Identification of Discrete Choice Models for Bundles and Binary Games∗

Jeremy T. Fox  Natalia Lazzati
University of Michigan and NBER†  University of Michigan‡

March 2014

Abstract
We study nonparametric identification of single-agent discrete choice models for bundles (without requiring bundle-specific prices) and for 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. Taking advantage of these similarities, we provide new identification results. Potential games are particularly useful for identification in games of three or more players.

∗Thanks to Daniel Ackerberg, Tilman Borgers, Paul Grieco, Stephan Lauermann, Arthur Lewbel, Rosa Matzkin, Amilcar Menichini and participants at various seminars and conferences for helpful comments.
†Jeremy T. Fox, University of Michigan, Ann Arbor, MI 48109-1220, e-mail: jeremyfox@gmail.com.
‡Natalia Lazzati, University of Michigan, Ann Arbor, MI 48109-1220, e-mail: nlazzati@umich.edu.
1 Introduction

1.1 Motivation and Main Results

Overview Single-agent discrete choice models for bundles and binary games play an important role in empirical work in industrial organization and a variety of other fields. Most studies on single-agent discrete choice models assume the alternatives are mutually exclusive. That is, they model situations where an agent faces a set of options and acceptance of one of them excludes all the others. We assume instead that the agent can acquire any subset or bundle of the set of available alternatives. Under the bundles model, we show that single-agent discrete choice models and binary action games of complete information are quite similar from an identification perspective. Moreover, we prove the two models are indeed equivalent when attention is restricted to the class of potential games under a specific equilibrium selection rule. The equivalence between the two models is important because any identification progress with one of them can be used with the other. Taking advantage of this link, we provide new identification results for discrete choice models and games.

Bundles Model To the best of our knowledge, we are the first to provide nonparametric identification results for discrete choice models for bundles. Our formulation of the bundles model does not require independent variation of covariates at the bundle level. Gentzkow (2007a) estimates a parametric probit model where a consumer can elect to acquire no newspaper, a print newspaper, an online newspaper, or both. Gentzkow’s data show that many consumers purchase both the print and online newspapers. This behavior may occur because the two goods are complements in consumption or because consumers with high unobserved standalone values for the print newspaper have high unobserved valuations for the online version as well. Our identification approach allows us to distinguish unobserved, correlated preferences from true interdependence of the alternatives in utility. The distinction between true complementarities (or substitutabilities) in demand from correlated, unobserved heterogeneous preferences for the goods is essential for policy-making, such as optimal pricing. Moreover, in the case of two alternatives, our model allows for heterogeneous interaction effects, measuring unobserved differences across agents in the degree of complementarities or substitutability. When the alternatives are substitutes, we can identify the distribution of heterogeneous interaction effects.

Games In addition to providing insights about multinomial choice models, our findings also contribute to the literature of static, binary choice games of complete information.¹ This set

¹See, for example, Heckman (1978), Bjorn and Vuong (1984), Bresnahan and Reiss (1991a, 1991b), Berry
of games includes games of product adoption with social interactions, entry games, and labor force participation games, among others. In binary games, interaction effects capture the strategic interdependence of player decisions. As in the single-agent bundles model, these effects are often hard to distinguish from unobserved, heterogeneous payoffs that are correlated across players, as these two sources of interdependence have similar observable implications. For example, similarities in smoking behavior among friends may be due to peer effects or due to correlated tastes for smoking. Likewise, firms may tend to enter geographic markets in clusters either because the competitive effects of entry are small or because of the unobservable profitability of certain markets. Though it is well known that not allowing for the correlation of unobservables can lead to biases in the estimates of interaction effects, much of the empirical literature on games assumes that unobservables are independent of each other. By contrast, our identification approach can distinguish and separately identify these two sources of interdependence. In addition, for two-player games, our model allows for heterogeneous interaction effects, measuring differences across plays of the game in the degree of strategic interactions. When the game is of strategic substitutes, we can identify the distribution of heterogeneous interaction effects.

**Identification and Connection Between the Models**

The identification strategy we propose relies on exclusion restrictions at the level of each alternative (for the single-agent model) and at the level of each player (for games). For the discrete choice model of bundles, these excluded regressors enter the standalone payoff of each alternative additively separably but do not affect the interaction terms. For games, each excluded regressor similarly affects the standalone payoff of one player but not the others. This approach allows us to nonparametrically identify aspects of the structure of both models: the utilities of all bundles as functions of covariates and the players’ subutilities of choosing each action as functions of regressors and the actions of other players as well as, in both models, the distribution of potentially correlated unobservables, including heterogeneous interaction effects. These sets of results rely on a large support condition on the excluded regressors. We then show that the homogenous standalone utilities and interaction effects can be recovered without the large support assumptions. This second set of results extends the work of Kline (2013b) on semiparametric games to the bundles model by taking advantage of the connection between the two models.

