Identification and estimation of discrete outcome models with latent special covariates*

Nail Kashaev †

nkashaev@uwo.ca

First version: November 13, 2018

This version: September 10, 2020

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 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. Using the proposed constructive identification technique, I also provide an asymptotically normal estimator of the finite-dimensional parameters of the model.

JEL classification numbers: C50, C57

Keywords: Discrete outcome models, multinomial choice, random coefficients, special covariate, games of complete information, identification at infinity, bundles

*I would like to thank Roy Allen, Victor H. Aguiar, Tim Conley, Nirav Mehta, Salvador Navarro, David Rivers, and Bruno Salcedo for useful comments and discussions.

†Department of Economics, University of Western Ontario.

1. Introduction

This paper studies identification and estimation of discrete outcome models with covariates that have bounded support. Often *some* latent variables in these models have full support (i.e. supported on the whole Euclidean space). Under common restrictions on the distribution of these unobservables, I constructively identify it and show how these latent variables can be used to construct special covariates (i.e., observables with full support) to nonparametrically (partially)identify the distribution of all the other unobservables. I apply the proposed method to three well-known models: multinomial choice models with random coefficients, bundles models, and finite games of complete information. My identification technique is constructive and leads to an asymptotically normal estimator of the finite-dimensional parameters of the model.

The results of this paper rest on two commonly used assumptions. First, I assume existence of excluded covariates that affect the distribution over outcomes via a latent index. Using variation in these excluded covariates I can identify the distribution of the index. Second, I assume that the distribution of the index is sufficiently "rich". As a result, I show how to identify the distribution over outcomes conditional on the realization of the observed covariates and the latent index nonparametrically. Since the latent index often has full support, I can treat the latent index as an observed covariate with full support and apply *any* identification technique that requires existence of such covariates to identify the rest of the model parameters (e.g., the distribution of other latent variables).

The latent index has different interpretations in different settings. For instance, in the random coefficients model, one of the random coefficients can be treated as the latent index. In analysis of finite games (e.g., entry 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.[1]

[1] Completeness of a family of distributions is a well-known concept 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. (2007b), Hu and Schennach (2008), Andrews (2011),

I provide two nonnested identification results. The first result uses one of the most popular parametrizations in applied work - Gaussian distribution of the latent index. The second result does not make any parametric assumptions about the distribution of latent variables. It, however, imposes more restrictions on the support of observables.

I contribute to the discrete outcome literature in several respects. I show how existing results that use full-support-excluded covariates with monotonicity restrictions[2] can be directly used in environments with bounded covariates. Formally, I demonstrate that my setting inherits all identifying properties of the setting with special covariate. I also contribute to the literature on semiparametric models by showing that common parametric restrictions can be used instead of covariates that have full support. This paper is also related to the literature on identification of finite-dimensional parameters in discrete outcome models with bounded covariates.[3] The main difference from that literature is that in my framework the distribution of latent variables (e.g., the random intercept) can be nonparametrically identified even if these latent variables have full support, but covariates are bounded.

My approach is complementary to existing methods. In situations where the researcher is not sure whether covariates have full support and is willing to impose mild restrictions because of tractability or data limitations, my approach can provide an additional reassurance of identification. Also, the results in this paper provide a more solid econometric foundation to the models with at least one normally distributed random coefficient.[4]

This paper also contributes to the literature on partially identified models. One of the common methods to construct confidence intervals in these models is test inversion.[5] However, constructing confidence intervals by test inversion requires checking a

Darolles et al. (2011), and d'Haultfoeuille (2011).

[2]See, for example, Manski (1985, 1988), Heckman (1990), Matzkin (1992), Ichimura and Thompson (1998), Lewbel (1998, 2000), Tamer (2003), Matzkin (2007), Berry and Haile (2009), Bajari et al. (2010), Gautier and Kitamura (2013), Gautier and Hoderlein (2015), Fox and Gandhi (2016), Dunker et al. (2017), Fox and Lazzati (2017), Fox et al. (2018), and Fox (2020)

[3]E.g., Magnac and Maurin (2007), Chen et al. (2016), and Kline (2016).

[4]See Fox et al. (2012) for the treatment of the random coefficients multinomial logit model.

[5]See, for instance, Chernozhukov et al. (2007a) and Andrews and Soares (2010)

large number of points in the parameter space and obtaining critical values for each of these points either by bootstrap, subsampling, or simulation methods. In most of the applications, the problem becomes computationally intractable. My general identification result, which requires a preliminary identification of a finite-dimensional parameter, implies that in many partially identified models (e.g., finite games) 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 much lower dimension.[6] This finding can lead to substantial computational gains in constructing confidence sets for partially identified parameters in moment inequalities or likelihood models. After conditioning on this finite-dimensional parameter of a lower dimension, the model becomes pointidentified and the profiled objective function (e.g., the log-likelihood function) has a unique global optimum. Thus, the researcher can potentially use critical values from the Gaussian or the chi-squared distribution instead of using bootstrap or simulations.[7]

The paper is organized as follows. Section 2 provides a motivating example. In Section 3, I describe the setting. Section 4.1 specializes the results from Section 3 for widely used normally distributed latent variables. Section 4.2 provides a fully nonparametric identification result. In Sections 5, I apply the results from Section 4 to three different discrete outcome models. Section 6 shows how to estimate the finite-dimensional parameters of the model and provides an empirical illustration. Section 7 concludes. All proofs can be found in Appendix A.

[6] For example, in a two-player entry game with *many* covariates the model is identified up to a three-dimensional parameter.

[7] I leave the formal treatment of this problem for future work.

2. Motivating Example: Binary Choice

Consider a simple single-agent binary choice problem.[8] A utility-maximizing agent has to decide whether she is buying a product ($y = 1$) or not buying it ($y = 0$). The utility of alternative $y = 0$ is normalized to 0. The utility of option $y = 1$ is

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

where $\mathbf{z}_{2,1} \in Z_{2,1} \subseteq \mathbb{R}$ is a product-specific observed bounded characteristic that is different from zero with positive probability; $\beta_0 + \beta_1 \mathbf{z}_1 + \mathbf{e}_1$ is an individual specific random coefficient that captures agent's taste for $\mathbf{z}_{2,1}$; $\mathbf{z}_1$ is an individual-specific observed bounded taste shifter; $\mathbf{e}_1$ and $\mathbf{g}_1$ are supported on $\mathbb{R}$ unobserved taste and utility shocks that are independent from each other and from $\boldsymbol{z} = (\mathbf{z}_1, \mathbf{z}_{2,1})^\mathsf{T}$. The objective of the econometrician is to recover the c.d.f. of $\mathbf{g}_1$, $F_{\mathbf{g}_1}$, and the parameters β_0 and β_1 from the observed distribution of choices $\mathbf{y}$ and covariates $\mathbf{z}$.

Let $\mathbf{v}_1 = \mathbf{z}_{2,1}(\beta_0 + \beta_1 \mathbf{z}_1 + \mathbf{e}_1)$. Note that

$$\Pr(\mathbf{y} = 0 | \mathbf{v}_1 = v_1) = F_{\mathbf{g}_1}(-v_1).$$

Hence, even if $\mathbf{z}$ has a bounded support, since $\mathbf{v}_1$ is supported on $\mathbb{R}$, we can identify $F_{\mathbf{g}_1}(g_1)$ for all $g_1 \in \mathbb{R}$ if we can identify $\Pr(\mathbf{y} = 0 | \mathbf{v}_1 = v_1)$ for all $v_1 \in \mathbb{R}$. In other words, if we can identify $\Pr(\mathbf{y} = 0 | \mathbf{v}_1 = \cdot)$ the above binary choice model 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

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

[8] Throughout the paper, deterministic vectors and functions are denoted by lower-case regular font Latin letters (e.g., x) and 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.

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

$$\{v_1 \,:\, v_1 = (\beta_0 + \beta_1 z_1 + e_1) z_{2,1}, e_1 \in \mathbb{R}, z \in Z\} = \mathbb{R}.$$

That is, we can treat $\mathbf{v}_1$ as *observed* special covariate.

To identify $\Pr(\mathbf{y} = 0 | \mathbf{v}_1 = \cdot)$ note that, since $\mathbf{y}$ and $\mathbf{z}$ are independent conditional on $\mathbf{v}_1$ (i.e., $\mathbf{z}$ affects the distribution of $\mathbf{y}$ only via $\mathbf{v}_1$), the observed $\Pr(\mathbf{y} = 0 | \mathbf{z} = z)$ has to satisfy the following integral equation

$$\Pr(\mathbf{y} = 0 | \mathbf{z} = z) = \int_{\mathbb{R}} \Pr(\mathbf{y} = 0 | \mathbf{v}_1 = v_1) dF_{\mathbf{v}_1|\mathbf{z}}(v_1|z)$$

for all z in the support. Since variation in $\mathbf{z}$ does not affect $\Pr(\mathbf{y} = 0 | \mathbf{v}_1 = \cdot)$, we can use this variation to identify $\Pr(\mathbf{y} = 0 | \mathbf{v}_1 = \cdot)$. In other words, if the family distributions $\{F_{\mathbf{v}_1|\mathbf{z}}(\cdot|z) \,:\, z \in Z\}$ is sufficiently "rich", then the above integral equation has a unique solution and we can identify $\Pr(\mathbf{y} = 0 | \mathbf{v}_1 = \cdot)$ if we can identify $F_{\mathbf{v}_1|\mathbf{z}}$. 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" (see Section 3 for a formal definition). Completeness assumption can be satisfied in many settings. For instance, it is satisfied if $\mathbf{e}_1$ is normally distributed.

To identify $F_{\mathbf{v}_1|\mathbf{z}}$ we need to identify β_0, β_1, and the distribution of $\mathbf{e}_1$, $F_{\mathbf{e}_1}$. I provide two approaches for identification of these object. The first approach assumes that $\mathbf{e}_1$ is normally distributed, but imposes minimal restrictions on the support of $\mathbf{z}$. The second approach does not impose parametric assumptions on $F_{\mathbf{e}_1}$, but requires the support of $\mathbf{z}_2$ to contain zero with an open neighborhood. Both approaches use partial derivatives of $\Pr(\mathbf{y} = 0 | \mathbf{z} = z)$ with respect to z to identify β_0 and β_1. To identify $F_{\mathbf{e}_1}$, in the second approach, I show how to recover all moments of $\mathbf{e}_1$.

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

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 indexes $\mathbf{v} \in V \subseteq \mathbb{R}^{d_v}$. 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 $v \in V$.

Assumption 1 is an exclusion restriction that requires covariates $\mathbf{z}$ to affect distribution over some outcomes only via the distribution of the latent vector $\mathbf{v}$. 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 restriction on the latent variable whose distribution is affected by the excluded covariates $\mathbf{z}$.

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

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]

Assume that

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

is known (or can be consistently estimated) for every excluded outcome $y \in Y^*$ and $x \in X$. The conditional distribution of other outcomes does not need to be known.[11]

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

Proof. 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

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

This implies that the nonzero function $h(y, \cdot) - h_0(y, \cdot)$ integrates to 0 for all $z \in Z'$. The

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

[11] For instance, to identify the multinomial choice model analyzed in Section 5.1, one only needs to know the conditional probability of choosing the default option.

latter contradicts to Assumption 2. The fact that the choice of y and w was arbitrary completes the proof. ■

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 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, then all identification techniques that require existence of covariates with full support can be applied (e.g., Fox and Gandhi (2016), Fox (2020) in the context of multinomial choice models with random coefficients and Bajari et al. (2010) in the context of games).

