Identification of semiparametric discrete outcome models with bounded covariates*

Nail Kashaev †
nkashaev@uwo.ca

This version: November 2018

Abstract Identification of discrete outcome models is often established by using special covariates that have full support. This paper shows how these identification results can be extended to a large class of commonly used semiparametric discrete outcome models when all covariates are bounded. I apply the proposed methodology to multinomial choice models, bundles models, and finite games of complete information.
Keywords: Discrete outcome, multinomial choice, random coefficients, games of complete information

*I would like to thank Roy Allen, Victor H. Aguiar, Tim Conley, Salvador Navarro, David Rivers, and Bruno Salcedo for useful comments and discussions.
†Department of Economics, University of Western Ontario.

1. Introduction

Covariates that have full support (i.e., supported on the whole Euclidean space) provide an elegant and powerful approach for establishing identification in many discrete outcome models.[1] Existence of such covariates is often necessary for nonparametric identification of distributions of latent variables.[2] "Identification-at-infinity" arguments are based on existence of covariates with full support.[3] However, finding such covariates in applied work is often problematic.

In this paper, I show that common parametric assumptions about the distribution of *some* unobservables (e.g. one normally distributed random coefficient in a model with multiple random coefficients) can fully restore the identification power of covariates with full support, even if covariates are in fact bounded. I provide two results connecting semiparametric discrete outcome models with covariates that have bounded support and (non)parametric discrete outcome models with special covariates that have full support. The first result is general and can be applied to a large class of semiparametric models. However, it requires a preliminary identification of a finite-dimensional parameter by using some auxiliary arguments. The second result does not require any extra identification steps, and uses one of the most popular parametrizations in applied work – the Gaussian distribution. I apply the proposed approaches to three well-known models: multinomial choice models with random coefficients, bundles models, and finite games of complete information.

The results of this paper rest on two commonly used assumptions. First, I assume existence of excluded (special) covariates that affect the distribution over outcomes via a latent index. Second, I impose commonly used parametric restrictions on the distribution of this index (e.g., the Gumbel or the Gaussian distribution). If the distribution of the index is sufficiently "rich", then I show how to identify the

[1]See, for example, Manski (1985), Matzkin (1992), Ichimura and Thompson (1998), Lewbel (1998), Lewbel (2000), Tamer (2003), Matzkin (2007), Berry and Haile (2009), Bajari et al. (2010), Gautier and Kitamura (2013), Fox and Gandhi (2016), Dunker et al. (2017), Fox and Lazzati (2017), and Fox et al. (2018).

[2]Full support assumption is not necessary if it is assumed that latent variables have bounded support. In this case the support of the special covariates have to be large enough to cover the support of unobservables.

[3]For example, Manski (1988), Heckman (1990), and Tamer (2003).

distribution over outcomes conditional on the realization of the observed covariates and the latent index. Since the index distribution is usually assumed to have full support, I can treat the latent index as observed covariate with full support and apply *any* identification technique that requires existence of such covariates.

The index has different interpretation in different settings. For instance, in random coefficients model one of the random coefficients can be treated as the latent index. In games, the role of the index is played by a component of random utilities corresponding to different outcomes. "Richness" of the latent index distribution is formalized by a notion of bounded completeness.[4]

This paper also contributes to the literature on partially identified models. I show that in many partially identified models the identified sets are "thin" in the following sense. The model parameters (including infinite-dimensional ones) are identified up to a finite-dimensional parameter of a lower dimension. This finding may lead to substantial computational gains in constructing confidence sets for partially identified parameters (e.g. Chen et al. (2011)) and sheds some light on the properties of identified sets in these models.

The paper is organized as follows. Section 2 provides a motivating example. In Section 3 I describe the setting and derive general identification results. Section 4 specializes the results from Section 3 for widely used normally distributed latent variables. In Sections 5 I apply the result from Section 3 to three different discrete outcome models. Section 6 concludes. All proofs can be found in Appendix A.

[4]Completeness of a family distribution is a well-known concept both in the Statistics and Econometrics literature. See for example Mattner et al. (1993), Newey and Powell (2003), Chernozhukov and Hansen (2005), Blundell et al. (2007), Chernozhukov et al. (2007), Hu and Schennach (2008), Andrews (2011), Darolles et al. (2011), and d'Haultfoeuille (2011).

2. Motivating Example: Binary Choice

Consider a simple single agent binary choice problem.[5] A utility maximizing agent has to choose $y \in Y = \{0, 1\}$. The utility of alternative $y = 0$ is normalized to 0. The utility of option $y = 1$ is

$$\mathbf{z}_{2,1}\mathbf{v}_1 + \mathbf{g}_1. \tag{1}$$

Random variables $\mathbf{v}_1$ and $\mathbf{g}_1$ represent the random slope coefficient corresponding to covariate $\mathbf{z}_{2,1} \in Z_{2,1} \subseteq \mathbb{R}$ and the random intercept, respectively.

The objective of the econometrician is to recover the c.d.f. of $\mathbf{g}_1$, $F_{\mathbf{g}_1}$, from observed distribution of choices $\mathbf{y}$ and covariates $\mathbf{z}$. Typically it is assumed that $\mathbf{v}_1 = 1$ a.s., and $\mathbf{z}_{2,1}$ and $\mathbf{g}_1$ are independent. In this case the main identifying conditions is

$$\Pr(\mathbf{y} = 1 | \mathbf{z}_{2,1} = z_{2,1}) = 1 - F_{\mathbf{g}_1}(-z_{2,1}),$$

for all $z_{2,1} \in Z_{2,1}$. Hence, if $\mathbf{g}_1$ is supported on $\mathbb{R}$, then nonparametric identification of its distribution can be achieved if and only if $\mathbf{z}_{2,1}$ has full support. Next I will show how another commonly used assumption about $\mathbf{v}_1$ can lead to identification of $F_{\mathbf{g}_1}$ even if all observed covariates have bounded support.

Assume that there exists a covariate $\mathbf{z}_1 \in Z_1 \subseteq \mathbb{R}$ such that for $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_{2,1})^\mathsf{T} \in Z = Z_1 \times Z_{2,1}$

$$\mathbf{v}_1 | (\mathbf{z} = z) \sim N(\beta_0 + \beta_1 z_1, 1), \quad \forall z \in Z,$$

where $\beta_1 \neq 0$ and $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. For this section only I assume that $\beta = (\beta_0, \beta_1)^\mathsf{T}$ is known or can be identified. Later on I will show how one can pointidentify β. Assuming that $\mathbf{g}_1$ is independent

[5] Throughout the paper, deterministic vectors and functions are denoted by lower-case regular font Latin letters (e.g., x), random objects by bold letters (e.g., $\mathbf{x}$). Capital letters are used to denote supports of random variables (e.g., $\mathbf{x} \in X$). I denote the support of a conditional distribution of $\mathbf{x}$ conditional on $\mathbf{z} = z$ by X_z. Also, given a family $x = (x_k)_{k \in K}$ and a particular index value $k \in K$, I use the notation x_{-k} for $(x_j)_{j \in K \setminus \{k\}}$. $F_{\mathbf{x}}(\cdot)$ $(f_{\mathbf{x}}(\cdot))$ and $F_{\mathbf{x}|\mathbf{z}}(\cdot|z)$ $(f_{\mathbf{x}|\mathbf{z}}(\cdot|z))$ denote the c.d.f. (p.d.f.) of $\mathbf{x}$ and $\mathbf{x}|\mathbf{z} = z$, respectively.

of $\mathbf{z}$ and $\mathbf{v}_1$, we can rewrite (1) as

$$\mathbf{z}_{2,1}(\beta_0 + \beta_1 \mathbf{z}_1 + \mathbf{e}_1) + \mathbf{g}_1,$$

where $\mathbf{e}_1$ is a standard normal random variable that is independent of $\mathbf{z}$.[6]

