Identification of Treatment Effects  
in a Triangular System of Equations

Sung Jae Jun, Ž Joris Pinkse, Haiqing Xu, and Neşe Yildiz

this version: May 2014

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.
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 restrictive support conditions 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. Kane and Rouse (1995) investigates labor market returns to community college attendance and four–year college education. In the context of marketing Lambrecht, Seim, and Tucker (2011) considers the multi–stage nature of the adoption process of online banking services, where interruptions in the initial sign–up stage and in the later stage of regular use are the treatments of interest. We further note that the double hurdle model of Cragg (1971) is a special case of our model, albeit that the identification methods developed here are of limited use in Cragg’s specification.

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. 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 1d’Haultfoeuille and Février (2011) also uses a recursion scheme for the purpose of identification, but both their method and their model is different from ours.
system approach, first used in Jun, Pinkse, and 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

\[
\begin{align*}
    y &= g(\alpha(x, s, d), \epsilon), \\
    s &= \sum_{j=1}^{\eta_s} \mathbb{I}\{v > m_j(w, d)\}, \\
    d &= \sum_{j=1}^{\eta_d} \mathbb{I}\{u > p_j(z)\},
\end{align*}
\]

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.

We now make several model assumptions. Let \( \mathcal{U} = (0, 1] \).

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

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

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

Both \(s\) and \(d\) 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.\(^2\)

\(^2\)We thank Elie Tamer for pointing this out.
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 the ones studied in Vytalil and Yildiz (2007, VY) and Jun, Pinkse, and Xu (2012, JPX). It is also more general than the double hurdle model of Cragg (1971), equations (5) and (6), albeit that our matching strategy for identification is of limited usefulness there.\footnote{Indeed, let \(s, d\) be binary, let \(x = w\), and let \(u, v, \epsilon\) be independent uniform \((0, 1)\). Define \(g(\alpha, \epsilon) = \Phi^{-1}\{1 - (1 - \epsilon) \Phi(\alpha / \sigma)\}\) + \(\alpha\). Then, for parameter vectors \(\hat{\beta}, \hat{\beta}\), and scale parameter \(\sigma\), letting \(\alpha(w, s, d) = \Phi(-w^\top \hat{\beta})\) if \(sd = 0\) and \(\alpha(w, 1, 1) = \Phi(-w^\top \tilde{\beta} / \sigma)\), \(\alpha(w, s, d) = -\infty\) if \(sd = 0\) and \(\alpha(w, 1, 1) = \Phi(-w^\top \tilde{\beta} / \sigma)\), \(\alpha(w, s, d) = \Phi(-w^\top \hat{\beta})\) otherwise, reproduces the likelihoods in equations (5) and (6), of Cragg (1971). We note however that our matching strategy will explicitly require that \(x\) and \(w\) can be varied separately.}

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 \(u\) and \(v\) are allowed to be dependent and that the support of \((u, v)\) given \(\epsilon\) need not be \(U^2\).

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 expectations;\footnote{Under additive separability of the error term, both types of monotonicity are satisfied.} a similar assumption can be found in VY. 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 \(s\) and \(d\) are treatment variables. Let \(y_{sd} = g(\alpha(x, s, d), \epsilon)\). Thus, \(y_{sd} = y\) if \((s, d) = (s, d)\), but if \((s, d) \neq (s, d)\) then \(y_{sd}\) is the value \(y\) would have taken if the same individual had \(s = s, d = d\). So \(y_{sd}\) 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

\[
\psi(x^*, s^*, d^*) = \mathbb{E}(y_{s^*d^*}|x = x^*) = \mathbb{E}g(\alpha(x^*, s^*, d^*), \epsilon),
\]

where \(x^*, s^*, d^*\) are chosen by the researcher. We obtain identification of \(m_s(w, d)\) as a byproduct. So, \(\psi(x^*, s^*, d^*)\) is the average of the counterfactual outcome when the treatments are exogenously fixed at \(s^*\) and \(d^*\), conditional on \(x = x^*\). For instance, \(\psi(1, 1, 1)\) could be the counterfactual mean earnings of a male worker \((x = 1)\) if he had both a college degree \((d = 1)\) and received on the job training \((s = 1)\), or it could be the counterfactual mean birth weight for an infant if her mother had a normal gestation length \((s = 1)\) and smoked \((d = 1)\).

