A Note on Identification of Discrete Choice Models for Bundles and Binary Games[*]

Jeremy T. Fox

Rice University and NBER[†]

Natalia Lazzati

UC Santa Cruz[‡]

September 2016
Forthcoming: *Quantitative Economics*

Abstract

We study nonparametric identification of single-agent discrete choice models for bundles (without requiring bundle-specific prices) and of binary games of complete information. We show these two models are quite similar from an identification standpoint. Moreover, they are mathematically equivalent when we restrict attention to the class of potential games and impose a specific equilibrium selection mechanism in the data generating process. We provide new identification results for the two related models.

1 Introduction

This paper provides identification results for both single-agent discrete choice *bundle* models and *binary games* of complete information. We establish a tight connection between these models from the perspective of algebraic arguments for identification and further show that the models are indeed mathematically equivalent when attention is restricted to the class of potential games under a specific equilibrium selection rule. This equivalence result allows researchers to extrapolate any further identification progress in one of these models to the other.

We focus on the case of two goods and two players and extend the analysis to the case of three or more goods and players in the online appendix. The approach we provide is nonparametric and allows us to recover the payoff relevant functions, the distribution of heterogeneous interaction effects, and the distribution of good- or player-specific unobservables. These structural features of the models are needed to make counterfactual predictions that do not directly derive from choice data. In the single-agent choice model, our results allow us, for example, to forecast purchasing decisions for bundle discounts that were not previously offered. In the case of games, we can point identify (under an assumed equilibrium selection rule) or set identify (without an equilibrium selection rule) equilibrium choices under different processes for matching players.

Our identification strategy crucially relies on exclusion restrictions. In the single-agent choice model, each excluded variable enters the standalone payoff of a single good additively. For games, each excluded variable similarly affects the standalone payoff of one player but not the payoff of the other player. We do not impose exclusion restrictions either for a bundle or for players' joint action profiles. The excluded variables play a dual role: identifying the sign of the interaction effects as well as the joint distribution of the unobservables. Indeed, we show in the online appendix, via a simple example, that the joint distribution of unobservables cannot be uniquely recovered without the excluded variables, at least in our general formulation of the model. These results clarify the critical, if not surprising, need for choice-specific exclusion restrictions, something that has been exploited in some applications (e.g., Liu, Chintagunta, and Zhu 2010) but is absent in others (e.g., Kretschmer, Miravete and Pernias 2012 and Hartman 2010).

For the bundles model, applying previous results on nonparametric identification requires treating the bundle as just another option in a multinomial choice model (Thompson 1989, Matzkin 1993, Lewbel 2000). For two goods, this amounts to having three exclusion restrictions: one variable that shifts standalone payoffs for each binary choice variable and another one that shifts the payoffs of the bundle itself (Athey and Stern 1998). Our identification results only

require exclusion restrictions at the level of the two binary variables. We allow for correlation in the unobservable components of the standalone payoffs and unobserved heterogeneity in the magnitude of the interaction effect. Our nonparametric results follow empirical papers that have estimated parametric bundles models while recognizing the intuitive gain in identification from exclusion restrictions, such as Gentzkow (2007a). Sher and Kim (2014) establish a weak identification concept for a bundles model without data on the market shares of bundles.

There is a growing literature on the semi or nonparametric identification of two-player binary games of complete information, without restricting the equilibrium selection rule. The basic result is Berry and Tamer (2006, Result 4) and Ciliberto and Tamer (2009, Theorem 2).[1] We show that their algebraic steps in identification proofs are very closely linked to the ones for the bundles model. We further show the proofs are identical for the case of potential games with the equilibrium selection rule based on potential maximizers. In addition, without imposing an equilibrium selection rule, we identify the distribution of unobserved heterogeneity in the interaction effects, which is not present in either of these two papers. In an entry game, we allow the reduction in profits from monopoly to duopoly to vary across observationally identical markets. Allowing heterogeneous interaction effects also appeared in simultaneously circulated papers by Kline (2015) and Dunker, Hoderlein and Kaido (2013).

Berry and Tamer (2006, Result 4) and subsequent papers assume the researcher knows the sign of the interaction effects: the game is known to be either of strategic substitutes or complements. Imposing the structure of potential games and its associated equilibrium selection rule, we exploit the equivalence to the bundles model to identify the sign of the interaction effect. This result is useful in models where offsetting effects mean that the sign is hard to anticipate (for example, the crimes model of Ballester, Calvó-Armengol and Zenou 2006).

The online appendix includes results for the case of three or more goods and players that do not closely extend the cited papers. We do not present a nonparametric or a semiparametric estimator; sieve maximum likelihood is one alternative (Chen, Tamer and Torgovitsky 2011). All proofs are in the appendix.

---

[1]Yildiz (2007) considers identification in cases where only equilibria in mixed strategies exist.

2 Model of Bundles and Games

2.1 Identification Results for a General Model

An agent chooses the values of two binary variables $a = (a_1, a_2) \in \{(0,0), (1,0), (0,1), (1,1)\}$ in order to maximize

$$U(a, W, Z, \varepsilon, \eta) = \sum_{i=1,2} (u_i(W) + Z_i + \varepsilon_i) \cdot a_i + \eta \cdot v(W) \cdot a_1 \cdot a_2, \tag{1}$$

where $Z_i \in \mathbb{R}$ *only* affects the utility associated to the binary variable i, $W \in \mathbb{R}^k$ is a vector of explanatory variables different from $Z = (Z_1, Z_2)$, and $(\varepsilon, \eta) = (\varepsilon_1, \varepsilon_2, \eta) \in \mathbb{R}^3$ is a vector that captures heterogeneous effects with distribution $F_{\varepsilon, \eta | W, Z}$. The vector (ε, η) is observed by the agent but not by the econometrician.

