(\eta_d+1)$ different values.[3] The last two equations in (1) could then be replaced by a single equation of the form $\boldsymbol{r} = \sum_{j=1}^{\eta_r} \mathbb{1}\{\boldsymbol{\xi} > n_j(\boldsymbol{w},\boldsymbol{z})\}$. There are two problems with doing this. First, since the threshold crossing form of the new equation is based on an arbitrary ordering, the structural interpretation of $\boldsymbol{u}$ and $\boldsymbol{v}$ would be lost. For example, this approach would make it impossible to do a counterfactual analysis about the choices of individuals with a high unobserved proclivity to attend college ($\boldsymbol{u}$ high) but who dislike job training programs ($\boldsymbol{v}$ low). Second, although it would still be possible to identify average treatment effects like those in (2) using the alternative approach, our matching procedure would require more stringent support restrictions on $(\boldsymbol{w},\boldsymbol{z})$, because the two functions m and p would be reduced to a single function n. Therefore, (1) is more interesting than a single endogenous regressor model.

3. Description

We now proceed with providing a broad and rough description of our identification strategy. As noted the object that we wish to identify is

$$\psi(x^*,s^*,d^*) = \mathbb{E}g\{\alpha(x^*,s^*,d^*),\boldsymbol{\epsilon}\}. \tag{4}$$

The objects that we can identify directly from the data are

$$\begin{cases} \delta(x,s,d,w,z) = \mathbb{E}\{\boldsymbol{y}\mathbb{1}(\boldsymbol{s}=s)\mathbb{1}(\boldsymbol{d}=d)|\boldsymbol{x}=x,\boldsymbol{w}=w,\boldsymbol{z}=z\} \\ \qquad = \mathbb{E}\Big(g\{\alpha(x,s,d),\boldsymbol{\epsilon}\}\mathbb{1}\{(\boldsymbol{u},\boldsymbol{v}) \in \big(p_d(z),p_{d+1}(z)\big] \times \big(m_s(w,d),m_{s+1}(w,d)\big]\Big), \\ \mathbb{P}(\boldsymbol{s}=s,\boldsymbol{d}=d|\boldsymbol{w}=w,\boldsymbol{z}=z) = \mathbb{P}\Big\{(\boldsymbol{u},\boldsymbol{v}) \in \big(p_d(z),p_{d+1}(z)\big] \times \big(m_s(w,d),m_{s+1}(w,d)\big]\Big\}, \end{cases} \tag{5}$$

where x,s,d,w,z can be varied at will (subject to support restrictions). If $\mathscr{S}_{\boldsymbol{uv}}$ is the support of $(\boldsymbol{u},\boldsymbol{v})$ and

$$\kappa(A,a) = \mathbb{E}\big[g(a,\boldsymbol{\epsilon})\mathbb{1}\{(\boldsymbol{u},\boldsymbol{v}) \in A\}\big], \tag{6}$$

then we would like to identify $\kappa\{\mathscr{S}_{\boldsymbol{uv}},\alpha(x^*,s^*,d^*)\}$ given that we can identify $\kappa\{A,\alpha(x,s,d)\}$ for rectangles $A = \big(p_d(z),p_{d+1}(z)\big] \times \big(m_s(w,d),m_{s+1}(w,d)\big]$ for combinations of x,s,d,w,z in the support. Please note that here and in the remainder of this paper a rectangle is defined as the Cartesian product of two half–open $(\cdot,\cdot]$ intervals.

[3] We thank Elie Tamer for pointing this out.

A few things should be apparent from (6). First, if we can find a sequence $\{z_n, w_n\}$ such that

$$\lim_{n\to\infty} p_{d^*}(z_n) = 0, \quad \lim_{n\to\infty} p_{d^*+1}(z_n) = 1, \quad \lim_{n\to\infty} m_{s^*}(w_n, d^*) = 0, \quad \lim_{n\to\infty} m_{s^*+1}(w_n, d^*) = 1,$$

then identification of (4) obtains since

$$\lim_{n\to\infty} \mathbb{E}\{\boldsymbol{y}\mathbb{1}(\boldsymbol{s} = s^*)\mathbb{1}(\boldsymbol{d} = d^*)|\boldsymbol{x} = x^*, \boldsymbol{w} = w_n, \boldsymbol{z} = z_n\}$$
$$= \mathbb{E}\Big[g\{\alpha(x^*, s^*, d^*), \boldsymbol{\epsilon}\}\mathbb{1}\{(\boldsymbol{u}, \boldsymbol{v}) \in \mathscr{U}^2\}\Big] = \psi(x^*, s^*, d^*).$$

Such an identification–at–infinity argument is undesirable since it generally makes inefficient use of the data (Khan and Tamer, 2010) and imposes extreme support restrictions.

Further, suppose that $\kappa(A, a)$ and $\kappa(A^*, a)$ are identified. If A, A^* are disjoint then $\kappa(A \cup A^*, a) = \kappa(A, a) + \kappa(A^*, a)$ is identified, also. Likewise, if A is contained in A^* then $\kappa(A^* - A, a) = \kappa(A^*, a) - \kappa(A, a)$ is identified, also. Such union and difference operations can be repeated for different A–sets. For instance, if $\kappa(A, a), \kappa(A^*, a), \kappa(\bar{A}, a)$ are identified, A is contained in A^* and $(A^* - A) \cap \bar{A} = \emptyset$ then $\kappa\{(A - A^*) \cup \bar{A}, a\}$ is identified.

Other than through an identification–at–infinity argument like the one described above, these two basic operations are insufficient to identify $\psi(x^*, s^*, d^*)$. For instance, if $s^* = d^* = 0$ then the rectangles A all originate at $(0, 0)$ and none of them extend beyond $\big(\sup_z p_1(z), \sup_w m_1(w, 0)\big)$ such that $\kappa\{\mathscr{S}_{\boldsymbol{uv}}, \alpha(x^*, s^*, d^*)\}$ cannot be identified.

Now suppose that for some $A, x, s, d, \tilde{x}, \tilde{s}, \tilde{d}$, $\kappa\{A, \alpha(\tilde{x}, \tilde{s}, \tilde{d})\}$ and $\kappa\{A, \alpha(x, s, d)\}$ are both identified. Then assumption C implies that $\kappa\{A, \alpha(\tilde{x}, \tilde{s}, \tilde{d})\} = \kappa\{A, \alpha(x, s, d)\}$ is then equivalent to $\alpha(\tilde{x}, \tilde{s}, \tilde{d}) = \alpha(x, s, d)$. Once it is known that $\alpha(\tilde{x}, \tilde{s}, \tilde{d}) = \alpha(x, s, d)$, identification of $\kappa\{A^*, \alpha(x, s, d)\}$ for any set A^* implies identification of $\kappa\{A, \alpha(\tilde{x}, \tilde{s}, \tilde{d})\}$. Again, these operations can be repeated for different combinations of $(\tilde{x}, \tilde{s}, \tilde{d})$, (x, s, d), and A, and combined with the intersection and union operations described above.

In the remainder of this paper these procedures are formally expressed in terms of a Dynkin system and their power is illustrated using some concrete examples. To get a whiff of the basic premise, consider the following rudimentary example which exploits only a few features of the proposed methodology and corresponds to the simplest possible meaningful case, i.e. $\eta_d = \eta_s = 1$. In particular, the example assumes that the joint support is simply the product of the marginal supports, which is unnecessary as will become apparent later in this paper.