The function \(\psi\) can be used to obtain many, but not all, causal effects of interest. Recall the dual binary treatment example involving college education and on the job training. Consider exogenously changing \(d\) and fixing \(s\) at a specified value \(x^*\). Then one can identify the ceteris paribus effect of a change in college education status on earnings for a male worker with job training, i.e. \(\psi(1, s^*, 1) - \psi(1, s^*, 0)\). We call this an \textit{average partial treatment effect}. Alternatively, we can define \textit{average joint treatment effects} by looking...
at the causal effects on earnings for male workers of exogenously changing both college education and job training status, i.e. $\psi(1, 1, 1) - \psi(1, 0, 0)$. One can aggregate up such effects across sexes, or indeed across job training statuses, e.g. $\mathbb{E}\{\psi(1, \tilde{s}, 1) - \psi(1, \tilde{s}, 0)\}$, where $\tilde{s}$ is drawn from a suitable job training status distribution.

But there are also effects of potential interest that cannot be identified using $\psi$. For instance, if the policy maker can only influence college education decisions, but not job training decisions directly, then an object of interest would be the effect of exogenously changing $d$ on a male worker’s mean earnings leaving $s$ to adjust according to the preferences of the worker and his employer, i.e. the parameter

$$\mathbb{E}g\{\alpha(x, s_1, 1), \epsilon\} - \mathbb{E}g\{\alpha(x, s_0, 0), \epsilon\},$$

where $s_d$ is the counterfactual value of $s$ when $d$ is exogenously fixed at $d$. We call this parameter an average total treatment effect. The total treatment effect can be decomposed into a direct effect and an indirect effect; these are further discussed in appendix B.

The fact that there are several causal parameters of potential interest arises both because there are multiple endogenous treatment variables and because of the triangular nature of the model. However, we do not believe that one parameter is generally more important than others but the purpose and context of the policy question of interest should be taken into account. As explained in appendix B, identification of causal parameters like (3) can be established by the matching method developed in this paper. Therefore, we focus on the identification of $\psi$ (and $m_s$) in the main text to highlight the idea of matching while we show in appendix B that the identification of (3) can be obtained by the same methods.

3. Description

We now proceed with providing a broad and rough description of our identification strategy. Please recall from (2) that

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

The objects that we can identify directly from the data are

$$\begin{align*}
\delta(x, s, d, w, z) &= \mathbb{E}\{y I(s = s)I(d = d)|x = x, w = w, z = z\} \\
&= \mathbb{E}\{g\{\alpha(x, s, d), \epsilon\}I((u, v) \in (p_d(z), p_{d+1}(z)) \times (m_s(w, d), m_{s+1}(w, d))\}, \\
\mathbb{P}(s = s, d = d|w = w, z = z) &= \mathbb{P}\{(u, v) \in (p_d(z), p_{d+1}(z)) \times (m_s(w, d), m_{s+1}(w, d))\}.
\end{align*} \tag{5}$$

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

$$\kappa(A, a) = \mathbb{E}[g(a, \epsilon)I((u, v) \in A)], \tag{6}$$
then we would like to identify $\kappa \{\delta_{uv}, \alpha(x^*, s^*, d^*)\}$ given that we can identify $\kappa \{A, \alpha(x, s, d)\}$ for rectangles $A = (p_d(z), p_{d+1}(z)] \times (m_z(w, d), m_{z+1}(w, d))$ 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 ($\cdot, ]$) intervals.

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

$$\lim_{n \to \infty} \mathbb{E}\{y 1(s = s^*) 1(d = d^*)|x = x^*, w = w_n, z = z_n\} = \mathbb{E}\left[\sup_{u} \{\alpha(x^*, s^*, d^*)\}, \epsilon 1((u, v) \in \mathcal{U}_2)\right] = \psi(x^*, s^*, d^*).$$

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 $\kappa(A, a), \kappa(A^*, a), \kappa(A, a)$ are identified, $A$ is contained in $A^*$ and $(A^* - A) \cap A = \emptyset$ then $\kappa((A - A^*) \cup A, a)$ 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 $(\sup_z p_1(z), \sup_w m_1(w, 0))$ such that $\kappa \{\delta_{uv}, \alpha(x^*, s^*, d^*)\}$ cannot be identified.

Now suppose that for some $A, x, s, d, \bar{x}, \bar{s}, \bar{d}$, $\kappa \{A, \alpha(\bar{x}, \bar{s}, \bar{d})\}$ and $\kappa \{A, \alpha(x, s, d)\}$ are both identified. Then assumption C implies that $\kappa \{A, \alpha(\bar{x}, \bar{s}, \bar{d})\} = \kappa \{A, \alpha(x, s, d)\}$ is then equivalent to $\alpha(\bar{x}, \bar{s}, \bar{d}) = \alpha(x, s, d)$. Once it is known that $\alpha(\bar{x}, \bar{s}, \bar{d}) = \alpha(x, s, d)$, identification of $\kappa \{A^*, \alpha(x, s, d)\}$ for any set $A^*$ implies identification of $\kappa \{A, \alpha(\bar{x}, \bar{s}, \bar{d})\}$. Again, these operations can be repeated for different combinations of $(\bar{x}, \bar{s}, \bar{d}), (x, s, d)$, 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 following quantities are identified directly from the data.