As we just mentioned, the restrictions we use and the identification strategies we follow for the two seemingly unrelated models are quite similar. This similarity stems from the fact

that almost identical key algebraic equations involving outcome probabilities and cumulative
distribution functions appear in the identification arguments for both bundles model and games.
In particular, the two identification strategies rely on previous knowledge of the sign of the
interaction terms. The major difference between the identification approaches is that in the
bundles model the sign of the interaction terms can be recovered from the empirical evidence,
while in the case of games this is not possible because of multiple equilibria and we therefore
assume the signs are known by the econometrician.

As mentioned at the beginning, potential games are at the heart of a formal equivalence
between identification of discrete choice models for bundles and binary games. Monderer and
Shapley (1996) define potential games as those games that admit an exact potential function.
This potential function is defined over the set of action profiles of the players such that the
function’s maximizer is a pure strategy Nash equilibrium for the associated game. When a
game admits a potential function, the function is uniquely defined up to an additive constant.
Based on this observation, Monderer and Shapley argue that the maximizer of the potential is
an equilibrium refinement. Recent theoretical and experimental work has provided economic
justification to this equilibrium selection rule.\(^2\) Contributing to this growing body of literature,
we show that if the game admits a potential function and players coordinate on the potential
maximizer, then the binary game is formally equivalent (from an econometric standpoint) to
the discrete choice model for bundles.

1.2 Related Literature

There is a small, mostly empirical literature on the demand for bundles of discrete goods when
those goods may have positive or negative interaction effects in utility (e.g., Athey and Stern
2012).\(^3\) To our knowledge, we are the first to give conditions for nonparametric identification
in models of demand for bundles without relying on bundle-specific covariates.\(^4\) We extend

\(^2\) See Ui (2001) for a theoretical justification of the selection rule based on potential maximizers and Van
Huyck, Battalio and Beil (1990), Goeree and Holt (2005) and Chen and Chen (2011) for experimental evidence.

\(^3\) Hendel (1999) and Dubé (2004) study models of demand for bundles where the bundles do not have interaction
effects. For other specifications of discrete choice demand with interactions between goods, see Manchanda,

\(^4\) Kim and Sher (2013) study identification where the researcher observes only the market share for each good,
not market shares for each bundle. They assume that the true distribution of heterogeneity represents a finite
number of consumers that is the same finite group across markets. They prove so-called generic identification,
where counterexamples to identification are proved to have probability zero, if a distribution is placed over
possible parameter values. This definition of identification is weak because, in other contexts, counterexamples
to identification for parameters that researchers intuitively think are fundamentally unidentified may have
identification techniques from the literature on multinomial choice models where only one good can be selected (e.g., Thompson 1989, Matzkin 1993, Lewbel 2000).

The sufficient conditions that we propose to identify the distribution of the unobservables for models with three or more goods are economically similar to some of the requirements imposed in tangentially related identification papers. For example, Berry, Gandhi and Haile (2013) provide a sufficient condition under which a system of demand functions (with continuous outcomes such as shares) can be inverted in unobservables. Their condition applies to goods that are (connected) substitutes, and is somewhat related to our result on demand when goods are substitutes. Matzkin (2008) studies the inversion of simultaneous equations in models with continuous outcomes, imposing quasi-concavity. We suggest a discrete notion of concavity.

We argue that identification in discrete choice models of bundles is quite similar to identification in binary games. For games of two players, our results extend those of Tamer (2003) and Berry and Tamer (2006, Result 4). Without assuming any equilibrium selection rule, both results, theirs and ours, require knowledge of the signs of the interactions terms. One advantage of our identification argument with respect to theirs is that we do not rely on identification at infinity to recover any of the unknown objects in the model. By identification at infinity, we mean recovering the interaction effects of one player by choosing values for the excluded regressors of all other players in the range where all these other players have dominant strategies. Though conceptually simple, this identification technique is empirically unattractive because it relies mainly on extreme regressor values that are unlikely to appear in any finite data set. In other econometric contexts such as sample selection models, identification at infinity has been shown to result in slow rates of convergence (under certain tail thickness properties for the regressor and unobservables) for estimators based on the identification arguments (e.g., Andrews and Schafgans 1998 and Khan and Tamer 2010). Though none of our results use identification at infinity, the first set of results we provide rely on the large support of the excluded regressors to identify the joint distribution of the unobservables. Using the work of Kline (2013b) on semiparametric games, we also show that part of the model structure can be recovered without the large support assumption. Another advantage of our paper over Berry and Tamer (2006, Result 4) is that we incorporate heterogeneous interaction effects.