Example 1. Consider figure 1 and suppose for now that the m_s functions are identified and that the joint support of the covariates equals the product of their marginal supports. Then the

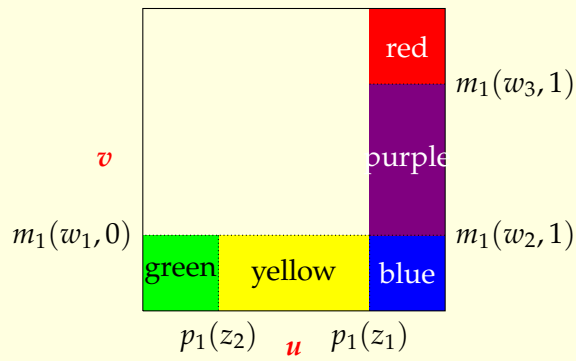


FIGURE 1. Simple matching procedures

following quantities are identified directly from the data.

$$\begin{cases} \delta(x^*, 0, 0, w_1, z_2) &= \kappa\{\text{green}, \alpha(x^*, 0, 0)\}, \\ \delta(x^*, 0, 0, w_1, z_1) &= \kappa\{\text{green+yellow}, \alpha(x^*, 0, 0)\}, \\ \delta(x_1, 0, 1, w_2, z_1) &= \kappa\{\text{blue}, \alpha(x_1, 0, 1)\}, \\ \delta(x_1, 0, 1, w_2, z_2) &= \kappa\{\text{blue+yellow}, \alpha(x_1, 0, 1)\}, \\ \delta(x_1, 0, 1, w_3, z_1) &= \kappa\{\text{blue+purple}, \alpha(x_1, 0, 1)\}, \\ \delta(x_2, 1, 1, w_3, z_1) &= \kappa\{\text{red}, \alpha(x_2, 1, 1)\}, \\ \delta(x_2, 1, 1, w_2, z_1) &= \kappa\{\text{red+purple}, \alpha(x_2, 1, 1)\}. \end{cases} \tag{7}$$

Subtracting the first and third lines in (7) from the second and fourth lines, respectively, yields $\kappa\{\text{yellow}, \alpha(x^*, 0, 0)\}$ and $\kappa\{\text{yellow}, \alpha(x_1, 0, 1)\}$, which are equal if and only if $\alpha(x^*, 0, 0) = \alpha(x_1, 0, 1)$. Likewise, subtracting the third and sixth lines in (7) from the fifth and seventh lines allows one to verify whether $\alpha(x_1, 0, 1) = \alpha(x_2, 1, 1)$.

Once values x_1, x_2, x_3 are found such that $\alpha(x^*, 0, 0) = \alpha(x_1, 0, 1) = \alpha(x_2, 1, 1) = \alpha(x_3, 1, 0)$, $\kappa\{\mathscr{S}_{\boldsymbol{uv}}, \alpha(x^*, 0, 0)\}$ can be computed as (for instance) the sum of the quantities $\delta(x^*, 0, 0, w_1, z_1)$, $\delta(x_1, 0, 1, w_2, z_1)$, $\delta(x_2, 1, 1, w_2, z_1)$, and $\delta(x_3, 1, 0, w_1, z_1)$. □

The above explanation presumes that we know the rectangles A, which in turn presumes that the p_d, m_s functions are identified. Although $p_d(z) = \mathbb{P}(\boldsymbol{d} < d | \boldsymbol{z} = z)$ is trivially identified, identification of the m_s functions requires some work. Indeed, the procedure for identifying the m_s functions is analogous to, but simpler than, the procedure for identifying $\psi(x^*, s^*, d^*)$ once the m_s functions have been identified.

Indeed, if

$$\theta(V, m) = \mathbb{P}(\boldsymbol{u} \in V, \boldsymbol{v} < m), \tag{8}$$

then

$$\theta\{\,(p_d(z), p_{d+1}(z)\,], m_s(w,d)\}, \tag{9}$$

is trivially identified for all s, d, w, z provided that $(w, z) \in \mathscr{S}_{\boldsymbol{wz}}$. Once again union and intersection operations can be used and combined. For any set V and any $s, d, w, \tilde{s}, \tilde{d}, \tilde{w}$, equality of $\theta\{V, m_{\tilde{s}}(\tilde{w}, \tilde{d})\}$ and $\theta\{V, m_s(w,d)\}$ is equivalent to $m_{\tilde{s}}(\tilde{w}, \tilde{d}) = m_s(w,d)$ and such equality implies that for any V^*, identification of $\theta\{V^*, m_{\tilde{s}}(\tilde{w}, \tilde{d})\}$ implies identification of $\theta\{V^*, m_s(w,d)\}$.

4. Identification of m

We now establish the identification of $m_{s^*}(w^*, d^*)$ formally. Define[4]

$$\theta(V, m) = \mathbb{P}(\boldsymbol{u} \in V, \boldsymbol{v} \leq m), \qquad V \subset \mathscr{U},\ m \in \mathscr{U}.$$

Further let $\mathscr{S}_{\boldsymbol{z}}(w)$ be the support of $\boldsymbol{z}$ conditional on $\boldsymbol{w} = w$ and define

$$\mathscr{V}(d, w) = \{(p_d(z), p_{d+1}(z)] : z \in \mathscr{S}_{\boldsymbol{z}}(w)\}, \quad d = 0, \ldots, \eta_d. \tag{10}$$

Then $\theta\{V, m_s(w,d)\}$ is identified when $V \in \mathscr{V}(d, w)$ because

$$\theta\{(p_d(z), p_{d+1}(z)], m_s(w,d)\} = \mathbb{P}(\boldsymbol{s} < s, \boldsymbol{d} = d | \boldsymbol{w} = w, \boldsymbol{z} = z). \tag{11}$$

We need $\boldsymbol{z}$ to belong to the union in (10) to ensure that the right hand side in (11) is defined.

We now show that $\theta\{V, m_s(w,d)\}$ is identified for a much broader class of sets than $\mathscr{V}(d, w)$.

Definition 1. $\mathscr{D}^*(d, s, w)$ is the collection $\mathscr{D}^*_\infty(d, s, w)$ in the following iterative scheme. Let $\mathscr{D}^*_0(d, s, w) = \mathscr{V}(d, w)$. Then, for all $t \geq 0$, $\mathscr{D}^*_{t+1}(d, s, w)$ consists of all sets A^* such that at least one of the following conditions is satisfied.

(i) $A^* \in \mathscr{D}^*_t(d, s, w)$;

(ii) $\exists A_1, A_2 \in \mathscr{D}^*_t(d, s, w) : A_1 \subset A_2,\ \mu(A_2 - A_1) > 0,\ A^* = A_2 - A_1$;

(iii) $\exists A_1, A_2 \in \mathscr{D}^*_t(d, s, w) : A_1 \cap A_2 = \emptyset,\ \mu(A_1 \cup A_2) > 0,\ A^* = A_1 \cup A_2$;

(iv) $\exists (\bar{d}, \bar{s}, \bar{w}) : m_s(d, w) = m_{\bar{s}}(\bar{d}, \bar{w}),\ \mathscr{D}^*_t(d, s, w) \cap \mathscr{D}^*_t(\bar{d}, \bar{s}, \bar{w}) \neq \emptyset,\ A^* \in \mathscr{D}^*_t(\bar{d}, \bar{s}, \bar{w})$. □

The conditions in definition 1 are similar to those in Jun, Pinkse, and Xu (2012). Note that $\{\mathscr{D}^*_t(d, s, w) : t = 0, 1, \cdots\}$ is an increasing sequence of collections, such that $\mathscr{D}^*(d, s, w)$ is the infinite union of $\mathscr{D}^*_t(d, s, w)$'s.[5] Note further that $\mathscr{D}^*(d, s, w)$ is indexed by s, w as well as d. If $\mathscr{S}_{\boldsymbol{z}}(w)$ is the same for all w values then the argument pursued in this section is simpler, but such

[4]We use $\subset$ as a generic symbol for subset, where some other authors might distinguish between proper and nonproper subsets.