In expression (1), the utility of selecting $a = (0,0)$ is normalized to zero. The term $u_i(W) + Z_i + \varepsilon_i$ is the standalone utility of assigning value 1 only to binary variable i, $a_i = 1$ and $a_{-i} = 0$.[2] Finally, the utility of assigning value 1 to both binary variables, $a = (1,1)$, is the sum of the standalone utilities plus the interaction term $\eta \cdot v(W)$. For each $W = w$, the binary variables are *complements* if $\eta \cdot v(w) \geq 0$ and *substitutes* otherwise.[3] We assume $\eta \in \mathbb{R}_+$, so that η modifies the magnitude but not the sign of the interaction effect, which therefore is given by the sign of $v(W)$.[4] The pair $\varepsilon_1, \varepsilon_2$ captures heterogeneity or idiosyncratic shocks at the level of each binary variable, and η reflects heterogeneity at the bundle level.

We use data on i.i.d. chosen actions and explanatory variables (a, w, z) to identify aspects of the unknown functions $(u_1(W), u_2(W), \nu(W), F_{\varepsilon, \eta | W, Z})$. Our model is nonparametric as the objects of interest are arbitrary, possibly discontinuous, functions of W.[5] Note that i.i.d. sampling allows us to nonparametrically identify the conditional choice probabilities $\Pr(a \mid w, z)$ with no further assumptions. Our results rely on the next restrictions.

---

[2] Z_i enters the utility with a positive sign. We can identify the sign of Z_i by seeing whether the marginal probability of selecting variable i increases with Z_i. If the sign of Z_i should be negative, redefine Z_i appropriately.

[3] When applied to the bundles model, Gentzkow (2007b) shows that this definition of complements and substitutes is equivalent to the usual definitions based on price responses.

[4] We let $\eta > 0$ instead of $\eta \geq 0$ because $\eta = 0$ is made redundant by the possibility that $v(W) = 0$. If we had an additive $\eta + v(W)$, under similar assumptions to ours we could identify the distribution of η as long as the support of η ensures that $\eta + v(W)$ always has the same sign conditional on W. Likewise, say η has two components, η_a and η_b where the total interaction effect is $\eta_a + \eta_b \cdot v(W)$. We could identify $v(W)$ if we similarly restrict the support of η_a so that $\eta_a + \eta_b \cdot v(W)$ always has the same sign conditional on W and we also assume $E[\eta_a \mid W] = 0$ and $E[\eta_b \mid W] = 1$.

[5] We do not explore the case where W instead multiplies random coefficients, as in the follow-on work of Dunker, Hoderlein and Kaido (2014).

A1: $Z_i \mid W, Z_{-i}$ has support on all $\mathbb{R}$ for $i = 1, 2$.

A2: (I) $F_{\varepsilon,\eta|W,Z} = F_{\varepsilon,\eta|W}$; (II) $\mathrm{E}(\varepsilon \mid W) = (0,0)$; and (III) $\mathrm{E}(\eta \mid W) = 1$.

A3: $\varepsilon \mid W$ has an everywhere positive Lebesgue density on its support.

The exclusion restrictions, captured by Z_1 and Z_2, play a dual role in our analysis: They are key to identifying the sign of the interaction effects, as reflected in the proof of Lemma 1 below. They are also needed to recover the joint distribution of unobservables, as captured by the non-identification result in Section C of the online appendix.

The large support restriction, A1, is a common requirement used in various ways in the literature on binary and multinomial choice models.[6] It allows us to recover the tails of the distribution of unobservables $F_{\varepsilon,\eta|W}$ without restricting its support; we show below that with mild changes we can still identify part of the model structure without A1.

The independence assumption A2(I) allows us to trace $F_{\varepsilon,\eta|W}$ using variation in Z. Assumptions A2(II) and A2(III) provide location normalizations for ε and η and rule out omitted variable bias from W in identifying $(u_1(\cdot), u_2(\cdot), \nu(\cdot))$; (ε, η) can be heteroskedastic with respect to W.[7] Assumption A3 gives probability zero to tie events.

Our main result requires initial identification of the sign of the interaction effect.

Lemma 1. *If A2(I) and A3 hold and the support of Z contains at least two points, then the sign of $v(w)$ is identified for each $W = w$.*

The proof of Lemma 1 builds on monotone comparative statics methods for stochastic models and crucially relies on the excluded variables Z_1, Z_2. The intuition for the identification of the sign of the interaction effect is simple. When the interaction effect is positive, the probability of selecting $a_1 = 1$ increases with the value of the excluded explanatory variable Z_2, as the latter makes alternative 2 more valuable. Because the opposite holds when the interaction effect is negative, whether the alternatives are complements or substitutes can be inferred from the data. Using Lemma 1, the following theorem states our identification result.

Theorem 1. *Under A1–3, $\left((u_i(\cdot))_{i=1,2}, v(\cdot)\right)$ is identified. If, in addition, η and ε are independent conditional on $W = w$, then $F_{\varepsilon|w}$ and $F_{\eta|w}$ are identified.*

---

[6]See, e.g., Manski (1975, 1985, 1988), Thompson (1989), Matzkin (1992, 1993, 2007), Ichimura and Thompson (1998), Lewbel (1998, 2000), Berry and Haile (2010), Fox and Gandhi (2016), and Chen et. al (forthcoming).

[7]If $E(\varepsilon \mid W)$ and $E(\eta \mid W)$ are non-constant functions of $W = w$, the values of $u_1(W)$, $u_2(W)$ and $v(W)$ produced by the proof strategy of Theorem 1 below will suffer from omitted variable bias. If, in this case, $E(\varepsilon \mid T) = (0,0)$ and $E(\eta \mid T) = 1$ for some "instruments" T, one could correctly identify $u_1(W)$, $u_2(W)$ and $v(W)$ under additional assumptions on (T, W) and a slight modification of the proof strategy for Theorem 1. See Lewbel (2000) for a treatment of omitted variable bias in discrete choice.

Remark 1. When the binary choice variables are substitutes, $F_{\varepsilon|w}$ can be identified from the data even if η and ε are not independently distributed.

For each $W = w$, our identification strategy proceeds by first learning the sign of $v(w)$ via Lemma 1. The reason is that for each sign of the interaction effect there are two action profiles, the identity of which are sign dependent, that allow us to recover all the objects of interest. When the choice variables are substitutes, so that $v(w) < 0$, we use the action profile $(0,0)$ to recover $u_1(\cdot)$, $u_2(\cdot)$ and $F_{\varepsilon|W}$ and rely on $(1,1)$ to identify $v(\cdot)$ and $F_{\eta|W}$. When the variables are complements, we use the action profiles $(0,1)$ and $(1,0)$ to recover the objects of interest.