We know of two simultaneous, independent papers on the identification of binary games of complete information: Kline (2013a) and Dunker, Hoderlein and Kaido (2013). Neither paper discusses the link with single-agent discrete choice models with bundles or potential games. For the two player case, Kline (2013a) presents identification results that, like ours, do not
rely on identification at infinity or on an equilibrium selection mechanism. In the case of three or more players, he uses identification at infinity to turn the game into a two-player game. Identification at infinity does not recover the joint distribution of unobservables in games of three or more players unless extra assumptions are imposed on the copula of the multivariate distribution. Dunker et al. (2013) study the two-player case and allow for random coefficients on all regressors, not just interaction effects, in a semiparametric specification and provide a computationally simple estimator for the density of random coefficients.

The parametric identification and estimation of discrete choice games of complete information has been recently studied by Bajari, Hong and Ryan (2010), Beresteanu, Molchanov and Molinari (2011), Ciliberto and Tamer (2009), and Galichon and Henry (2010). Our results for games depart from this literature in that our model is nonparametric and we show point rather than set identification. Given the natural trade-offs, our approach is more restrictive in that we focus on the binary actions case (e.g., do or do not enter, do or do not smoke, etc.). Binary and ordered games of incomplete information have received a lot of attention.5 As Bajari et al. (2010) explain, the challenges involved in complete information games are quite different. One advantage of the complete information setup is that players do not have ex-post regret: they do not prefer to take a different action after seeing the actions of their rivals. This feature is particularly important in settings where the action of each player has longstanding consequences, such as entry. Another advantage of complete information with correlated unobservables over incomplete information games (without unobserved heterogeneity) is that the unobservables can reflect unmeasured characteristics of, for instance, the geographic market that are observed by the players.

We do not formally explore estimation because establishing consistency for particular nonparametric estimators is somewhat straightforward conceptually once identification is established. Consider first models where the data generating process lacks multiple equilibria, such as discrete choice with bundles and potential games where players coordinate on the potential maximizer. Under the full independence of observable regressors and unobservables, one could apply simulated maximum likelihood with sieve-based approaches for modeling the distribution of the unobservables and the also infinite-dimensional standalone utilities and interaction effects (Chen 2007, Chen, Tamer and Torgovitsky 2011). However, the results of Khan and Nekipelov (2012) on rates of convergence deserve some attention. For semiparametric discrete games of complete information, Khan and Nekipelov establish that there is no estimator of the

interaction effects that always converges at the parametric rate. They argue that the reason for the slow rates of convergence is the binary nature of the other player’s action in a given player’s best response, not the presence of multiple equilibria. Therefore, we also conjecture that there exists no estimator that converges at the parametric rate for semiparametric models of the demand for bundles and potential games. In a recent working paper, Zhou (2013) studies semiparametric estimation of binary games of complete information under a symmetry restriction on the unobservables, which results in a parametric rate of convergence for the interaction term.

Our identification results for two-player binary games allow multiple equilibria; standard maximum likelihood is inconsistent even when our model is point identified. Modified likelihood estimators from the literature on multiple equilibria can be adapted to our nonparametric case (Bresnahan and Reiss 1991b, Berry 1992, Tamer 2003).

Section 2 and 3 present our identification results for the two-alternative bundles model and for two-player games, respectively. Section 4 introduces potential games and shows the formal equivalence between the bundles model and games. Section 5 extends our identification results to the case of three or more alternatives and players. Section 6 concludes, and all proofs and some basic results on comparative statics are collected in the Appendices.

2 Two-Alternative Bundles Model

2.1 Bundles Model Structure

Consider an agent that faces two binary alternatives, \{1, 2\}. The only difference with standard models in multinomial choice is that these alternatives are not mutually exclusive. The choice set is therefore \{(0, 0), (1, 0), (0, 1), (1, 1)\}, where we use \(a_i = 1\) to indicate that alternative \(i\) is selected and \(a_i = 0\) to indicate that \(i\) is not selected. We define \(a = (a_1, a_2)\).

If the agent selects only alternative \(i\), her standalone payoff is \(u_i(X) + \varepsilon_i\) for \(i = 1, 2\), where \(X \in \mathcal{X} \subseteq \mathbb{R}^k\) is a vector of observable covariates and \(\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2\) indicates a random vector observed by the agent but not by the econometrician. We let \(\varepsilon\) have the bivariate distribution \(F_{\varepsilon|X}\). If the decision maker chooses both alternatives, her utility is

\[
    u_1(X) + \varepsilon_1 + u_2(X) + \varepsilon_2 + \eta \cdot v(X) ,
\]