[5]Please note that this is the infinite union of collections of sets, not the collection of infinite unions of sets. To see the difference, consider that $\mathscr{U} = \cup_{n=1}^{\infty}(1/n, 1]$ but $\mathscr{U} \notin \cup_{n^*=1}^{\infty}\{(1/n, 1]\}_{n=1}^{n^*}$. It is the latter concept that is used here.

support restrictions are undesirable because it excludes the possibility that $\boldsymbol{w}, \boldsymbol{z}$ have elements in common and it also precludes the situation in which certain combinations of $(\boldsymbol{w}, \boldsymbol{z})$ values cannot occur.

All elements of $\mathscr{D}^*$ are defined in terms of (combinations of) the unknown p_d and m_s functions. Hence, each element can be thought of as an unknown parameter. In lemma 1 we show that all elements in $\mathscr{D}^*$ are identified. Subsequently, we obtain a condition that is sufficient for identification of $m_{s^*}(w^*, d^*)$.

Lemma 1. *Suppose that assumptions A and B are satisfied.*

(i) For all $(d, s, w) \in \mathscr{S}_{\boldsymbol{dsw}}$, every $V \in \mathscr{D}^*(d, s, w)$ is identified;

(ii) $\theta\{V, m_s(w, d)\}$ is identified whenever $(d, s, w) \in \mathscr{S}_{\boldsymbol{dsw}}$ and $V \in \mathscr{D}^*(d, s, w)$.

Proof. See appendix A. □

Assumption D. $\mathscr{U} \in \mathscr{D}^*(d^*, s^*, w^*)$.

Since $\{\mathscr{D}_t^*(d^*, s^*, w^*) : t = 0, 1, 2, \cdots\}$ is an increasing sequence of collections of sets and $\boldsymbol{d}, \boldsymbol{s}$ take finitely many values, assumption D is equivalent to the requirement that there exists a finite T such that $\mathscr{U} \in \mathscr{D}_T^*(d^*, s^*, w^*)$, in contrast to e.g. d'Haultfœuille and Février (2011).

Theorem 1. *If assumptions A, B and D are satisfied then, $m_{s^*}(w^*, d^*)$ is identified.*

Proof. See appendix A. □

Assumption D involves conditions on the support of $\boldsymbol{z}$; the class $\mathscr{D}^*(d, s, w)$ is mostly determined by the amount of variation available in $\boldsymbol{z}$ given $\boldsymbol{d} = d, \boldsymbol{s} < s, \boldsymbol{w} = w$. Assumption D is satisfied in various situations. Assumption D is satisfied when there exists a disjoint partition of $\mathscr{U}$ such that every element of the partition belongs to $\mathscr{D}^*(d^*, s^*, w^*)$. Even though $\mathscr{V}(d^*, s^*, w^*)$ will generally not contain such a partition, $\mathscr{D}^*(d^*, s^*, w^*)$ usually does contain $\mathscr{U}$.

Indeed, suppose that $\mathscr{D}^*(d^*, s^*, w^*) \cap \mathscr{D}^*(\bar{d}, \bar{s}, \bar{w}) \neq \emptyset$ for some $(\bar{d}, \bar{s}, \bar{w}) \in \mathscr{S}_{\boldsymbol{dsw}}$. Then, by (iv) in definition 1, $m_{s^*}(w^*, d^*) = m_{\bar{s}}(\bar{w}, \bar{d})$ implies that $\mathscr{D}^*(d^*, s^*, w^*) = \mathscr{D}^*(\bar{d}, \bar{s}, \bar{w})$. Therefore, not only $\mathscr{V}(d^*, w^*)$ but also $\mathscr{V}(\bar{d}, \bar{w})$ should be taken into account, which is particularly useful when $d^* \neq \bar{d}$. This reasoning suggests a simple sufficient condition, which we state as a corollary.

Corollary 1 (Sufficient conditions)**.** *Suppose that there exists a sequence $\{(s_j, w_j) \in \mathscr{S}_{\boldsymbol{sw}} : j = 0, 1, \cdots, \eta_d\}$ such that $m_{s_j}(w_j, j) = m_{s^*}(w^*, d^*)$ for all $j = 0, 1, \cdots, \eta_d$. Further, suppose that*

$$\forall j = 1, \ldots, \eta_d - 1 : \inf_{z \in \mathscr{S}_{\boldsymbol{z}}(w_{j+1})} p_{j+1}(z) \leq \sup_{z \in \mathscr{S}_{\boldsymbol{z}}(w_j)} p_j(z), \tag{12}$$

where each p_j is a continuous function and z is continuously distributed. Then, $m_{s^*}(w^*, d^*)$ is identified. □

Please note that corollary 1 only imposes restrictions on the relationship between p_d and p_{d+1} for all values of d. For instance, we do not require there to be a direct relationship between p_d and p_{d+2}. Indeed, the matching procedure can be *chained* in the sense that we can first establish equality of $m_{s_0}(w_0, 0)$ to $m_{s_1}(w_1, 1)$, then uncover that $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1) = m_{s_2}(w_2, 2)$, and so on.

To illustrate corollary 1, consider the following example.

Example 2 (Ordered response)**.** Suppose that for all d, z and some β_0 and $-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_{\eta_d+1} = \infty$, $p_d(z) = \Phi(\gamma_d + \beta^\intercal z)$, as would be the case in an ordered probit model. This is one of the least favorable cases for our procedure since for all z, z^* and $d = 1, \ldots, \eta_d$, $p_d(z) < p_d(z^*) \Rightarrow p_{d+1}(z) \leq p_{d+1}(z^*)$ and $p_{d-1}(z) \leq p_{d-1}(z^*)$.

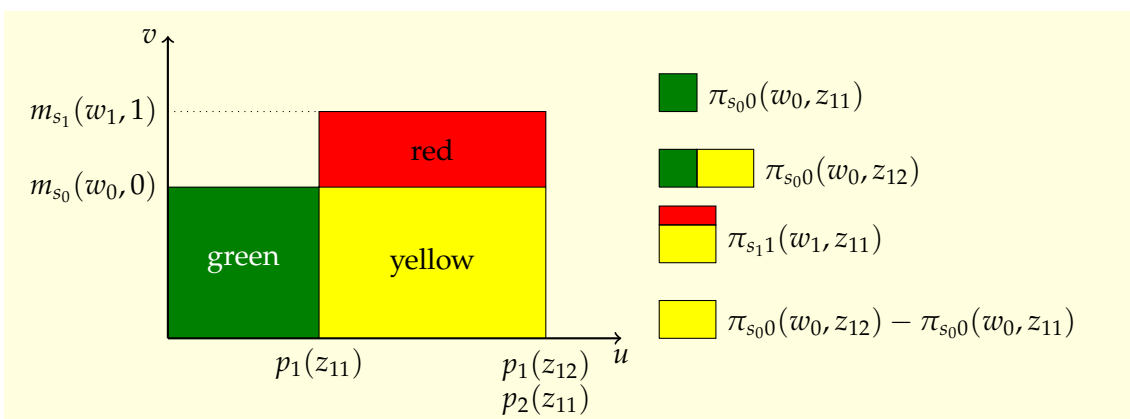
So the conditions of corollary 1 are satisfied if

$$\sup_{z, z^* \in \mathscr{S}_z} \beta^\intercal (z - z^*) \geq \max_{d=1,\ldots,\eta_d - 1} (\gamma_{d+1} - \gamma_d). \qquad \square$$