At this point I can only establish identification of h_0 up to $F_{\mathbf{v}|\mathbf{x}}$. In Section 4.1, I show how one can identify $F_{\mathbf{v}|\mathbf{x}}$ and, thus, h_0, when $F_{\mathbf{v}|\mathbf{x}}$ is assumed to be the Gaussian distribution, without any additional restrictions on h_0 and with minimal support restrictions on covariates. In Section 4.2, I relax the normality assumption on $F_{\mathbf{v}|\mathbf{x}}$. This weakening of the parametric assumptions comes with a cost: I have to impose some smoothness conditions on h_0 and extra support restrictions on covariates. In Section 5.3, I show how identification of h_0 up to $F_{\mathbf{v}|\mathbf{x}}$ can be used to characterize the identified set of a partially identified game of complete information.

4. Identification

4.1. Gaussian Distribution

The parameter h_0 can be identified with any known 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 – a Gaussian distribution.

Assumption 3 (i) The latent $\mathbf{v} = (\mathbf{v}_i)_{i=1,\ldots,d_v}$ satisfies

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

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

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

(iii) The support of $\mathbf{z}$ conditional on $\mathbf{w} = w$, Z_w, contains an open ball for every $w \in W$;

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

Assumption 3(i) is motivated by random coefficient models. The covariate $\mathbf{z}_{2,i}$ can be interpreted as choice-specific characteristic. The random coefficient $[\beta_{0,i}(\mathbf{w}) + \beta_{1,i}(\mathbf{w})\mathbf{z}_{1,i} + \mathbf{e}_i]$ captures agent-specific heterogeneity in tastes. The only support restriction is imposed on $\mathbf{z}$ (Assumption 3(iii)). It implies that the excluded covariates are not discrete and that there are no overlapping components between excluded covariates, and between $\mathbf{w}$ and $\mathbf{z}$. Non of the covariates are assumed to have unbounded support or to take values in a neighborhood of zero. Assumptions 3(i)-(iii) are sufficient for Assumption 2 since the family of normal distributions indexed by the mean is complete as long as the parameter space for the mean contains an open ball.[12]

Assumptions 3(iv) 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_0 in utility

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

indexes v. For instance, in multinomial choice models the probability of choosing an outside option is decreasing in mean utilities of other choices.

We need one more technical assumption before we proceed to the result. 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 $x = ((z_{1,i}, z_{1,-i})^\top, z_2^\top, w^\top)^\top$

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

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 I fix all covariates but one, then the probability of observing one excluded outcome conditional on covariates is neither an affine nor an exponential function of the nonfixed covariate. Assumption 4 is not very restrictive since it rules out only some exponential and linear probability models. Moreover, it is testable.

Let $\beta_0(\cdot) = \{\beta_{0,i}(\cdot)\}_{i=1}^{d_v}$ and $\beta_1(\cdot) = \{\beta_{1,i}(\cdot)\}_{i=1}^{d_v}$.

Proposition 4.1 *Suppose that Assumptions 1, 3, and 4 hold. Then h_0, β_0, and β_1 are identified.*

Proposition 4.1 establishes identification of h_0 and $F_{\mathbf{v}|\mathbf{x}}$ for normally distributed latent variables. Assumption 2 is implied by the assumptions of Proposition 4.1 and does not need to be directly imposed. Identification of β_0 and β_1 is constructive and leads to a simple estimation procedure (see Section 6.1). It is important to note that Proposition 4.1 does not require differentiability or even continuity of h_0, thus, allowing for discrete or degenerate unobserved heterogeneity in the model (e.g., $\mathbf{g}_1$ in Section 2 can be a constant or discrete random variable).

The proof of the identification of β_0 and β_1 uses the multiplicative structure of $z_{1,i}$ and $z_{2,i}$, and properties of the standard normal p.d.f. Informally, note that

$$\mathbf{v}_i = \beta_{0,i}(\mathbf{w})\mathbf{z}_{2,i} + \beta_{1,i}(\mathbf{w})\mathbf{z}_{1,i}\mathbf{z}_{2,i} + \mathbf{e}_i\mathbf{z}_{2,i}.$$

Since $z_{1,i}$ and $z_{2,i}$ can be moved independently, I can use variation in $z_{1,i}$ while keeping $z_{1,i}z_{2,i}$ by varying $z_{2,i}$ to identify $\beta_{0,i}(w)$. Then, by varying $z_{2,i}$, I can identify $\beta_{1,i}(w)$.

4.2. Nonparametric Identification

In this section the normality assumption is relaxed. However, I will impose additional restrictions on h_0 and on the support of covariates.

Assumption 5 (i) The latent $\mathbf{v} = (\mathbf{v}_i)_{i=1,\ldots,d_v}$ satisfies

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

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

(ii) Conditional on $\mathbf{w} = w$, $\{\mathbf{e}_i\}_{i=1,\ldots,d_v}$ are mean-zero and variance-one independent random variables that are independent of $\mathbf{z}$ for all $w \in W$;

(iii) The distribution of $\mathbf{e}_i$ conditional on $\mathbf{w} = w$ can be identified from $\kappa \leq \infty$ moments $\mathbb{E}\left[\mathbf{e}_i^l | \mathbf{w} = w\right]$, $l = 1, \ldots, \kappa$, for all $i = 1, \ldots, d_v$ and $w \in W$;

(iv) $h_0(y^*, \cdot, w)$ has bounded derivatives up to order κ and $\partial^l_{v_i^l} h_0(y^*, \cdot, w)|_{v=0} \neq 0$ for all $l \leq \kappa$, all $i = 1, \ldots, d_v$, and all $w \in W$;

(v) For every $w \in W$ the support of $\mathbf{z}$ conditional on $\mathbf{w} = w$, Z_w, contains z^* with an open neighborhood such that $z^*_{2,i} = 0$ for all $i = 1, \ldots, d_v$;

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

Assumptions 5(i) and (v) coincide with Assumptions 3(i) and (iv). Assumption 5(ii) is a standard location and scale normalization of the distribution of $\mathbf{e}_i$.[13] It restricts $\mathbf{e}_i$ conditional on $\mathbf{w} = w$ to have a finite expectation and a strictly positive variance

[13] Since h_0 is nonparametric function, we need to put a scale a location normalization on v.

for all w. Assumption 5(iii) is standard and allows to identify the distribution of $\mathbf{e}$ nonparametrically from moments of $\mathbf{e}$ (see, for instance, Fox et al., 2012 and Lewbel and Pendakur, 2017 for sufficient conditions). Note that κ does not have to be finite.

Proposition 4.2 *If Assumptions 1 and 5 hold, then β_0 and $F_{\mathbf{e}|\mathbf{x}}$ are identified. If, moreover, Assumption 2 is satisfied, then h_0 is also identified.*

The proof of Proposition 4.2 is similar to the proof of Theorem 11 in Fox et al. (2012). The main difference is that, instead of parametric restrictions, Proposition 4.2 uses interaction between $\mathbf{z}_{1,i}$ and $\mathbf{z}_{2,i}$. The main drawback of Proposition 4.2 is that it requires $\mathbf{z}_{2,i}$ to fall into a neighborhood of zero with positive probability.

5. Applications

In this section I show how the results from Sections 3 and 4 can be used to in different discrete outcome models. In particular, in Sections 5.1 and 5.2, I provide two sets of results (based on Propositions 4.1 and 4.2) for multinomial choice models and bundles models. In Section 5.3, I use results from Section 3 to describe the identified set for partially-identified payoff parameters in binary games of complete information.

5.1. Multinomial Choice

Consider the following random coefficients model motivated by Nevo (2001). The agent has to choose between J inside goods (e.g., different brands of cereals) 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

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

The random coefficient $\left[\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1\right]$ represents individual specific heterogeneous tastes associated with product characteristic $\mathbf{z}_{2,y}$ (e.g., fiber content). The covariate $\mathbf{z}_1$ is observed individual-specific taste shifter (e.g., age or income). The latent random vector $\mathbf{g} = (\mathbf{g}_y)_{y\in Y\setminus\{0\}}$ captures other sources of unobserved heterogeneity (e.g., the random coefficients that interact with $\mathbf{w}$).

The observed covariates are $\mathbf{x} = (\mathbf{z}_1, \mathbf{z}_2^\top, \mathbf{w}^\top)^\top$, where $\mathbf{z}_2 = (\mathbf{z}_{2,y})_{y\in Y\setminus\{0\}}$. The vector of covariates $\mathbf{w}$ may include the rest of product/agent characteristics. Importantly, I will impose no restrictions on the dependence structure between $\mathbf{g}$ and $\mathbf{w}$. Assume that the agents are utility-maximizers.[14]

Assumption 6 (i) $\beta_1(w) \neq 0$ for all $w \in W$;

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

(iii) Random shocks $\mathbf{g}$ are conditionally independent of $\mathbf{z}$ conditional on $\mathbf{w} = w$ for all $w \in W$. 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;

(iv) For every $w \in W$, there exists $(z_1^*, z_2^{*\top})^\top$ such that it is contained in Z_w with some open neighborhood, $z_{2,y}^* > 0$ or $z_{2,y}^* < 0$ for all $y \in Y$, and

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

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

is neither an exponential nor an affine function.

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.[15] However, I do not impose sign restrictions on $[\beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1]$. 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 $[\beta_0(\mathbf{w})+\beta_1(\mathbf{w})\mathbf{z}_1+\mathbf{e}_1]$ can be positive (negative) with probability that is arbitrarily close to 1.

Assumption 6(iii) together with utility-maximizing behavior guarantees that Assumption 1 is satisfied. Assumptions 6(ii)-(iii) are the only restriction on $\mathbf{g}$. I allow $\mathbf{g}$ to be constant, discrete, or continuous random variable with unknown support. Assumption 6(iv) is a testable restriction. It implies that one can find z_2^* such that Assumption 2 is satisfied for $v = z_{2,1}^*(\beta_0(w) + \beta_1(w)z_1 + e_1)$, and implies Assumption 4. The sign restrictions on components of z_2^* are only needed to identify the sign of $\beta_1(w)$.

Assumption 6 is sufficient to apply Proposition 4.1. The following assumption is needed if one does not want to assume normality of $\mathbf{e}_1$.

Assumption 7 (i) $\beta_1(w) \neq 0$ for all $w \in W$;

(ii) Conditional on $\mathbf{w} = w$, $\mathbf{e}_1$ is a mean-zero and variance-one random variable that is independent of $\mathbf{z}$ for all $w \in W$;

(iii) The distribution of $\mathbf{e}_1$ conditional on $\mathbf{w} = w$ can be identified from $\kappa \leq \infty$ moments $\mathbb{E}\left[\mathbf{e}_1^l|\mathbf{w} = w\right]$, $l = 1, \ldots, \kappa$, for all $w \in W$;

(iv) Random shocks $\mathbf{g}$ are conditionally independent of $\mathbf{z}$ conditional on $\mathbf{w} = w$ for all $w \in W$.