\[
\begin{align*}
\delta(x^*, 0, 0, w_1, z_2) &= \kappa\{\text{green}, \alpha(x^*, 0, 0)\}, \\
\delta(x^*, 0, 0, w_1, z_1) &= \kappa\{\text{green+yellow}, \alpha(x^*, 0, 0)\}, \\
\delta(x_1, 0, 1, w_2, z_1) &= \kappa\{\text{blue}, \alpha(x_1, 0, 1)\}, \\
\delta(x_1, 0, 1, w_2, z_2) &= \kappa\{\text{blue+yellow}, \alpha(x_1, 0, 1)\}, \\
\delta(x_1, 0, 1, w_3, z_1) &= \kappa\{\text{blue+purple}, \alpha(x_1, 0, 1)\}, \\
\delta(x_2, 1, 1, w_3, z_1) &= \kappa\{\text{red}, \alpha(x_2, 1, 1)\}, \\
\delta(x_2, 1, 1, w_2, z_1) &= \kappa\{\text{red+purple}, \alpha(x_2, 1, 1)\}.
\end{align*}
\] (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 $\alpha(x^*, 0, 0) = \alpha(x_1, 0, 1) = \alpha(x_2, 1, 1) = \alpha(x_3, 1, 0)$, $\kappa\{\delta_{uv}, \alpha(x^*, 0, 0)\}$ can be computed as (for instance) the sum of the quantities $\delta(x^*, 0, 0, w_1, z_1), \delta(x_1, 0, 1, w_2, z_1), \delta(x_2, 1, 1, w_3, z_1)$, 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 $p_d(z) = P(d < d|z = z)$ 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

\[ \theta(V, m) = P(\mathbf{u} \in V, \mathbf{v} < m), \] (8)
then
\[ \theta \{(p_d(z), p_{d+1}(z), m_s(w, d)) \}, \]  
(9)
is trivially identified for all \( s, d, w, z \) provided that \((w, z) \in \mathcal{S}_{wz} \). 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 \( \theta \{ V, m_{\tilde{z}}(\tilde{w}, \tilde{d}) \} \) and \( \theta \{ V, m_s(w, d) \} \) is equivalent to \( m_{\tilde{z}}(\tilde{w}, \tilde{d}) = m_s(w, d) \) and such equality implies that for any \( V^* \), identification of \( \theta \{ V^*, m_{\tilde{z}}(\tilde{w}, \tilde{d}) \} \) 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\(^5\)
\[ \theta(V, m) = \mathbb{P}(u \in V, v \leq m), \quad V \subset \mathcal{U}, \ m \in \mathcal{U}. \]
Further let \( \mathcal{S}_z(w) \) be the support of \( z \) conditional on \( w = w \) and define
\[ \mathcal{V}(d, w) = \{(p_d(z), p_{d+1}(z)) : z \in \mathcal{S}_z(w) \}, \quad d = 0, \ldots, \eta_d. \]  
(10)
Then \( \theta \{ V, m_s(w, d) \} \) is identified when \( V \in \mathcal{V}(d, w) \) because
\[ \theta \{(p_d(z), p_{d+1}(z)), m_s(w, d)\} = \mathbb{P}(s < s, d = d | w = w, z = z). \]  
(11)
We need \( z \) 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 \( \mathcal{V}(d, w) \).

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

(i) \( A^* \in \mathcal{D}_t^*(d, s, w); \)
(ii) \( \exists A_1, A_2 \in \mathcal{D}_t^*(d, s, w) : A_1 \subset A_2, \mu(A_2 - A_1) > 0, \ A^* = A_2 - A_1; \)
(iii) \( \exists A_1, A_2 \in \mathcal{D}_t^*(d, s, w) : A_1 \cap A_2 = \emptyset, \mu(A_1 \cup A_2) > 0, \ A^* = A_1 \cup A_2; \)
(iv) \( \exists (\tilde{d}, \tilde{s}, \tilde{w}) : m_s(d, w) = m_{\tilde{z}}(\tilde{d}, \tilde{w}), \mathcal{D}_t^*(d, s, w) \cap \mathcal{D}_{\tilde{t}}^*(\tilde{d}, \tilde{s}, \tilde{w}) \neq \emptyset, \ A^* \in \mathcal{D}_{\tilde{t}}^*(\tilde{d}, \tilde{s}, \tilde{w}). \]

The conditions in definition 1 are similar to those in Jun, Pinkse, and Xu (2012). Note that \( \{ \mathcal{D}_t^*(d, s, w) : t = 0, 1, \cdots \} \) is an increasing sequence of collections, such that \( \mathcal{D}^*(d, s, w) \) is the infinite union of \( \mathcal{D}_t^*(d, s, w) \)’s.\(^6\)

Note further that \( \mathcal{D}^*(d, s, w) \) is indexed by \( s, w \) as well as \( d \). If \( \mathcal{S}_z(w) \) is the same for all \( w \) values then

\(^{5}\)We use \( \subset \) as a generic symbol for subset, where some other authors might distinguish between proper and nonproper subsets.