To illustrate the idea of theorem 1, we provide the following two fairly concrete examples. Let

$$\begin{aligned} \pi_{sd}(w, z) = \mathbb{P}(s < s, d = d | w = w, z = z) &= \mathbb{P}\{p_d(z) < u \leq p_{d+1}(z),\ v \leq m_s(w, d)\} \\ &= \theta\{(p_d(z), p_{d+1}(z)], m_s(w, d)\}, \end{aligned} \tag{13}$$

which is identified provided that $z \in \mathscr{S}_z(w)$.



Here: $\pi_{s_0 0}(w_0, z_{12}) - \pi_{s_0 0}(w_0, z_{11}) < \pi_{s_1 1}(w_1, z_{11})$, so $m_{s_0}(w_0, 0) < m_{s_1}(w_1, 1)$.

FIGURE 2. Verifying whether $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1)$.

Example 3 (Uncovering that $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1)$)**.** We verify whether $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1)$ for some candidate pair (s_1, w_1). Our approach is described below and illustrated in figure 2, which assumes the existence of values z_{11}, z_{12} such that $p_1(z_{12}) = p_2(z_{11})$. It should be apparent from figure 2 that $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1)$ if and only if the measure of the red area is zero.

The measures of the yellow area, the yellow plus the green area, and the yellow plus the red area are identified directly from the data. The measure of the yellow area can then be learned as (yellow+green) − green and finally the measure of the red area as (yellow+red) − yellow.

The formal identification argument is as follows. First,

$$\mathscr{D}_0^*(0, s_0, w_0) \supset \big\{ \big(0, p_1(z_{11})\big], \big(0, p_1(z_{12})\big] \big\}, \quad \mathscr{D}_0^*(1, s_1, w_1) \supset \big\{ \big(p_1(z_{11}), p_2(z_{11})\big] \big\}.$$

Using (i) and (ii) of definition 1 it follows that $V = \big(p_1(z_{11}), p_1(z_{12})\big] = \big(p_1(z_{11}), p_2(z_{11})\big] \in \mathscr{D}_1^*(0, s_0, w_0) \cap \mathscr{D}_1^*(1, s_1, w_1)$. Thus,

$$\begin{cases} \theta\big\{V, m_{s_0}(w_0, 0)\big\} & = \pi_{s_0 0}(w_0, z_{12}) - \pi_{s_0 0}(w_0, z_{11}), \\ \theta\big\{V, m_{s_1}(w_1, d_1)\big\} & = \pi_{s_1 1}(w_1, z_{11}), \end{cases}$$

are both identified; they are equal if and only if $m_{s_1}(w_1, 1) = m_{s_0}(w_0, 0)$. □

In example 3 it is implicitly assumed that $z_{11}, z_{12} \in \mathscr{S}_z(w_0)$ and that $z_{11} \in \mathscr{S}_z(w_1)$. However, theorem 1 does not require this. Indeed, if there exist z_{110}, z_{111} such that $p_1(z_{110}) = p_1(z_{111})$, $p_1(z_{12}) = p_2(z_{111})$, and both $z_{110}, z_{12} \in \mathscr{S}_z(w_0)$ and $z_{111} \in \mathscr{S}_z(w_1)$ then we can match $\pi_{s_0 0}(w_0, z_{12}) - \pi_{s_0 0}(w_0, z_{110})$ with $\pi_{s_1 1}(w_1, z_{111})$ to obtain $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1)$.

Example 4 (Verifying that $m_{s_1}(w_1, 1) = m_{s_2}(w_2, 2)$)**.** We now turn to the task of verifying that $m_{s_1}(w_1, 1) = m_{s_2}(w_2, 2)$ once $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1)$ has been established. The procedure is illustrated in figure 3 and described below, which presumes the existence of z_{21}, z_{22} for which $p_3(z_{22}) = p_2(z_{21})$.

Again the question is whether the measure of the red area equals zero. Pink, orange, and yellow are directly identified, which allows us to deduce (pink + orange). Further, (pink + orange + yellow + red) $= \pi_{s_0 0}(w_0, z_{21}) + \pi_{s_1 1}(w_1, z_{21})$ is identified, and hence so is (yellow + red), which in turn implies the identification of red.

Formally, it follows from example 3 that $\mathscr{D}_t^*(0, s_0, w_0) = \mathscr{D}_t^*(1, s_1, w_1)$ for all $t \geq 2$. So for sufficiently large t, $V = \big(p_2(z_{22}), p_2(z_{21})\big] \in \mathscr{D}_t^*(0, s_0, w_0)$. But since $V = \big(p_2(z_{22}), p_3(z_{22})\big] \in \mathscr{D}_2^*(0, s_2, w_2)$, equality of $m_{s_0}(w_0, 0)$ and $m_{s_2}(w_2, 2)$ can be verified using the set V. □

Once we have ascertained that $m_{s_0}(w_0, 0) = m_{s_1}(w_1, 1) = m_{s_2}(w_2, 2)$, we can identify

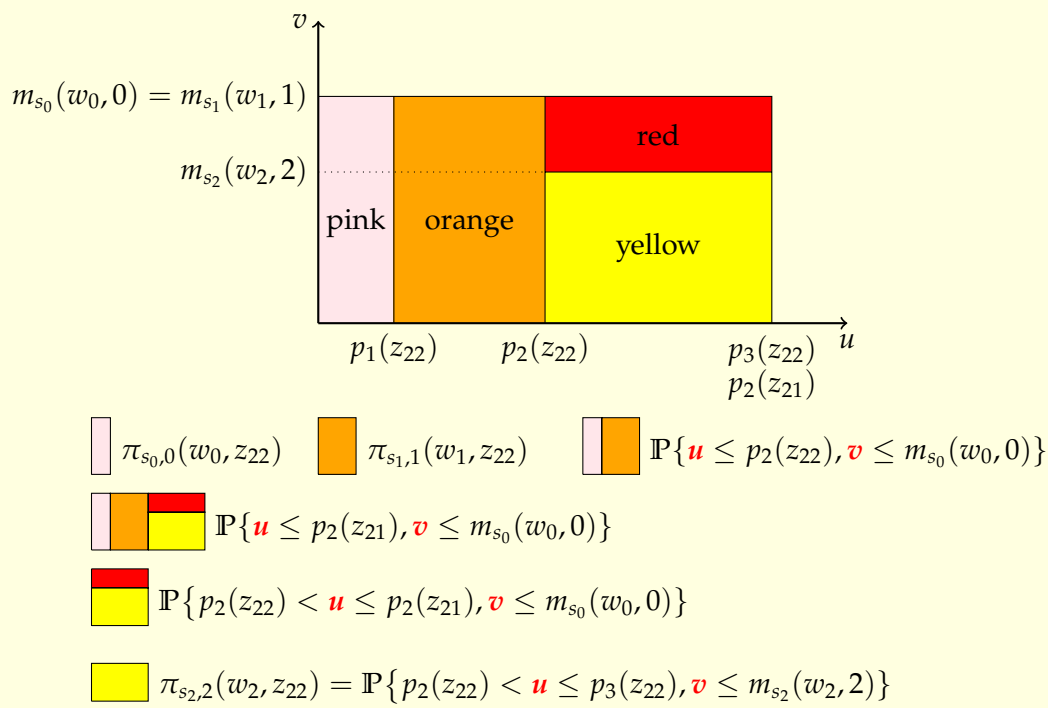


FIGURE 3. Verifying whether $m_{s_1}(w_1,1) = m_{s_2}(w_2,2)$ given that $m_{s_0}(w_0,0) = m_{s_1}(w_1,1)$.