We finally show that we can recover part of the model structure without the large support assumption A1. To do so, we replace A2(II) with a location normalization that is not as sensitive to the tails of the unobservables. The proof of this result builds on Kline (2016), whose analysis is for a semiparametric two-player game. For each $W = w$, let $u_1(w)$, $u_2(w)$, $u_1(w) + v(w)$ and $u_2(w) + v(w)$ all lie in a bounded interval $\Theta \subseteq \mathbb{R}$. If the interval varies with w, let Θ be the union of all such w-specific intervals.

Theorem 2. *Under the next restrictions,* $\left((u_i(\cdot))_{i=1,2}, v(\cdot) \right)$ *is identified.*

A1′: $Z_i \mid W, Z_{-i}$ has support on a superset of Θ for $i = 1, 2$.

A2′: (I) $F_{\varepsilon|W,Z} = F_{\varepsilon|W}$; (II) $\varepsilon \mid W$ has mode $(0,0)$; (III) $F_{\varepsilon|W}$ is C^2; and (IV) $\eta = 1$.

A3′: $\varepsilon \mid W$ has an everywhere positive Lebesgue density on its support.

2.2 Two-Goods Bundle Model

In this specification of the model, each of the binary variables represents a good. The consumer selects the combination of goods $a = (a_1, a_2) \in \{0,1\}^2$ that maximizes

$$U(a, W, p, \varepsilon, \eta) = \sum_{i=1,2} (u_i(W) - p_i + \varepsilon_i) \cdot a_i + \eta \cdot v(W) \cdot a_1 \cdot a_2,$$

where $u_i(W) - p_i + \varepsilon_i$ is the quasi-linear utility from consuming good i, p_i is the price of the good, and $\eta \cdot v(W)$ is the extra utility the consumer gets if she acquires the two goods together. This model is identical to the previous one if we treat prices as the excluded variables.

Corollary 1. *If $Z = (-p_1, -p_2)$, identification of the bundles model follows from Theorem 1.*

We next cover different uses of our result. Bundling, the strategy of offering two or more products as a specially priced package, is widely used by firms and has been the focus of a large economic literature. Under mixed bundling, in addition to offering the two individual goods, a separate price p_3 is offered for the bundle $(1,1)$. Predicting demand for a bundle that was not previously offered requires structural identification of the model, as choice probabilities $\Pr(a \mid w, p_1, p_2)$ per se are not enough to infer the counterfactual effects on sales of different p_3. Moreover, Venkatesh and Mahajan (2009) explain that one lesson from the bundling literature is that optimal strategies are very sensitive to the details of the model. In particular, optimal bundle discounts $p_1 + p_2 - p_3$ depend on the correlation between the standalone unobservables ε_1 and ε_2 as well as on the values of the interaction effects $\eta \cdot v(W)$. For this reason, some recent papers are proposing new ways to estimate the copula of the unobservables (e.g., Letham, Sun, and Sheopuri 2014). We flexibly identify these relevant features.

Like in other multinomial choice models, we can compute welfare differences across prices. Let $a^*(w, p, \varepsilon, \eta)$ and $U^*(p) = U(a^*, w, p, \varepsilon, \eta)$ be the optimal choice and maximized utility of a consumer that faces a price vector p. As our structural model can be used to calculate the joint distribution $G(U^*(p'), U^*(p''))$ of maximized utilities under prices p' and p'', we can identify any treatment effect for moving from p' to p''. For example, the median of $U^*(p'') - U^*(p')$ gives the median utility loss in monetary terms of moving from p' to p''.

2.3   Two-Player Game

Consider a binary choice game of complete information. Each player $i = 1, 2$ chooses an action $a_i \in \{0, 1\}$. A player's payoff from action 0 is normalized to 0, and the payoff of action 1 is

$$U_{1,i}(a_{-i}, W, Z_i, \varepsilon_i, \eta_i) = u_i(W) + Z_i + \varepsilon_i + \eta_i \cdot v_i(W) \cdot a_{-i}. \tag{2}$$

In (2), $u_i(W) + Z_i + \varepsilon_i$ is the standalone value of action 1 and $\eta_i \cdot v_i(W)$ is the effect that the choice of the other player has on player i. Thus, $(\eta_1, \eta_2) \in \mathbb{R}_+^2$ has two dimensions. Finally, Z_1 and Z_2 are excluded variables at the level of each player. In an entry game, Z_i might represent the distance between a geographic market and the headquarters of chain i.

A pair $a^* = (a_1^*, a_2^*)$ is a pure strategy Nash equilibrium for w, z and $\varepsilon_1, \varepsilon_2, \eta_1, \eta_2$ if, for $i = 1, 2$,

$$a_i^* = \begin{cases} 1 & \text{if } u_i(w) + z_i + \varepsilon_i + \eta_i \cdot v_i(w) \cdot a_{-i}^* > 0, \\ 0 & \text{if } u_i(w) + z_i + \varepsilon_i + \eta_i \cdot v_i(w) \cdot a_{-i}^* < 0, \\ 1 \text{ or } 0 & \text{otherwise.} \end{cases}$$

The econometrician observes the distribution of equilibrium choices, $\Pr\left(a \mid w, z\right)$, for a cross-section of independent games that share the same structure $\left(\left(u_i\left(\cdot\right), v_i\left(\cdot\right)\right)_{i=1,2}, F_{\varepsilon_1, \varepsilon_2, \eta_1, \eta_2 | W}\right)$.

Though this model involves strategic interactions across two agents, its identification is quite similar, in terms of the algebra in the proofs, to the previous model. The proof of Part 1 of Theorem 3 below shows that this is true even if we do not impose an equilibrium selection rule. Part 2 of the theorem shows that the two models are indeed mathematically equivalent if we restrict attention to the class of potential games and use a specific equilibrium selection rule motivated by theoretical and empirical findings.

