Dummy Endogenous Variables in Weakly Separable Multiple Index Models without Monotonicity

Songnian Chen — HKUST

Shakeeb Khan — Boston College

Xun Tang — Rice University

April 1, 2020

Abstract

We study the identi cation and estimation of treatment e ect parameters in weakly separable models. In their seminal work, Vytlacil and Yildiz (2007) showed how to identify and estimate the average treatment e ect of a dummy endogenous variable when the outcome is weakly separable in a single index. Their identi cation result builds on a monotonicity condition with respect to this single index. In comparison, we consider similar weakly separable models with multiple indices, and relax the monotonicity condition for identi cation. Unlike Vytlacil and Yildiz (2007), we exploit the full information in the distribution of the outcome variable, instead of just its mean. Indeed, when the outcome distribution function is more informative than the mean, our method is applicable to more general settings than theirs; in particular we do not rely on their monotonicity assumption and at the same time we also allow for multiple indices. To illustrate the advantage of our approach, we provide examples of models where our approach can identify parameters of interest whereas existing methods would fail. These examples include models with multiple unobserved disturbance terms such as the Roy model and multinomial choice models with dummy endogenous variables, as well as potential outcome models with f9iA1ell as vm

where $P(z) \equiv E(D|Z = z)$. The only term that is not directly identifiable on the right-hand side of (2.1) is

$$E(Y_1|D = 0, X = x, Z = z) = E[g(v(x, 1), \varepsilon)|U > P(z)].$$

The main idea behind our approach follows that of Vytlacil and Yildiz (2007), which is to find some $\tilde{x} \in S_0$ such that

$$v(x, 1) = v(\tilde{x}, 0) \tag{2.2}$$

so that

$$E(Y|D = 0, X = \tilde{x}, Z = z) = \quad E(Y_0$$

the assumption that $v(X,D)\in\mathbb{R}$ is a single index and that for any $(x,\bar{x})\in(S_1\cap S_0)$, $E[g(v(x,1),\varepsilon)|U=u]=E[g(v(\bar{x},0),\varepsilon)|U=u]$ if and only if $v(x,1)=v(\bar{x},0)$. There are two shortcomings with this approach. First, it requires the condition (Assumption 4) that $E[g(v(x,d),\varepsilon)|U=p]$ is a strictly monotonic function of $v(x,d)$. Second, when $v(x,d)$ is a vector of multiple indices instead of a single index, their approach breaks down. In comparison, we achieve the same purpose by matching conditional distributions $F_{g|p}(\cdot;v(x,1))$ and $F_{g|p}(\cdot;v(\bar{x},0))$. As we show in Section 3, in several important applications, the outcome Y is either discrete (e.g. multinomial choices), or multi-dimensional with both discrete and continuous components (e.g., potential outcomes determined by a Roy model). In either cases, the latent index function $v(.)$ is vector-valued and the monotonicity condition in Vytlacil and Yildiz (2007) is not satisfied.

3 Examples

We now present several examples in which the latent indices are multi-dimensional. In the first and third example, the monotonicity condition in Vytlacil and Yildiz (2007) is not satisfied; in the second example, the identification requires a generalization of the monotonicity condition into an invertibility condition in higher dimensions.

Example 1. (Heteroskaedastic shocks in outcome) Consider a triangular system where a continuous outcome is determined by double indices $v(X,D)\equiv(v_1(X,D),v_2(X,D))$:

$$Y=g(v(X,D),\varepsilon)=v_1(X,D)+v_2(X,D)\varepsilon \text{ for } D\in\{0,1\}.$$

The selection equation determining the actual treatment is the same as (1.2). In this case the concept of monotonicity in $v\in\mathbb{R}^2$ is not well-defined, so the procedure proposed in Vytlacil and Yildiz (2007) is not suitable here[3]. Nevertheless, we can apply the method in Section 2 to identify the average treatment effect by using the distribution of outcome to find pairs of x and $\bar{x}$ such that $v(x,1)=v(\bar{x},0)$. Assume the range of $v_2(\cdot)$ is positive. To see the necessity in Assumption A4, note that

$$\begin{aligned}F_{g|u}(y;v(x,d)) &= E[1\{v_1(x,d)+v_2(x,d)\varepsilon\le y\}|U=u]\\ &= F_{\varepsilon|u}\left(\frac{y-v_1(x,d)}{v_2(x,d)}\right)\end{aligned}$$