The key step is to show that we can identify

$$\Pr(\mathbf{y} = 1 | \mathbf{z} = z, \mathbf{v}_1 = v_1)$$

for all $v_1 \in \mathbb{R}$ and $z \in Z$. If we can identify $\Pr(\mathbf{y} = 1|\mathbf{z} = \cdot, \mathbf{v}_1 = \cdot)$, then we can analyze the model as if $\mathbf{v}_1$ is observed. In this case $\mathbf{z}_{2,1}\mathbf{v}_1$ becomes an observed special covariate with *full* support since $\mathbf{v}_1$ is supported on $\mathbb{R}$. By using variation in $\mathbf{v}_1$ we can recover $F_{\mathbf{g}_1}$ even if the support of $\mathbf{z}$ is bounded. In other words, the above binary choice model with just one normally distributed random slope coefficient with bounded covariates, in terms of identification features, is equivalent to the binary choice model where the utility from choosing $y = 1$ is equal to

$$\mathbf{r}_1 + \mathbf{g}_1, \tag{2}$$

where $\mathbf{r}_1$ and $\mathbf{g}_1$ are independent, and $\mathbf{r}_1$ is observed covariate supported on

$$\{r_1 \in \mathbb{R} \,:\, r_1 = v_1 z_{2,1}, v_1 \in \mathbb{R}, z_{2,1} \in Z_{2,1}\} = \mathbb{R}.$$

It is left to show that $\Pr(\mathbf{y} = 1|\mathbf{z} = z, \mathbf{v}_1 = \cdot)$ can be identified. First, note that under independence condition

$$\Pr(\mathbf{y} = 1|\mathbf{z}_1 = z_1, \mathbf{z}_{2,1} = z_{2,1}, \mathbf{v}_1 = v_1) = \Pr(\mathbf{y} = 1|\mathbf{z}_{2,1} = z_{2,1}, \mathbf{v}_1 = v_1).$$

Second, note that for any fixed $z_{2,1}$ we have the following integral equation

$$\Pr(\mathbf{y} = 1|\mathbf{z}_1 = z_1, \mathbf{z}_{2,1} = z_{2,1}) = \int_{\mathbb{R}} \Pr(\mathbf{y} = 1|\mathbf{z}_{2,1} = z_{2,1}, \mathbf{v}_1 = v)\phi(v - (\beta_0 + \beta_1 z_1))dv,$$

for all $z_1 \in Z_1$. Since variation in z_1 does not affect $\Pr(\mathbf{y} = 1|\mathbf{z}_{2,1} = z_{2,1}, \mathbf{v}_1 = \cdot)$, we can use this variation to identify it. In other words, if the family of normal

[6] Assumption that the variance of $\mathbf{e}_1$ is 1 is a scale normalization.

distributions $\{\phi(\cdot - \beta_0 - \beta_1 z_1) \ : \ z_1 \in Z_1\}$ is sufficiently "rich", then the integral equation has a unique solution. For discrete distributions the "richness" condition is usually characterized by the rank condition. For continuous distributions the "richness" condition is associated with a notion of "completeness". In our example, the "completeness" condition is satisfied since the family of $\{\phi(\cdot - \beta_0 - \beta_1 z_1) : z_1 \in Z_1\}$ is complete if $\mathrm{int}(Z_1) \neq \emptyset$.[7] Since the choice of $z_{2,1}$ was arbitrary, we recover $\Pr(\mathbf{y} = 1|\mathbf{z}_{2,1} = \cdot, \mathbf{v}_1 = \cdot)$ and thus can work with well-known model (2).

Two important assumptions needed for the result to hold are: (i) existence of a parametric index $\mathbf{v}_1$ such that the choices are affected by some excluded covariates only through the index; and (ii) the distribution of the index conditional on excluded covariates is sufficiently rich (complete). The rest of the paper generalizes these key assumptions to environments with multiple agents and outcomes, and establishes identification of β.

3. General Model

Each instance of the environment is characterized by an endogenous outcome $\mathbf{y}$ from a known finite set Y, a vector of observed exogenous characteristics $\mathbf{x} \in X \subseteq \mathbb{R}^{d_x}$, $d_x < \infty$, that can be partitioned into $x = (z^\intercal, w^\intercal)^\intercal$, and a vector of unobserved structural variables $\mathbf{v} \in V \subseteq \mathbb{R}^{d_v}$.[8] It is assumed that the econometrician observes the joint distribution of $(\mathbf{y}, \mathbf{x}^\intercal)^\intercal$.

Assumption 1 (Exclusion Restrictions) There exist $Y^* \subseteq Y$ and $h_0 : Y^* \times W \times V \to [0, 1]$, such that

$$\Pr(\mathbf{y} = y|\mathbf{z} = z, \mathbf{w} = w, \mathbf{v} = v) = h_0(y, w, v),$$

for all $y \in Y^*$, $x = (z^\intercal, w^\intercal)^\intercal \in X$, and $v \in V$.

[7] $\mathrm{int}(Z_1)$ denotes the interior of Z_1.

[8] My analysis allows for countable sets of outcomes, but for the exposition purposes I focus on finite sets.

Assumption 1 is an exclusion restriction that requires covariates $\mathbf{z}$ to affect distribution over some outcomes only via the distribution of the latent $\mathbf{v}$. Note that the exclusion restriction does not need to be imposed on all outcomes. For instance, in single agent decision models one can identify the payoff parameters by observing only the probability of choosing the outside option (e.g., Thompson (1989) and Lewbel (2000)). Assumption 1 does not rule out existence of other latent variables (different from $\mathbf{v}$) since exclusion restrictions are imposed on the distribution over outcomes conditional on $\mathbf{x} = x$ and $\mathbf{v} = v$.[9]

The next assumption is a parametric restriction on the latent variable whose distribution is affected by the excluded covariates $\mathbf{z}$.

Assumption 2 (Bounded completeness) For every $w \in W$, there exists $Z' \subseteq Z_w$ such that the family of distributions $\left\{F_{\mathbf{v}|\mathbf{z},\mathbf{w}}(\cdot|z, w), z \in Z'\right\}$ is boundedly complete. That is,

$$\forall z \in Z',\ \int_V g(t) dF_{\mathbf{v}|\mathbf{z},\mathbf{w}}(t|z, w) = 0 \implies g(\mathbf{v}) = 0 \text{ a.s..}$$

Completeness assumptions have been widely used in econometric analysis. Completeness is typically imposed on the distribution of observables (e.g. Newey and Powell (2003)). However, many commonly used parametric restrictions on the distribution of unobservables imply Assumption 2. For instance, it is satisfied for the Gaussian distribution and the Gumbel distribution. [10]

$$\mu(y|x) = \Pr(\mathbf{y} = y|\mathbf{x} = x)$$

is known for every $y \in Y^*$ and $x \in X$.

Proposition 3.1 *Under Assumptions 1 and 2, h_0 is identified from μ up to $F_{\mathbf{v}|\mathbf{x}}$.*

Proposition 3.1 implies that under exclusion restrictions if one assumes that the latent variable has a known distribution belonging to a boundedly complete family, then one can work with the model as if the realizations of latent variables are observed in the data since we can identify $h_0(y, w, \cdot) = \Pr(\mathbf{y} = y|\mathbf{w} = w, \mathbf{v} = \cdot)$. Thus, if I

[9] I consider a model with unobserved heterogeneity that is not fully captured by $\mathbf{v}$ in Section 5.1.

[10] For testability of the completeness assumptions see Canay et al. (2013).

know or can identify $F_{\mathbf{v}|\mathbf{x}}$ (see Section 4), for identification I can interpret *latent* variables ($\mathbf{v}$) as *observed* covariates. If these latent variables have full support (e.g. normal errors), then all identification techniques that require existence of covariates with full support can be applied (e.g., Fox and Gandhi (2016) in the context of random coefficients model and Bajari et al. (2010) in the context of games). In other words, I can transform a model with covariates that have bounded support into the model with covariates that have full support, and then use existing methods to identify different objects of interest.

The following example demonstrates how knowing h_0 can help to identify some underlying aspects of the model.

Example 3.1 Suppose that $Y = Y^* = \{0, 1\}$ and h_0 is identified.[11] Assume that $\gamma(\mathbf{w}) + \mathbf{v}$ represents the utility agent gets from choosing $y = 1$, where γ is some unknown function of w. Assume, moreover, that the utility from choosing $y = 0$ is normalized to zero. The identified h_0 is consistent with the utility maximizing behavior if and only if

$$h_0(1, w, v) = \mathbb{1}\,(\,\gamma(w) + v \geq 0\,)$$

for almost all $v \in V$ and $w \in W$. Thus, if $V = \mathbb{R}$, then one can test for utility maximizing behavior, and can identify $\gamma(\cdot)$ if the agent maximizes utility.