\(^{6}\)Khan and Nekipelov’s impossibility result is not related to the issues of identification at infinity in the previous literature on games.
where $\eta \cdot v(X)$ is the interaction effect, the change in utility for consuming both alternatives. For each $X = x$, we say that the alternatives are complements if $\eta \cdot v(x) \geq 0$ and that they are substitutes otherwise. The function $v(X)$ allows the interaction effect to depend on covariates. The non-negative random variable $\eta \in \mathbb{R}_+$ represents heterogeneity in the interaction effect across agents and has the univariate distribution $F_{\eta|X}$; allowing heterogeneity in the interaction effect is an advantage of our model over the previous literature on, as we argue below, the related topic of games. The distributions $F_{\varepsilon|X}$ and $F_{\eta|X}$ are the corresponding marginals of the joint distribution $F_{\varepsilon,\eta|X}$ of $(\varepsilon_1, \varepsilon_2, \eta)$. We finally normalize to 0 the utility of selecting the bundle $(0, 0)$.

The agent chooses the set of alternatives $a$ to maximize her utility

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

(1)

where $1(.)$ is the standard indicator function. The purpose of our analysis is to use data on chosen bundles and covariates $(a, x)$ to point identify aspects of the structure of the model $((u_i)_{i \leq 2}, v, F_{\varepsilon,\eta|X})$. Our identification approach is nonparametric as all the unknowns are functions. Identifying $F_{\varepsilon,\eta|X}$ in addition to the unknown functions $((u_i)_{i \leq 2}, v)$ would allow the computation of marginal effects of changing elements of $X$ on bundle choices and hence elasticities.\(^7\) The distribution $F_{\varepsilon,\eta|X}$ would also be needed to compute choice probabilities under counterfactuals that involve changing the unknown functions $((u_i)_{i \leq 2}, v)$.

### 2.2 Identification of the Bundles Model

Our identification strategy relies on excluded regressors. That is, we need a pair of regressors each of which enters the standalone utility of one alternative without affecting the utility of the other one. Specifically, we assume $X = (W, Z)$, where $W$ is a sub-vector of standard covariates and $Z$ represents the two excluded regressors. We define $Z = (Z_1, Z_2)$, where $Z_i$ is a scalar for $i = 1, 2$. We let

$$u_i(X) = u_i(W) + Z_i \text{ for } i = 1, 2 \text{ and } v(X) = v(W).$$

\(^7\)Under stricter semiparametric assumptions on $((u_i)_{i \leq 2}, v, F_{\varepsilon,\eta|X})$, choice probabilities and elasticities can be computed outside of the support of $X$ used in identification. As we do not allow for the endogeneity of regressors $X$, marginal effects and elasticities can also be identified directly from the data on choice probabilities for realizations of $X$ inside the support used for identification.

\(^8\)The equation $u_i(X) = u_i(W) + Z_i$ uses the overloaded notation $u_i$ for conciseness and should be interpreted as follows: a function $u_i(X)$ of all the elements of $X$ is equal to the sum of another function $u_i(W)$ of only the elements of $W$ and the value of $Z_i$. We directly identify $u_i(W)$, whose identification then identifies $u_i(X)$.
We next provide a set of further identifying assumptions and then discuss each of them. Let $Z_{-i}$ be the excluded regressor for the alternative other than $i$.

**B1:** $Z_i \mid W, Z_{-i}$ has support on all $\mathbb{R}$ for $i = 1, 2$.

**B2:** (i) $F_{\varepsilon,\eta \mid X} = F_{\varepsilon,\eta \mid W}$; (ii) $E(\varepsilon \mid W) = (0, 0)$; and (iii) $E(\eta \mid W) = 1$.

**B3:** $\varepsilon \mid W$ has an everywhere positive Lebesgue density on its support.

The exclusion restrictions are key to identifying the joint distribution of unobservables separately from the interaction effect. The intuition is simple: $Z_i$ only affects the value of alternative $-i$ via the interaction effect. Consequently, interaction effects must be present if changes in the realization of $Z_i$ correspond to changes in the marginal probability of selecting the other alternative. The large support restriction, B1, is a standard requirement used in various ways in the literature on binary and multinomial choice models.\(^9\) In identification arguments where the object of interest includes the distribution of unobservables, the large support restriction is used to identify the tails of the distribution of unobservables. Identifying the full distribution of some linear functions of $\varepsilon$ and $\eta$ is key to our identification strategy for the homogeneous parameters $((u_i)_{i \leq 2}, v)$, as the means of $\varepsilon \mid W$ and $\eta \mid W$ are sensitive to the masses in the tails and we rely on identification of the means to recover standalone payoffs $u_1$ and $u_2$ and the interaction effect $v$.\(^{10}\) Below we present a different identification strategy for the homogeneous parameters $((u_i)_{i \leq 2}, v)$ that does not rely on the large support assumption.