\(^{6}\)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 \( \mathcal{U} = \bigcup_{n=1}^\infty \{(1/n, 1]\) but \( \mathcal{U} \neq \bigcup_{n=1}^\infty \{(1/n, 1]\} \). It is the latter concept that is used here.
the argument pursued in this section is simpler, but such support restrictions are undesirable because it excludes the possibility that \( w, z \) have elements in common and it also precludes the situation in which certain combinations of \( (w, z) \) values cannot occur.

All elements of \( D^* \) 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 \( D^* \) 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 \( (d, s, w) \in S_{dsw} \), every \( V \in D^*(d, s, w) \) is identified;

(ii) \( \theta(V, m_s(w, d)) \) is identified whenever \( (d, s, w) \in S_{dsw} \) and \( V \in D^*(d, s, w) \).

**Proof.** See appendix A.

**Assumption D.** \( U \in D^*(d^*, s^*, w^*) \).

Since \( \{D_t^*(d^*, s^*, w^*) : t = 0, 1, 2, \cdots \} \) is an increasing sequence of collections of sets and \( d, s \) take finitely many values, assumption D is satisfied when there exists a finite \( T \) such that \( U \in D_T^*(d^*, s^*, w^*) \).

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

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

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

\[
\forall j = 1, \ldots, n_d - 1: \inf_{z \in S_z(w_{j+1})} p_{j+1}(z) \leq \sup_{z \in S_z(w_j)} p_j(z), \tag{12}
\]

where each \( p_j \) is a continuous function and \( z \) is continuously distributed. Then, \( m_{s_j}(w^*, d^*) \) 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 \( \beta_0 \) and \(-\infty = \gamma_0 < \gamma_1 < \cdots < \gamma_{\eta_d+1} = \infty, p_d(z) = \Phi(\gamma_d + \beta^T z) \), 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

\[
\sup_{z, z^* \in S_z} \beta^T (z - z^*) \geq \max_{d = 1, \ldots, \eta_d - 1} (\gamma_{d+1} - \gamma_d). \quad \square
\]

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

\[
\pi_{sd}(w, z) = \Pr(s < s, d = d| w = w, z = z) = \Pr\{ p_d(z) < u \leq p_{d+1}(z), v \leq m_s(w, d) \} = \theta\{ (p_d(z), p_{d+1}(z), m_s(w, d)) \}, \quad (13)
\]

which is identified provided that \( z \in S_z(w) \).

**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,

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

Using (i) and (ii) of definition 1 it follows that

\[ V = \left( p_1(z_{11}), p_1(z_{12}) \right) = \left( p_1(z_{11}), p_2(z_{12}) \right) \in \mathcal{D}_1^* (0, s_0, w_0) \cap \mathcal{D}_1^* (1, s_1, w_1). \]

Thus,

\[
\begin{align*}
\theta \{ V, m_{s_0}(w_0, 0) \} &= \pi_{s_0,0}(w_0, z_{12}) - \pi_{s_0,0}(w_0, z_{11}), \\
\theta \{ V, m_{s_1}(w_1, 1) \} &= \pi_{s_1,1}(w_1, z_{11}),
\end{align*}
\]

are both identified; they are equal if and only if \( m_{s_1}(w_1, 1) = m_{s_0}(w_0, 0) \). \[\square\]

In example 3 it is implicitly assumed that \( z_{11}, z_{12} \in \delta_z(w_0) \) and that \( z_{11} \in \delta_z(w_1) \). 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 \( z_{110}, z_{12} \in \delta_z(w_0) \) and \( z_{111} \in \delta_z(w_1) \) 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) = $\pi_{s_00}(w_0, z_{21}) + \pi_{s_11}(w_1, z_{21})$ is identified, and hence so is (yellow + red), which in turn implies the identification of red.

Formally, it follows from example 3 that $D^*_t(0, s_0, w_0) = D^*_t(1, s_1, w_1)$ for all $t \geq 2$. So for sufficiently large $t$, $V = \{p_2(z_{22}), p_2(z_{21})\} \in D^*_t(0, s_0, w_0)$. But since $V = \{p_2(z_{22}), p_3(z_{22})\} \in D^*_2(0, s_2, w_2)$,

\[\text{equality of } m_{s_0}(w_0, 0) \text{ and } m_{s_2}(w_2, 2) \text{ can be verified using the set } V.\]

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

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

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

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 $\psi$

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 $A \subset D_{uv}$,

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

The role of $\kappa$ is similar to that of the function $\theta$ 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 $\kappa$.