Proposition 3.1 establishes identification of h_0 only up to $F_{\mathbf{v}|\mathbf{x}}$. Typically $F_{\mathbf{v}|\mathbf{x}}$ is known up to a finite-dimensional parameter. In Section 4 I show how one can identify this parameter and thus h_0 if $F_{\mathbf{v}|\mathbf{x}}$ is assumed to be the Gaussian distribution. In Section 5.3 I show how Proposition 3.1 can be used to characterize the identified set of a partially identified game of complete information.

[11] For binary outcome models assuming that both outcomes satisfy the exclusion restriction is equivalent to assuming that just one of them does.

4. Gaussian Distribution

Proposition 3.1 can be applied with any known parametric distribution $F_{\mathbf{v}|\mathbf{x}}$ as long as the family of the distributions generated by the variation in excluded covariates is complete. The most prominent example of such families is the exponential family of distributions. In this section I specialize the results from the previous section to probably one of the most common parametrization in applied work – Gaussian errors. I show how to identify $F_{\mathbf{v}|\mathbf{x}}$ for such models. The following assumption is sufficient for Assumption 2 to hold.

Assumption 3 (i) The latent $\mathbf{v}$ satisfies

$$\mathbf{v}_i = \mathbf{z}_{2,i}[\beta_{0,i}(\mathbf{w}) + \beta_{1,i}(\mathbf{w})\mathbf{z}_{1,i} + \mathbf{e}_i] \quad \text{a.s.}$$

where $\beta_{0,i}(\cdot)$ and $\beta_{1,i}(\cdot)$, $i = 1, \ldots, d_v$, are some unknown measurable functions such that $\beta_{1,i}(w) \neq 0$ for all $w \in W$;

(ii) $\{\mathbf{e}_i\}_{i=1,\ldots,d_v}$ are independent identically distributed standard normal random variables that are independent of $\mathbf{x}$;

(iii) For every $w \in W$ the support of $\mathbf{z}|(\mathbf{w} = w)$, Z_w, contains an open ball;

(iv) The sign of either $\beta_{0,i}(w)$ or $\beta_{1,i}(w)$ is known for every $w \in W$ and i.

Assumption 3(i) is motivated by random coefficient models. The covariate $\mathbf{z}_{2,i}$ can be interpreted as choice ("product") specific characteristic. The random coefficients $[\beta_{0,i}(\mathbf{w}) + \beta_{1,i}(\mathbf{w})\mathbf{z}_{1,i} + \mathbf{e}_i]$ captures agent specific heterogeneity in tastes. The only support restriction is imposed on $\mathbf{z}$ (Assumption 3(iii)). Non of the covariates are assumed to have full or unbounded support. Assumptions 3(i)-(iii) are sufficient for Assumption 2 since the family of normal distributions indexed by the mean parameter is complete as long as the parameter space for the mean parameter contains an open ball.[12]

[12] Any distribution from the exponential family of distributions (e.g., the Gumbel distribution) would be sufficient for Assumption 2 to hold.

Assumptions 3(iv) is a normalization. It requires that either the sign of the marginal effect of $z_{1,i}$ or the sign of the intercept (as long at it is not equal to zero) are known (or can be identified). In discrete outcome models with almost surely unique equilibrium (e.g. multinomial choice) the sign of $\beta_{1,i}(\cdot)$ can often be identified because of monotonicity of h in utility indexes v. For instance, in multinomial choice models the probability of choosing an outside option is decreasing in mean utilities of other choices. Without additional restrictions the sign can not be identified as the following example demonstrates.

Example 4.1 Suppose that Assumptions 1 and 3(i)-(iii) are satisfied with $d_v = 1$ and $x = (z_{11}, z_{21})^\intercal$. Take $y \in Y^*$ and note that

$$\begin{aligned}
\mu(y|x) &= \int_{-\infty}^{+\infty} h_0(y, v_1)\phi(v_1/z_{2,1} - \beta_{0,1} - \beta_{1,1}z_{1,1})dv_1 = \\
&= \int_{-\infty}^{+\infty} h_0(y, v_1)\phi(-v_1/z_{2,1} + \beta_{0,1} + \beta_{1,1}z_{1,1})dv_1 = \\
&= \int_{-\infty}^{+\infty} h_0(y, -v_1)\phi(v_1/z_{2,1} + \beta_{0,1} + \beta_{1,1}z_{1,1})dv_1 = \\
&= \int_{-\infty}^{+\infty} h_0(y, -v_1)\phi(v_1/z_{2,1} - (-\beta_{0,1}) - (-\beta_{1,1})z_{1,1})dv_1,
\end{aligned}$$

where the second equality follows from the symmetry of $\phi(\cdot)$ and the third equality follows from the change of variables. Hence, if h_0 and $(\beta_{0,1}, \beta_{1,1})^\intercal$ can generate the data, then $\tilde{h}_0$ and $(-\beta_{0,1}, -\beta_{1,1})^\intercal$ such that $\tilde{h}_0(y, v) = h_0(y, -v)$ for all y and v can generate the data too.

Proposition 3.1 only identifies h_0 up to the distribution of unobservables. In particular, under Assumption 3, h_0 is identified up to $\{\beta_{0,i}(\cdot), \beta_{1,i}(\cdot)\}_{i=1}^{d_v}$. The following assumption allows us to identify the distribution of unobservables and thus h_0.

Let $z_{1,-i} = (z_{1,k})_{k \neq i}$. For a fixed $y^* \in Y^*$, $z_{1,-i}$ and z_2, let $\eta : Z_{1,i|w,z_{1,-i},z_2} \to [0, 1]$ be such that for $x = ((z_{1,i}, z_{1,-i})^\intercal, z_2^\intercal, w^\intercal)^\intercal$

$$\eta(z_{1,i}) = \mu(y^*|x).$$

Assumption 4 For every $w \in W$ and $i = 1, 2, \ldots, d_y$, there exists $y^* \in Y^*$ and $z_{2,i} \in Z_{2,i|w} \setminus \{0\}$ such that $\eta(\cdot)$ is neither an exponential nor an affine function of

$z_{1,i}$.

Assumption 4 means that if we fix all covariates but one, then the probability of observing one excluded outcome conditional on covariates is neither affine nor exponential function of the non-fixed covariate. Assumption 4 is not very restrictive since it rules out only some exponential and linear probability models. Moreover, it is testable.

Proposition 4.1 *Suppose that Assumptions 1, 3, and 4 hold. Then h_0 and $\{\beta_{0,i}(\cdot), \beta_{1,i}(\cdot)\}_{i=1}^{d_v}$ are identified.*

Proposition 4.1 establishes identification of h_0 and $F_{\mathbf{v}|\mathbf{x}}$ for normally distributed latent variables. Next I will show how one can use h_0 in different discrete outcome models.

5. Applications

5.1. Multinomial Choice

Consider the following random coefficients model. The agent has to choose between J inside goods and an outside option of no purchase. That is, $y \in Y = \{0, 1, \ldots, J\}$. I normalize the utility from alternative $y = 0$ to 0. The random utility from choosing an alternative $y \neq 0$ is of the form

$$\mathbf{u}_y = \mathbf{z}_{2,y}\Big[\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1\Big] + \mathbf{g}_y. \tag{3}$$

The random coefficient $\Big[\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1\Big]$ represents individual specific heterogeneous tastes associated with product characteristic $\mathbf{z}_{2,y}$. The latent random vector $\mathbf{g} = (\mathbf{g}_y)_{y \in Y\setminus\{0\}}$ captures other sources of unobserved heterogeneity which are different from $\Big[\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1\Big]$. The observed covariates are $\mathbf{x} = (\mathbf{z}_1, \mathbf{z}_2^{\mathsf{T}}, \mathbf{w}^{\mathsf{T}})^{\mathsf{T}}$, where $\mathbf{z}_2 = (\mathbf{z}_{2,y})_{y \in Y \setminus \{0\}}$. Assume that the agents are utility maximizers.[13]

Assumption 5 (i) $\mathbf{e}_1$ is an independent of $(\mathbf{g}^\top, \mathbf{x}^\top)^\top$ standard normal random variable;

(ii) $\beta_1(\mathbf{w}) \neq 0 \quad$ a.s.;

(iii) For every $w \in W$ the support of $\mathbf{z}|\mathbf{w} = w$, Z_w, contains an open ball.