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

(v) $F_{\mathbf{g}|\mathbf{w}}(\cdot|w)$ has bounded derivatives up to order κ and $\partial^l_{g^l_y} F_{\mathbf{g}|\mathbf{w}}(\cdot|w)|_{g=0} \neq 0$ for all $l \leq \kappa$, all $y \in Y$, and all $w \in W$;

(vi) For every $w \in W$ the support of $\mathbf{z}$ conditional on $\mathbf{w} = w$, Z_w, contains z^* with an open neighborhood such that $z^*_{2,y} = 0$ for all $y \in Y$;

(vii) Assumption 2 is satisfied with $\boldsymbol{\tau} = \beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1$ instead of $\mathbf{v}$.

Assumption 7 requires the conditional distribution $F_{\mathbf{g}|\mathbf{w}}$ to be sufficiently smooth in the neighborhood of zero (e.g., $\mathbf{g}$ cannot have finite support).

Let $\mathcal{T}_w$ be the support of the random coefficient $\boldsymbol{\tau} = \beta_0(\mathbf{w})+\beta_1(\mathbf{w})\mathbf{z}_1+\mathbf{e}_1$ conditional on $\mathbf{w} = w$. Note that if $\mathbf{e}_1$ conditional on $\mathbf{w} = w$ has full support, then $\mathcal{T}_w = \mathbb{R}$

Proposition 5.1 *Suppose either Assumption 6 or Assumption 7 holds. Then*

(i) β_0, β_1, and $F_{\mathbf{e}_1|x}$ 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

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

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 $z_2/\left\|z_2\right\|$ can cover is sufficiently rich and the support of $\mathbf{e}_1$ conditional on $\mathbf{w} = w$ is $\mathbb{R}$, 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, Gautier and Hoderlein, 2015, Fox and Gandhi, 2016, and Fox, 2020) can be applied. For instance, if $E_{1|w} = \mathbb{R}$ (i.e. $\mathbf{e}_1$ conditional on $\mathbf{w} = w$ has full support) and $\{z_2/\|z_2\| : z_2 \in Z_{2|w}\}$ is equal to a unit sphere in $\mathbb{R}^J$ for every w, then $R_w = \mathbb{R}^J$ and I can nonparametrically identify $F_{\mathbf{g}|\mathbf{w}}$.[16] Note that, under Assumption 7(vi), $\mathbf{z}_2$ covers all directions, hence, if $E_{1|w} = \mathbb{R}$, then $R_w = \mathbb{R}^J$.

The following corollaries combine Proposition 5.1 with existing results that use special covariates.

Corollary 5.2 (Theorem 2, Fox and Gandhi, 2016) *For all $y \neq 0$ let $\mathbf{g}_y = \boldsymbol{\theta}_y(\mathbf{w})$, where $\boldsymbol{\theta}_y$ is a random function such that its realization θ_y is a map from W to $\mathbb{R}$. Suppose*

(i) Assumption 6 holds;

(ii) $R_w = \mathbb{R}^J$ for all $w \in W$;

(iii) $\boldsymbol{\theta} = (\boldsymbol{\theta}_y)_{y\neq 0}$ and $\mathbf{w}$ are independent;

(iv) The support of $\boldsymbol{\theta}$, Θ, satisfies Assumption 4 in Fox and Gandhi (2016);

then β_0, β_1, and the distribution of $\boldsymbol{\theta}$ are identified.

Corollary 5.3 (Theorem 1, Fox, 2020) *For all $y \neq 0$ let $\mathbf{g}_y = \boldsymbol{\theta}^\top \mathbf{w}_y + \boldsymbol{\epsilon}_y$, where $\boldsymbol{\theta}$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_y)_{y\in Y\setminus\{0\}}$ are random coefficients, and $\mathbf{w}_y$ is the vector of product-y-specific variable covariates. Suppose*

(i) Assumption 7 holds;

(ii) $R_w = \mathbb{R}^J$ for all $w \in W$;

(iii) $(\boldsymbol{\theta}^\top, \boldsymbol{\epsilon}^\top)^\top$ and $\mathbf{w} = (\mathbf{w}_y)_{y\in Y\setminus\{0\}}$ are independent;

(iv) The support of $\mathbf{w}$ contains an open ball of dimensionality of $\mathbf{w}$;

(v) $(\boldsymbol{\theta}^\top, \boldsymbol{\epsilon}^\top)^\top$ has finite absolute moments and its distribution is uniquely determined by its moments;

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

then β_0*,* β_1*, and the distribution of* $(\mathbf{e}_1, \boldsymbol{\theta}^\intercal, \boldsymbol{\epsilon}^\intercal)^\intercal$ *are identified.*

To the best of my knowledge, Corollary 5.3 is the first result that establishes nonparametric identification of the whole distribution of the random coefficients in the multinomial choice environments without assuming existence of special covariates. Fox et al. (2012) and Allen and Rehbeck (2020) also allow for bounded covariates. However, they either do not identify the distribution of the random intercept, $\mathbf{g}_y$, (Allen and Rehbeck, 2020), or impose parametric restrictions on it (Fox et al., 2012).

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, $y = (0,1,0,1,0,\ldots,0)^\intercal$ corresponds to the case when the agent purchased a bundle of goods 2 and 4. I normalize the utility from "not buying", $y = 0 \in \mathbb{R}^J$, to 0. The random utility from choosing an alternative $y \neq 0$ is of the form

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

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 5.4 *Suppose either Assumption 6 or Assumption 7 holds. Then*

(i) β_0*,* β_1*, and* $F_{\mathbf{e}_1|x}$ *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 $\mathbf{r} = (\mathbf{r}_j)_{j=1,\dots,J}$ *is an observed covariate independent of* $\mathbf{g} = (\mathbf{g}_y)_{y \in Y \setminus \{0\}}$ *conditional on* $\mathbf{w}$ *with the conditional support*

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

The following corollary combines Proposition 5.4 with Theorem 1 in Fox and Lazzati (2017).

Corollary 5.5 (Theorem 1, Fox and Lazzati, 2017) *Let* $J = 2$ *and*

$$\mathbf{g}_{(1,0)} = \theta_1(\mathbf{w}) + \boldsymbol{\epsilon}_1,$$
$$\mathbf{g}_{(0,1)} = \theta_2(\mathbf{w}) + \boldsymbol{\epsilon}_2,$$
$$\mathbf{g}_{(1,1)} = \mathbf{g}_{(1,0)} + \mathbf{g}_{(0,1)} + \boldsymbol{\xi}\theta_3(\mathbf{w}),$$

where $\theta_i(\cdot)$, $i = 1, 2, 3$, *are some unknown functions, and* $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \boldsymbol{\xi})^\mathsf{T} \in \mathbb{R}^2 \times \mathbb{R}_+$. *Suppose*

(i) *Assumption 6 or Assumption 7 holds;*

(ii) $R_w = \mathbb{R}$ *for all* w*;*

(iii) $(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2)^\mathsf{T} | \mathbf{w} = w$ *has an everywhere positive Lebesgue density on its support for all* $w \in W$*;*

(iv) $\mathbb{E}\left[\, \boldsymbol{\epsilon}_i | \mathbf{w} = w \,\right] = 0$ *and* $\mathbb{E}\left[\, \boldsymbol{\xi} | \mathbf{w} = w \,\right] = 1$ *for all* $w \in W$ *and* $i = 1, 2$,

then $\theta_i(\cdot)$, $i = 1, 2, 3$, *and the c.d.fs* $F_{\boldsymbol{\epsilon}_i | \mathbf{w}}$, $i = 1, 2$, *and* $F_{\boldsymbol{\xi} | \mathbf{w}}$ *are identified.*

5.3. Binary games of complete information

In the multinomial choice and the bundles models, I can establish identification of the objects of interest without requiring covariates with full support. The example 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 $i \in I$, where $\|I\|$ is the number of elements in set I. Every player must choose $y_i \in \{0, 1\}$. Thus, the outcome space is $Y = \{0, 1\}^{\|I\|}$.[17] 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 $(\mathbf{y}, \mathbf{x}^\top)^\top$, where $\mathbf{x} = (\mathbf{z}^\top, \mathbf{w}^\top)^\top \in X$ 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 8 (i) Assumption 1 is satisfied with $\mathbf{v} = (\beta_{0,i}(\mathbf{w})\mathbf{z}_i + \mathbf{e}_i)_{i \in I}$;

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

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

is at least 2.

Assumption 8(i) implies that excluded covariates affect the distribution of some out-

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

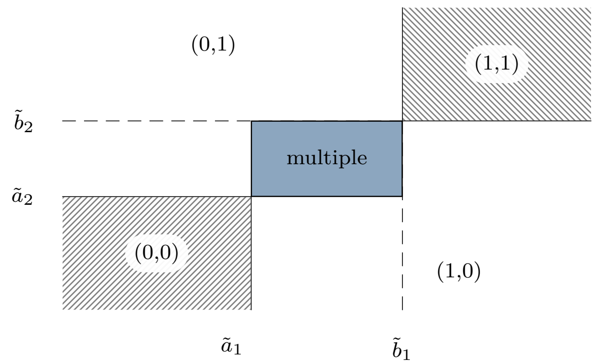

Figure 1 – Predictions of Nash equilibrium for different realizations of v when $\delta_{0,1,2} < 0$ and $\delta_{0,2,1} < 0$ in a two player game.

comes via payoffs only. The index $\mathbf{v}$ is multidimensional and differs from the one in Section 4 – there is no action specific covariate $\mathbf{z}_{2,i}$. Assumption 8(ii) imposes restrictions on Y^*. If one thinks of an entry game where $y_i = 1$ corresponds to the entry decision, the outcomes in Assumption 8(ii) have the following interpretation. The outcome $(0)_{k\in I}$ corresponds to the market where nobody enters. The outcome $\left(\mathbb{1}\left(k \in \{i,j\}\right)\right)_{k\in I}$ corresponds to the market where only players i and j enter. Similarly, $\left(\mathbb{1}\left(k = i\right)\right)_{k\in I}$ 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 9 (i) The shocks $\{\mathbf{e}_i\}_{i\in I}$ are i.i.d. standard normal random variables and are independent of $\mathbf{x}$;

(ii) $\beta_{0,i}(w) \neq 0$ for all $i \in I$ and $w \in W$;

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

Assumption 9 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 assume that agents play Nash equilibria (both in pure and mixed strategies). 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 $\tilde{a}_i, \tilde{b}_i$, $i = 1, 2$, 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 uniquely recover the threshold values $\tilde{a}_i, \tilde{b}_i$, $i = 1, 2$. Since $\tilde{a}_i = -\alpha_{0,i}$, and $\tilde{b}_i = -\alpha_{0,i} - \delta_{0,i,j}$, $i, j = 1, 2$, α_0 and δ_0 are identified. Note that in the multiplicity region $v \in (\tilde{a}_1, \tilde{b}_1) \times (\tilde{a}_2, \tilde{b}_2)$ there are three Nash equilibria, and players can play mixtures of these equilibria as long as their behavior is consistent with Assumption 8.