Note that we do not require an excluded regressor to enter the interaction effect, i.e., a scalar $Z_3$ such that $v(X) = v(W) + Z_3$. If we had such a regressor, then identification of the bundles model would be equivalent to identification of a multinomial choice model with three inside alternatives (instead of two) and the outside option of selecting no alternative (e.g., Thompson 1989, Matzkin 1993, Lewbel 2000).

---


\(^{10}\)This mean independence condition allows flexible heteroskedasticity in $W$. However, the means of $\varepsilon$ and $\eta$ are naturally sensitive to the weight in the tails of $\varepsilon$ and $\eta$. Consequently, under Assumption B2(ii) the rate of convergence of an assumed intercept in a parametric portion of the model (for example, $u_i(w)$ is specified parametrically with an intercept) may be less than the standard parametric rate and related to the thicknesses of the tails of $\varepsilon$ and $Z$. This mean independence condition and its sensitivity to the support of $Z$ (relative to the support of $\varepsilon$ in a world without $\eta$) has been examined in the case of semiparametric binary and multinomial choice models (e.g., Lewbel 1998, 2000; Khan and Tamer 2010; Dong and Lewbel 2014). Additional assumptions, such as tail symmetry in the distribution of $\varepsilon$, may remove some of these identification and estimation issues (Magnac and Maurin 2007). Other sets of identifying assumptions from the literature on binary and multinomial choice could be modified for our models.
Assumption B2(i) requires the vector \((\varepsilon_1, \varepsilon_2, \eta)\) to be independent of the large support regressors \(Z\). For any \(W = w\), the mean of \(\varepsilon\) is otherwise not separately identified from \(u_i (w)\), which subsumes the role of any intercept. Assumptions B2(ii) and B2(iii) provide location normalizations for \(\varepsilon\) and \(\eta\) and rule out omitted variable bias in the identification of \(u_i (W)\) and \(v (W)\). Assumption B3 gives probability zero to tie events; it rules out with probability 1 that an agent is indifferent between two bundles.

Under B2(i) and B3, the conditional probability of observing bundle \(a\) given the observables \(X = x\) is captured by
\[
\Pr (a \mid x) = \int 1 \left[ a \in \arg \max_{\hat{a}} U (\hat{a}, x, \varepsilon, \eta) \right] dF_{\varepsilon, \eta | w}. \tag{2}
\]

We next describe the main identification results.

The approach we use to identify the model structure requires prior identification of the sign of the interaction effect.

**Lemma 1.** If B2(i) holds and \(Z\) takes on support in an open set, then the sign of \(v (w)\) is identified for each \(W = w\).

The proof of Lemma 1 builds on monotone comparative statics methods and therefore relies on the concepts of stochastic dominance, supermodularity, and increasing differences, which we describe in Appendix B for completeness. The intuition for the sign identification is simple: when the interaction effect is positive, the probability of selecting alternative 1 increases with the excluded regressor \(Z_2\), as this excluded regressor makes alternative 2 more valuable. Because the opposite holds when the interaction effect is negative, whether the goods are complements or substitutes can be inferred from the data. Using Lemma 1, the following theorem states identification of aspects of the model.

**Theorem 1.** Under B1–3, \(\{(u_i)_{i \leq 2}, v\}\) is identified. In addition:

- If \(v (w) < 0\), then \(F_{\varepsilon \mid W}\) is identified. Moreover, if \(\eta\) and \(\varepsilon\) are independent conditional on \(W\), then \(F_{\eta \mid W}\) is also identified.

- If \(v (w) \geq 0\) and \(\eta = 1\), then \(F_{\varepsilon \mid W}\) is identified.

\(^{11}\)If \(E(\varepsilon \mid W)\) and \(E(\eta \mid W)\) are nonconstant functions of \(W = w\), the values of \(u_1 (W), u_2 (W)\) and \(v (W)\) produced by the proof strategy for Theorem 1 below will suffer from omitted variable bias. If, in this case, \(E(\varepsilon \mid T) = (0, 0)\) and \(E(\eta \mid T) = (1, 1)\) 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.
For each $W = w$, our identification strategy proceeds by first learning the sign of $v(w)$ using Lemma 1. The reason is that for each sign of the interaction effect there are two bundles (the identity of which is sign dependent) that share the following feature: **whenever unilateral changes of each alternative decrease utility, then the bundle maximizes the overall payoff.** As we explain next, this feature facilitates identification.