Similarly to the existing treatment of random coefficients model, I assume that the random coefficients in front of $\mathbf{z}_{2,y}$ are the same for each alternative y.[14] However, I do not impose sign restrictions on $\left[\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1\right]$. Note that

$$\Pr(\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1 \geq 0 | \mathbf{x} = x) = \Phi(\beta_0(w) + \beta_1(w)z_1),$$

where $\Phi(\cdot)$ is the standard normal c.d.f. Thus, since there are no restrictions on $\beta_0(\cdot)$, the random coefficient $\left[\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1\right]$ can be positive (negative) with probability that is arbitrarily close to 1.

Assumption 6 (i) Random shocks $\mathbf{g}$ are conditionally independent of $\mathbf{z}$ conditional on $\mathbf{w} = w$. That is, for all $x = (z^\top, w^\top)^\top \in X$

$$F_{\mathbf{g}|\mathbf{z},\mathbf{w}}(\cdot|z, w) = F_{\mathbf{g}|\mathbf{w}}(\cdot|w),$$

where $F_{\mathbf{g}|\mathbf{z},\mathbf{w}}(\cdot|z, w)$ and $F_{\mathbf{g}|\mathbf{w}}(\cdot|w)$ are conditional c.d.fs of $\mathbf{g}|\mathbf{z} = z, \mathbf{w} = w$ and $\mathbf{g}|\mathbf{w} = w$, respectively;

(ii) For every $w \in W$, there exists $z_2 \neq 0$ such that it is contained in $Z_{2|w}$ with some open neighborhood and $z_{2,y} = z_{2,y'}$ for any $y, y' \in Y \setminus \{0\}$;

(iii) For every $w \in W$ there exists $z_2 \in Z_{2|w}$ such that

$$\Pr(\mathbf{y} = 0 | \mathbf{z}_1 = \cdot, \mathbf{z}_2 = z_2, \mathbf{w} = w)$$

[13] The results below hold even if utility maximizing behavior is not assumed as long as Assumption 1 is satisfied for at least one outcome.

[14] The model extends to the case when the random slope coeffcient is choice specific as in Assumption 3. But in this case one would need to find more excluded covariates.

is neither exponential nor affine function.

Assumption 6(i) together with utility maximizing behavior guarantees that Assumption 1 is satisfied. Assumption 6(i) is the only restriction on $\mathbf{g}$. I allow $\mathbf{g}$ to be discrete or continuous random variable with unknown support. Assumptions 6(ii)-(iii) are testable restrictions. Assumptions 6(ii) imply that one can find $z_{2,1} = z_{2,2} = \cdots = z_{2,J} \neq 0$ such that Assumption 2 is satisfied for $v = z_{2,1}(\beta_0(w) + \beta_1(w)z_1 + e_1)$. Assumption 6(iii) is sufficient for Assumption 4 to hold.

Proposition 5.1 *Suppose Assumptions 5 and 6 hold. Then*

(i) $\beta_0(\cdot)$ *and* $\beta_1(\cdot)$ *are identified;*

(ii) The above model inherits all identifying properties of the following random coefficients model:

$$\mathbf{u}_y = \mathbf{r}_y + \mathbf{g}_y, \quad y \neq 0,$$
$$\mathbf{u}_y = 0, \quad y = 0,$$

where $\mathbf{r} = (\mathbf{r}_y)_{y \in Y \setminus \{0\}}$ *is an observed covariate independent of* $\mathbf{g} = (\mathbf{g}_y)_{y \in Y \setminus \{0\}}$ *conditional on* $\mathbf{w}$ *with the conditional support*

$$R_w = \left\{ r \in \mathbb{R}^J \, : \, r = \lambda z_2, \; \lambda \in \mathbb{R}, \; z_2 \in Z_{2|w} \right\}.$$

Proposition 5.1 implies that the original random coefficient model can be represented in the "special-covariate-with-full-support" framework without assuming existence of such covariates. Moreover, if the set of directions that $z_2 / \|z_2\|$ can cover is sufficiently rich, then $R_w = \mathbb{R}^J$ and all the identification results that require existence of special covariates with full support (e.g., Lewbel (2000), Berry and Haile (2009), and Fox and Gandhi (2016)) can be applied. For instance, if $\{z_2/\|z_2\| \, : \, z_2 \in Z_{2|w}\}$ is equal to a unit sphere in $\mathbb{R}^J$ for every w, then $R_w = \mathbb{R}^J$ and I can nonparametrically identify $F_{\mathbf{g}|\mathbf{w}}$.[15]

[15] In general, I can identify $F_{\mathbf{g}|\mathbf{w}}(\cdot|w)$ over R_w only.

5.2. Bundles

Consider the following bundles model motivated by Gentzkow (2007), Dunker et al. (2017), and Fox and Lazzati (2017). There are J goods and the agent can purchase any bundle consisting of these goods. The vector y describes the purchasing decision of the agent. That is, $y \in Y = \{0,1\}^J$. For instance, $y = (0,1,0,1,0,\ldots,0)^\mathsf{T}$ corresponds to the case when the agent purchased a bundle of goods 2 and 4. I normalize the utility from "not buying", $y = 0$, to 0. The random utility from choosing an alternative $y \neq 0$ is of the form

$$\mathbf{u}_y = (\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1)\sum_{j=1}^{J} y_j \mathbf{z}_{2,j} + \mathbf{g}_y. \tag{4}$$

Although the model (4) looks similar to the model (3), there is one important difference: there is no bundle specific covariate since $z_{2,j}$ affects not only the utility from buying good j alone, but also every bundle that includes it.

Proposition 5.2 *Suppose Assumptions 5 and 6 hold. Then*

(i) $\beta_0(\cdot)$ *and* $\beta_1(\cdot)$ *are identified;*

(ii) The above model inherits all identifying properties of the following bundles model:

$$\mathbf{u}_y = \sum_{j=1}^{J} y_j \mathbf{r}_j + \mathbf{g}_y \quad ,$$
$$\mathbf{u}_0 = 0.$$

where $\mathbf{r} = (\mathbf{r}_j)_{j=1,\ldots,J}$ *is an observed covariate independent of* $\mathbf{g} = (\mathbf{g}_y)_{y \in Y \setminus \{0\}}$ *conditional on* $\mathbf{w}$ *with the conditional support*

$$R_w = \left\{ r \in \mathbb{R}^J : r = \lambda z_2, \lambda \in \mathbb{R},\ z_2 \in Z_{2|w} \right\}.$$

Note that for $J = 2$ if one assumes that

$$\mathbf{g}_{(1,0)} = f_1(\mathbf{w}) + \boldsymbol{\epsilon}_1,$$

$$\mathbf{g}_{(0,1)} = f_2(\mathbf{w}) + \boldsymbol{\epsilon}_2,$$
$$\mathbf{g}_{(1,1)} = \mathbf{g}_{(1,0)} + \mathbf{g}_{(0,1)} + \xi f_3(\mathbf{w}),$$

where $f_i(\cdot)$, $i = 1, 2, 3$, are some unknown functions, and $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\xi})^\intercal \in \mathbb{R}^2 \times \mathbb{R}_+$, then model (3) is equivalent to the one in Fox and Lazzati (2017), and one can apply their Theorem 1 to identify $f_i(\cdot)$, $i = 1, 2, 3$, and the distributions of $\boldsymbol{\epsilon}_i|\mathbf{w}$, $i = 1, 2$, and $\boldsymbol{\xi}|\mathbf{w}$.

5.3. Binary games of complete information

In the multinomial choice and the bundles models I am able to establish identification of the objects of interest without requiring covariates with full support. The example that is considered in this section is different: the model is not pointidentified. However, the Lebesgue measure of the identified set is zero. In particular, all parameters of the model are identified up to a finite-dimensional parameter of lower-dimension.

There are $\|I\| < \infty$ players indexed by $i \in I$. Every player must choose $y_i \in \{0, 1\}$. Thus, the outcome space is $Y = \{0, 1\}^{\|I\|}$.[16] Players i's payoff from choosing action y_i when the other agents are choosing y_{-i} is given by

$$\pi_{0i}(y) = \left[\alpha_{0,i}(\mathbf{w}) + [\beta_{0,i}(\mathbf{w})\mathbf{z}_i + \mathbf{e}_i] + \sum_{j \in I \setminus \{i\}} \delta_{0,i,j}(\mathbf{w}) y_j\right] y_i,$$

where $\mathbf{e}_i$, $i \in I$, are observed by players but unobserved by the econometrician shocks; $\alpha_{0,i}(\cdot)$, $\beta_{0,i}(\cdot)$ and $\delta_{0,i,j}(\cdot)$ are unknown functions. The econometrician observes a joint distribution of $(\mathbf{y}, \mathbf{x}^\intercal)^\intercal$, where $\mathbf{x} = (\mathbf{z}^\intercal, \mathbf{w}^\intercal)^\intercal \in X$ with $\mathbf{z} = (\mathbf{z}_i)_{i \in I}$, is a vector of observed covariates. Let $\beta_0(\cdot) = (\beta_{0,i}(\cdot))_{i \in I}$, $\alpha_0(\cdot) = (\alpha_{0,i}(\cdot))_{i \in I}$, and $\delta_0(\cdot) = (\delta_{0,i,j}(\cdot))_{i \neq j \in I}$.