In game theory, a potential function is a real-valued function defined on the space of pure strategy action profiles of the players such that the change in any player's payoff from a unilateral deviation is equal to the change in the potential function. Potential games are games that admit such a function. Monderer and Shapley (1996) show that our game is a potential game if and only if $\eta_1 \cdot v_1\left(W\right) = \eta_2 \cdot v_2\left(W\right)$. Letting this term be $\eta \cdot v\left(W\right)$, it follows by Ui (2000) that the **potential function** of the game is identical to the overall utility in (1)

$$U\left(a, W, Z, \varepsilon, \eta\right) = \sum_{i=1,2} \left(u_i\left(W\right) + Z_i + \varepsilon_i\right) \cdot a_i + \eta \cdot v\left(W\right) \cdot a_1 \cdot a_2;$$

we just need to interpret the single agent as a fictitious planner choosing the actions of both players.

The assumption of equal interaction effects is often maintained in the empirical literature when players are anonymous, so that players' indices have no particular meaning. In an entry game, equal interaction effects means that the reduction in profits from monopoly to duopoly is the same for both players.

Potential functions were first used in economic theory to show existence of Nash equilibria in pure strategies. Finite potential games always have a pure strategy equilibrium. In addition, Monderer and Shapley (1996) show that when the game admits a potential function, this function is uniquely defined up to an additive constant. Thus, it offers an equilibrium refinement. Subsequent work studied whether the selection rule based on the potential maximizers is economically meaningful. Lab experiments studying the so-called minimum effort games have shown that observed choices are consistent with the maximization of objects close to the potential function of the game (Van Huyck, Battalio and Beil 1990, Goeree and Holt 2005, and Chen and Chen 2011). These results are remarkable because this class of games often has a large number of equilibria. Further, Ui (2001) shows that if the potential maximizer is unique, then this equilibrium is robust in the sense of Kajii and Morris (1997).

We next provide two sets of identifying conditions: In Part 1 we do not impose an equilibrium selection rule, but we do assume the econometrician knows the sign of the interaction effects. An advantage of the potential game specification (with its associated equilibrium selection rule) in Part 2 is that we can infer the sign of the interaction effects from the data.

Theorem 3. *Suppose that A1–3 hold with $E(\eta_1, \eta_2 \mid W) = (1,1)$ instead of A2(III).*

1. *If, for each $W = w$, $sign(v_1(w)) = sign(v_2(w))$ and the econometrician knows the sign, then $(u_i(\cdot), v_i(\cdot))_{i=1,2}$ is identified. If, in addition, (η_1, η_2) and ε are independent conditional on W, then $F_{\varepsilon|W}$, $F_{\eta_1|W}$ and $F_{\eta_2|W}$ are also identified.*

2. *If $\eta_1 \cdot v_1(W) = \eta_2 \cdot v_2(W) = \eta \cdot v(W)$ and players coordinate on the potential maximizer, then identification of the game follows from Theorem 1.*

Remark 2. In Part 1, when the strategic interactions are negative, $F_{\varepsilon|w}$ can be identified from the data even if (η_1, η_2) and ε are not conditionally independent. In addition, we can also identify the joint distribution $F_{\eta_1, \eta_2|W}$.

We next elaborate on the meaning of an equilibrium selection rule based on potential maximizers. When $\eta \cdot v(w) < 0$ and the game has multiple equilibria, the equilibrium set is $\{(0,1), (1,0)\}$. In this case, $(0,1)$ maximizes the potential if $u_2(w) + z_2 + \varepsilon_2 > u_1(w) + z_1 + \varepsilon_1$, while $(1,0)$ is the maximizer otherwise. This equilibrium selection rule predicts that the player choosing action 1 is the one with the highest stand-alone value. In an entry game, the most profitable entrant enters in the region where the identity of the entrant is otherwise indeterminate. Alternatively, when $\eta \cdot v(w) > 0$ and the game has multiple equilibria, the equilibrium set is $\{(0,0), (1,1)\}$. In this case $(1,1)$ maximizes the potential if

$$(-u_1(w) - z_1 - \varepsilon_1 - \eta \cdot v(w))(-u_2(w) - z_2 - \varepsilon_2 - \eta \cdot v(w)) > (u_1(w) + z_1 + \varepsilon_1)(u_2(w) + z_2 + \varepsilon_2),$$

while $(0,0)$ is the maximizer otherwise. In this second case, the potential maximizer is the less risky equilibrium of Harsanyi and Selten (1988). The lab experiments cited above have shown that the less risky equilibrium is selected even when another equilibrium gives higher payoffs to both players, so that the latter equilibrium Pareto dominates the less risky equilibrium.

Part 1 of Theorem 3 gives identification results for the binary game without imposing an equilibrium selection rule. Instead, the researcher needs to know the sign of the interaction effects. There are many games where theoretical arguments are compatible with both negative and positive interaction effects (e.g., the crimes model of Ballester, Calvó-Armengol and Zenou

2006). In these cases, the potential specification coupled with the selection rule based on potential maximizers offers an alternative identification strategy.

Recovering the structure of the game point identifies counterfactuals involving a known equilibrium selection rule and set identifies counterfactuals that do not impose an equilibrium selection rule. We can evaluate the impact on equilibrium choices of changing the equilibrium selection rule. In a coordination game, this could help us to estimate how much the players are losing by not coordinating on the Pareto optimal equilibrium. Also, we can examine the impact on action profile probabilities from altering the underlying matching process across players by increasing the correlation in the joint distribution of observables or unobservables, while keeping their marginals fixed —e.g., we could predict choices under assortative matching (Graham, Imbens and Ridder 2014 study a similar exercise). Finally, as explained before for bundles, discrete choice models can be used to calculate quantiles of welfare differences across different values of the explanatory variables W, Z. These true counterfactuals or welfare estimates require knowledge of the entire structure of the game and not just the choice probabilities in the data, $\Pr(a \mid w, z)$.

A Proofs

A.1 Proof of Lemma 1

We first show that, for each $W = w$, the sign of $\eta \cdot v(w)$ has different observable implications. We then recover the sign of the interaction effect. Section B of the online appendix reviews some of the concepts we use in this proof.