When the alternatives are substitutes so that $v(w) \leq 0$, then the two bundles that share this feature are $(0, 0)$ and $(1, 1)$. To show our claim, note that $(0, 0)$ maximizes the overall utility if

$$0 \geq \epsilon_1 + u_1(w) + z_1, \quad 0 \geq \epsilon_2 + u_2(w) + z_2,$$

and

$$0 \geq \epsilon_1 + u_1(w) + z_1 + \epsilon_2 + u_2(w) + z_2 + \eta \cdot v(w).$$

Because $\eta \cdot v(w) \leq 0$, the third inequality is implied by the first two inequalities. Therefore, $(0, 0)$ is optimal whenever it is not utility increasing for the agent to individually incorporate in its bundle any of the alternatives. It follows that the observed probability of choosing neither good, conditional on the covariates, is given by

$$\Pr ((0, 0) | x) = \Pr (\epsilon_1 + u_1(w) \leq -z_1, \epsilon_2 + u_2(w) \leq -z_2 | x). \tag{3}$$

We therefore use the probabilities $\Pr ((0, 0) | x)$ to identify the CDF of $(\epsilon_1 + u_1(w), \epsilon_2 + u_2(w))$ at all points in the support of $-Z$. With full support on the excluded regressors $Z$ and the mean independence assumption on the unobservables, we identify $u_1, u_2$ and the bivariate distribution $F_{\epsilon|w}$. Note that $v$ and $\eta$ do not enter (3).

We use the bundle $(1, 1)$ to recover $v(W)$ and $F_{\eta|w}$. When the alternatives are substitutes, $(1, 1)$ is an optimal bundle if the next inequalities are simultaneously satisfied

$$\epsilon_1 + u_1(w) + z_1 + \eta \cdot v(w) \geq 0, \quad \epsilon_2 + u_2(w) + z_2 + \eta \cdot v(w) \geq 0,$$

and

$$\epsilon_1 + u_1(w) + z_1 + \epsilon_2 + u_2(w) + z_2 + \eta \cdot v(w) \geq 0.$$

Because $\eta \cdot v(w) \leq 0$, the third inequality is implied by the first two. Therefore, $(1, 1)$ is optimal whenever it is not utility increasing for the agent to individually discard from this bundle either of the alternatives. It follows that the observed probability of choosing both alternatives, conditional on the covariates, is given by

$$\Pr ((1, 1) | x) = \Pr (\epsilon_1 + u_1(w) + \eta \cdot v(w) \geq -z_1, \epsilon_2 + u_2(w) + \eta \cdot v(w) \geq -z_2 | x). \tag{4}$$

We then recover $v$ and the univariate distribution $F_{\eta|w}$ by using the full support of $Z$, knowledge of $u_1, u_2$ and $F_{\epsilon|w}$ from the previous step, and properties of characteristic functions under the
independence of $\varepsilon$ and $\eta$. If $\varepsilon$ and $\eta$ are dependent, then this argument does not identify the trivariate distribution $F_{\varepsilon,\eta|W}$, although the homogeneous functions $u_1$, $u_2$ and $v$ as well as $F_{\varepsilon|W}$ are identified even when $\varepsilon$ and $\eta$ are dependent. Unfortunately, the full set of counterfactuals and elasticities cannot be computed if the entire structure of the model, which includes $F_{\varepsilon,\eta|W}$ under dependence, is not identified.\footnote{The independence of $\varepsilon$ and $\eta$ suggests that the variance of the unobservable component of utility for bundles, $\varepsilon_1 + \varepsilon_2 + \eta$, will often (but not always because $\varepsilon_1$ and $\varepsilon_2$ can be correlated) be greater than the two variances of the unobservable components of the standalone utilities, $\varepsilon_1$ and $\varepsilon_2$. Most prior parametric empirical work on bundles assumption that $\eta \equiv 1$, so that our full identification result under the independence of $\varepsilon$ and $\eta$ still extends the literature.} If one wishes to use Theorem 1 to prove the consistency of point identified sieve maximum likelihood (Chen 2007), the researcher needs to impose that $\varepsilon$ and $\eta$ are independent. Alternatively, one can allow for dependence between $\varepsilon$ and $\eta$ and use set identified sieve maximum likelihood (Chen, Tamer and Torgovitsky 2011). The homogenous functions $u_1$, $u_2$ and $v$ will still be point identified.

When the alternatives are complements, then the two bundles that share the above feature are $(0, 1)$ and $(1, 0)$. To recover the homogeneous functions $((u_i)_{i \leq 2}, v)$, we use the observed probabilities of selecting these bundles. We do not need to restrict $F_{\varepsilon,\eta|W}$ to identify $((u_i)_{i \leq 2}, v)$.

Our argument for complements does not identify the univariate distribution of $\eta$ and we assume $\eta = 1$ when we wish to identify the bivariate distribution $F_{\varepsilon|W}$. Such an assumption is needed to prove the consistency of point identified sieve maximum likelihood (Chen 2007). An alternative to assuming $\eta = 1$ is set identified sieve maximum likelihood (Chen et al. 2011).

We will show later that the proof strategy we just discussed for the bundles model is very similar to the one we use to identify a binary game. This is surprising as single agent choice models and games are a priori quite different models.