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

Proposition 5.6 *Under Assumptions 8 and 9 if players behave according to Nash equilibrium (both in pure and mixed strategies), then α_0 and δ_0 are identified up to β_0.*

Proposition 5.6 states that if marginal effects of some excluded covariates on payoffs are known, then one can identify the rest of the payoff parameters. Assumption 9(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 5.6 are identified up to β_0 and V.

Full identification in this binary game can be achieved if one can identify β_0. The standard identification-at-infinity argument requires existence of player-action-specific covariates that have full support (unbounded from above *and* below).[18] 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).[19] 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).[20]

[18] Assuming Nash equilibrium in pure strategies, Kline (2016) establishes identification of payoff parameters without requiring unbounded covariates and parametric restrictions on the distribution of unobservables.

[19] See, for instance, Tamer (2003), Bajari et al. (2010), and Kashaev and Salcedo (2020).

[20] One may argue that it is always possible to transform covariate into a covariate with full support.

6. Estimation and Empirical Application

6.1. Estimation of β in the Multinomial Choice Model

Proposition 4.1 constructively identifies β_0 and β_1. In this section I use it to estimate these parameters. To simplify the presentation, I focus on the multinomial choice model with random coefficients with normally distributed $\mathbf{e}_1$ considered in Section 5.1.[21] Moreover, I will assume that there are no nonexcluded covariates $\mathbf{w}$ (i.e., $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are constant functions). Note that, even though β_0 and β_1 are finite-dimensional parameters and the distribution of $\mathbf{e}_1$ is assumed to be known, the model is still semiparametric since the distribution of $\mathbf{g}$ is not parametric.

The first ingredient of the estimator is a nonparametric estimator of $p_0(z) = \Pr(\mathbf{y} = 0|\mathbf{z} = \cdot)$, $\hat{p}_0(\cdot)$. Any consistent and smooth enough estimator $\hat{p}_0(\cdot)$ will deliver consistent estimators of $\beta = (\beta_1, \beta_0)^\intercal$.[22] For concreteness, I will work with the series estimator based on products of powers of components of z (polynomial regressions). That is, given a sample of i.i.d. observations $\left\{\mathbf{y}^{(i)}, \mathbf{z}^{(i)}\right\}_{i=1}^n$, define

$$\hat{p}_0(z) = \psi^K(z)^\intercal \left(\Psi^\intercal \Psi\right)^- \sum_{i=1}^n \psi^K\left(\mathbf{z}^{(i)}\right) \mathbb{1}\left(\mathbf{y}^{(i)} = 0\right),$$

where $\psi^K(\cdot)$ is a vector of orthonormal basis functions based on products of powers of components of z, $\Psi = \left(\psi^K\left(\mathbf{z}^{(1)}\right), \psi^K\left(\mathbf{z}^{(2)}\right), \ldots, \psi^K\left(\mathbf{z}^{(n)}\right)\right)^\intercal$, and $\left(\Psi^\intercal \Psi\right)^-$ is the Moore-Penrose generalized inverse. I assume that the sum of powers of components of z in ψ^K is monotonically increasing in K.

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.

[21] Proposition 4.2 also provides a constructive identification for β_0 and β_1. However, Assumption 7(vi) fails to hold in my illustrative application presented in Section 6.2. Additionally, Proposition 4.2 uses derivatives of identifiable functions evaluated at a single point, thus, most likely, leading to a consistent estimator with nonparametric rate of convergence.

[22] The normality of $\mathbf{e}_1$ implies that $p_0(\cdot)$ has continuous derivatives of any order. See Appendix A.5 for details.

The sign of β_1 can be trivially estimated from $\hat{p}_0$ since

$$\text{sign}(\beta_1) = \text{sign}\left(p_0((z_1', z_2^\intercal)^\intercal) - p_0((z_1, z_2^\intercal)^\intercal)\right) \text{sign}(z_{2,y^*})\text{sign}(z_1' - z_1)$$

if $z_2 \geq 0$ or $z_2 \leq 0$ with $z_{2,y^*} \neq 0$. Hence, for simplicity I assume that $\beta_1 > 0$.

Given the nonparametric power series estimator $\hat{p}_0$, let

$$\hat{\beta}_1 = \sqrt{\frac{\sum_{i=1}^n \hat{p}_{111}\left(\mathbf{z}^{(i)}\right)\hat{p}_1\left(\mathbf{z}^{(i)}\right) - \hat{p}_{11}\left(\mathbf{z}^{(i)}\right)^2}{\sum_{i=1}^n \hat{p}_{12}\left(\mathbf{z}^{(i)}\right)\hat{p}_1\left(\mathbf{z}^{(i)}\right) - \hat{p}_2\left(\mathbf{z}^{(i)}\right)\hat{p}_{11}\left(\mathbf{z}^{(i)}\right) - \hat{p}_1\left(\mathbf{z}^{(i)}\right)^2}},$$

$$\hat{\beta}_0 = \hat{\beta}_1 \frac{\sum_{i=1}^n \hat{p}_2\left(\mathbf{z}^{(i)}\right) - \mathbf{z}_1^{(i)}\hat{p}_1\left(\mathbf{z}^{(i)}\right)}{\sum_{i=1}^n \hat{p}_1\left(\mathbf{z}^{(i)}\right)} - \frac{1}{\hat{\beta}_1}\frac{\sum_{i=1}^n \hat{p}_{11}\left(\mathbf{z}^{(i)}\right)}{\sum_{i=1}^n \hat{p}_1\left(\mathbf{z}^{(i)}\right)},$$

where

$$\hat{p}_1(z) = \partial_{z_1}\hat{p}_0(z), \quad \hat{p}_{11}(z) = \partial^2_{z_1^2}\hat{p}_0(z), \quad \hat{p}_{111}(z) = \partial^3_{z_1^3}\hat{p}_0(z),$$
$$\hat{p}_2(z) = \sum_{y=1}^J z_{2,y}\partial_{z_{2,y}}\hat{p}_0(z), \quad \hat{p}_{12}(z) = \partial_{z_1}\hat{p}_2(z).$$

Note that $\hat{\beta}$ is essentially a nonlinear function of sample averages of different derivatives of estimated $\hat{p}_0$. Following Newey (1994) and Newey (1997), in order to achieve $\sqrt{n}$-consistency and asymptotic normality of the proposed estimator, I will have to establish existence of the Reisz representer of a particular directional derivative. Let

$$\bar{v}_1(z) = -\Big[4p_{1111}(z)f_{\mathbf{z}}(z) + 8p_{111}(z)\partial_{z_1}f_{\mathbf{z}}(z) + 5p_{11}(z)\partial^2_{z_1^2}f_{\mathbf{z}}(z) + p_1(z)\partial^3_{z_1^3}f_{\mathbf{z}}(z)\Big]/f_{\mathbf{z}}(z),$$
$$\bar{v}_2(z) = \Big[\beta_1\{(1-J)f_{\mathbf{z}}(z) + z_1\partial_{z_1}f_{\mathbf{z}}(z) - \sum_y z_{2,y}\partial_{z_{2,y}}f_{\mathbf{z}}(z)\} - \partial^2_{z_1^2}f_{\mathbf{z}}(z)\Big]/f_{\mathbf{z}}(z),$$
$$\bar{v}(z) = (\bar{v}_1(z), \bar{v}_2(z))^\intercal,$$

where $f_{\mathbf{z}}$ is the p.d.f. of $\mathbf{z}$, and p_1, p_{11}, p_{111}, and p_{1111} are first, second, third, and forth derivatives of p_0 with respect to z_1, respectively.

Assumption 10 (i) The support of $\mathbf{z}$, Z, is a Cartesian product of compact connected nonsingleton intervals in $\mathbb{R}$.

(ii) $f_{\mathbf{z}}$ is bounded away from zero on the interior of Z;

(iii) $f_{\mathbf{z}}(\cdot)$, $\partial_{z_1} f_{\mathbf{z}}(\cdot)$, $\partial_{z_{2,y}} f_{\mathbf{z}}(\cdot)$, and $\partial^2_{z_1^2} f_{\mathbf{z}}(\cdot)$ equal to zero at the boundary of Z for all y;

(iv) $\mathbb{E}\left[\bar{v}(\mathbf{z})\bar{v}(\mathbf{z})^{\mathsf{T}}\right]$ is finite and nonsingular;

Assumptions 10(i)-(ii) are standard in the literature on nonparametric estimation of conditional expectations. Similarly to the average derivative estimator of Powell et al. (1989), in order to achieve $\sqrt{n}$-consistency the estimator I need to impose restrictions on the behavior of $f_{\mathbf{z}}$ on the boundary of its support. Since Powell et al. (1989) work with the first derivative they only require $f_{\mathbf{z}}$ to vanish on the boundary. My estimator involves derivatives up to order 3, thus, leading to Assumption 10(iii). Assumption 10(iv) is the mean-square continuity condition that requires the variance of the score function of $\mathbf{z}$ (i.e log $f_{\mathbf{z}}$) and derivatives of it to be finite.

The following proposition establishes asymptotic normality of my estimator and is based on Theorem 6 of Newey (1997). Denote

$$G = \begin{pmatrix} 2\beta_1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbb{E}\left[p_{12}(\mathbf{z})p_1(\mathbf{z}) - p_2(\mathbf{z})p_{11}(\mathbf{z}) - p_1(\mathbf{z})^2\right] & 0 \\ 0 & \beta_1 \mathbb{E}\left[p_1(\mathbf{z})\right] \end{pmatrix}^{-1},$$

Proposition 6.1 *If (i) $\left\{\mathbf{y}^{(i)}, \mathbf{z}^{(i)}\right\}_{i=1}^{n}$ are i.i.d.; (ii) Assumptions 6 and 10 are satisfied, and Assumption 6(iv) is satisfied for all $z^* \in Z$; (iii) $K^6/n \to_{n\to\infty} 0$, then*

$$\sqrt{n}(\hat{\beta} - \beta) \to_d \mathrm{N}(0, V),$$

where $V = G\mathbb{E}\left[\bar{v}(\mathbf{z})\bar{v}(\mathbf{z})^{\mathsf{T}} p_0(\mathbf{z})(1 - p_0(\mathbf{z}))\right] G^{\mathsf{T}}$.

In the proof of Proposition 6.1 I also provide a consistent estimator of the asymptotic variance matrix V that is based on the estimator proposed in Newey (1997).

6.2. Illustrative Empirical Application

To illustrate the proposed estimation procedure I analyze margarine purchasing decisions of households using the multinomial choice model with normally distributed $\mathbf{e}_1$. The data set is a cross-section of 242 purchasing decisions from Springfield, MO (Benoit et al., 2016). Every observation contains only information on the household annual income, which I use as the agent-specific covariate z_1, agent choices (y), and product-specific prices p_y.[23] There are 5 brands: Generic ($y = 0$), Blue Bonnet ($y = 1$), House Brand ($y = 2$), Shed Spread ($y = 3$), and Fleischmann's ($y = 4$).[24] I model the utility from purchasing every brand as

$$\tilde{\mathbf{u}}_y = (\beta_0 + \beta_1 \mathbf{z}_1 + \mathbf{e}_1)\mathbf{p}_y + \tilde{\mathbf{g}}_y.$$

The coefficient $\beta_0 + \beta_1 \mathbf{z}_1$ can be thought of as the average marginal utility with respect to price.