Step 1 If $\eta \cdot v(w) \geq 0$, $U(a, w, z, \varepsilon, \eta)$ is supermodular in (a_1, a_2). In addition, it has increasing differences in both (a_1, z_1) and (a_2, z_1). By A3, the maximizer is unique with probability 1. Let

$$a^*(w, z, \varepsilon, \eta) \equiv (a_1^*(w, z, \varepsilon, \eta), a_2^*(w, z, \varepsilon, \eta)) \equiv \arg\max \left\{ U(a, w, z, \varepsilon, \eta) : (a_1, a_2) \in \{0, 1\}^2 \right\}.$$

It follows by Topkis' theorem that $a^*(w, z, \varepsilon, \eta)$ increases (in the coordinatewise order) in z_1 with probability 1. By A2(I), the unobservables ε, η are independent of z_1. Thus, for all $z_1' > z_1$ and every upper set U in $\{0, 1\}^2$, we have that

$$\Pr(a^*(w, z_1', z_2, \varepsilon, \eta) \in U \mid w, z_1', z_2) \geq \Pr(a^*(w, z_1, z_2, \varepsilon, \eta) \in U \mid w, z_1, z_2).$$

That is, the random vector $a^*(w, z, \varepsilon, \eta) \mid w, z_1, z_2$ increases with respect to first order stochastic

10

dominance in z_1. Because stochastic dominance is preserved under marginalization, in the data, $\Pr(a_2 = 1 \mid w, z_1, z_2)$ increases in z_1 (Müller and Stoyan 2002, Theorem 3.3.10, p. 94). Similarly, we can show that $\Pr(a_1 = 1 \mid w, z_1, z_2)$ increases in z_2.

Step 2 If $\eta \cdot v(w) \leq 0$, $U(a, w, z, \varepsilon, \eta)$ is supermodular in $(a_1, -a_2)$. In addition, $U(a, w, z, \varepsilon, \eta)$ has increasing differences in both (a_1, z_1) and $(-a_2, z_1)$. By A3, the maximizer is unique with probability 1. Then, by Topkis' theorem, $(a_1^*(w, z, \varepsilon, \eta), -a_2^*(w, z, \varepsilon, \eta))$ increases in z_1 with probability 1. Following similar arguments to those in Step 1, we get that $\Pr(a_2 = 1 \mid w, z_1, z_2)$ decreases in z_1. Similarly, $\Pr(a_1 = 1 \mid w, z_1, z_2)$ decreases in z_2.

Step 3 By A3, if $\eta \cdot v(w) \neq 0$, then $\Pr(a_2 = 1 \mid w, z_1, z_2)$ and $\Pr(a_1 = 1 \mid w, z_1, z_2)$ are not constant as functions of z_1 and z_2. It follows from Steps 1 and 2 that, for each $W = w$, the sign of $\eta \cdot v(w)$ is identified from available data. $\qquad \square$

A.2 Proof of Theorem 1

This proof relies on the sign of the interaction effect that we can recover using Lemma 1. It also relies on three key results regarding characteristic functions (φ). Let $\varepsilon_1, \varepsilon_2, \eta$ be three random variables. In addition, let us assume that ε_1 and ε_2 are, separately, independent of η. Then, for each $\theta, \theta_1, \theta_2 \geq 0$, the next three equalities hold

1. $\varphi_{\varepsilon_i + \eta}(\theta) = \varphi_{\varepsilon_i}(\theta) \varphi_\eta(\theta)$ for $i = 1, 2$

2. $\varphi_{\varepsilon_1 + \eta, \varepsilon_2 + \eta}(\theta_1, \theta_2) = \varphi_{\varepsilon_1, \varepsilon_2}(\theta_1, \theta_2) \varphi_\eta(\theta_1 + \theta_2)$

3. $\varphi_{\varepsilon_1, \varepsilon_2 + \eta}(\theta_1, \theta_2) = \varphi_{\varepsilon_1, \varepsilon_2}(\theta_1, \theta_2) \varphi_\eta(\theta_2)$ and $\varphi_{\varepsilon_1 + \eta, \varepsilon_2}(\theta_1, \theta_2) = \varphi_{\varepsilon_1, \varepsilon_2}(\theta_1, \theta_2) \varphi_\eta(\theta_1)$.

Substitutes $(\eta \cdot v(w) \leq 0)$ Under A3, the probability of selecting neither of the two binary variables is

$$\Pr(\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq -z_2, \varepsilon_1 + u_1(w) + \varepsilon_2 + u_2(w) + \eta \cdot v(w) \leq -z_1 - z_2 \mid w, z).$$

Because $\eta \cdot v(w) \leq 0$, the third inequality above is implied by the first two. Thus,
$$\Pr((0,0) \mid w, z) = \Pr(\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq -z_2 \mid w, z).$$

11

Define $\alpha = (\alpha_1, \alpha_2) \equiv (\varepsilon_1 + u_1(w), \varepsilon_2 + u_2(w))$. By A2(I), the vector $(\varepsilon_1, \varepsilon_2, \eta)$ and hence ε are independent of Z. For an arbitrary point of evaluation $\alpha^\star = (\alpha_1^\star, \alpha_2^\star)$, we have

$$F_{\alpha|w}(\alpha^\star) = \Pr(\alpha_1 \leq \alpha_1^\star, \alpha_2 \leq \alpha_2^\star \mid w) = \Pr(\alpha_1 \leq -z_1, \alpha_2 \leq -z_2 \mid w, z) = \Pr((0,0) \mid w, z),$$

for excluded variable choices $z_1 = -\alpha_1^\star$ and $z_2 = -\alpha_2^\star$. Therefore, the variation in Z from A1 and $\Pr((0,0) \mid w, z)$ identifies the CDF of α, for each $W = w$.