[3] For this particular design, the approach proposed in Vuong and Xu (2017) should be valid. But it will not be for a slightly modified model, such as $Y=v_1(X,D)+(e_2+v_2(X,D)\,e_1)$, whereas ours will be.

for $d = 0; 1$. If the CDF of ε is increasing over $\mathbb{R}$, then for all y and $x \in S_1$ and $\tilde{x} \in S_0$,

$$F_{\varepsilon|u}(y; v(x; 1)) = F_{\varepsilon|u}(y; v(\tilde{x}; 0))$$

if and only if

$$\frac{y - v_1(x; 1)}{v_2(x; 1)} = \frac{y - v_1(\tilde{x}; 0)}{v_2(\tilde{x}; 0)}.$$

Differentiating with respect to y yields

$$v_2(x; 1) = v_2(\tilde{x}; 0)$$

which in turn implies

$$v_1(x; 1) = v_1(\tilde{x}; 0).$$

The sufficiency in Assumption A-4 is straight-forward.

Example 2. (Multinomial potential outcome) Consid ... ngular system where the outcome is multinomial. The multinomial response mode ... ng and rich history ... both applied and theoretical econometrics. Recent examples i ... iparametric litera ... include Lee (1995), Ahn, Powell, Ichimura, and Ruud (2017), S ... , and Song (... Pakes and Porter (2014), Khan, Ouyang, and Tamer (2019). But ... e work here ... of those papers allow for dummy endogenous variables or potent ... mes.

$$Y = g(v(X; D); \varepsilon) = \arg \max_{j=0,1,\ldots,J} y_{j;D}$$

where

$$y_{j;D} = v_j(X; D) + \varepsilon_j \quad \text{for } j = 1, 2, \ldots, J; \; y_{0;D}$$

By Ruud (2000) and Ahn, Powell, Ichimura, and Ruud (2017), the mapping from $v \in \mathbb{R}^J$ to $(F_{g|u}(j; v) : j \le J) \in \mathbb{R}^J$ is smooth and invertible provided that $\varepsilon \in \mathbb{R}^J$ has non-negative density everywhere. This implies Assumption A-4.

Example 3 . (Potential outcome from the Roy model) Consider a treatment effect model with an endogenous binary treatment D and with the potential outcome determined by a latent Roy model. The Roy model has also been studied extensively from both applied and theoretical perspectives. See for example the literature survey in Heckman and E.Vytlacil (2007) and the seminal paper in Heckman and Honoré (1990).

Here the observed outcome consists of two pieces: a continuous measure $Y = DY_1 + (1 - D)Y_0$ and a discrete indicator $W = DW_1 + (1 - D)W_0$ for $d = 0, 1$. These potential outcomes are given by

$$Y_d = \max_{j \in \{a, b\}} y_{j,d} \quad \text{and} \quad W_d = \arg\max_{j \in \{a, b\}} y_{j,d}$$

where a and b index potential outcomes realized in different sectors, with

$$y_{j,d} = v$$

This would allow us to recover the right hand side of (3.1) as

$$\Pr\{Y_0 \quad y; W_0 = a \mid X = \bar{x}; Z = z; D = 0\}.$$

To find such a pair of $(x; \bar{x})$, define $h_{d;W}(x; p; p'); h_{d;W}(x; p$

Example 4. (Potential outcome with random coe cients) Random coe cient models are prominent in both the theoretical and applied econometrics literature. They permit a exible way to allow for conditional heteroscedasticity and unobserved heterogeneity. See, for example Hsiao and Pesaran (2008) for a survey. Here we consider a treatment e ect model where the potential outcome is determined through random coe cients:

$$Y = DY_1 + (1 \quad D)Y_0 \text{ where } Y_d = (\quad _d + X^0 \quad _d) \text{ for } d = 0; 1$$

and the binary endogenous treatment D is determined as in the selection equation (1.2). The random intercepts $\quad _d$ 2 R and the random vectors of coe cients $\quad _d$ are given by

$$\quad _d = \quad _d(X) + \quad _d \text{ and } \quad _d = \quad _d(X) + "_d$$

where for any x and d 2 f 0; 1g., ($_d(x)$; $_d(x)$) 2 R^{K+1} is a vector of constant parameters while $\quad _d$ 2 R and $"_d$ 2 R^K are unobservable noises.