If I treat the utility from consuming Generic brand as the baseline utility and subtract it from all utilities, the normalized utility from purchasing different brands for $y = 1, 2, 3, 4$ is

$$\mathbf{u}_y = (\beta_0 + \beta_1 \mathbf{z}_1 + \mathbf{e}_1)[\mathbf{p}_y - \mathbf{p}_0] + \tilde{\mathbf{g}}_y - \tilde{\mathbf{g}}_0$$

and the utility from purchasing Generic brand is 0. Hence, I can define $\mathbf{z}_{2,y} = \mathbf{p}_y - \mathbf{p}_0$ and $\mathbf{g}_y = \tilde{\mathbf{g}}_y - \tilde{\mathbf{g}}_0$, $y = 1, 2, 3, 4$, where $\mathbf{p}_0$ is the price of Generic margarine.

Assumption 6(iii) requires $\mathbf{g}$ and $\mathbf{z}$ to be independent conditionally on $\mathbf{w}$. Although, price $\mathbf{p}_y$ is probably correlated with unobserved part of the utility $\tilde{\mathbf{g}}_y$, the price difference $\mathbf{z}_y = \mathbf{p}_y - \mathbf{p}_0$ may be independent from $\mathbf{g}_y = \tilde{\mathbf{g}}_y - \tilde{\mathbf{g}}_0$.[25]

[23] Income and prices are measured in thousands of US dollars and US dollars, respectively.
[24] For specific details about the data set see Benoit et al. (2016).
[25] The choice of $\mathbf{z}$ is driven by the limitations of the data set.

The estimates of β_0 and β_1 are $\hat{\beta}_0 = -39.1$ (standard error= 43.8) and $\hat{\beta}_1 = -16.7 \times 10^{-3}$ (standard error= 3.97×10^{-6}). [26] As expected, the sign of β_0 is negative (although the coefficient is not significant at the 5 percent level). The coefficient in front of the income variable is negative and significant at the 5 percent level. However, the maximal value that $\hat{\beta}_1 \mathbf{z}_1$ can take in the sample is substantially smaller than $\hat{\beta}_0$ ($\max_i \mathbf{z}_1^{(i)} \hat{\beta}_1 / \hat{\beta}_0 = 0.055$, standard error= 0.062). The latter indicates that income does not affect marginal disutility from margarine price much.

7. Conclusion

This paper shows that commonly used exclusion restrictions and richness assumptions about the distribution of some unobservables may lead to full nonparametric identification in discrete outcome models even when covariates are bounded. The proposed identification framework extends the results from a large literature that uses special covariates with full support to environments where such full-support covariates are not available. It also leads to an asymptotically normal estimator of the finite-dimensional parameters of the model.

The partial identification result can substantially decrease the computational complexity of constructing confidence sets for partially identified parameters. For instance, the likelihood ratio statistic of Chen et al. (2011) is asymptotically chi-squared distributed after profiling β 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 the chi-squared distribution.

[26] I use the tensor product of the 4-th degree Chebyshev polynomials for z_1 and the 1-st degree Chebyshev polynomials for every $z_{2,y}$.

References

Allen, R. and Rehbeck, J. (2020). Identification of random coefficient latent utility models. *arXiv preprint arXiv:2003.00276*.

Andrews, D. W. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157.

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

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

Benoit, D. F., Van Aelst, S., and Van den Poel, D. (2016). Outlier-robust bayesian multinomial choice modeling. *Journal of Applied Econometrics*, 31(7):1445–1466.

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.

Brown, L. D. (1986). *Fundamentals of statistical exponential families with applications in statistical decision theory*, volume 9 of *Lecture notes – Monograph series*. Institute of Mathematical Statistics.

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, S., Khan, S., and Tang, X. (2016). Informational content of special regressors in heteroskedastic binary response models. *Journal of econometrics*, 193(1):162–182.

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., Hong, H., and Tamer, E. (2007a). Estimation and confidence regions for parameter sets in econometric models 1. *Econometrica*, 75(5):1243–1284.

Chernozhukov, V., Imbens, G. W., and Newey, W. K. (2007b). 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. (2020). A note on nonparametric identification of distributions of random coefficients in multinomial choice models. Technical report.

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., il Kim, K., Ryan, S. P., and Bajari, P. (2012). The random coefficients logit model is identified. *Journal of Econometrics*, 166(2):204–212.

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 Hoderlein, S. (2015). A triangular treatment effect model with random coefficients in the selection equation. *arXiv preprint arXiv:1109.0362*.

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. (2020). Discerning solution concepts for discrete games. *Journal of Business & Economic Statistics*.

Kline, B. (2016). The empirical content of games with bounded regressors. *Quantitative Economics*, 7(1):37–81.

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.

Lewbel, A. and Pendakur, K. (2017). Unobserved preference heterogeneity in demand using generalized random coefficients. *Journal of Political Economy*, 125(4):1100–1148.

Magnac, T. and Maurin, E. (2007). Identification and information in monotone binary models. *Journal of Econometrics*, 139(1):76–104.

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.

Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, pages 1349–1382.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of econometrics*, 79(1):147–168.

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

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, pages 1403–1430.

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 4.1

Note that h_0 is identified up to β_0 and β_1. Hence, I only need to show that β_0 and β_1 are identified.

Fix some $i \in \{1, 2, \ldots, d_v\}$, $z_{2,-i}$, $z_{1,-i}$, and w in the support. 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

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

After some rearrangements I get

$$\tilde{\eta}(z_{1,i}, z_{2,i}) = \int_{\mathbb{R}} F_0(t) \phi(t/z_{2,i} - \beta_0 - \beta_{1,i} z_{1,i}) dt, \tag{5}$$

where $\tilde{\eta}(z_{1,i}, z_{2,i}) = z_{2,i} \eta(z_{1,i}, z_{2,i})$.

Next, note that since $\phi''(x) = -\phi(x) - x\phi'(x)$ the following system of equations holds[27]

$$\begin{aligned}
\partial_{z_{1,i}} \tilde{\eta}(z_{1,i}, z_{2,i}) &= -\beta_{1,i}(w) \int F_0(t) \phi'(t/z_{2,i} - \beta_{0,i}(w) - \beta_{1,i}(w) z_{1,i}) dt, \\
\partial^2_{z^2_{1,i}} \tilde{\eta}(z_{1,i}, z_{2,i}) &= \beta_1^2 \int F_0(t) \phi''(t/z_{2,i} - \beta_{0,i}(w) - \beta_{1,i}(w) z_{1,i}) dt \\
&= -\beta_1^2 \tilde{\eta}(z_{1,i}, z_{2,i}) - \beta_{1,i}(w)(\beta_{0,i}(w) + \beta_{1,i}(w) z_{1,i}) \partial_{z_{1,i}} \tilde{\eta}(z_{1,i}, z_{2,i}) - \\
&- \beta_{1,i}(w)^2 \int t F_0(t) \phi'(t/z_{2,i} - \beta_{0,i}(w) - \beta_{1,i}(w) z_{1,i}) dt / z_{2,i}.
\end{aligned}$$

[27] I can differentiate under the integral sign since (i) h_0 being a bounded function implies that F_0 is a bounded function, (ii) all derivatives of the standard normal p.d.f. are bounded functions.

Moreover,

$$\partial_{z_{2,i}}\tilde{\eta}(z_{1,i},z_{2,i}) = -\int F_0(t)t\phi'(t/z_{2,i}-\beta_{0,i}(w)-\beta_{1,i}(w)z_{1,i})dt/z_{2,i}^2.$$

Hence,

$$\partial^2_{z^2_{1,i}}\tilde{\eta}(z_{1,i},z_{2,i}) = -\beta_1^2\tilde{\eta}(z_{1,i},z_{2,i}) - \beta_{1,i}(w)(\beta_{0,i}(w)+\beta_{1,i}(w)z_{1,i})\partial_{z_{1,i}}\tilde{\eta}(z_{1,i},z_{2,i}) + \beta_{1,i}(w)^2 z_{2,i}\partial_{z_{2,i}}\tilde{\eta}(z_{1,i},z_{2,i}).$$

Equivalently

$$\frac{\beta_{0,i}(w)}{\beta_{1,i}(w)} = \frac{z_{2,i}\partial_{z_{2,i}}\tilde{\eta}(z_{1,i},z_{2,i}) - \tilde{\eta}(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\tilde{\eta}(z_{1,i},z_{2,i})} - z_{1,i} - \frac{\partial^2_{z^2_{1,i}}\tilde{\eta}(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\tilde{\eta}(z_{1,i},z_{2,i})}\frac{1}{\beta_{1,i}(w)^2}.$$

Replacing $\tilde{\eta}(z_{1,i},z_{2,i})$ by $z_{2,i}\eta(z_{1,i},z_{2,i})$ I get

$$\frac{\beta_{0,i}(w)}{\beta_{1,i}(w)} = \frac{z_{2,i}\partial_{z_{2,i}}\eta(z_{1,i},z_{2,i}) - z_{1,i}\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})} - \frac{\partial^2_{z^2_{1,i}}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}\frac{1}{\beta_{1,i}(w)^2}. \tag{6}$$

Thus, $\beta_{0,i}(w)/\beta_{1,i}(w)$ is identified up to $\beta_{1,i}(w)^2$. Differentiating the last equation with respect to $z_{1,i}$ leads to the following equation:

$$\frac{1}{\beta_{1,i}(w)^2} = \partial_{z_{1,i}}\left[\frac{z_{2,i}\partial_{z_{2,i}}\eta(z_{1,i},z_{2,i}) - z_{1,i}\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}\right] \Big/ \partial_{z_{1,i}}\left[\frac{\partial^2_{z^2_{1,i}}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}\right]. \tag{7}$$

Hence, if

$$\partial_{z_{1,i}}\left[\frac{\partial^2_{z^2_{1,i}}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}\right] \neq 0 \tag{8}$$

for *some* $z_{1,i}$ and $z_{2,i}$, then $\beta_{1,i}(w)^2$ is identified. Suppose this is not the case. That is, for all $z_{1,i}$ and $z_{2,i}$

$$\partial_{z_{1,i}}\left[\frac{\partial^2_{z^2_{1,i}}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}\right] = 0.$$

Equivalently,

$$\partial^2_{z^2_{1,i}}\left[\log(\partial_{z_{1,i}}\eta(z_{1,i}, z_{2,i}))\right] = 0$$

for all $z_{1,i}$ and $z_{2,i}$. The latter would imply that either

$$\eta(z_{1,i}, z_{2,i}) = K_1(z_{2,i})e^{K_3(z_{2,i})z_{1,i}} + K_2(z_{2,i})$$

or

$$\eta(z_{1,i}, z_{2,i}) = K_1(z_{2,i})z_{1,i} + K_2(z_{2,i})$$

for some functions $K_i(\cdot)$, $i = 1, 2, 3$. Since it is assumed that $\eta(\cdot, z_{2,i})$ is neither an exponential nor an affine function, I can conclude that for some $z_{1,i}$ and $z_{2,i}$ Equation (8) is satisfied. Thus, $\beta_{1,i}(w)^2$ is identified (hence, $|\beta_{1,i}(w)|$ is also identified). Hence, I identify $\beta_{0,i}(w)/\beta_{1,i}(w)$. If $\beta_{0,i}(w)/\beta_{1,i}(w) = 0$, then the sign of $\beta_{1,i}(w)$ is identified from Assumption 3(iv). If $\beta_{0,i}(w)/\beta_{1,i}(w) \neq 0$, then the sign of either $\beta_{1,i}(w)$ or $\beta_{0,i}(w)$ is identified from Assumption 3(iv). Knowing the sign of, say, $\beta_{0,i}(w)$ and $\beta_{0,i}(w)/\beta_{1,i}(w)$ identifies $\beta_{1,i}(w)$ and $\beta_{0,i}(w)$. Since the choice of i and w was arbitrary I can identify β_0 and β_1.