In many applications, the large support requirement on the excluded regressors $Z_1$, $B_1$, might not be satisfied. When the alternatives refer to goods and we use prices as excluded regressors, $B_1$ requires the observation of extreme price values that are unlikely to appear in the data. We next show that we can recover part of the model structure without large support restrictions if we replace $B_2(\Omega)$ with a location normalization that is not as sensitive to the tails of the unobservables as the one we used earlier. The proof of the following result builds on Kline (2013b), whose analysis is for a semiparametric two-player game. One of the main ideas we want to convey is the similarity of identification in bundles and games, and so the extension of Kline’s result to bundles is in line with this message.

For each realization $W = w$, let the scalars $u_1(w)$, $u_2(w)$, $u_1(w) + v(w)$ and $u_2(w) + v(w)$ all lie in a bounded interval $\Theta \subseteq \mathbb{R}$. If the interval varies with $w$, let $\Theta$ be the union of all such $w$-specific intervals.
Theorem 2. Under the next restrictions, \(\left((u_i)_{i \leq 2}, v\right)\) is identified.

**B1':** \(Z_i | W, Z_{-i}\) has support on a superset of \(\Theta\) for \(i = 1, 2\).

**B2':** (i) \(F_{\varepsilon|X} = F_{\varepsilon|W}\); (ii) \(\varepsilon | W\) has mode \((0, 0)\); and (iii) \(F_{\varepsilon|W}\) is \(C^2\).

**B3':** \(\varepsilon | W\) has an everywhere positive Lebesgue density on its support.

**B4':** \(\eta = 1\).

Note that B1' is not a true limited support condition because it only restricts the required support of the excluded regressors when the researcher constrains the parameter spaces for \(u_1(W), u_2(W)\) and \(v(W)\) before examining the data. It is, nevertheless, weaker than B1. Assumption B4' rules out heterogeneous effects \(\eta\) in the interaction term for the cases of both substitutes and complements. Although not stated in the theorem, the proof indicates that we can identify \(u_1(W)\) and \(u_2(W)\) for the case of substitutes without any restrictions on \(\eta\) or its dependence with \(\varepsilon\), as \(\eta\) does not enter \(\Pr (0, 0 | x)\), equation (3), for the case of substitutes.

From equation (3), one can further see that the distribution of \((\varepsilon_1, \varepsilon_2)\) will be identified in the support of \((-Z_1 + u_1(W), -Z_2 + u_2(W))\). Therefore, as \(u_1(W)\) and \(u_2(W)\) are identified, the bivariate CDF of \(\varepsilon\) will be identified at some points.

For the case of substitutes, the algebra in the proof can be easily modified to identify \(v(W)\) under heterogeneous interaction effects \(\eta\) if \(\eta\) enters additively separably from \(v(W)\), as in \(v(W) + \eta\). \(\eta\)'s support is such that the sign \(v(W) + \eta\) never changes given a realization of \(W\), and we use the assumption that the random variable \((\varepsilon_1 + \eta, \varepsilon_2 + \eta) | W\) has a mode of \((0, 0)\). For the case of complements, a slightly less elegant assumption about the mode of a linear sum of \(\varepsilon\) and \(\eta\) can be used to identify \(u_1(W), u_2(W)\) and \(v(W)\) under the presence of the additively separable version of \(\eta\).

Zhou (2013) studies identification and estimation of binary games of complete information under assumption B1' while adding the assumption that the distribution of the vector \(\varepsilon\) is so-called centrally symmetric. Compared to the approach of Kline (2013b), Zhou’s symmetry assumption results in a faster rate of convergence, namely the standard, parametric \(\sqrt{n}\) rate, for the intercept terms in parametric versions of \(u_1(W), u_2(W)\) and \(v(W)\).

### 2.3 Testability

Our bundles model has testable implications. As the proof of Lemma 1 indicates, the choice probability of alternative 1, the sum of the probabilities of \((1, 0)\) and \((1, 1)\), is monotonically
increasing in $Z_1$. Likewise, the choice probability of alternative 2, the sum of the probabilities of $(0, 1)$ and $(1, 1)$, is monotone with a possibly unknown direction in the excluded regressor for the other alternative, $Z_1$. These monotonicity restrictions on how sums of choice probabilities vary with $Z$ can be tested.

In addition, the model is overidentified. Consider the case of substitutes. The constructive identification arguments in the proof of Theorem 1 use data on only the outcomes $(0, 0)$ and $(1, 1)$. There are two other outcomes, $(1, 0)$ and $(0, 1)$, that were not used in the identification argument for the case of substitutes. The model completely determines the choice probabilities $Pr((1, 0) \mid X)$ and $Pr((0, 1) \mid X)$ for all $X = x$. The choice probabilities sum to 1 for each $x$, so for each $x$ there is a single overidentification restriction involving the choice probabilities not used for identification. An analogous argument applies to the case of complements.