The following two assumptions are sufficient for Assumptions 1 and 2.

Assumption 7 (i) Assumption 1 is satisfied with $\mathbf{v} = (\beta_{0,i}(\mathbf{w})\mathbf{z}_i + \mathbf{e}_i)_{i \in I}$;

[16] I work with binary action spaces for ease of exposition. The result can be extended to multiple players and actions games.

(ii) For every $i, j \in I$, $i \neq j$, the cardinality of

$$\left\{ \Big(0\Big)_{k \in I},\ \Big(\mathbb{1}\,(\,k \in \{i, j\}\,)\Big)_{k \in I},\ \Big(\mathbb{1}\,(\,k = i\,)\Big)_{k \in I},\ \Big(\mathbb{1}\,(\,k = j\,)\Big)_{k \in I} \right\} \bigcap Y^*$$

is at least 2.

Assumption 7(i) implies that excluded covariates affect the distribution of some outcomes via payoffs only. Assumption 7(ii) imposes restrictions on the set of those outcomes. If one thinks of an entry game where $y_i = 1$ corresponds to the entry decision, the outcomes in Assumption 7(ii) have the following interpretation. The outcome $\Big(0\Big)_{k \in I}$ corresponds to the market where nobody enters. The outcome $\Big(\mathbb{1}\,(\,k \in \{i, j\}\,)\Big)_{k \in I}$ corresponds to the market where only players i and j enter. Similarly, $\Big(\mathbb{1}\,(\,k = i\,)\Big)_{k \in I}$ means that only players i enters. Note that although the cardinality of Y is $2^{\|I\|}$, the cardinality of Y^* can be as low as $\|I\| + 1$.

Assumption 8 (i) The shocks $\{\mathbf{e}_i\}_{i \in I}$ are i.i.d. standard normal random variables and independent of $\mathbf{x}$;

(ii) $\beta_{0,i}(\mathbf{w}) \neq 0$ a.s. for all $i \in I$;

(iii) For every $w \in W$, the support of $\mathbf{z}|\mathbf{w} = w$, Z_w, contains an open ball.

Assumption 8 is a parametric restriction on the distribution of payoffs. It is satisfied by the parametrization used in Bajari et al. (2010) and Ciliberto and Tamer (2009).

I consider three complete information equilibrium/solution concepts: (i) minimax play, (ii) "collusive" behavior, and (iii) rationalizability. Under minimax solution concept every player picks the action that maximizes her minimum possible payoff. Under "collusive" behavior players play the outcome that maximizes the sum of individual profits. Under rationalizability players play any action that survives iterated strict dominance elimination. The most commonly used solution concept of Nash equilibrium is nested within rationalizability. Hence, all identification results that are derived under rationalizability are valid for Nash equilibrium (both in pure and mixed strategies).

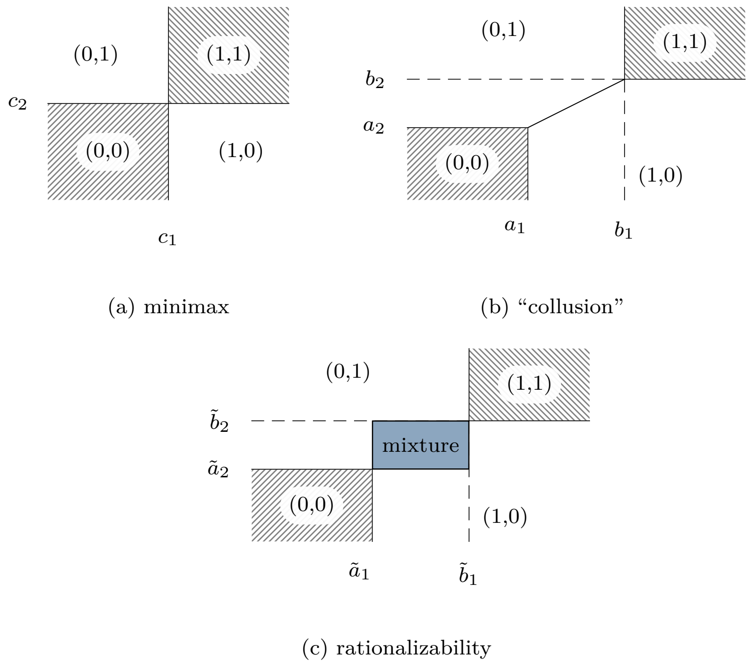

Figure 1 – Predictions for different realizations of v when $\delta_{0,1,2} < 0$ and $\delta_{0,2,1} < 0$ in a two player game.

These three solution concepts have different implications for identification. For instance, in the two player case, if players display "collusive" behavior, then there is no hope of separate identification of "competition" effects $\delta_{0,i,j}$. At best, one can identify $\delta_{0,1,2} + \delta_{0,2,1}$. Fortunately, it is possible to distinguish between these three models up to $\beta_0(\cdot)$ since under Assumptions 7 and 8 I am able to nonparametrically recover h_0 up $\beta_0(\cdot)$.

Figure 1 illustrates predictions in a two player binary game for different realizations of v when $\delta_{0,1,2} < 0$ and $\delta_{0,2,1} < 0$. The thresholds a_i, b_i, c_i, $i = 1, 2$, are determined by $\alpha_{0,i}$ and $\delta_{0,i,j}$, $i, j = 1, 2$. If one can identify h_0 for two outcomes, say $(0, 0)$ and $(1, 1)$, then one can distinguish between all three equilibrium concepts as long as the following three conditions are satisfied.

$$(a_1, a_2)^\intercal \neq (b_1, b_2)^\intercal \quad \text{(minimax vs. collusion)}, \tag{5}$$

$$(\tilde{a}_1, \tilde{a}_2)^\intercal \neq (\tilde{b}_1, \tilde{b}_2)^\intercal \quad \text{(minimax vs. rationalizability)}, \tag{6}$$

$$\frac{\tilde{b}_2 - \tilde{a}_2}{\tilde{b}_1 - \tilde{a}_1} \neq \frac{b_2 - a_2}{b_1 - a_1} \quad \text{(rationalizability vs. collusion).} \tag{7}$$

Conditions (5)-(7) hold if $\delta^2_{0,1,2}(w) \neq \delta^2_{0,2,1}(w)$ for some $w \in W$. Thus, if I can identify the correct solution concept, then I can recover the threshold values. For instance, if I know that agents play rationalizable strategies, then I can identify $\tilde{a}_i, \tilde{b}_i$, $i = 1, 2$. Since $\tilde{a}_i = -\alpha_{0,i}$, and $\tilde{b}_i = -\alpha_{0,i} - \delta_{0,i,j}$, $i, j = 1, 2$, I also can identify α_0 and δ_0.

The above intuition generalizes to games with more than two players, more than two actions, and without any sign restrictions on $\delta_{0,i,j}$. The following result establishes identification in binary games.

Proposition 5.3 *Under Assumptions 7 and 8*

(i) If there exist $i, j \in I$, $i \neq j$, such that $\Pr(\delta^2_{0,i,j}(\mathbf{w}) \neq \delta^2_{0,j,i}(\mathbf{w})) > 0$, then one can determine whether players behave according to minimax play, "collusive" behavior, or rationalizability up to $\beta_0(\cdot)$;

(ii) If players behave according to rationalizability, then $\alpha_0(\cdot)$ and $\delta_0(\cdot)$ are identified up to $\beta_0(\cdot)$;

(iii) If players behave according to "collusive" behavior, then $\alpha_0(\cdot)$ and $\{\delta_{0,i,j}(\cdot) + \delta_{0,j,i}(\cdot)\}_{i \neq j \in I}$ are identified up to $\beta_0(\cdot)$;

(iv) If players behave according to minimax play, then $\{\alpha_{0,i}(\cdot) + \min\{\delta_{0,i,j}(\cdot), 0\}\}_{i \neq j \in I}$ are identified up to $\beta_0(\cdot)$.