Note that for identification of $\beta_{1,i}(w)$ and $\beta_{0,i}(w)$ I do not need to exclude all exponential functions of z_1, since instead of differentiating Equation (6) with respect to $z_{1,i}$ I can differentiate it with respect to $z_{2,i}$. For the identification result to hold it suffices to exclude functions of the form

$$\eta(z_{1,i}, z_{2,i}) = K_1(z_{2,i})e^{K_2 z_{1,i}} + K_3(z_{2,i})$$

or

$$\eta(z_{1,i}, z_{2,i}) = K_1(z_{2,i})z_{1,i} + K_3(z_{2,i}),$$

where $K_1(\cdot)$ and $K_2(\cdot)$ are some functions of z_2, and K_3 is a constant.

I conclude the proof by noting that from Equations (6) and (7) it follows that

$$\beta_{1,i}^2(w) = \frac{\partial^3_{z_{1,i}^3}\eta(z_{1,i},z_{2,i})\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i}) - \partial^2_{z_{1,i}^2}\eta(z_{1,i},z_{2,i})[\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})]^2}{z_{2,i}[\partial^2_{z_{1,i},z_{2,i}}\eta(z_{1,i},z_{2,i})\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i}) - \partial_{z_{2,i}}\eta(z_{1,i},z_{2,i})\partial^2_{z_{1,i}^2}\eta(z_{1,i},z_{2,i})] - [\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})]^2},$$

$$\beta_{0,i}(w) = \frac{z_{2,i}\partial_{z_{2,i}}\eta(z_{1,i},z_{2,i}) - z_{1,i}\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}\beta_{1,i}(w) - \frac{\partial^2_{z_{1,i}^2}\eta(z_{1,i},z_{2,i})}{\partial_{z_{1,i}}\eta(z_{1,i},z_{2,i})}\frac{1}{\beta_{1,i}(w)}.$$

Hence, if $\eta(\cdot, z_{2,i})$ is neither an exponential nor an affine function for all $z_{2,i}$, then I can construct an average derivative type estimator. This structure is exploited in Proposition 6.1.

A.2. Proof of Propositions 4.2

Note that h_0 is identified up to β_0, β_1, and $F_{\mathbf{e}|\mathbf{x}}$. Hence, I only need to show that β_0, β_1, and $F_{\mathbf{e}|\mathbf{x}}$ are identified.

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

$$F_0((\beta_{0,i}(w) + \beta_{1,i}(w)z_{1,i} + e_i)z_{2,i}) = \int h_0(y^*, w, (\beta_{0,1}(w) + \beta_{1,1}(w)z_{1,1} + e_1)z_{2,1}, \ldots, (\beta_{0,d_v}(w) + \beta_{1,d_v}(w)z_{1,d_v} + e_{d_v})z_{2,s_v})dF_{\mathbf{e}_{-i}|\mathbf{x}}(e_{-i}|x),$$

where $F_{\mathbf{e}_{-i}|\mathbf{x}}(e_{-i}|x) = \prod_{k\neq i} F_{\mathbf{e}_k|\mathbf{x}}(e_k|x)$, and $\eta(z_{1,i}, z_{2,i}) = \mu(y^*|z, w)$.

Assumptions 1 implies that

$$\eta(z_{1,i}, z_{2,i}) = \int F_0((\beta_{0,i}(w) + \beta_{1,i}(w)z_{1,i} + e_i)z_{2,i})dF_{\mathbf{e}_i|x}(e|x).$$

Next, since $\mathbf{e}_i$ and $\mathbf{z}$ are independent conditional on $\mathbf{w}$ and $h_0(y^*, w, \cdot)$ is κ-times differentiable with bounded derivatives, the dominated convergence theorem implies that

$$\partial^l_{z_{1,i}^l}\eta(z_{1,i}, z_{2,i}) = \beta_{1,i}(w)^l z_{2,i}^l \int \partial^l_{t^l} F_0((\beta_{0,i}(w) + \beta_{1,i}(w)z_{1,i} + e_i)z_{2,i})dF_{\mathbf{e}_i|w}(e|w)$$

for any $l \leq \kappa$. Hence, since derivatives of $h_0(y^*, w, \cdot)$ are bounded, applying the dominated convergence theorem again we get that

$$\lim_{z_{2_i}\to 0}\frac{\partial^l_{z^l_{1,i}}\eta(z_{1,i},z_{2,i})}{z^l_{2,i}}=\beta_{1,i}(w)^l\int\partial^l_{t^l}F_0(0)dF_{\mathbf{e}_i|w}(e|w)=\beta_{1,i}(w)^l\partial^l_{t^l}F_0(0),$$

and, thus, $\beta_{1,i}(w)^l\partial^l_{t^l}F_0(0)$ is identified for any $l\leq\kappa$. Similarly note that, since $h_0(y^*,w,\cdot)$ has bounded derivatives,

$$\partial^l_{z^l_{2,i}}\eta(z_{1,i},0)=\int\partial^l_{t^l}F_0(0)(\beta_{0,i}(w)+\beta_{1,i}(w)z_{1,i}+e_i)^l dF_{\mathbf{e}_i|x}(e|x) \tag{9}$$

for every $l\leq\kappa$. Hence, since $\mathbb{E}\left[\,\mathbf{e}_i|\mathbf{w}=w\,\right]=0$ and $\beta_{1,i}(w)\partial_t F_0(0)$ is identified,

$$\beta_{0,i}(w)\partial_t F_0(0)=\partial_{z_{2,i}}\eta(z_{1,i},0)-\beta_{1,i}(w)\partial_t F_0(0)z_{1,i}$$

is also identified. Thus, we can identify the ratio $\beta_{0,i}(w)/\beta_{1,i}(w)$ and learn the sign of $\beta_{1,i}(w)$ from Assumption 5(v). For $l=2$, since $\mathbb{E}\left[\,\mathbf{e}_i|\mathbf{w}=w\,\right]=0$ and $\mathbb{E}\left[\,\mathbf{e}^2_i|\mathbf{w}=w\,\right]=1$, we also can derive that

$$\partial^2_{z^2_{2,i}}\eta(z_{1,i},0)=\int\partial^2_{t^2}F_0(0)(\beta_{0,i}(w)+\beta_{1,i}(w)z_{1,i}+e_i)^2dF_{\mathbf{e}_i|x}(e|x)=\partial^2_{t^2}F_0(0)\left[(\beta_{0,i}(w)+\beta_{1,i}(w)z_{1,i})^2+1\right].$$

Hence,

$$\partial^2_{z^2_{2,i}}\eta(z_{1,i},0)=\beta_{1,i}(w)^2\partial^2_{t^2}F_0(0)\left[(\beta_{0,i}(w)/\beta_{1,i}(w)+z_{1,i})^2+1/\beta_{1,i}(w)^2\right].$$

As a result, since we identified $\beta_{0,i}(w)/\beta_{1,i}(w)$ and $\beta_{1,i}(w)^2\partial^2_{t^2}F_0(0)$ in the previous steps,

$$1/\beta_{1,i}(w)^2=\frac{\partial^2_{z^2_{2,i}}\eta(z_{1,i},0)}{\beta_{1,i}(w)^2\partial^2_{t^2}F_0(0)}-(\beta_{0,i}(w)/\beta_{1,i}(w)+z_{1,i})^2$$

is identified. Since the already identified the sign of $\beta_{1,i}(w)$ and $\beta_{0,i}(w)/\beta_{1,i}(w)$ we can identify $\beta_{0,i}(w)$ and $\beta_{1,i}(w)$. Moreover, I identify $\partial^l_{t^l}F_0(0)$ since $\beta_{1,i}(w)^l\partial^2_{t^l}F_0(0)$ for all $l\leq\kappa$.

To identify all moments of $\mathbf{e}_i$ up to order κ I use Equation (9) to derive the following recursive equation

$$\mathbb{E}\left[\mathbf{e}_i^l|\mathbf{w}=w\right] = \frac{\partial^l_{z^l_{2,i}}\eta(z_{1,i},0)}{\partial^l_{t^l}F_0(0)} - \sum_{k=1}^{l}\binom{l}{k}(\beta_{0,i}(w)+z_{1,i})^k\mathbb{E}\left[\mathbf{e}_i^{l-k}|\mathbf{w}=w\right].$$

Hence, by Assumption 5(iii), I also can identify $F_{\mathbf{e}_i|\mathbf{w}}(\cdot|w)$. Since the choice of i and w was arbitrary and conditional on $\mathbf{w}=w$ the random variables $\{\mathbf{e}_i\}_{i=1}^{d_v}$ are independent, I identify β_0, β_1, and $F_{\mathbf{e}|\mathbf{x}}$.

A.3. Proof of Propositions 5.1 and 5.4

(i). Under Assumption 6.(iv) or Assumption 7.(vi) there exists z_2^* with some open neighbourhood such that $z^*_{2,y'} = \lambda_{y'} z^*_{2,1}$ for all $y, y' \in Y$ with $\min_{y'} \lambda_{y'} > 0$. Let

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

Note that since $\mathbf{e}_1$ and $\mathbf{z}$ are independent conditional on $\mathbf{w}$ we have that for $x = (z_1^*, z_2^{*\intercal}, w^\intercal)^\intercal$

$$\mu(0|x^*) = \int_{\mathbb{R}} F_{\mathbf{g}|\mathbf{w}}(-z^*_{2,1}(\beta_0(w)+\beta_1(w)z_1^*+e_1),\ldots,-\lambda_J z^*_{2,1}(\beta_0(w)+\beta_1(w)z_1^*+e_1)|w)dF_{\mathbf{e}_1|\mathbf{w}}(e_1|w).$$

Thus, I can identify the sign of $\beta_1(w)$ since $F_{\mathbf{g}|\mathbf{w}}(\cdot|w)$ is weakly monotone.

Assumption 1 is satisfied for $Y^* = \{0\}$ and for $h(0,w,v) = F_{\mathbf{g}|\mathbf{w}}(v, \lambda_2 v, \cdot, \lambda_J v|w)$. Assumption 4 is implied by Assumption 6(iv).