$$\theta\{(0,p_3(z_{22})], m_{s_0}(w_0,0)\} =$$

$$\theta\{(0,p_1(z_{22})], m_{s_0}(w_0,0)\} + \theta\{(p_1(z_{22}),p_2(z_{22})], m_{s_1}(w_1,1)\} + \theta\{(p_2(z_{22}),p_3(z_{22})], m_{s_2}(w_2,2)\},$$

since $(0,p_3(z_{22})] = (0,p_1(z_{22})] \cup (p_1(z_{22}),p_2(z_{22})] \cup (p_2(z_{22}),p_3(z_{22})]$.

When the support of z and w is the Cartesian product of the marginals (as in these examples), assumption D is reduced to the requirement that p_d has sufficient variability and z sufficiently rich support, as in corollary 1

5. Identification of ψ

We now turn to the identification of the main object of interest, i.e. $\psi^* = \psi(x^*, s^*, d^*)$, for which we use the fact that the m function is identified.

Recall from (6) that for $A \subset \mathscr{S}_{uv}$,

$$\kappa(A,a) = \mathbb{E}\big[g(a,\epsilon)\mathbb{1}\{(u,v) \in A\}\big].$$

The role of κ is similar to that of the function θ in section 4. Indeed, if A is a set of positive measure then by assumption C, $\kappa(A,a) = \kappa(A,\tilde{a})$ if and only if $a = \tilde{a}$. We start with the identification of κ.

Let $\mathscr{S}_{\boldsymbol{wz}}(x)$ be the support of $(\boldsymbol{w}, \boldsymbol{z})$ conditional on $\boldsymbol{x} = x$. We define $\mathscr{M}$ to be the collection of (d, s, w) triples for which $m_s(w, d)$ and $m_{s+1}(w, d)$ are both identified. Formally, we let

$$\mathscr{M}^*(s) = \begin{cases} \{(d, w) : \mathscr{U} \in \mathscr{D}^*(d, 1, w)\}, & s = 0, \\ \{(d, w) : \mathscr{U} \in \mathscr{D}^*(d, s, w) \cap \mathscr{D}^*(d, s+1, w)\}, & 1 \leq s \leq \eta_s - 1, \\ \{(d, w) : \mathscr{U} \in \mathscr{D}^*(d, \eta_s, w)\}, & s = \eta_s, \end{cases}$$

$$\mathscr{M} = \{(d, s, w) : (d, w) \in \mathscr{M}^*(s)\},$$

and

$$\mathscr{K}(x, s, d) = \Big\{ \big(p_d(z), p_{d+1}(z)\big] \times \big(m_s(w, d), m_{s+1}(w, d)\big] : (w, z) \in \mathscr{S}_{\boldsymbol{wz}}(x) \text{ and } (d, s, w) \in \mathscr{M} \Big\}.$$

So by theorem 1 $\mathscr{K}(x, s, d)$ is a collection of nonempty rectangles whose corner points are all identified under assumptions A and B. Moreover, for $K = \big(p_d(z), p_{d+1}(z)\big] \times \big(m_s(w, d), m_{s+1}(w, d)\big]$, $\kappa\{K, \alpha(x, s, d)\}$ is identified, because

$$\kappa\{K, \alpha(x, s, d)\} = \mathbb{E}\{\boldsymbol{y}\mathbb{1}(\boldsymbol{d} = d)\mathbb{1}(\boldsymbol{s} = s) | \boldsymbol{x} = x, \boldsymbol{w} = w, \boldsymbol{z} = z\}. \tag{14}$$

We now extend $\mathscr{K}(x, s, d)$ to a larger class of sets K for which the identification of $\kappa\{K, \alpha(x, s, d)\}$ obtains.

Definition 2. $\mathscr{D}(x, s, d)$ is the collection $\mathscr{D}_\infty(x, s, d)$ in the following iterative scheme. Let $\mathscr{D}_0(x, s, d) = \mathscr{K}(x, s, d)$. Then for all $t \geq 0$, $\mathscr{D}_{t+1}(x, s, d)$ consists of all sets A^* such that at least one of the following four conditions is satisfied.

(i) $A^* \in \mathscr{D}_t(x, s, d)$;

(ii) $\exists A_1, A_2 \in \mathscr{D}_t(x, s, d) : A_1 \subset A_2,\ \mu(A_2 - A_1) > 0,\ A^* = A_2 - A_1$;

(iii) $\exists A_1, A_2 \in \mathscr{D}_t(x, s, d) : A_1 \cap A_2 = \emptyset,\ \mu(A_1 \cup A_2) > 0,\ A^* = A_1 \cup A_2$;

(iv) $\exists (\tilde{x}, \tilde{s}, \tilde{d}) : \alpha(\tilde{x}, \tilde{s}, \tilde{d}) = \alpha(x, s, d), \mathscr{D}_t(x, s, d) \cap \mathscr{D}_t(\tilde{x}, \tilde{s}, \tilde{d}) \neq \emptyset, A^* \in \mathscr{D}_t(\tilde{x}, \tilde{s}, \tilde{d})$. □

The collection $\mathscr{D}(x, s, d)$ (like $\mathscr{D}^*(d, s, w)$) consists of sets defined in terms of the unknown p_d, m_s, α functions, such that $\mathscr{D}(x, s, d)$ can be interpreted as a set of unknown parameters.

Lemma 2. *Suppose that assumptions A to C are satisfied.*

(i) For all $(x^*, s^*, d^*) \in \mathscr{S}_{\boldsymbol{xsd}}$, every $K \in \mathscr{D}(x^*, s^*, d^*)$ is identified;

(ii) $\kappa\{K, \alpha(x^*, s^*, d^*)\}$ is identified whenever $(x, s, d) \in \mathscr{S}_{\boldsymbol{xsd}}$ and $K \in \mathscr{D}(x^*, s^*, d^*)$.

Assumption E. $\mathscr{U}^2 \in \mathscr{D}(x^*, d^*, s^*)$.

Like for assumption D, assumption E equivalently requires that there be a finite T such that $\mathscr{U}^2 \in \mathscr{D}_T(x^*, d^*, s^*)$.

Theorem 2. *Suppose that assumptions A to C and E are satisfied. Then ψ^* is identified.*

Our method for identifying ψ is similar to our method for identifying m described in section 4: $\mathscr{D}(x, s, d)$ is now generated from a collection of rectangles, not a collection of intervals. Further, if we can ascertain that $\alpha(x^*, s^*, d^*) = \alpha(\bar{x}, \bar{s}, \bar{d})$ then $\mathscr{D}(x^*, s^*, d^*) \cap \mathscr{D}(\bar{x}, \bar{s}, \bar{d}) \neq \emptyset$ implies that the two collections in fact coincide. This is particularly helpful when $s^* \neq s$ and $d^* \neq \bar{d}$.