Under A3, the probability of selecting both items is

$$\Pr((1,1) \mid w, z) = \Pr(\varepsilon_1 + u_1(w) + \varepsilon_2 + u_2(w) + \eta \cdot v(w) \geq -z_1 - z_2,$$
$$\varepsilon_1 + u_1(w) + \eta \cdot v(w) \geq -z_1, \varepsilon_2 + u_2(w) + \eta \cdot v(w) \geq -z_2 \mid w, z).$$

Because $\eta \cdot v(w) \leq 0$, the first inequality above is implied by the last two. Thus,
$$\Pr((1,1) \mid w, z) = \Pr(\varepsilon_1 + u_1(w) + \eta \cdot v(w) \geq -z_1, \varepsilon_2 + u_2(w) + \eta \cdot v(w) \geq -z_2 \mid w, z).$$

Define $\beta = (\beta_1, \beta_2) \equiv (-\varepsilon_1 - u_1(w) - \eta \cdot v(w), -\varepsilon_2 - u_2(w) - \eta \cdot v(w))$. Given A2(I) and using the previous logic, we can use variation in Z and $\Pr((1,1) \mid w, z)$ to get $F_{\beta|w}$. By A2(II) and A2(III), ε and η have mean $(0,0)$ and 1, respectively, conditional on $W = w$. Thus,

$$E[(\alpha_1, \alpha_2) \mid w] = (u_1(w), u_2(w))$$
$$E[(\beta_1, \beta_2) \mid w] = (-u_1(w) - v(w), -u_2(w) - v(w))$$
$$E[(\alpha_1, \alpha_2) \mid w] + E[(\beta_1, \beta_2) \mid w] = (-v(w), -v(w)).$$

One can see that $((u_i)_{i \leq 2}, v)$ is identified for $W = w$.

Once $((u_i)_{i \leq 2}, v)$ is identified, we can obtain the distribution of $(\varepsilon_1, \varepsilon_2)$ from the one of α using a simple location shift. Therefore, we identify $F_{\varepsilon|W}$. Likewise, we can move from the distribution of $\beta \mid W = w$ to the distribution of $(-\varepsilon_1 - \eta \cdot v(w), -\varepsilon_2 - \eta \cdot v(w))$ and, by a known multiplicative change of variables, the distribution of $(\varepsilon_1 + \eta \cdot v(w), \varepsilon_2 + \eta \cdot v(w))$. Assuming that η and ε are independent conditional on W, $F_{\eta \cdot v(w)|w}$ is identified from knowledge of $F_{\varepsilon|w}$, given the second result we mentioned above for characteristic functions. Finally, a known, multiplicative change of variables, given that $v(w)$ is known, identifies $F_{\eta|w}$.

Complements $(\eta \cdot v(w) \geq 0)$ Under A3, the probability of selecting only binary variable 1 is

$$\Pr(-\varepsilon_1 - u_1(w) \leq z_1, \varepsilon_2 + u_2(w) + \eta \cdot v(w) \leq -z_2, -\varepsilon_1 - u_1(w) + \varepsilon_2 + u_2(w) \leq z_1 - z_2 \mid w, z).$$

Because $\eta \cdot v(w) \geq 0$, the third inequality above is implied by the first two. Thus,
$$\Pr((1,0) \mid w,z) = \Pr(-\varepsilon_1 - u_1(w) \leq z_1, \varepsilon_2 + u_2(w) + \eta \cdot v(w) \leq -z_2 \mid w,z).$$

Let $\alpha = (\alpha_1, \alpha_2) \equiv (-\varepsilon_1 - u_1(w), \varepsilon_2 + u_2(w) + \eta \cdot v(w))$. By A2(I), $(\varepsilon_1, \varepsilon_2, \eta)$ is independent of Z. Thus, for an arbitrary point of evaluation $\alpha^\star = (\alpha_1^\star, \alpha_2^\star)$ of the CDF $F_{\alpha \mid w}$,

$$F_{\alpha \mid w}(\alpha^\star) = \Pr(\alpha_1 \leq \alpha_1^\star, \alpha_2 \leq \alpha_2^\star \mid w) = \Pr(\alpha_1 \leq z_1, \alpha_2 \leq -z_2 \mid w,z) = \Pr((1,0) \mid w,z)$$

for choices $z_1 = \alpha_1^\star$ and $z_2 = -\alpha_2^\star$. Therefore, the variation in Z from A1 and $\Pr((1,0) \mid w,z)$ identifies the CDF of α, for each $W = w$. Applying the same logic, we can use variation in Z and $\Pr((0,1) \mid w,z)$ to recover $F_{\beta \mid w}$, where the random vector β is $\beta = (\beta_1, \beta_2) \equiv (\varepsilon_1 + u_1(w) + \eta \cdot v(w), -\varepsilon_2 - u_2(w))$. By A2(II) and A2(III), ε has $(0, 0)$ mean and η has a mean of 1 conditional on $W = w$. Therefore,

$$\mathrm{E}\left[(\alpha_1, \alpha_2) \mid w\right] = (-u_1(w), u_2(w) + v(w))$$
$$\mathrm{E}\left[(\beta_1, \beta_2) \mid w\right] = (u_1(w) + v(w), -u_2(w))$$
$$\mathrm{E}\left[(\alpha_1, \alpha_2) \mid w\right] + \mathrm{E}\left[(\beta_1, \beta_2) \mid w\right] = (v(w), v(w)).$$

One can see that u_1, u_2 and v are identified for this $W = w$.

Once $\left((u_i)_{i \leq 2}, v\right)$ is identified we can recover, using a multiplicative change of variables, the distribution of $(\varepsilon_1, \varepsilon_2 + \eta \cdot v(w))$ from the distribution of α. Therefore we can also identify the distribution of $\varepsilon_2 + \eta \cdot v(w)$. In addition, we can recover, using a multiplicative change of variables, the distribution of $(\varepsilon_1 + \eta \cdot v(w), \varepsilon_2)$ from the distribution of β. Thus we can also identify the distribution of ε_2. Knowing the distributions of $\varepsilon_2 + \eta \cdot v(w)$ and ε_2, we can use the first result we mentioned above for characteristic functions to identify $F_{\eta \cdot v(w) \mid w}$. As before, a known, multiplicative change of variables, given that $v(w)$ is known, identifies $F_{\eta \mid w}$.