Let $\delta_{wz}(x)$ be the support of $(w, z)$ conditional on $x = x$. We define $\mathcal{M}$ 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

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

and

\[ \mathcal{K}(x, s, d) = \left\{(p_d(z), p_{d+1}(z)) \times (m_x(w, d), m_{x+1}(w, d)) : (w, z) \in S^x_{w,z}(x) \text{ and } (d, s, w) \in \mathcal{M}\right\}. \]

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

\[ \kappa\{K, \alpha(x, s, d)\} = \mathbb{E}\{y^1(d = d) \mathbb{I}(s = s)|x = x, w = w, z = z\}. \quad (14) \]

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

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

1. \( A^* \in \mathcal{D}_t(x, s, d) \);
2. \( \exists A_1, A_2 \in \mathcal{D}_t(x, s, d) : A_1 \subset A_2, \mu(A_2 - A_1) > 0, A^* = A_2 - A_1 \);
3. \( \exists A_1, A_2 \in \mathcal{D}_t(x, s, d) : A_1 \cap A_2 = \emptyset, \mu(A_1 \cup A_2) > 0, A^* = A_1 \cup A_2 \);
4. \( \exists (\bar{x}, \bar{s}, \tilde{d}) : \alpha(\bar{x}, \bar{s}, \tilde{d}) = \alpha(x, s, d), \mathcal{D}_t(x, s, d) \cap \mathcal{D}_t(\bar{x}, \bar{s}, \tilde{d}) \neq \emptyset, A^* \in \mathcal{D}_t(\bar{x}, \bar{s}, \tilde{d}) \). \qed

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

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

(i) For all \( (x^*, s^*, d^*) \in S_{xsd}, \) every \( K \in \mathcal{D}(x^*, s^*, d^*) \) is identified.

(ii) \( \kappa\{K, \alpha(x^*, s^*, d^*)\} \) is identified whenever \( (x, s, d) \in S_{xsd} \) and \( K \in \mathcal{D}(x^*, s^*, d^*) \).

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

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

**Theorem 2.** Suppose that assumptions A to C and E are satisfied. Then \( \psi^* \) is identified.

Our method for identifying \( \psi \) is similar to our method for identifying \( m \) described in section 4: \( \mathcal{D}(x, s, d) \) is now generated from a collection of rectangles, not a collection of intervals. Further, if we can ascertain that \( \alpha(x^*, s^*, d^*) = \alpha(\bar{x}, \bar{s}, \tilde{d}) \) then \( \mathcal{D}(x^*, s^*, d^*) \cap \mathcal{D}(\bar{x}, \bar{s}, \tilde{d}) \neq \emptyset \) implies that the two collections in fact coincide. This is particularly helpful when \( s^* \neq s \) and \( d^* \neq \tilde{d} \).
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 $\psi$ if $\eta_s = \eta_d = 2$](image)

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

The example is illustrated in figure 4, which depicts a situation in which $\psi^*$ 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 $\{x_{sd}\}$, $\mathcal{D}(x_{sd}, s, d)$ is the same for all values of $s, d$, which implies that $\mathcal{U}^2$ is an element of $\mathcal{D}(x_{sd}, s, d)$ for all $s, d$, which implies identification. From hereon we use the shorthand notation $\mathcal{D}(s, d)$ to mean $\mathcal{D}(x_{sd}, s, d)$.

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

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

Since $p_1(z_1) = p_1^*, p_2(z_1) = p_2^*, m_1(w_1, 1) = m_1^*$, and $m_2(w_1, 1) = m_2^*$ it follows that $K_{1214} \in \mathcal{D}(1, 1)$. Likewise, using $m_1(w_1, 1) = m_1^*$ and $m_1(w_2, 1) = m_2^*$ it follows that $K_{1201}, K_{1204} \in \mathcal{D}(0, 1)$, which implies that $K_{1214} = K_{1204} \cap K_{1201} \in \mathcal{D}(0, 1)$, also. So $K_{1214} \in \mathcal{D}(1, 1) \cap \mathcal{D}(0, 1)$ such that by the assumption on $\alpha$ made earlier in the example and condition (iv) of definition 2, $\mathcal{D}(1, 1) = \mathcal{D}(0, 1)$. 
We next show that $\mathcal{D}(0,0) = \mathcal{D}(1,0) = \mathcal{D}(2,0)$. Now, $K_{0145} \in \mathcal{D}(1,0)$ 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 $K_{0104}$, $K_{0105} \in \mathcal{D}(0,0)$ and hence that $K_{0145} = K_{0105} \cap K_{0104} \in \mathcal{D}(0,0)$. Likewise, $m_2(w_5,0) = m_4^* < m_5^* = m_2(w_3,0)$, such that $K_{0145} = K_{0146} - K_{0156} \in \mathcal{D}(2,0)$. Consequently, $K_{0145} \in \mathcal{D}(0,0) \cap \mathcal{D}(1,0) \cap \mathcal{D}(2,0)$ which (together with the assumption on $\alpha$ used in this example) implies that $\mathcal{D}(0,0) = \mathcal{D}(1,0) = \mathcal{D}(2,0)$.