Proposition 5.3 states that if marginal effects of some excluded covariates on payoffs are known, then one can identify the solution concept, together with some (or sometimes all) payoff parameters. Assumption 8(i) implies that the errors in payoffs are not correlated. I can allow for unknown variance covariance matrix V. In this case all objects in Proposition 5.3 are identified up to $\beta_0(\cdot)$ and V.

Full identification in this binary game can be achieved if one can identify $\beta_0(\cdot)$. The standard identification-at-infinity argument requires existence of player-action-specific covariates that have full support (unbounded from above *and* below). However, the full support assumption is only needed to separately identify the intercepts of the mean utilities (α_0 and δ_0). In contrast, in order to identify β_0 one only needs to have player-action-specific covariates with unbounded support (e.g., unbounded from above only).[17] In other words, full identification can be achieved in a substantially bigger set of applications (e.g., prices or income can potentially take arbitrary large positive values, but cannot be negative).[18]

6. Conclusion

This paper shows that commonly used exclusion restrictions and parametric assumptions about the distribution of some unobservables may lead to identification in discrete outcome models. The proposed identification framework allows one to extend the results from a large literature that uses special covariates with full support to environments where such full-support covariates are not available.

The partial identification result can substantially decrease computational complexity of constructing confidence sets for partially identified parameters. For instance, the likelihood ratio statistic of Chen et al. (2011) is asymptotically χ^2 distributed after profiling β_0 under the null hypothesis, since the model in this case is identified. Thus, there is no need to use bootstrap and one can take critical values from χ^2 distribution.

References

Andrews, D. W. K. (2011). Examples of L2-complete and boundedly-complete distributions. Discussion Paper 1801, Cowles Foundation.

[17] See, for instance, Tamer (2003), Bajari et al. (2010), and Kashaev and Salcedo (2017).

[18] One may argue that it is always possible to transform any covariate into a covariate with full support. For instance, one can always treat logarithm of income or price as a covariate. Thus, when price goes to 0 the logarithm of price goes to $-\infty$. However, in order to interpret linear parametrization of a payoff function in this case, one would need to explain why prices that are close to zero would lead to extremely negative (or positive) profits.

Bajari, P., Hong, H., and Ryan, S. P. (2010). Identification and estimation of a discrete game of complete information. *Econometrica*, 78(5):1529–1568.

Berry, S. T. and Haile, P. A. (2009). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Technical report, National Bureau of Economic Research.

Blundell, R., Chen, X., and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669.

Canay, I. A., Santos, A., and Shaikh, A. M. (2013). On the testability of identification in some nonparametric models with endogeneity. *Econometrica*, 81(6):2535–2559.

Chen, X., Tamer, E., and Torgovitsky, A. (2011). Sensitivity analysis in semiparametric likelihood models. *Yale University and Northwestern University*.

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

Chernozhukov, V., Imbens, G. W., and Newey, W. K. (2007). Instrumental variable estimation of nonseparable models. *Journal of Econometrics*, 139(1):4–14.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828.

Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565.

d'Haultfoeuille, X. (2011). On the completeness condition in nonparametric instrumental problems. *Econometric Theory*, 27(3):460–471.

Dunker, F., Hoderlein, S., Kaido, H., et al. (2017). Nonparametric identification of random coefficients in endogenous and heterogeneous aggregate demand models. Technical report, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Fox, J. T. and Gandhi, A. (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. *The RAND Journal of Economics*, 47(1):118–139.

Fox, J. T. and Lazzati, N. (2017). A note on identification of discrete choice models for bundles and binary games. *Quantitative Economics*, 8(3):1021–1036.

Fox, J. T., Yang, C., and Hsu, D. H. (2018). Unobserved heterogeneity in matching games. *Journal of Political Economy*, 126(4):1339–1373.

Gautier, E. and Kitamura, Y. (2013). Nonparametric estimation in random coefficients binary choice models. *Econometrica*, 81(2):581–607.

Gentzkow, M. (2007). Valuing new goods in a model with complementarity: Online newspapers. *American Economic Review*, 97(3):713–744.

Heckman, J. (1990). Varieties of selection bias. *The American Economic Review*, 80(2):313–318.

Hu, Y. and Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216.

Ichimura, H. and Thompson, T. S. (1998). Maximum likelihood estimation of a binary choice model with random coefficients of unknown distribution. *Journal of Econometrics*, 86(2):269–295.

Kashaev, N. and Salcedo, B. (2017). Identification of solution concepts for discrete games. Working paper.

Lewbel, A. (1998). Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica*, pages 105–121.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, 97(1):145–177.

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of econometrics*, 27(3):313–333.

Manski, C. F. (1988). Identification of binary response models. *Journal of the American statistical Association*, 83(403):729–738.

Mattner, L. et al. (1993). Some incomplete but boundedly complete location families. *The Annals of Statistics*, 21(4):2158–2162.

Matzkin, R. L. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica: Journal of the Econometric Society*, pages 239–270.

Matzkin, R. L. (2007). Heterogeneous choice. *Econometric Society Monographs*, 43:75.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, 70(1):147–165.

Thompson, T. S. (1989). Identification of semiparametric discrete choice models.

A. Proofs

A.1. Proof of Proposition 3.1

Fix some $y \in Y^*$ and $w \in W$ (for brevity I will drop w in the notation below). Under Assumption 1 I have the following integral equation

$$\forall z \in Z \ : \ \mu(y|z) = \int_V h(y^*, v) dF_{v|z}(v|z).$$

Suppose that there exists h with $h(y^*, v) \neq h_0(y^*, v)$ for all v in some nonzero measure set V' such that

$$\forall z \in Z \ : \ \mu(y|z) = \int_V h(y^*, v) dF_{v|z}(v|z) = \int_V h_0(y^*, v) dF_{v|z}(v|z).$$

Which implies that the nonzero function $h(y,\cdot)-h_0(y,\cdot)$ integrates to 0 for all $z \in Z'$. The latter contradicts to Assumption 2. The fact that the choice of y and w was arbitrary completes the proof.

A.2. Proof of Proposition 4.1

Note that Proposition 3.1 implies that h_0 is identified up to $\{\beta_{0,i}(\cdot), \beta_{1,i}(\cdot)\}_{i=1}^{d_v}$. Hence, I only need to show that $\{\beta_{0,i}(\cdot), \beta_{1,i}(\cdot)\}_{i=1}^{d_v}$ is identified.

Fix some w, $i \in \{1, 2, \ldots, d_v\}$, $z_{2,-i}$, $z_{1,-i}$, and take y^* from Assumption 4. To simplify notation let $F_0 : \mathbb{R} \to \mathbb{R}$ and $\eta : \mathbb{R}^2 \to \mathbb{R}$ such that

$$F_0(v_i) = \int_{\mathbb{R}^{d_v-1}} h_0(y^*, w, v) \prod_{k \neq i} \frac{\phi\left(v_k/z_{2,k} - \beta_{0,k}(w) - \beta_{1,k}(w) z_{1,k}\right)}{z_{2,k}} dv_k,$$

where $\phi(\cdot)$ is the standard normal p.d.f., and $\eta(z_{1,i}, z_{2,i}) = \mu(y^*|z, w)$.

Assumptions 1 and 3 imply that

$$\eta(z_{1,i}, z_{2,i}) = \int_{\mathbb{R}} F_0(v_i) \frac{\phi(v_i/z_{2,i} - \beta_{0,i} - \beta_{1,i} z_{1,i})}{z_{2,i}} dv_i,$$

After dropping index i from the notation and some rearrangements I get

$$\tilde{\eta}(z_1, z_2) = \int_{\mathbb{R}} F_0(t) \phi(t/z_2 - \beta_0 - \beta_1 z_1) dt, \tag{8}$$

where $\tilde{\eta}(z_1, z_2) = z_2 \eta(z_1, z_2)$.