Finally, given that we know the distributions of $(\varepsilon_1, \varepsilon_2 + \eta \cdot v(w))$ and $\eta \cdot v(w)$, we can use the third result we mentioned above for characteristic functions to recover $F_{\varepsilon \mid W}$. $\qquad \square$

A.3   Proof of Theorem 2

This proof relies on the sign of the interaction effect, which is identified by Lemma 1.

Substitutes $(\eta \cdot v(w) \leq 0)$ Under A3$'$, given the proof of Theorem 1,
$$\Pr((0,0) \mid w,z) = \Pr(\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq -z_2 \mid w,z).$$

By A2$'$(III), we can get the density f of $F_{\varepsilon_1 + u_1(w), \varepsilon_2 + u_2(w) \mid W}$ (in the support of Z) as follows

$$f_{\varepsilon_1+u_1(w),\varepsilon_2+u_2(w)|W}(-z_1,-z_2) = \partial^2 \Pr\left((0,0) \mid w,z\right)/\partial z_1 \partial z_2 =$$

$$\partial^2 \Pr\left(\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq -z_2 \mid w,z\right)/\partial z_1 \partial z_2.$$

By A2′(II), this expression has mode $(u_1(w), u_2(w))$. By A1′, the cross partial is maximized at z_1' and z_2' whenever $u_1(w) = -z_1'$ and $u_2(w) = -z_2'$. Thus $(u_i)_{i\leq 2}$ can be identified from the data.

Under A2′(IV) and A3′, given the proof of Theorem 1,

$$\Pr\left((1,1) \mid w,z\right) = \Pr\left(\varepsilon_1 + u_1(w) + v(w) \geq -z_1, \varepsilon_2 + u_2(w) + v(w) \geq -z_2 \mid w,z\right).$$

Thus, by A2′(III) we can obtain the density of $F_{\varepsilon_1+u_1(w)+v(w),\varepsilon_2+u_2(w)+v(w)|W}$ (in the support of Z) by taking cross-partial derivatives of the survival function

$$f_{\varepsilon_1+u_1(w)+v(w),\varepsilon_2+u_2(w)+v(w)|W}(-z_1,-z_2) = \partial^2 \Pr\left((1,1) \mid w,z\right)/\partial z_1 \partial z_2 =$$

$$\partial^2 \Pr\left(\varepsilon_1 + u_1(w) + v(w) \geq -z_1, \varepsilon_2 + u_2(w) + v(w) \geq -z_2 \mid w,z\right)/\partial z_1 \partial z_2.$$

By A2′(II), this expression has mode $(u_1(w) + v(w), u_2(w) + v(w))$. By A1′ the cross partial derivative is maximized at $-z_1', -z_2'$ when $u_1(w) + v(w) = -z_1'$ and $u_2(w) + v(w) = -z_2'$. It follows that v can be recovered from the data.

Complements $(\eta \cdot v(w) \geq 0)$ Under A2′(IV) and A3′, given Theorem 1's proof, we get

$$\Pr\left((1,0) \mid w,z\right) = \Pr\left(-\varepsilon_1 - u_1(w) \leq z_1, \varepsilon_2 + u_2(w) + v(w) \leq -z_2 \mid w,z\right).$$

Thus, by A2′(III), we can get the density of $F_{-\varepsilon_1-u_1(w),\varepsilon_2+u_2(w)+v(w)|W}$ (in the support of Z) by

$$f_{-\varepsilon_1-u_1(w),\varepsilon_2+u_2(w)+v(w)|W}(z_1,-z_2) =$$

$$-\partial^2 \Pr\left(-\varepsilon_1 - u_1(w) \leq z_1, \varepsilon_2 + u_2(w) + v(w) \leq -z_2 \mid w,z\right)/\partial z_1 \partial z_2.$$

By assumptions A1′ and A2′(II), this expression has mode $(-u_1(w), u_2(w) + v(w))$. By A1′ the cross partial is maximized at z_1' and $-z_2'$ when $-u_1(w) = z_1'$ and $u_2(w) + v(w) = -z_2'$. Thus u_1 and $u_2 + v$ can be identified from the data.

We can then use $\Pr\left((0,1) \mid w,z\right)$ to recover u_2 and $u_1 + v$. Thus, $\left((u_i)_{i\leq 2}, v\right)$ is identified.$\square$

A.4   Proof of Theorem 3

Part 2 follows directly from Theorem 1. We next show, avoiding repeating all the intermediate steps, that the proof for Part 1 is almost identical to the one of Theorem 1.

Submodular Game ($\eta_1 \cdot v_1(w) \leq 0$ and $\eta_2 \cdot v_2(w) \leq 0$) Under A3, when the game is submodular and $(0,0)$ is an equilibrium, it is unique (e.g., Tamer 2003). The same is true regarding $(1,1)$. The probability that both players choose action 0, $\Pr((0,0) \mid w, z)$, is identical to the probability of selecting $(0,0)$ when the binary variables are substitutes. In addition, the probability that both players choose action 1 simplifies to

$$\Pr((1,1) \mid w,z) = \Pr(\varepsilon_1 + u_1(w) + \eta_1 \cdot v_1(w) \geq -z_1, \varepsilon_2 + u_2(w) + \eta_2 \cdot v_2(w) \geq -z_2 \mid w, z).$$

This expression differs from the probability of selecting $(1,1)$ in the proof of Theorem 1 in that we have $\eta_1 \cdot v_1(w)$ and $\eta_2 \cdot v_2(w)$ instead of just $\eta \cdot v(w)$. Given the independent variation of Z_1 and Z_2, the extra term does not introduce any difficulty in the identification strategy.