Either by Proposition 4.1 or Proposition 4.2 β_0, β_1, and $F_{\mathbf{e}_1|x}$ are identified. If one uses Proposition 4.1, then $F_{\mathbf{e}_1|\mathbf{x}}$ is identified since it is assumed to be standard normal. Note that in Proposition 4.1 we used derivatives of $\eta(z_{1,i}, z_{2,i})$ in order to identify βs. In the multinomial choice model

$$\eta(z_1^*, z^*_{2,1}) = \mu(0|x^*),$$

where $x^* = (z_1^*, (\lambda_y z_{2,1}^*))_y^\intercal, w^\intercal)^\intercal$ and $\lambda_1 = 1$. As a result,

$$\partial_{z_{2,1}^*}\eta(z_1^*, z_{2,1}^*) = \sum_y \lambda_y \partial_{z_{2,y}^*}\mu(0|x^*).$$

Since $\lambda_y = z_{2,y}^*/z_{2,1}^*$ we get that

$$z_{2,1}^*\partial_{z_{2,1}^*}\eta(z_1^*, z_{2,1}^*) = \sum_y z_{2,y}^*\partial_{z_{2,y}^*}\Pr(\mathbf{y} = 0|\mathbf{x} = x).$$

Hence, if Assumption 6(iv) is satisfied not just for one z^* but for all z, then

$$\beta_1^2(w) = \frac{\partial^3_{z_{1,i}^3}\mu(0|x)\partial_{z_1}\mu(0|x) - \partial^2_{z_1^2}\mu(0|x)[\partial_{z_1}\mu(0|x)]^2}{\sum_y z_{2,y}\partial^2_{z_1,z_{2,y}}\mu(0|x)\partial_{z_1}\mu(0|x) - \sum_y z_{2,y}\partial_{z_{2,y}}\mu(0|x)\partial^2_{z_1^2}\mu(0|x)] - [\partial_{z_1}\mu(0|x)]^2},$$

$$\beta_0(w) = \frac{\sum_y z_{2,y}\partial_{z_{2,y}}\mu(0|x) - z_1\partial_{z_1}\mu(0|x)}{\partial_{z_1}\mu(0|x)}\beta_1(w) - \frac{\partial^2_{z_1^2}\mu(0|x)}{\partial_{z_1}\mu(0|x)}\frac{1}{\beta_1(w)} \tag{10}$$

for all x.

(ii). Since β_0, β_1, and $F_{\mathbf{e}_1|\mathbf{x}}$ are identified, I can define the index $\boldsymbol{\tau}$. Let

$$\boldsymbol{\tau} = \beta_0(\mathbf{w}) + \beta_1(\mathbf{w})\mathbf{z}_1 + \mathbf{e}_1.$$

Note that $F_{\boldsymbol{\tau}|\mathbf{x}}$ constitutes a boundedly complete family either because of Assumption 7(vii) or by normality of $\mathbf{e}_1$ and continuity of $\mathbf{z}_1$ (Brown, 1986). Hence, since

$$\begin{aligned}\Pr(\mathbf{y} = 0|\mathbf{x} = x) &= \int_{\mathbb{R}} F_{\mathbf{g}|\mathbf{w}}(-z_{2,1}\tau, \ldots, -z_{2,J}\tau|w)dF_{\mathbf{e}_1|w}(\tau - \beta_0(w) + \beta_1(w)z_1|w) = \\ &= \int_{\mathbb{R}} \tilde{g}(z_2, w, \tau)dF_{\mathbf{e}_1|w}(\tau - \beta_0(w) + \beta_1(w)z_2|w)\end{aligned}$$

and Assumptions 1 is satisfied I can identify

$$\tilde{g}(z_2, w, \tau) = F_{\mathbf{g}|\mathbf{w}}(-z_{2,1}\tau, \ldots, -z_{2,J}\tau|w)$$

for all z_2, w, τ. Note that since τ can take any value in

$$\mathcal{T}_w = \{\tau \,:\, e_1 + \beta_1(w)z_1 + \beta_0(w), e_1 \in E_{1|w}, z_1 \in Z_{1|w}\}$$

for any direction $-z_2/\left\|z_2\right\|$ in the support of $\mathbf{z}_2$ conditional on $\mathbf{w} = w$, I can recover $F_{\mathbf{g}|\mathbf{w}}(g|w)$ for any g such that $g = -z_2\tau/\left\|z_2\right\|$ for some $\tau \in \mathcal{T}_w$. That is, I identify $F_{\mathbf{g}|\mathbf{w}}(\cdot|w)$ over the set

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

A.4. Proof of Proposition 5.6

First, I fix some $w \in W$ and for notation simplicity I drop dependence on w. Since Assumptions 1 and 2 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 payoffs of any player arbitrary small ("close" to $-\infty$). Hence, under Nash solution concept 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

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

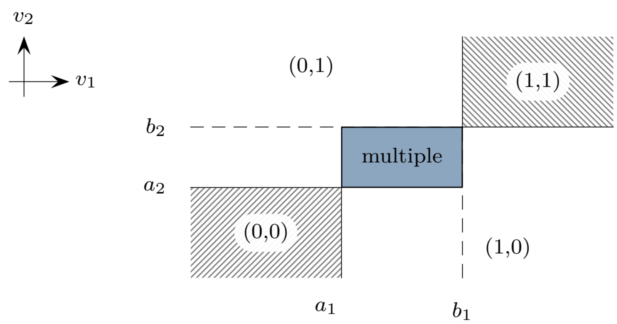

Figure 2 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} < 0$ and $\delta_{0,2,1} < 0$.

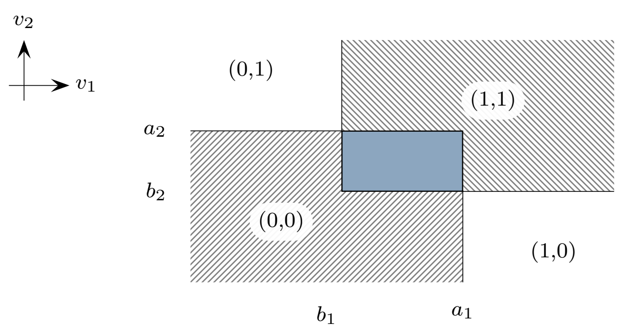

Figure 3 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} > 0$ and $\delta_{0,2,1} > 0$.

Moreover, by Assumption 8(ii), in this two player game at least two outcomes from the set

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

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 identify the parameters of the model I analyze the predictions about outcomes $(0,0)$ and $(1,1)$. Without loss of generality let $i = 1$ and $j = 2$. 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 2-5, where $a_1 = -\alpha_{0,1}$, $a_2 = -\alpha_{0,2}$, $b_1 = -\alpha_{0,1} - \delta_{0,1,2}$, and $b_2 = -\alpha_{0,2} - \delta_{0,2,1}$.

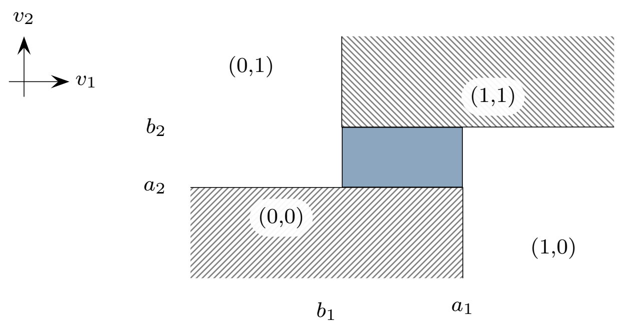

Figure 4 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} > 0$ and $\delta_{0,2,1} < 0$.

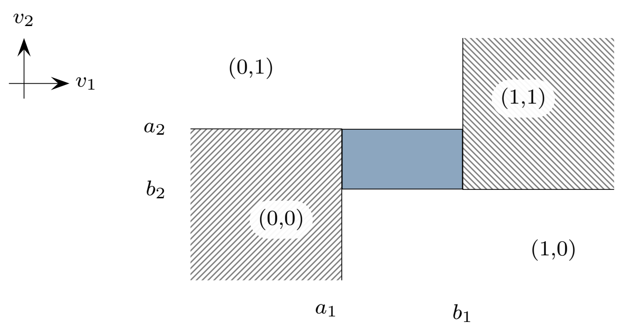

Figure 5 – Rationalizable correspondences for different realizations of v when $\delta_{0,1,2} < 0$ and $\delta_{0,2,1} > 0$.

Thus, I can determine the signs of the $\delta_{0,1,2}$ and $\delta_{0,2,1}$. Moreover, the identified threshold values values a_i, b_i, $i = 1, 2$, uniquely identify the rest of the parameters. The result then follows from the fact that the choice of players i, j, and nonexcluded covariate $w \in W$ was arbitrary.

A.5. Proof of Proposition 6.1

To simplify the notation, I will focus on the binary choice case.

Step 1. In this step I make several observations about p_0 and its derivatives. By definition $0 \leq h_0(v) \leq 1$ for all v and

$$p_0(z) = \int_{\mathbb{R}} h_0((\beta_0 + \beta_1 z_1 + e_1)z_{2,1})\phi(e_1)de_1 = \int_{\mathbb{R}} h_0(v)\phi(v/z_{2,1} - \beta_1 z_1 - \beta_0)dv/z_{2,1}.$$

Hence, p_0 is continuously differentiable of any order. Moreover, $p_0(z) = 0$ if and only if $h(v) = 0$ for all v. The latter means that probability of picking the outside option conditional on $\mathbf{z} = z$ and $\mathbf{e}_1 = e_1$ equals to 0 for all e. Since $\mathbf{g}_1$ is independent of $\mathbf{z}$ and $\mathbf{e}_1$, this implies that

$$\mathbf{g}_1 \geq -z_{2,1}(\beta_0 + \beta_1 z_1 + e_1)$$

with probability 1 for all e_1. The latter is not possible since $\mathbf{e}_1$ has full support. Thus, $p_0(z) > 0$ for all z. Similarly, one can show that $p_0(z) < 1$ for all z.

Next consider $p_1(z) = \partial_{z_1} p_0(z)$. Since $\partial_t \phi(t) = -t\phi(t)$,

$$\begin{aligned} p_1(z) &= \beta_1 \int_{\mathbb{R}} h_0(v)(v/z_{2,1} - \beta_1 z_1 - \beta_0)\phi(v/z_{2,1} - \beta_1 z_1 - \beta_0)dv/z_{2,1} \\ &= \beta_1 \int_{\mathbb{R}} h_0((\beta_0 + \beta_1 z_1 + e_1)z_{2,1})e_1\phi(e_1)de_1. \end{aligned}$$

Hence, since $0 \leq h_0(v) \leq 1$ for all v, I get that for some $C_1 < \infty$

$$\sup_z |p_1(z)| \leq \beta_1 \int_{\mathbb{R}} |e_1| \phi(e_1)de_1 \leq C_1.$$

Similarly, note that $p_2(z) = z_{2,1}\partial_{z_{2,1}} p_0$ and

$$p_2(z)/z_{2,1} = -p_0(z)/z_{2,1}^2 - \int_{\mathbb{R}} h_0(v)(v/z_{2,1} + \beta_1 z_1 - \beta_0)v\phi(v/z_{2,1} - \beta_1 z_1 - \beta_0)dv/z_{2,1}^3.$$

Hence, given bounded support for z, I can conclude that $\sup_z |p_2(z)|$ is also finite. Repeating the above steps one can show that all higher order partial derivatives of p_0 are bounded. The bound might not be the same for all derivatives, but for any finite collection of them there is a uniform bound.