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

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

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

In the above example it was shown that $\mathcal{D}(s,d)$ was the same for all values of $s$, $d$. This is not necessary for the identification of $\psi^*$. Indeed, all that is required is that $U^2 \in \mathcal{D}(s^*,d^*,d^*)$; 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 $\mathcal{K}$–sets includes $U^2$ 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 $\psi^*$ 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 $(w_1, z_1)$ belongs to $\delta_{w,z}$ 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. Sketch of an estimation procedure

Below follows a sketch of a simple estimation procedure of $\psi^* = \psi(x^*, s^*, d^*)$. This procedure is provided to demonstrate how $\psi^*$ can be estimated, but in order to keep the sketch simple we will make several assumptions which are much stronger than those made in the identification portion of this paper. For instance, we shall assume that the joint support of $(w, z)$ is the Euclidean product of the marginal supports, that $s, d$ only take the values 0, 1, 2, and that there is sufficient variation in $z, p_d(z)$ to allow for the matches used. More complicated procedures can be devised that exploit some salient features of this paper (such as chaining) and lift such restrictions, but such procedures are beyond the scope of this paper, which primarily deals with identification. In earlier work (Jun, Pinkse, and Xu, 2012) we provide rigorous results for an estimation procedure that does not impose a joint support assumption albeit in a considerably simpler model than the one considered here.

We will moreover not be assuming the use of any particular nonparametric methodology. Most objects to be estimated can be expressed as conditional expectations (or probabilities), sometimes with estimated regressors. Some of these conditional expectations are then integrated with respect to one of the conditioning variables à la Linton and Härdle (1996). There are numerous important details in the theoretical development and empirical implementation of such methods, but these can by now be considered to be well–established and elaborate discussions thereof are available at various places in the literature. Hence we do not discuss them here. Whenever an object is estimable by standard nonparametric methodology (ENPM) we will so indicate.

6.1. Estimation of $m$. We commence our discussion with the estimation of $m_s(w, 1)$. Please note that

$$m_s(w, 1) = \sum_{d=0}^{2} \mathbb{E} \lambda_{sd}(w, z),$$

where $\lambda_{sd}(w, z) = \mathbb{P}\{s < s, d = d | J_d(s, w) = J_1(s, w), z = z\}$ with

$$J_d(s, w) = \mathbb{P}\{p_1(z) < u \leq p_2(z), v \leq m_s(w, d)\}.$$  

Once estimates of $J_0, J_1, J_2$ are available $\lambda_{s0}, \lambda_{s1}, \lambda_{s2}$ are ENPM and $m_s(w, 1)$ can then be estimated by integrating out over $z$ in the spirit of Linton and Härdle (1996).

Now, $J_1(s, w) = \int \mathbb{P}(d = 1, s < s | w = w, z = z) dF_z(z)$, which is ENPM. For the estimation of $J_0, J_2$ it is helpful to introduce $\zeta_{sdj}(w, p) = \mathbb{P}\{s < s, d = d | w = w, p_j(z) = p\}$ which is ENPM given
that \( p_d(z) = \mathbb{P}(d < d|z = z) \). Since
\[
\mathcal{J}_d(s, w) = \begin{cases} 
\mathbb{E}\xi_{s01}\{w, p_2(z)\} - \mathbb{E}\xi_{s01}\{w, p_1(z)\}, & d = 0, \\
\mathbb{E}\xi_{s22}\{w, p_1(z)\} - \mathbb{E}\xi_{s22}\{w, p_2(z)\}, & d = 2,
\end{cases}
\]
they too are ENPM.