Supermodular Game ($\eta_1 \cdot v_1(w) \geq 0$ and $\eta_2 \cdot v_2(w) \geq 0$) Under A3, when $(1,0)$ is an equilibrium, with probability 1 it is unique (e.g., Tamer 2003). The probability that player 1 selects action 1 and player 2 action 0 simplifies to

$$\Pr((1,0) \mid w,z) = \Pr(-\varepsilon_1 - u_1(w) \leq z_1, \varepsilon_2 + u_2(w) + \eta_2 \cdot v_2(w) \leq -z_2 \mid w, z).$$

This expression differs from the probability of selecting only the first variable in the proof of Theorem 1 in that now we have $\eta_2 \cdot v_2(w)$ instead of $\eta \cdot v(w)$. It follows by simple inspection that this difference does not introduce any additional difficulty in the identification strategy, except for the fact that we need to use a slightly different property of characteristic functions for independent variables, namely,

$$\varphi_{\varepsilon_1 + \eta_1, \varepsilon_2 + \eta_2}(\theta_1, \theta_2) = \varphi_{\varepsilon_1, \varepsilon_2}(\theta_1, \theta_2)\, \varphi_{\eta_1, \eta_2}(\theta_1, \theta_2).$$

$\square$

References

[1] Athey, S., and S. Stern (1998). An empirical framework for testing theories about complementarity in organizational design. Working paper.

[2] Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's who in networks. Wanted: The key player. Econometrica 74, 1403–1417.

[3] Berry, S., and P. Haile (2010). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Yale University working paper.

[4] Berry, S., and E. Tamer (2006). Identification in models of oligopoly entry. Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, vol. 2.

[5] Chen, R., and Y. Chen (2011). The potential of social identity for equilibrium selection. American Economic Review 101, 2562–2589.

[6] Chen, S., S. Khan and X. Tang (forthcoming). Informational content of special regressors in heteroskedastic binary response models. Journal of Econometrics.

[7] Chen, X., E. Tamer, and A. Torgovitsky (2011). Sensitivity analysis in semiparametric likelihood models. Working paper.

[8] Ciliberto, F., and E. Tamer (2009). Market structure and multiple equilibria in airline markets. Econometrica 77, 1791–1828.

[9] Dunker, F., S. Hoderlein, and H. Kaido (2013). Random coefficients in static games of complete information. Working paper.

[10] Dunker, F., S. Hoderlein, and H. Kaido (2014). Nonparametric identification of endogenous and heterogeneous aggregate demand models: Complements, bundles and the market level. Working paper.

[11] Fox, J., and A. Gandhi (2016). Nonparametric identification and estimation of random coefficients in multinomial choice models. RAND Journal of Economics 47, 118–139.

[12] Gentzkow, M. (2007a). Valuing new goods in a model with complementarity: Online newspapers. American Economic Review 97, 713-744.

[13] Gentzkow, M. (2007b). Valuing new goods in a model with complementarity: Online newspapers. University of Chicago working paper.

[14] Goeree, J., and C. Holt (2005). An experimental study of costly coordination. Games Economics Behavior 51, 349–364.

[15] Graham, B., G. Imbens, and G. Ridder (2014). Complementarity and aggregate implications of assortative matching: A nonparametric analysis. Quantitative Economics 5, 29–66.

[16] Harsanyi, J., and S. Selten (1988). A general theory of equilibrium selection in games. MIT Press.

[17] Hartman, W. (2010). Demand estimation with social interactions and the implications for targeted marketing. Journal Marketing Science 29, 585–601.

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

[19] Kajii, A., and S. Morris (1997). Refinements and higher order beliefs: A unified survey. Northwestern University working paper.

[20] Kline, B. (2015). Identification of complete information games. Journal of Econometrics 189, 117–131

[21] Kline, B. (2016). The empirical content of games with bounded regressors. Quantitative Economics 7, 37–81.

[22] Kretchsmer, T., E. Miravete, and J. Pernias (2012). Competitive pressure and the adoption of complementary innovations. American Economic Review 102, 1540–1570.

[23] Letham, B., W. Sun, and A. Sheopuri (2014). Latent variable copula inference for bundle pricing from retail transaction data. Proceedings of the 31 st International Conference on Machine Learning, Beijing, China. JMLR: W&CP volume 32.

[24] Lewbel, A. (1998). Semiparametric latent variable model estimation with endogenous or mismeasured regressors. Econometrica 66, 105–121.

[25] Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables. Journal of Econometrics 97, 145–177.

[26] Liu, H., P. Chintagunta, and T. Zhu (2010). Complementarities and the demand for home broadband internet services. Marketing Science 29, 701–720.

[27] Manski, C. (1975) Maximum score estimation of the stochastic utility model of choice. Journal of Econometrics 3, 205–228.

[28] Manski, C. (1985) Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. Journal of Econometrics 27, 313–333.

[29] Manski, C. (1988) Identification of binary response models. Journal of Econometrics 83, 729–738.

[30] Matzkin, R. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. Econometrica 60, 239–270.

[31] Matzkin, R. (1993). Nonparametric identification and estimation of polychotomous choice models. Journal of Econometrics 58, 137–168.

[32] Matzkin, R. (2007). Heterogeneous choice. Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society.

[33] Monderer, D., and L. Shapley (1996). Potential games. Games Economic Behavior 14, 124–143.

[34] Müller, A., and D. Stoyan (2002). Comparison methods for stochastic models and risks. John Wiley & Sons, Inc.

[35] Sher, I., and K. Kim (2014). Identifying combinatorial valuations from aggregate demand. Journal of Economic Theory 153, 428–458.

[36] Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. Review of Economic Studies 70, 147–165.

[37] Thompson, T.S. (1989). Identification of semiparametric discrete choice models. University of Minnesota working paper.

[38] Topkis, D. (1998). Supermodularity and complementarity. Princeton University Press.

[39] Ui, T. (2000). A Shapley value representation of potential games. Games Economic Behavior 31, 121–135.

[40] Ui, T. (2001). Robust equilibria of potential games. Econometrica 69, 1373–1380.

[41] Van Huyck, J., R. Battalio, and R. Beil (1990). Tactic coordination games, strategic uncertainty, and coordination failure. American Economic Review 80, 234–248.

[42] Venkatesh, R., and V. Mahajan (2009). The design and pricing of bundles: A review of normative guidelines and practical approaches. Handbook of Pricing and Research in Marketing, Ed. by V. Rao.

[43] Yildiz, M. (2007). Wishful thinking in strategic environments. Review of Economic Studies 74, 319–344.