As in Vytlacil and Yildiz (2007), assume some elements in Z in the selection equation are excluded from X. We allow the vector of unobservable terms ($_1$; $_0$; $_0$; $_1$; U) to be arbitrarily correlated. We also assume that

$$(X; Z) \; ? \; (\; _1; \; _0; \; _0; \; _1; U), \qquad (4.1)$$

with the marginal distribution of U normalized to standard uniform, so that (Z) is directly identi ed as $P(Z) \quad E(D j Z =$

while the second term is counterfactual and can be written as

$$
\begin{aligned}
& {}_0(x;y;p) \quad E[1\{U \quad P\}1\{ {}_1 + X' {}_1 \quad y\}|X = x; P = p] \\
= \; & E[1\{U \quad p\}1\{ {}_1(x) + {}_1 + x'({}_1(x) + "_1) \quad y\}] \\
= \; & \int_p^1 \Pr\{ {}_1 + x' {}_1 \quad y \quad {}_1(x) \quad x' {}_1(x)|U = u\}du.
\end{aligned}
$$

For any p on the support of P given $X = x$, de ne

$$
\begin{aligned}
& h_1(x;y;p) \quad E[D1\{Y \quad y\}|X = x; P = p] \\
= \; & E[1\{U < P\}1\{ {}_1 + X' {}_1 \quad y\}|X = x; P = p] = E[1\{U < p\}1\{ {}_1 + x' {}_1 \quad y\}] \\
= \; & \int_0^p \Pr\{ {}_1 + x' {}_1 \quad y \quad {}_1(x) \quad x' {}_1(x)|U = u\}du;
\end{aligned}
$$

where the second equality uses (4.1). Likewise, under (4.1) we have:

$$
\begin{aligned}
& h_0(x;y;p) \quad E[(1 \quad D)1\{Y \quad y\}|X = x; P = p] \\
= \; & \int_p^1 \Pr\{ {}_0 + x' {}_0 \quad y \quad {}_0(x) \quad x' {}_0(x)|U_i = u\}du.
\end{aligned}
$$

Assume[4]

$F_{({}_1; {}_1)}$ XjU2ua= and can

because of (4.3). Thus the counterfactual$_0(x; y; p)$ would be identi ed as $h_0(x; t(x; y); p)$.

It remains to show that for each pair $(x; y)$ we can uniquely recover $t(x; y)$ using quantities that are identi able in the data-generating process. To do so, we de ne two auxiliary functions as follows: for $p_1 > p_2$ on the support of P given $X = x$, let

$$h_1(x; y; p_1; p_2) \quad h_1(x; y; p_1) \quad h_1(x; y; p_2)$$
$$= \int_{p_2}^{p_1} \Pr\{ \ _1 + x^0 \ _1 < y \quad _1(x) \quad x^0 \ _1(x) | U = u\} du;$$

and

$$h_0(x; y; p_1; p_2) \quad h_0(x; y; p_2) \quad h_0(x; y; p_1)$$
$$= \int_{p_2}^{p_1} \Pr\{ \ _0 + x^0 \ _0 < y \quad _0(x) \quad x^0 \ _0(x) | U = u\} du.$$

Suppose $_d + x^0 \ _d$ is continuously distributed over R for all values of x conditional on all u 2 $[0; 1]$. Then for any xed pair $(x; y)$ and $p_1 < p_2$,

$$h_1(x; y; p_1; p_2) = h_0(x; t(x; y); p_1; p_2)$$

if and only if

$$t(x; y) = y \quad _1(x)$$

t

De ne a measure of distance between

Now we describe an estimation procedure for the distributional treatment e ect in Example 4, where we had a model with random coe cients. In this case, the parameter of interest is for a chosen value of the scalar y,

$$_2(y) = \Pr\{Y_1 \quad y\}.$$

First, for xed values of y and $p_1 > p_2$, we propose to estimate $t(x; y)$ as

$$\hat{t}(x; y; p_1; p_2) = \arg\min_t (h_1(x; y; p_1; p_2) \quad h_0(x; t; p_1; p_2))^2$$

and then average over values of $p_1; p_2$:

$$\hat{}(x; y) = \frac{1}{n(n \quad 1)} \sum_{i \neq j} I[P_i > P_j]\hat{t}(x; y; P_i; P_j)$$

An infeasible estimator for the parameter $_2(y)$, which assumes $t(x; y)$ is known, would be

$$\hat{}_2(y) = \frac{1}{n}\sum_{i=1}^{n} (D_i 1\{Y_i \quad y\} + (1 \quad D_i)1\{Y_i \quad t(X_i; y)\}).$$

In practice, for feasible estimation, one needs to replace $t(x; y)$ by its estimator $\hat{}(x; y)$.