We now consider a simple example that illustrates the basics of the machinery developed above. The example is limited relative to the theoretical results in several respects, which we discuss after the example.

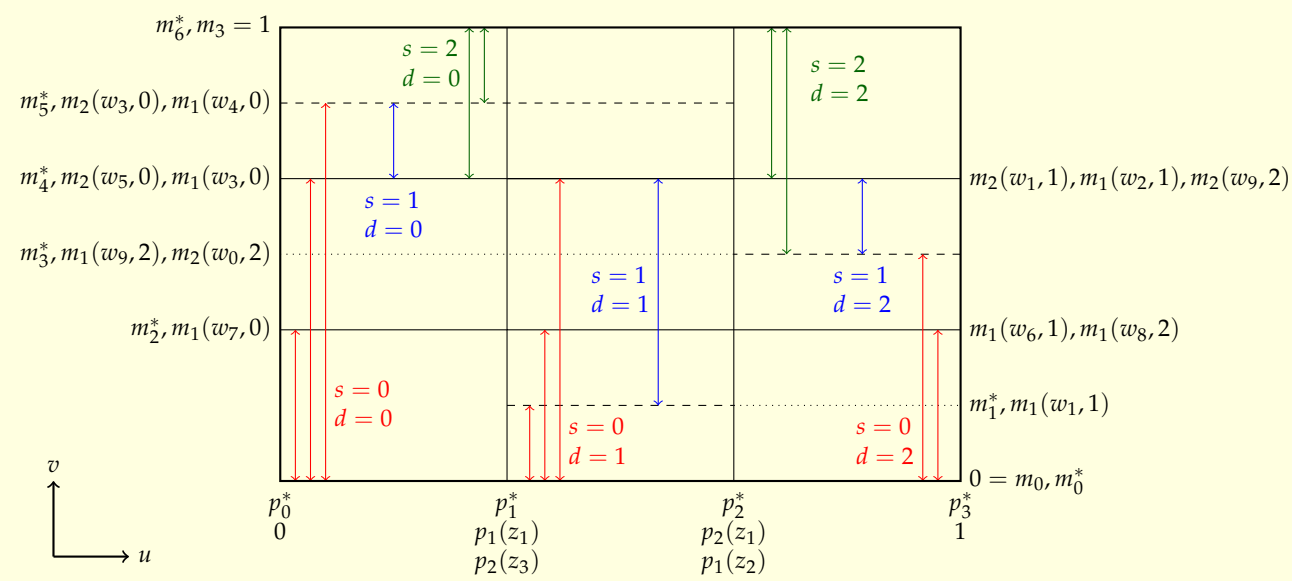


FIGURE 4. Identification of ψ if $\eta_s = \eta_d = 2$

Example 5. We will focus on the simplest interesting case, i.e. $\eta_s = \eta_d = 2$ with covariate support $\mathscr{S}_{xwz} = \mathscr{S}_x \times \mathscr{S}_w \times \mathscr{S}_z$. Because of the absence of support restrictions we will use $\mathscr{K}(s, d)$ instead of $\mathscr{K}(x, s, d)$ in this example. Identification of p_d is trivial and identification of m_s was discussed in section 4, so the discussion below starts from the point at which identification of p_d and m_s has already been established.

The example is illustrated in figure 4, which depicts a situation in which ψ^* is identified for all values of x^*, s^*, d^* provided that $\alpha(x, s^*, d^*)$ varies sufficiently as a function of x. In the discussion below we assume that there exists a $\{x_{sd}\}$ such that $\alpha(x_{sd}, s, d)$ is the same for all values of s and

d, such that the existence of the w, z combinations in figure 4 is sufficient. We show that for such $\{x_{sd}\}$, $\mathscr{D}(x_{sd}, s, d)$ is the same for all values of s, d, which implies that $\mathscr{U}^2$ is an element of $\mathscr{D}(x_{sd}, s, d)$ for all s, d, which implies identification. From hereon we use the shorthand notation $\mathscr{D}(s, d)$ to mean $\mathscr{D}(x_{sd}, s, d)$.

We start by showing that $\mathscr{D}(1,1) = \mathscr{D}(0,1)$ if $\alpha(x_{11}, 1, 1) = \alpha(x_{01}, 0, 1)$. Let

$$K_{hrij} = \left(p_h^*, p_r^*\right] \times \left(m_i^*, m_j^*\right], \quad h = 0,1;\ r = h+1, \ldots, 2;\ i = 0, \ldots, 5;\ j = i+1, \ldots, 6.$$

Since $p_1(z_1) = p_1^*$, $p_2(z_1) = p_2^*$, $m_1(w_1, 1) = m_1^*$, and $m_2(w_1, 1) = m_4^*$ it follows that $K_{1214} \in \mathscr{D}(1,1)$. Likewise, using $m_1(w_1, 1) = m_1^*$ and $m_1(w_2, 1) = m_4^*$ it follows that $K_{1201}, K_{1204} \in \mathscr{D}(0,1)$, which implies that $K_{1214} = K_{1204} \cap K_{1201} \in \mathscr{D}(0,1)$, also. So $K_{1214} \in \mathscr{D}(1,1) \cap \mathscr{D}(0,1)$ such that by the assumption on α made earlier in the example and condition (iv) of definition 2, $\mathscr{D}(1,1) = \mathscr{D}(0,1)$.

We next show that $\mathscr{D}(0,0) = \mathscr{D}(1,0) = \mathscr{D}(2,0)$. Now, $K_{0145} \in \mathscr{D}(1,0)$ because $m_1(w_3, 0) = m_4^* < m_5^* = m_2(w_3, 0)$. Further, $m_1(w_3, 0) = m_4^* < m_5^* = m_1(w_4, 0)$ implies that $K_{0104}, K_{0105} \in \mathscr{D}(0,0)$ and hence that $K_{0145} = K_{0105} \cap K_{0104} \in \mathscr{D}(0,0)$. Likewise, $m_2(w_5, 0) = m_4^* < m_5^* = m_2(w_3, 0)$, such that $K_{0145} = K_{0146} - K_{0156} \in \mathscr{D}(2,0)$. Consequently, $K_{0145} \in \mathscr{D}(0,0) \cap \mathscr{D}(1,0) \cap \mathscr{D}(2,0)$ which (together with the assumption on α used in this example) implies that $\mathscr{D}(0,0) = \mathscr{D}(1,0) = \mathscr{D}(2,0)$.

Given that $m_1(w_6, 1) = m_2^*$ it follows that $K_{1202} \in \mathscr{D}(0,1)$. Likewise, using w_7, $K_{0102}, K_{0202} \in \mathscr{D}(0,0)$ and hence $K_{1202} = K_{0102} \cap K_{0202} \in \mathscr{D}(0,0)$, also. Repeating the same argument for w_8 results in $\mathscr{D}(0,0) \cap \mathscr{D}(0,1) \cap \mathscr{D}(0,2) \neq \emptyset$ and hence $\mathscr{D}(0,0) = \mathscr{D}(0,1) = \mathscr{D}(0,2) = \mathscr{D}(1,0) = \mathscr{D}(1,1) = \mathscr{D}(2,0)$.