Step 2. Combining bounds for derivatives form Step 1, the uniform weak law of large numbers, and consistency of $\hat{p}_0$, I can deduce that

$$\frac{1}{n}\sum_{i=1}^{n}\hat{p}_{111}\left(\mathbf{z}^{(i)}\right)\hat{p}_1\left(\mathbf{z}^{(i)}\right)-\hat{p}_{11}\left(\mathbf{z}^{(i)}\right)^2 \to_p \mathbb{E}\left[\,p_{111}(\mathbf{z})p_1(\mathbf{z})-p_{11}(\mathbf{z})^2\,\right],$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{p}_{12}\left(\mathbf{z}^{(i)}\right)\hat{p}_1\left(\mathbf{z}^{(i)}\right)-\hat{p}_2\left(\mathbf{z}^{(i)}\right)\hat{p}_{11}\left(\mathbf{z}^{(i)}\right)-\hat{p}_1\left(\mathbf{z}^{(i)}\right)^2 \to_p \mathbb{E}\left[\,p_{12}(\mathbf{z})p_1(\mathbf{z})-p_2(\mathbf{z})p_{11}(\mathbf{z})-p_1(\mathbf{z})^2\,\right],$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{p}_2\left(\mathbf{z}^{(i)}\right)-\mathbf{z}_1^{(i)}\hat{p}_1\left(\mathbf{z}^{(i)}\right) \to_p \mathbb{E}\left[\,p_2(\mathbf{z})-\mathbf{z}_1p_1(\mathbf{z})\,\right],$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{p}_{11}\left(\mathbf{z}^{(i)}\right) \to_p \mathbb{E}\left[\,p_{11}(\mathbf{z})\,\right],$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{p}_1\left(\mathbf{z}^{(i)}\right) \to_p \mathbb{E}\left[\,p_1(\mathbf{z})\,\right].$$

Thus, Equation (10) and the continuous mapping theorem imply that $\hat{\beta} \to_p \beta$.

Step 3. Consider

$$\mathcal{G}_n = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \hat{p}_{111}\left(\mathbf{z}^{(i)}\right)\hat{p}_1\left(\mathbf{z}^{(i)}\right)-\hat{p}_{11}\left(\mathbf{z}^{(i)}\right)^2 \\ \beta_1^2\left[\hat{p}_2\left(\mathbf{z}^{(i)}\right)-\mathbf{z}_1^{(i)}\hat{p}_1\left(\mathbf{z}^{(i)}\right)\right]-\hat{p}_{11}\left(\mathbf{z}^{(i)}\right) \end{pmatrix}.$$

To prove asymptotic normality of $\mathcal{G}_n$ I will use Theorem 6 in Newey (1997). The data is assumed to be i.i.d., the outcome variable is finite and p_0 is bounded and bounded away from 0. Hence, Assumptions 1 and 4 from Newey (1997) are satisfied. Assumption 8 in Newey (1997) is assumed. Assumption 9 in Newey (1997) follows from Step 1. Finally, consider $a(p_0) = (a_1(p_0), a_0(p_0))^\mathsf{T}$ with

$$a_1(p_0) = \mathbb{E}\left[\,p_{111}(\mathbf{z})p_1(\mathbf{z})-p_{11}(\mathbf{z})^2\,\right],$$
$$a_2(p_0) = \mathbb{E}\left[\,\beta_1^2[p_2(\mathbf{z})-\mathbf{z}_1p_1(\mathbf{z})]-p_{11}(\mathbf{z})\,\right].$$

The directional derivative of a at p_0 in direction g_0 is then $D(g_0) = (D_1(g_0), D_2(g_0))^\mathsf{T}$ with

$$D_1(g_0) = \mathbb{E}\left[\,p_{111}(\mathbf{z})g_1(\mathbf{z})+g_{111}(\mathbf{z})p_1(\mathbf{z})-2p_{11}(\mathbf{z})g_{11}(\mathbf{z})\,\right],$$
$$D_2(g_0) = \mathbb{E}\left[\,\beta_1^2[g_2(\mathbf{z})-\mathbf{z}_1g_1(\mathbf{z})]-g_{11}(\mathbf{z})\,\right].$$

Applying integration by parts several times and using the fact that $f_{\mathbf{z}}$ and its partial derivatives vanish at the boundary of the support of $\mathbf{z}$ (Assumption 10(iii)), I get

$$\begin{aligned}
\mathbb{E}\left[\, p_{111}(\mathbf{z})g_1(\mathbf{z})\,\right] &= -\mathbb{E}\left[\, \partial_{z_1}[p_{111}(\mathbf{z})f_{\mathbf{z}}(\mathbf{z})]g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right], \\
\mathbb{E}\left[\, p_1(\mathbf{z})g_{111}(\mathbf{z})\,\right] &= -\mathbb{E}\left[\, \partial^3_{z_1^3}[p_1(\mathbf{z})f_{\mathbf{z}}(\mathbf{z})]g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right], \\
\mathbb{E}\left[\, p_{11}(\mathbf{z})g_{11}(\mathbf{z})\,\right] &= \mathbb{E}\left[\, \partial^2_{z_1^2}[p_{11}(\mathbf{z})f_{\mathbf{z}}(\mathbf{z})]g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right], \\
\mathbb{E}\left[\, z_1 g_1(\mathbf{z})\,\right] &= -\mathbb{E}\left[\, (f_{\mathbf{z}}(\mathbf{z}) + \mathbf{z}_1\partial_{z_1}f_{\mathbf{z}}(\mathbf{z}))g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right], \\
\mathbb{E}\left[\, g_{11}(\mathbf{z})\,\right] &= \mathbb{E}\left[\, \partial^2_{z_1^2}f_{\mathbf{z}}(\mathbf{z})g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right], \\
\mathbb{E}\left[\, g_2(\mathbf{z})\,\right] &= -\mathbb{E}\left[\, (f_{\mathbf{z}}(\mathbf{z}) + \mathbf{z}_{2,1}\partial_{z_2}f_{\mathbf{z}}(\mathbf{z}))g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right].
\end{aligned}$$

As a result,

$$\begin{aligned}
D_1(g_0) &= -\mathbb{E}\left[\, \{4p_{1111}(\mathbf{z})f_{\mathbf{z}}(\mathbf{z}) + 8p_{111}(\mathbf{z})\partial_{z_1}f_{\mathbf{z}}(\mathbf{z}) + 5p_{11}(\mathbf{z})\partial^2_{z_1^2}f_{\mathbf{z}}(\mathbf{z}) + p_1(\mathbf{z})\partial^3_{z_1^3}f_{\mathbf{z}}(\mathbf{z})\}g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right], \\
D_2(g_0) &= \mathbb{E}\left[\, \{\beta_1^2[\mathbf{z}_1\partial_{z_1}f_{\mathbf{z}}(\mathbf{z}) - \mathbf{z}_{2,1}\partial_{z_2}f_{\mathbf{z}}(\mathbf{z})] - \partial^2_{z_1^2}f_{\mathbf{z}}(\mathbf{z})\}g_0(\mathbf{z})/f_{\mathbf{z}}(\mathbf{z})\,\right].
\end{aligned}$$

Hence,

$$D(g_0) = \mathbb{E}\left[\, \bar{v}(\mathbf{z})g_0(\mathbf{z})\,\right].$$

Moreover, $\bar{v}$ is continuously differentiable and $\mathbb{E}\left[\, \bar{v}(\mathbf{z})\bar{v}(\mathbf{z})^{\mathsf{T}}\,\right]$ is finite and nonsigular (Assumption 10(iv)). Hence, Assumption 7 in Newey (1997) is also satisfied, thus, by Theorem 6 in Newey (1997)

$$\sqrt{n}\,(\mathcal{G}_n - \mathcal{G}) \to_d N(0, \tilde{V}),$$

where

$$\mathcal{G} = \mathbb{E}\begin{bmatrix} p_{111}(\mathbf{z})p_1(\mathbf{z}) - p_{11}(\mathbf{z})^2 \\ \beta_1^2\,[p_2(\mathbf{z}) - \mathbf{z}_1 p_1(\mathbf{z})] - p_{11}(\mathbf{z}) \end{bmatrix}$$

and

$$\tilde{V} = \mathbb{E}\left[\, \bar{v}(\mathbf{z})\bar{v}(\mathbf{z})^{\mathsf{T}}p_0(\mathbf{z})(1 - p_0(\mathbf{z}))\,\right].$$

Moreover, I can construct a consistent estimator of $\tilde{V}$ using Theorem 6 in Newey (1997). In particular, let $\hat{a}(\hat{p}_0)$ be a sample counterpart of $a(p_0)$ and

$$\begin{aligned}
\hat{\gamma} &= \left(\Psi^\mathsf{T}\Psi\right)^{-} \sum_{i=1}^{n} \psi^K\left(\mathbf{z}^{(i)}\right) \mathbb{1}\left(\mathbf{y}^{(i)} = 0\right), \\
\hat{A} &= \partial_\gamma \hat{a}(\psi^K(z)^\mathsf{T}\hat{\gamma}), \\
\hat{Q} &= \Psi^\mathsf{T}\Psi/n, \\
\hat{\Sigma} &= \sum_{i=1}^{n} \psi^K\left(\mathbf{z}^{(i)}\right) \psi^K\left(\mathbf{z}^{(i)}\right)^\mathsf{T} \left[\mathbb{1}\left(\mathbf{y}^{(i)} = 0\right) - \hat{p}_0\left(\mathbf{z}^{(i)}\right)\right]^2 /n.
\end{aligned}$$

Then

$$\hat{\tilde{V}} = \hat{A}^\mathsf{T}\hat{Q}^{-}\hat{\Sigma}\hat{Q}^{-}\hat{A} \to_p \tilde{V}.$$

Step 4. Combining Step 2 with the continuous mapping theorem, Slutsky's theorem, and the Delta method, implies that

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to_d \begin{pmatrix} 2\beta_1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbb{E}\left[\, p_{12}(\mathbf{z})p_1(\mathbf{z}) - p_2(\mathbf{z})p_{11}(\mathbf{z}) - p_1(\mathbf{z})^2 \,\right] & 0 \\ 0 & \beta_1 \mathbb{E}\left[\, p_1(\mathbf{z}) \,\right] \end{pmatrix}^{-1} N\left(0, \tilde{V}\right).$$

Step 5. Consistency of

$$\hat{V} = \hat{G}\hat{\tilde{V}}\hat{G}^\mathsf{T},$$

where

$$\hat{G} = \begin{pmatrix} 2\hat{\beta}_1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} n^{-1}\sum_{i=1}^n \hat{p}_{12}\left(\mathbf{z}^{(i)}\right)\hat{p}_1\left(\mathbf{z}^{(i)}\right) - \hat{p}_2\left(\mathbf{z}^{(i)}\right)\hat{p}_{11}\left(\mathbf{z}^{(i)}\right) - \hat{p}_1\left(\mathbf{z}^{(i)}\right)^2 & 0 \\ 0 & n^{-1}\hat{\beta}_1\sum_{i=1}^n \hat{p}_1\left(\mathbf{z}^{(i)}\right) \end{pmatrix}^{-1},$$

follows from consistency of $\hat{\beta}$, $\hat{\tilde{V}}$, Step 2, and the continuous mapping theorem.