6 Simulation Study

This section presents simulation evidence for the performance of the proposed estimation procedures described in Section 5, for both the Average Treatment E ect and the Distributional Treatment E ect. We report results for both our proposed estimator and that in Vytlacil and Yildiz (2007), for several designs. These include designs where the said monotonicity condition fails, and designs where the disturbance terms in the outcome equation are multidimensional.

Throughout all designs we model the treatment or dummy endogenous variable as

$$D = I[Z \quad U > 0]$$

where $Z; U$ are independent standard normal. We experiment with the following designs for the outcome

Design 1

$$Y = X + 0{:}5 \; D +$$

where X is standard normal, $(; U)$ are distributed bivariate normal, each with mean 0 and variance 1, with correlations of 0,0.25,0.5.

Design 2

$$Y = X + 0.5 \; D + (X + D)$$

where X is distributed standard normal, $(\; ; U)$ are distributed bivariate normal, each with mean 0 and variance 1, with correlations of 0,0.25,0.5.

Design 3

$$Y = (X + 0.5 \; D + \;)^2$$

where X is distributed standard normal, $(\; ; U)$ are distributed bivariate normal, each with mean 0 and variance 1, with correlations of 0,0.25,0.5.

We note that the monotonicity condition is satisfied in design 1 but fails in the other two designs. For each of these designs, we report results for estimating the parameter $E[Y_1]$, which denotes the expected value for potential outcome under treatment $D = 1$. The two estimators used in the simulation study were the one proposed in Section 5 and the method proposed in Vytlacil and Yildiz (2007). The summary statistics, scaled by the true parameter value, Mean Bias, Median Bias, Root Mean Squared Error, (RMSE), and Median Absolute Deviation (MAD) were evaluated for sample sizes of 100, 200, 400 for 401 replications. Results for each of these designs are reported in Tables 1 to 3 respectively. In implementing our procedure we assumed the propensity score function is known, and conducted next stage estimation using a nonparametric kernel estimator with normal kernel function, and a bandwidth of $n^{1/5}$

the sample size grows, which is expected, as the monotonicity condition rely on is satisfied in these designs. In this case, their approach has smaller standard errors largely due to the relative simpler structure of the infeasible version, but their biases persist even when the sample size increases.

For designs 2 and 3, where monotonicity is violated, the procedure proposed in Vytlacil and Yildiz (2007) does not perform well. In design 2 in Table 2 both the bias and RMSE are generally increasing with the sample size. Results for their estimator are better in design 3, but the bias hardly converges with the sample size and is much larger compared to our estimator.

We also simulate data from a model with dummy endogenous variable and potential outcomes determined by random coefficients. It is important to note that for this design, the original matching idea in Vytlacil and Yildiz (2007) does not apply. This is because different values of x lead to different distribution of the composite error $_d + x^0{}_d$. Our contribution in Section 4 is to propose a new approach based on matching different values of outcome y, rather than the regressors x. Based on the counterfactual framework discussed in Section 4, here the treatment variable D is modeled as the same way as the dummy endogenous variable above. Similarly the regressor X is standard normal. For both $Y_0; Y_1$ the random intercepts were modeled as constants (0 and 1, respectively) and the additive error terms were each standard normal. For the random slopes, the means were 1 and 2 respectively, and the additive error terms were also standard normal, independent of all other disturbance terms and each other. Here we use the procedure in Section 4 to estimate the parameter $_2 = P(Y_1 < y)$, where in the simulation we set $y = 1$. The same four summary statistics are reported for sample sizes 100,200,400, based on 401 replications. Results for this random coefficients design are reported in Table 4.