Finally, using w_9, w_0 it follows that $K_{2334} \in \mathscr{D}(1,2) \cap \mathscr{D}(2,2)$ and using w_8, w_9 it can be deduced that $K_{1224} \in \mathscr{D}(1,1) \cap \mathscr{D}(1,2)$, such that $\mathscr{D}(s,d)$ is identical for all s, d.

To see that $\mathscr{U}^2 \in \mathscr{D}(1,1)$ note that each of the nine rectangles with solid boundaries in figure 4 belongs trivially to some $\mathscr{D}(s,d)$ (e.g. $K_{1224} \in \mathscr{D}(1,1)$). Since the union of the nine rectangles is exactly $\mathscr{U}^2$ and $\mathscr{D}(s,d)$ is the same for all s, d, identification is hereby established. □

In the above example it was shown that $\mathscr{D}(s,d)$ was the same for all values of s, d. This is not necessary for the identification of ψ^*. Indeed, all that is required is that $\mathscr{U}^2 \in \mathscr{D}(x^*, s^*, d^*)$; it does not matter which combinations of (s,d) pairs are matched with each other, as long as the Dynkin system generated by the union of their $\mathscr{K}$–sets includes $\mathscr{U}^2$ as an element.

Example 5 is limited in several respects. First, the support of covariates was assumed to be the Cartesian product of the marginal supports and to be independent of s, d. With support

restrictions, the procedure to establish identification of ψ^* would be similar, but more care should be taken in the selection of w, z pairs to ensure that the support restrictions are satisfied. For instance, figure 4 of example 5 indicates that (w_j, z_1) belongs to $\mathscr{S}_{\boldsymbol{wz}}$ for a number of different values of j, but this condition can be relaxed in numerous ways.

Further, it was assumed that $\eta_s = \eta_d = 2$. With more than two categories the essence of the identification procedure does not change, but figure 4 would be messier. An essential ingredient of example 5 is that there are values of z_1, z_3 for which $p_1(z_1) = p_2(z_3)$ and likewise for m_s. This is analogous to corollary 1. It should be pointed out that with more than three categories ($\eta_d > 2$ or $\eta_s > 2$), it is not necessary for there to be a z_4–value for which $p_1(z_1) = p_3(z_4)$. Indeed, what is needed is for there to be a pair z_4, z_5 such that $p_2(z_4) = p_3(z_5)$. As mentioned earlier, such a chaining argument can be extended to any number of categories, i.e. one could obtain a set of sufficient conditions similar to those in corollary 1.

6. Inference

In a sample of finite size it may not be possible to find all exact matches even if these do exist in the 'population.' Further, the various estimable objects are not observed but should be estimated, which introduces estimation error.

There are two solutions to these problems, which we discuss making an implicit assumptions that at least one of the elements in each of $\boldsymbol{w}$, $\boldsymbol{z}$ has a continuous distribution.

One solution is to 'bin' covariate values into a finite number of groups and obtain a set valued estimate based on bounds; so even though there is point identification, the estimator is set valued.[6] The number of groups can increase with the sample size reducing the size of the set to a single point in the limit. The second possibility — which is more in the spirit of the identification method advocated in this paper — is to use smoothing. Full development of such procedures is beyond the scope of this paper.

Appendix A. Proofs

Proof of Lemma 1. We show both parts simultaneously and use mathematical induction. For all (d, s, w) any $V_0 \in \mathscr{D}_0^*(d, s, w)$ can be expressed as $V_0 = \big(p_d(z), p_{d+1}(z)\big]$ for some $z \in \mathscr{S}_{\boldsymbol{z}}(w)$, is hence identified, and satisfies $\theta\{V_0, m_s(w, d)\} = \mathbb{P}(\boldsymbol{s} < s, \boldsymbol{d} = d | \boldsymbol{w} = w, \boldsymbol{z} = z)$, which is hence also identified.

[6]The maximum score estimator has the same characteristic.

Now suppose that for arbitrary t and all (d, s, w), identification of $V_t, \theta\{V_t, m_s(w, d)\}$ has been established for all $V_t \in \mathscr{D}_t^*(d, s, w)$. We now establish identification of $\{V_{t+1}, \theta\{V_{t+1}, m_s(w, d)\}\}$ for any set $V_{t+1} \in \mathscr{D}_{t+1}^*(d, s, w)$ and any (d, s, w).

Since $V_{t+1} \in \mathscr{D}_{t+1}^*(d, s, w)$ it must be the set A^* in one of the four conditions in definition 1. We verify identification in each of the four cases. First (i). If $V_{t+1} \in \mathscr{D}_t^*(d, s, w)$ then identification of both objects is trivial. Now (ii). Since both V_{t+1} and $\theta\{V_{t+1}, m_s(w, d)\}$ are differences between two identified objects, they are identified, also. The argument is analogous for (iii).

Finally, (iv). We know that $V_{t+1} \in \mathscr{D}_t^*(\bar{d}, \bar{s}, \bar{w})$ where $\bar{d}, \bar{s}, \bar{w}$ are such that there exists a set $V^* \in \mathscr{D}_t^*(d, s, w) \cap \mathscr{D}_t^*(\bar{d}, \bar{s}, \bar{w})$. Since all sets in $\mathscr{D}_t^*(d, s, w)$ and $\mathscr{D}_t^*(\bar{d}, \bar{s}, \bar{w})$ are identified, the existence and identity of such a set V^* can be established. Further, $\theta\{V^*, m_s(w, d)\}$ and $\theta\{V^*, m_{\bar{s}}(\bar{w}, \bar{d})\}$ are both identified and equal if and only if $m_s(w, d) = m_{\bar{s}}(\bar{w}, \bar{d})$. Given that V_{t+1} belongs to $\mathscr{D}_t^*(\bar{d}, \bar{s}, \bar{w})$, it is identified and so is $\theta\{V_{t+1}, m_s(w, d)\}$ because it is known to equal $\theta\{V_{t+1}, m_{\bar{s}}(\bar{w}, \bar{d})\}$, which is identified. □

Proof of Theorem 1. It follows from the fact that $\theta\{\mathscr{U}, m_{s^*}(d^*, w^*)\} = m_{s^*}(d^*, w^*)$. □

Proof of Corollary 1. We use mathematical induction. Suppose that for some $1 \leq i \leq \eta_d$ it has been established that $\forall j < i : \mathscr{D}^*(0, s_0, w_0) = \mathscr{D}^*(j, s_j, w_j)$. By (12) there exist a z_1, z_2 for which $p_{i-1}(z_1) = p_i(z_2)$. Now,

$$\big(p_{i-1}(z_2), p_i(z_2)\big] \begin{cases} \in \mathscr{D}^*(i, s_i, w_i), \\ = \big(0, p_{i-1}(z_1)\big] - \big(0, p_{i-1}(z_2)\big] \in \mathscr{D}^*(0, s_0, w_0), \end{cases}$$

such that $\mathscr{D}^*(0, s_0, w_0) = \mathscr{D}^*(i, s_i, w_i)$. □

Proof of Lemma 2. The proof is very similar to, but somewhat more complicated than, that of lemma 1. We establish both parts simultaneously and again use mathematical induction. For all (x, s, d) any $K_0 \in \mathscr{D}_0(x, s, d)$ can be expressed as $K_0 = \big(p_d(z), p_{d+1}(z)\big] \times \big(m_s(w, d), m_{s+1}(w, d)\big]$ for some $(w, z) \in \mathscr{S}_{\boldsymbol{wz}}(x, s, d)$ for which $(d, s, w) \in \mathscr{M}$. K_0 is hence identified and satisfies

$$\kappa\big\{K_0, \alpha(x, s, d)\big\} = \mathbb{E}\big\{\boldsymbol{y}\mathbb{1}(\boldsymbol{d} = d)\mathbb{1}(\boldsymbol{s} = s) | \boldsymbol{x} = x, \boldsymbol{w} = w, \boldsymbol{z} = z\big\},$$

which is hence also identified.