Next, note that since $\phi''(x) = -\phi(x) - x\phi'(x)$ the following system of equations holds

$$\begin{aligned}
\partial_{z_1} \tilde{\eta}(z_1, z_2) &= -\beta_1 \int F_0(t) \phi'(t/z_2 - \beta_0 - \beta_1 z_1) dt, \\
\partial^2_{z_1^2} \tilde{\eta}(z_1, z_2) &= \beta_1^2 \int F_0(t) \phi''(t/z_2 - \beta_0 - \beta_1 z_1) dt = -\beta_1^2 \tilde{\eta}(z_1, z_2) - \beta_1(\beta_0 + \beta_1 z_1) \partial_{z_1} \tilde{\eta}(z_1, z_2) - \\
&\quad - \beta_1^2 \int t F_0(t) \phi'(t/z_2 - \beta_0 - \beta_1 z_1) dt / z_2;
\end{aligned}$$

Moreover,

$$\partial_{z_2}\tilde{\eta}(z_1, z_2) = -\int F_0(t)t\phi'(t/z_2 - \beta_0 - \beta_1 z_1)dt/z_2^2.$$

Hence,

$$\partial^2_{z_1^2}\tilde{\eta}(z_1, z_2) = -\beta_1^2\tilde{\eta}(z_1, z_2) - \beta_1(\beta_0 + \beta_1 z_1)\partial_{z_1}\tilde{\eta}(z_1, z_2) + \beta_1^2 z_2\partial_{z_2}\tilde{\eta}(z_1, z_2);$$

Equivalently

$$\frac{\beta_0}{\beta_1} = \frac{z_2\partial_{z_2}\tilde{\eta}(z_1, z_2) - \tilde{\eta}(z_1, z_2)}{\partial_{z_1}\tilde{\eta}(z_1, z_2)} - z_1 - \frac{\partial^2_{z^2}\tilde{\eta}(z_1, z_2)}{\partial_{z_1}\tilde{\eta}(z_1, z_2)}\frac{1}{\beta_1^2}.$$

Thus, β_0/β_1 is identified up to β_1^2. Moreover, the last equality implies that β_1^2 is identified if for some z_2 and z_1

$$\partial_{z_1}\left(\frac{\partial_{z_1^2}\tilde{\eta}(z_1, z_2)}{\partial_{z_1}\tilde{\eta}(z_1, z_2)}\right) \neq 0.$$

Suppose this is not the case. That is, for all z_2 and z_1

$$\partial^2_{z_1^2}\left(\log(\partial_{z_1}\tilde{\eta}(z_1, z_2))\right) = 0.$$

The latter would imply that either

$$\tilde{\eta}(z_1, z_2) = K_1(z_2)e^{K_2(z_2)z_1} + K_3(z_2)$$

or

$$\tilde{\eta}(c, z) = K_4(z_2)z_1 + K_3(z_2)$$

for some functions $K_i(\cdot)$, $i = 1, 2, 3, 4$. Since it is assumed that $\tilde{\eta}(\cdot, z_2) = z_2\eta(\cdot, z_2)$ is neither exponential nor affine function, I can conclude that β_1^2 is identified (hence, $|\beta_1|$ is also identified). Hence, I identify β_0/β_1. If $\beta_0/\beta_1 = 0$, then the sign of β_1 is identified from Assumption 3(iv). If $\beta_0/\beta_1 \neq 0$, then the sign of either β_1 or β_0 is identified from Assumption 3(iv). Knowing the sign of, say, β_0 and β_0/β_1 identifies β_1 and β_0.

A.3. Proof of Propositions 5.1 and 5.2

(i). Note that under Assumption 6.(ii) there exists z_2 with some open neighbourhood such that $z_{2,y} = z_{2,y'}$ for all $y, y' \in Y$. Let

$$\mathbf{v}_1 = -\mathbf{z}_{2,1}(\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1) \text{ a.s..}$$

Assumption 5 implies that Assumptions 3(i)-(iii) are satisfied for the above $\mathbf{v}_1$. Moreover,

$$\Pr(\mathbf{y} = 0|\mathbf{x} = x) = \int_{\mathbb{R}} F_{\mathbf{g}|\mathbf{w}}(-z_{2,1}(\beta_0(w)+\beta_1(w)z_1+e_1), \dots, -z_{2,1}(\beta_0(w)+\beta_1(w)z_1+e_1)|w)\phi(e_1)de_1$$

identifies the sign of $\beta_1(w)$ since $F_{\mathbf{g}|\mathbf{w}}(\cdot|w)$ is weakly monotone. Thus, Assumption 3(iv) is also satisfied.

Assumption 1 is satisfied for $Y^* = \{0\}$ and for $h(0, w, v) = F_{\mathbf{g}|\mathbf{w}}(v, v, \cdot, v|w)$. Assumption 4 is implied by Assumption 6. Hence, by Proposition 3.1 $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are identified.

(ii). Since $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are identified I can redefine the index v. Let

$$\mathbf{v} = \beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1 \text{ a.s..}$$

Note that

$$\begin{aligned}\Pr(\mathbf{y} = 0|\mathbf{x} = x) &= \int_{\mathbb{R}} F_{\mathbf{g}|\mathbf{w}}(-z_{2,1}v, \dots, -z_{2,J}v|w)\phi(v - \beta_0(w) + \beta_1(w)z_1)dv = \\ &= \int_{\mathbb{R}} h(z_2, w, v)\phi(v - \beta_0(w) + \beta_1(w)z_2)dv\end{aligned}$$

Since assumption of Proposition 3.1 are satisfied I pointidentify

$$h(z_2, w, v) = F_{\mathbf{g}|\mathbf{w}}(-z_{2,1}v, \dots, -z_{2,Y}v|w)$$

for all z_2, w, v. Note that since v can take any value in $\mathbb{R}$ for any direction $-z_2/\|z_2\|$ in the support of z_2 I can recover $F_{\mathbf{g}|\mathbf{w}}(g|w)$ for any g such that $g = -z_2 v/\|z_2\|$ for some $v \in \mathbb{R}$. That is, I identify $F_{\mathbf{g}|\mathbf{w}}(\cdot|w)$ over the set

$$R_w = \left\{ r \in \mathbb{R}^J \,:\, r = \lambda z_2,\ \lambda \in \mathbb{R},\ z_2 \in Z_{2|w} \right\}.$$

A.4. Proof of Proposition 5.3

First, I fix some $w \in W$ and for notation simplicity I drop dependence on w. Since assumptions of Proposition 3.1 are satisfied, I identify

$$\Pr(\mathbf{y} = y|\mathbf{v} = \cdot) = h_0(y, \cdot)$$

for all $y \in Y^*$. Since $v \in \mathbb{R}^{\|I\|}$ and v_i enters only payoffs of player i, I can make a payoff of any player arbitrary small ("close" to $-\infty$). Hence, under all solution concepts under consideration I can force any player to choose $y_i = 0$. Take any two players $i \neq j$ and consider

$$h_{0,i,j}(y, v_i, v_j) = \lim_{v_k \to -\infty, k \in I \setminus \{i,j\}} h_0(y, v)$$

for all $y \in Y^*$ and v_i, v_j. Note that $h_{0,i,j}$ corresponds to a two player binary game with payoffs

$$\left[\alpha_{0,i} + \delta_{0,i,j} y_j + v_i\right] y_i$$

and

$$\left[\alpha_{0,j} + \delta_{0,j,i} y_i + v_j\right] y_j.$$

Moreover, by Assumption 7(ii), in this two player game at least two outcomes from

$$\{(0,0), (1,0), (0,1), (1,1)\}$$

satisfy exclusion restrictions. Assume that $(0,0)$ and $(1,1)$ satisfy the exclusion restriction (the proof for any other case, e.g., $(0,0)$ and $(1,0)$, is almost the same).

To discriminate between solution concepts I analyze their predictions about outcomes $(0,0)$ and $(1,1)$. Without loss of generality let $i = 1$ and $j = 2$.

Case 1. Suppose that agents are behaving according to minimax play. Then

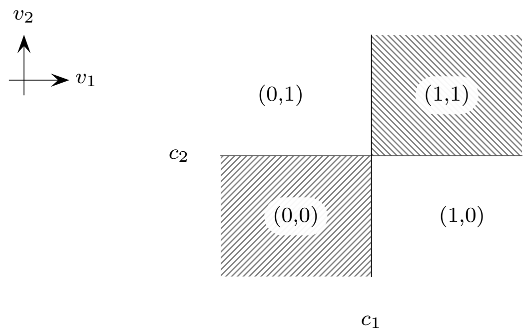

Figure 2 – Minimax correspondences for different values of v.