Finally, to obtain estimates of \( m_s(w, 0) \) and \( m_s(w, 2) \) one can simply estimate
\[
m_s(w, d) = \mathbb{E}\{m_s(w, 1)\} J_1(s, w) = J_d(s, w).
\]

6.2. **Estimation of \( \psi \).** We focus here on the estimation of \( \psi^* = \psi(x^*, s^*, d^*) \) for \( s^* = d^* = 1 \); other combinations of \( (s^*, d^*) \) can be handled analogously. Let \( \rho_{sd} = y \mathbb{1}(s = s,d = d) \). Please note that
\[
\psi^* = \sum_{s,d=0}^2 \mathbb{E} \nu_{sd}(x^*, w, z) \quad \text{with} \quad \nu_{sd}(x^*, w, z) = \mathbb{E}\{\rho_{sd}|\alpha(x, s, d) = \alpha(x^*, 1, 1), w = w, z = z\}.
\]
Naturally, \( \nu_{11}(x^*, w, z) \) is ENPM. For \( s \neq 1 \) and/or \( d \neq 1 \) other methods must be developed to estimate \( \nu_{sd}(x^*, w, z) \). We will focus on the case \( s = d = 0 \) where the other cases can be handled analogously and possibly (if \( s = s^* \) or \( d = d^* \)) more easily.

Let
\[
\kappa_{jt}^*(x, w, z) = \mathbb{E}\{\rho_{00}|x = x, p_1(z) = p_j(z), m_1(w, 0) = m_t(w, 1)\},
\]
which is ENPM. Define
\[
\mathcal{W}(x, w, z) = \{\kappa_{22}^*(x, w, z) - \kappa_{21}^*(x, w, z) - \kappa_{12}^*(x, w, z) + \kappa_{11}^*(x, w, z)\} - \mathbb{E}(\rho_{11}|x = x^*, w = w, z = z),
\]
which is ENPM. Then \( \mathcal{W}^*(x) = \mathbb{E}\mathcal{W}(x, w, z) = 0 \) is equivalent to \( \alpha(x, 0, 0) = \alpha(x^*, 1, 1) \). Finally, \( \nu_{00}(x^*, w, z) = \mathbb{E}\{\rho_{00}|\mathcal{W}^*(x) = 0, w = w, z = z\} \) is ENPM.

**Appendix A. Proofs**

**Proof of Lemma 1.** We show both parts simultaneously and use mathematical induction. For all \( (d, s, w) \) any \( V_0 \in \mathcal{D}^*_0(d, s, w) \) can be expressed as \( V_0 = (p_d(z), p_{d+1}(z)) \) for some \( z \in \mathcal{S}_2(w) \), is hence identified, and satisfies \( \theta\{V_0, m_s(w, d)\} = \mathbb{P}(s < s, d = d|w = w, z = z) \), which is hence also identified.

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 \( V_t \in \mathcal{D}^*_t(d, s, w) \). We now establish identification of \( \{V_{t+1}, \theta\{V_{t+1}, m_s(w, d)\}\} \) for any set \( V_{t+1} \in \mathcal{D}^*_{t+1}(d, s, w) \) and any \( (d, s, w) \).

Since \( V_{t+1} \in \mathcal{D}^*_{t+1}(d, s, w) \) 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 \( V_{t+1} \in \mathcal{D}^*_t(d, s, w) \) 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 $V_{t+1} \in D^*_t(\tilde{d}, \tilde{s}, \tilde{w})$ where $\tilde{d}, \tilde{s}, \tilde{w}$ are such that there exists a set $V^* \in D^*_t(d, s, w)$ and $D^*_t(\tilde{d}, \tilde{s}, \tilde{w})$ are identified, the existence and identity of such a set $V^*$ can be established. Further, $\theta\{V^*, m_s(w, d)\}$ and $\theta\{V^*, m\}$ are both identified and equal if and only if $m_s(w, d) = m\tilde{z}(\tilde{w}, \tilde{d})$. Given that $V_{t+1}$ belongs to $D^*_t(\tilde{d}, \tilde{s}, \tilde{w})$, it is identified and so is $\theta\{V_{t+1}, m_s(w, d)\}$ because it is known to equal $\theta\{V_{t+1}, m\tilde{z}(\tilde{w}, \tilde{d})\}$, which is identified.

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

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

$$
(p_i(z_1), p_i(z_2)) = (0, p_{i-1}(z_1)) - (0, p_{i-1}(z_2)) \in D^*(0, s_0, w_0),
$$

such that $D^*(0, s_0, w_0) = D^*(i, s_i, w_i)$.

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 $K_0 \in D_0(x, s, d)$ can be expressed as $K_0 = (p_d(x), p_{d+1}(x)) \times (m_s(w, d), m_{s+1}(w, d))$ for some $(w, z) \in S_{wz}(x, s, d)$ for which $(d, s, w) \in \mathcal{M}$. $K_0$ is hence identified and satisfies

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

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 $K_t \in D_t(x, s, d)$. We now establish identification of $\{K_{t+1}, \kappa\{K_{t+1}, \alpha(x, s, d)\}\}$ for any set $K_{t+1} \in D_{t+1}(x, s, d)$ and any $(x, s, d)$.