Now suppose that for arbitrary t and all (x, s, d) identification of $K_t, \kappa\{K_t, \alpha(x, s, d)\}$ has been established for all $K_t \in \mathscr{D}_t(x, s, d)$. We now establish identification of $\{K_{t+1}, \kappa\{K_{t+1}, \alpha(x, s, d)\}\}$ for any set $K_{t+1} \in \mathscr{D}_{t+1}(x, s, d)$ and any (x, s, d).

Since $K_{t+1} \in \mathscr{D}_{t+1}(x, s, d)$ it must be the set A^* in one of the four conditions in definition 2. We verify identification in each of the four cases. First (i). If $K_{t+1} \in \mathscr{D}_t(x, s, d)$ then identification of both objects is trivial. Now (ii). Since both K_{t+1} and $\kappa\{K_{t+1}, \alpha(x, s, d)\}$ are differences between two identified objects, they are identified, also. The argument is analogous for (iii).

Finally (iv). We know that $K_{t+1} \in \mathscr{D}_t(\bar{x}, \bar{s}, \bar{d})$, where $\bar{x}, \bar{s}, \bar{d}$ are such that there exists a set $K^* \in \mathscr{D}_t(x, s, d) \cap \mathscr{D}_t(\bar{x}, \bar{s}, \bar{d})$. Since all sets in $\mathscr{D}_t(x, s, d)$ and $\mathscr{D}_t(\bar{x}, \bar{s}, \bar{d})$ are identified, the existence and identity of such a set K^* can be established. Further, $\kappa\{K^*, \alpha(x, s, d)\}$ and $\kappa\{K^*, \alpha(\bar{x}, \bar{s}, \bar{d})\}$ are both identified and equal if and only if $\alpha(x, s, d) = \alpha(\bar{x}, \bar{s}, \bar{d})$ by assumption C. Given that K_{t+1} belongs to $\mathscr{D}_t(\bar{x}, \bar{s}, \bar{d})$, it is identified and so is $\kappa\{K_{t+1}, \alpha(x, s, d)\}$ because it is equal to $\kappa\{K_{t+1}, \alpha(\bar{x}, \bar{s}, \bar{d})\}$, which is identified. □

Proof of Theorem 2. When $\mathscr{S}_{uv} \subset K$, we have $\psi(x^*, s^*, d^*) = \kappa\{K, \alpha(x^*, s^*, d^*)\}$. Apply the previous theorem. □

References

Black, D., and J. Smith (2004): "How robust is the evidence on the effects of college quality? Evidence from matching," *Journal of Econometrics*, 121(1), 99–124.

Chernozhukov, V., and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245–261.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71(5), 1405–1441.

Chiburis, R. (2010): "Semiparametric bounds on treatment effects," *Journal of Econometrics*, 159(2), 267–275.

Dearden, L., J. Ferri, and C. Meghir (2002): "The effect of school quality on educational attainment and wages," *Review of Economics and Statistics*, 84(1), 1–20.

d'Haultfœuille, X., and P. Février (2011): "Identication of nonseparable models with endogeneity and discrete instruments," Discussion paper, CREST.

Flores, C., and A. Flores-Lagunes (2009): "Identification and estimation of causal mechanisms and net effects of a treatment under unconfoundedness," Discussion paper, IZA Discussion Paper.

Heckman, J. (2005): "The scientific model of causality," *Sociological methodology*, 35(1), 1–97.

Heckman, J., and E. Vytlacil (2007): "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation," *Handbook of econometrics*, 6, 4779–4874.

Imai, K., and D. Van Dyk (2004): "Causal inference with general treatment regimes," *Journal of the American Statistical Association*, 99(467), 854–866.

Imbens, G., and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77(5), 1481–1512.

Jun, S., J. Pinkse, and H. Xu (2011): "Tighter bounds in triangular systems," *Journal of Econometrics*, 161(2), 122–128.

——— (2012): "Discrete endogenous variables in weakly separable models," *Econometrics Journal*, 15(2), 288–312.

Khan, S., and E. Tamer (2010): "Irregular identification, support conditions, and inverse weight estimation," *Econometrica*, 78(6), 2021–2042.

Lechner, M. (2001): "Identification and estimation of causal effects of multiple treatments under the conditional independence assumption," *Econometric evaluation of labour market policies*, pp. 43–58.

Lewbel, A. (2007): "Endogenous selection or treatment model estimation," *Journal of Econometrics*, 141(2), 777–806.

Mourifié, I. (2012): "Sharp bounds on treatment effects," Discussion paper, Université de Montréal.

Pinkse, J. (2001): "Nonparametric regression estimation using weak separability," Discussion paper, UBC.

Shaikh, A., and E. Vytlacil (2011): "Partial identification in triangular systems of equations with binary dependent variables," *Econometrica*, 79(3), 949–955.

Vytlacil, E., and N. Yildiz (2007): "Dummy endogenous variables in weakly separable models," *Econometrica*, 75(3), 757–779.

Nomenclature

$\mathbb{1}$	indicator function, page 3
α	unknown function, page 3
$\boldsymbol{d}$	endogenous discrete regressor in second stage, page 3
$\mathscr{D}$	Dynkin system, page 13
$\mathscr{D}^*$	Dynkin system, page 8
δ	observable conditional expectation, page 5
$\boldsymbol{\epsilon}$	error, page 3
η_d	number of values $\boldsymbol{d}$ can take minus one, page 3

η_s	number of values $\boldsymbol{s}$ can take minus one, page 3
g	unknown function, page 3
$\mathscr{K}$	base collection of sets, page 13
κ	expectation, page 12
$\mathscr{M}$	intersection of Dynkin systems, page 12
m_s	unknown function, page 3
μ	Lebesgue measure, page 3
p_d	unknown function, page 3
κ	expectation over sets, page 5
π	conditional probability, page 10
ψ	treatment effect, page 4
$\boldsymbol{s}$	endogenous discrete regressor in third stage, page 3
$\mathscr{S}_{dsw}$	support of $\boldsymbol{d}, \boldsymbol{s}, \boldsymbol{w}$, page 9
$\mathscr{S}_{wz}$	support of $\boldsymbol{w}, \boldsymbol{z}$, page 12
$\mathscr{S}_z$	conditional support of $\boldsymbol{z}$, page 8
$\mathscr{S}_{uv}$	support of $(\boldsymbol{u}, \boldsymbol{v})$, page 3
$\mathcal{T}_d$	treatment effect, page 4
θ	probability in $(\boldsymbol{u}, \boldsymbol{v})$ space, page 7
$\boldsymbol{u}$	error, page 3
$\mathscr{U}$	$(0, 1]$, page 3
$\boldsymbol{v}$	error, page 3
$\mathscr{V}$	collection of sets in $\mathscr{U}$, page 8
$\boldsymbol{w}$	covariate, page 3
$\boldsymbol{x}$	covariate, page 3
$\boldsymbol{y}$	dependent variable, page 3
$\boldsymbol{y}_{sd}$	counterfactual value of $\boldsymbol{y}$, page 4
$\boldsymbol{z}$	covariate, page 3