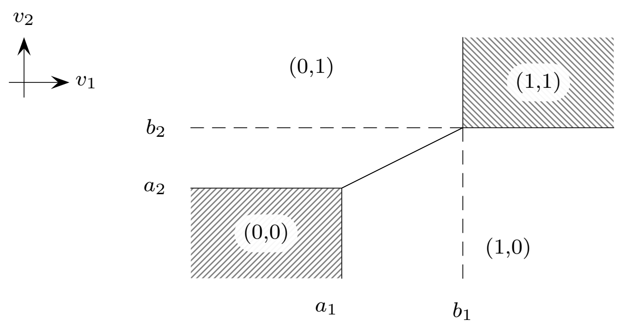

Figure 3 – "Collusive" correspondences for different realizations of v when $\delta_{0,1,2} + \delta_{0,2,1} < 0$.

for any $\alpha_{0,1}$, $\alpha_{0,2}$, $\delta_{0,1,2}$, and $\delta_{0,2,1}$, the predictions about outcomes $(0,0)$ and $(1,1)$ depending on the value of $(v_1, v_2)^\intercal \in \mathbb{R}^2$ can be depicted as in Figure 2, where $c_1 = -\alpha_{0,1} - \min\{\delta_{0,1,2}, 0\}$ and $c_2 = -\alpha_{0,2} - \min\{\delta_{0,2,1}, 0\}$

Case 2. Suppose that agents are behaving according to "collusive" solution concept. Then for any $\alpha_{0,1}$, $\alpha_{0,2}$, $\delta_{0,1,2}$, and $\delta_{0,2,1}$, the predictions about outcomes $(0,0)$ and $(1,1)$ depending on the value of $v \in \mathbb{R}^2$ can be depicted as in figures 3-4, where $a_1 = -\alpha_{0,1}$, $a_2 = -\alpha_{0,2}$, $b_1 = -\alpha_{0,1} - (\delta_{0,1,2} + \delta_{0,2,1})$, and $b_2 = -\alpha_{0,2} - (\delta_{0,1,2} + \delta_{0,2,1})$.

Case 3. Suppose that agents are playing rationalizable strategies. Then for any $\alpha_{0,1}$, $\alpha_{0,2}$, $\delta_{0,1,2}$, and $\delta_{0,2,1}$, the predictions about outcomes $(0,0)$ and $(1,1)$ depending on the value of $v \in \mathbb{R}^2$ can be depicted as in figures 5-8, where $a_1 = -\alpha_{0,1}$, $a_2 = -\alpha_{0,2}$,

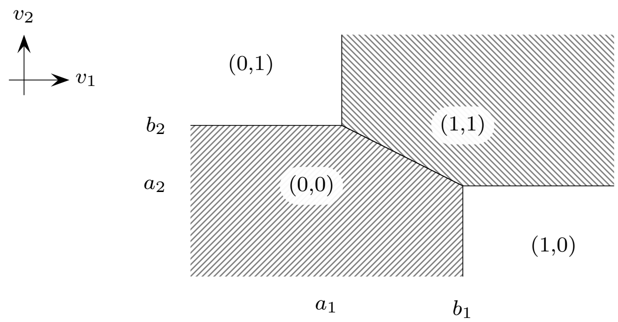

Figure 4 – "Collusive" correspondences for different realizations of v when $\delta_{0,1,2} + \delta_{0,2,1} > 0$.

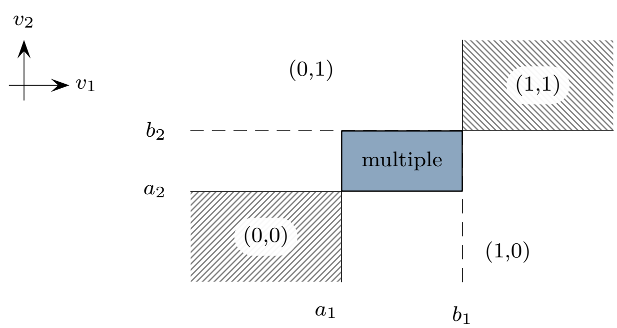

Figure 5 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} < 0$ and $\delta_{0,2,1} < 0$.

$b_1 = -\alpha_{0,1} - \delta_{0,1,2}$, and $b_2 = -\alpha_{0,2} - \delta_{0,2,1}$.

As a result, minimax is consistent with the data if and only if $h_{0,i,j}((0,0),\cdot)$ and $h_{0,i,j}((1,1),\cdot)$ match Figure 2. If $h_{0,i,j}((0,0),\cdot)$ and $h_{0,i,j}((1,1),\cdot)$ match Figure 7 or Figure 8, then only rationalizability can explain the data. There is still a possibility that $h_{0,i,j}((0,0),\cdot)$ and $h_{0,i,j}((1,1),\cdot)$ match Figure 3 and Figure 5, or Figure 4 and Figure 6. Hence, one might think that rationalizability and "collusive" behavior both can explain the data. However, since in Proposition 5.3(i) I assume that there exist two players such that $\delta^2_{0,i,j} \neq \delta^2_{0,j,i}$, the "multiplicity" region under rationalizability is never a square (the lengths of the region are $\|\delta_{0,1,2}\|$ and $\|\delta_{0,2,1}\|$). In contrast, under "collusive" behavior the "multiplicity" region is always a square (the lengths

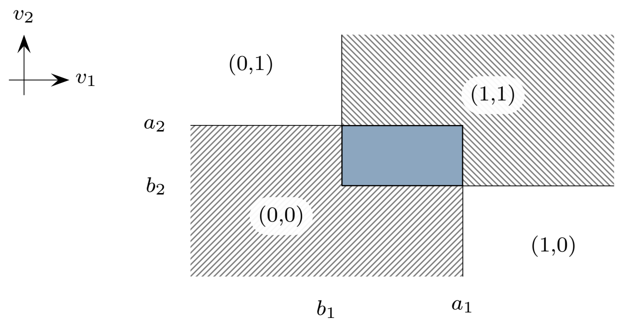

Figure 6 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} > 0$ and $\delta_{0,2,1} > 0$.

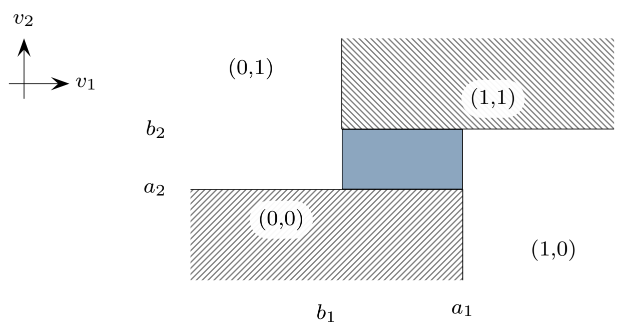

Figure 7 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} > 0$ and $\delta_{0,2,1} < 0$.

of the region are $\|\delta_{0,1,2} + \delta_{0,2,1}\|$ and $\|\delta_{0,1,2} + \delta_{0,2,1}\|$). Thus by analyzing these two "asymmetric" players I can determine the correct solution concept.

If I determined that the correct solution concept is minimax, then at most I can identify $\alpha_{0,1} + \min\{\delta_{0,1}, 0\}$ and $\alpha_{0,2} + \min\{\delta_{0,2}, 0\}$. If the correct solution concept is "collusive" behavior, then I can pointidentify $\alpha_{0,1}$, $\alpha_{0,2}$, and $\delta_{0,1,2} + \delta_{0,2,2}$. If agents play rationalizable strategies, then I can pointidentify all payoff parameters.

The result then follows from the fact that the choice of players i, j and $w \in W$ was arbitrary.

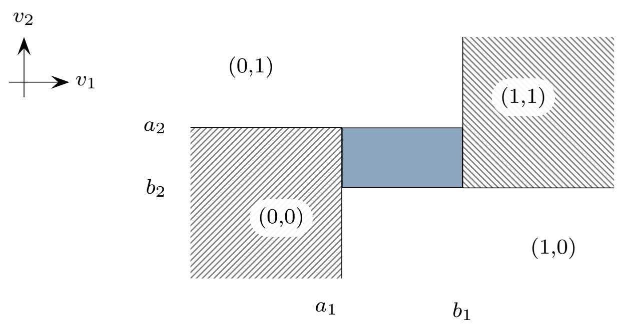

Figure 8 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} < 0$ and $\delta_{0,2,1} > 0$.