The estimator proposed in Section 5 performs well; but the bias and RMSE are much small at 400 observations compared to 100 and 200 observations, indicating convergence at the parametric rate.

Table 1

	CKT			VY		
v	0	1=4	1=2	0	1=4	1=2
n=100						

Table 4

CKT			
v	0	1 =4	1 =2
n=100			
MEAN BIAS	0.0109	-0.0086	0.0038
MEDIAN BIAS	0.0000	-0.0064	0.0126
RMSE	0.1011	0.0979	0.0955
MAD	0.0600	0.0648	0.0652
n=200			
MEAN BIAS	-0.0050	-0.0150	0.0095
MEDIAN BIAS	-0.0100	-0.0161	0.0029
RMSE	0.0669	0.0669	0.0665
MAD	0.0400	0.0454	0.0457
n=400			
MEAN BIAS	0.0012	-0.0132	0.0074
MEDIAN BIAS	0.0049	-0.0162	0.0077
RMSE	0.0501	0.0494	0.0495
MAD	0.0349	0.0325	0.0360

7 Conclusion

In this paper, we considered identification and estimation of nonseparable models with endogenous binary treatment. Existing approaches are based on a monotonicity condition, which is violated in models with multiple unobserved idiosyncratic shocks. Such models arise in many important empirical settings, including Roy models and multinomial choice models with dummy endogenous variables, as well as treatment effect models with random coefficients. We establish novel identification results for these models which are constructive and conducive to estimation procedures which are easy to compute and whose limiting distributional properties follow from standard large sample theorems. A simulation study indicates adequate finite sample performance of our proposed methods.

This paper leaves open areas for future research. Our method requires the selection of the
-439(AJ9ose)-329(limitino4ll)23ultiple study3ulour W,1(o)-2sortonrde(ds.)]TJ 1udy3ul(these3uln)-255(l

Auerbach, E. (2019): "Identification and Estimation of a Partially Linear Regression Model using Network Data," mimeograph, Northwestern University.

Chen, S., S. Khan, and X. Tang (2016): "On the Informational Content of Special Regressors in Heteroskedastic Binary Response Models," *Journal of Econometrics*, 193, 162–182.

Chen, S. H., and S. Khan (2014): "Semiparametric Estimation of Program Impacts on Dispersion of Potential Wages," *Journal of Applied Econometrics*, 29, 901–919.

Chernozhukov, V., and C. Hansen

Khan, S., A. Maurel, and Y. Zhang (2019): \Informational Content of Factor Structures in Simultaneous Discrete Response Models," Working Paper, Duke University.

Khan, S., F. Ouyang, and E. Tamer (2019): \Inference in Semiparametric Mutinomial Response Models," Boston College Working Paper.

Lee, L.-F. (1995): \Semiparametric Maximum Likelihood Estimation of Polychotomous and Sequential Choice Models," Journal of Econometrics, 65, 381{428.

Lewbel, A., D. Jacho-Chavez, and J. Encarnciono (2016): \Identi cation and estimation of semiparametric two-step models," Quantitative Economics, 7, 561{589.

Mourifi e, I. (2015): \Sharp Bounds on Treatment E ects in a Binary Triangular System," Journal of Econometrics, 187(1), 74{81.

Pakes, A., and J. Porter (2014): \Moment Inequalties for Multinomial Choice with Fixed E ects," Harvard University Working Paper.

Ruud, P. (2000): \Semiparametric estimation of discrete choice models," mimeograph, University of California at Berkeley.

Shaikh, A. M., and E. Vytlacil (2011): \Partial Identi cation in Triangular Systems of Equations with Binary Dependent Variables," Econometrica, 79(3), 949{955.

Shi, X., M. Shum, and W. Song (2018): \Estimating Semi-Parametric Panel Multinomial Choice Models using Cyclic Monotonicity," Econometrica, 86, 737{761.

Torgovitsky, A. (2015): \Identi cation of Nonseparable Models Using Instruments with Small Support," Econometrica, 3, 1185{1197.

Vuong, Q., and H. Xu (2017): \Counterfactual mapping and individual treatment e ects in nonseparable models with binary endogeneity," Quantitative Economics, pp. 589{610.

Vytlacil, E. J., and N. Yildiz (2007): \Dummy Endogenous Variables in Weakly Separable Models," Econometrica, 75, 757{779.