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

\[ \square \]

**Proof of Theorem 2.** When \( S_{uv} \subseteq K \), we have \( \psi(x^*, s^*, d^*) = \kappa \{ K, \alpha(x^*, s^*, d^*) \} \). Apply the previous theorem. \[ \square \]

**Appendix B. Extension**

As mentioned in section 2 it is possible to use the methodology developed in this paper to identify objects that are not based on \( \psi \). We now demonstrate how the average total treatment (3) can be identified. Consider the special case with binary \( s, d \). We discuss the identification of \( \bar{y}(x; d) \). Consider \( \bar{y}(x; 0; 0) \).

(18)

\[ \bar{y}(x; 1; 0) = \bar{y}(1, 1) - \bar{y}(1, 0). \]

We note that \( \bar{y} \) and \( \psi \) are different objects unless \( \psi \) and \( \epsilon \) are known to be independent.\(^7\) We now sketch an identification argument for \( \bar{y} \). Consider \( \mathbb{E} \left[ \mathbb{I}(d = 1) \{ \alpha(x, s, 0), \epsilon \} | x = x, w = w, z = z \right] \), which can be written as

\[ \mathbb{E} \left[ \mathbb{I}(u > p_1(z)) \mathbb{I}(v > m_1(w, 0)) \{ \alpha(x, 0, 0), \epsilon \} \right] + \mathbb{E} \left[ \mathbb{I}(u > p_1(z)) \mathbb{I}(v > m_1(w, 0)) \{ \alpha(x, 1, 0), \epsilon \} \right]. \]

The method developed in the paper explains how to find \( (x, w) \) and \( (\tilde{w}, \tilde{x}) \) such that \( \alpha(x, 1, 0) = \alpha(\tilde{x}, 1, 1) \) and \( m_1(w, 0) = m_1(\tilde{w}, 1) \). Identification of the second term in (18) then follows from the fact that it is equal to

\[ \mathbb{E} \left[ \mathbb{I}(u > p_1(z)) \mathbb{I}(v > m_1(\tilde{w}, 1)) \{ \alpha(\tilde{x}, 1, 1), \epsilon \} \right] = \mathbb{E} \left[ \mathbb{I}(s = 1) \mathbb{I}(d = 1) | z = z, w = w, x = \tilde{x} \right]. \]

The first term in (18) can be dealt with similarly.

Note that the causal effect in (17) can be decomposed into direct and indirect effects, \( \mathbb{E} \{ \alpha(x, s_1, 1), \epsilon \} - \mathbb{E} \{ \alpha(x, s_1, 0), \epsilon \} \) and \( \mathbb{E} \{ \alpha(x, s_0, 0), \epsilon \} - \mathbb{E} \{ \alpha(x, s_0, 0), \epsilon \} \). These objects can be identified using the above arguments dealing with (18).

\(^7\) If \( \epsilon \) and \( \psi \) are independent, then \( \bar{y}(1, d) = \mathbb{E}\psi(1, s, d) \).
References


**Nomenclature**

\( \mathbb{1} \) indicator function, page 3

\( \alpha \) unknown function, page 3

\( d \) endogenous discrete regressor in second stage, page 3

\( \mathcal{D} \) Dynkin system, page 13

\( \mathcal{D}^* \) Dynkin system, page 8

\( \delta \) observable conditional expectation, page 5

\( \epsilon \) error, page 3

\( \eta_d \) number of values \( d \) can take minus one, page 3

\( \eta_s \) number of values \( s \) can take minus one, page 3

\( g \) unknown function, page 3

\( \mathcal{K} \) base collection of sets, page 13

\( \kappa \) expectation, page 12

\( \mathcal{M} \) intersection of Dynkin systems, page 12

\( m_s \) unknown function, page 3

\( \mu \) Lebesgue measure, page 3

\( p_d \) unknown function, page 3

\( \kappa \) expectation over sets, page 5

\( \pi \) conditional probability, page 10

\( \psi \) treatment effect, page 4

\( \psi^* \) alternative treatment effect, page 19

\( s \) endogenous discrete regressor in third stage, page 3
\$s_{dsd}\$ support of \(d, s, w\), page 9

\$s_{wz}\$ support of \(w, z\), page 12

\$s_{z}\$ conditional support of \(z\), page 8

\$s_{uv}\$ support of \((u, v)\), page 3

\(\theta\) probability in \((u, v)\) space, page 7

\(u\) error, page 3

\(\mathcal{U}\) \((0, 1]\), page 3

\(v\) error, page 3

\(\mathcal{V}\) collection of sets in \(\mathcal{U}\), page 8

\(w\) covariate, page 3

\(x\) covariate, page 3

\(y\) dependent variable, page 3

\(y_{sd}\) counterfactual value of \(y\), page 4

\(z\) covariate, page 3