### 3 Two-Player Game

#### 3.1 Game Structure

Consider a simultaneous, binary choice game of complete information. The set of players is $\{1, 2\}$. Each player $i$ chooses an action $a_i$ from two possible alternatives $\{0, 1\}$. We denote by $X \in \mathcal{X} \subseteq \mathbb{R}^k$ a vector of observable state variables and let $\varepsilon = (\varepsilon_1, \varepsilon_2) \in \mathbb{R}^2$ and $\eta = (\eta_1, \eta_2) \in \mathbb{R}^2_+$ indicate two vectors of random terms that are observed by the players but not by the econometrician. The random vector $(\varepsilon_1, \varepsilon_2, \eta_1, \eta_2)$ is distributed according to $F_{\varepsilon, \eta \mid X}$, with respective bivariate distributions $F_{\varepsilon \mid X}$ and $F_{\eta \mid X}$. The payoff of player $i$ from choosing action 1 is given by

$$U_{1,i}(a_{-i}, X, \varepsilon_i, \eta_i) = u_i(X) + \varepsilon_i + \eta_i \cdot v_i(X) 1(a_{-i} = 1), \quad (5)$$

while the return from action 0, $U_{0,i}(a_{-i}, X, \varepsilon_i, \eta_i)$, is normalized to 0. The first element of (5), $u_i(X) + \varepsilon_i$, is the standalone value of action 1, and the interaction term $\eta_i \cdot v(X)$ captures the effect that the choice of the other player, $a_{-i}$, has on player $i$. Thus, the difference between $\varepsilon_i$ and $\eta_i$ is that $\varepsilon_i$ is an unobservable in standalone payoffs while $\eta_i$ is an unobservable in the interaction effects. In an entry game, $\varepsilon_i$ may be an unobservable in the fixed costs of entry and $\eta_i$ may be an unobservable affecting competition. We denote the game by $\Gamma(X, \varepsilon, \eta).$ \[14\]

\[13\]Another form of testability comes from the fact that the choice probability for $Pr((1, 1) \mid x)$ in (4) identifies the bivariate distribution of $(\varepsilon_1 + u_1(w) + \eta \cdot v(w) \geq -z_1, \varepsilon_2 + u_2(w) + \eta \cdot v(w) \geq -z_2)$ conditional on $W = w$. Taking the mean, conditional on $w$, of either marginal of this bivariate distribution will identify $v(w)$. For this reason, the function $v(w)$ is overidentified. $F_{\eta \mid W}$ is overidentified for a similar reason.

\[14\]The additive separability of $u_i(x) + \varepsilon_i$ and $1(a_{-i} = 1) \eta_i \cdot v(x)$ does not restrict the model; it is just a convenient way to write the payoffs.
A vector of decisions $a^* = (a_1^*, a_2^*)$ is a pure strategy Nash equilibrium for $X = x$ if, for $i = 1, 2$,$$
abla_i = \begin{cases} 
1 & \text{if } u_i(x) + \varepsilon_i + \eta_i \cdot v_i(x) 1(a_{-i} = 1) > 0, \\
0 & \text{if } u_i(x) + \varepsilon_i + \eta_i \cdot v_i(x) 1(a_{-i} = 1) < 0, \\
1 \text{ or } 0 & \text{otherwise.} 
\end{cases}$

We write $D(x, \varepsilon, \eta)$ for the equilibrium set (in pure strategies) of $\Gamma(x, \varepsilon, \eta)$. The conditions we impose for identification guarantee that $D(x, \varepsilon, \eta)$ is non-empty.

The econometrician observes the vector of covariates and choices, $((a_i)_{i \leq 2}, x)$, for a cross-section of independent two-player games that share the same structure $((u_i, v_i)_{i \leq 2}, F_{\varepsilon, \eta|X})$. We assume the observable choices in each play of the game correspond to a pure strategy Nash equilibrium. These independent games may be different geographic markets in an entry application or different pairs of friends in a peer effects model. The purpose of our analysis is to use data on selected actions and covariates to recover aspects of the structure of the model.\footnote{If the players are anonymous such that agent indices do not have common meaning across markets or plays of the game, then it is common in empirical work to assume $u_1 = u_2$, $v_1 = v_2$, and $F_{\varepsilon, \eta|X}$ is exchangeable in the two pairs of player-specific arguments $(\varepsilon_1, \eta_1)$ and $(\varepsilon_2, \eta_2)$.}

### 3.2 Identification of the Game

The identification strategy for the two-player game relies, as with the bundles model, on excluded regressors. We decompose the observables, $X = (W, Z)$, into regular covariates $W$ and excluded regressors $Z = (Z_1, Z_2)$, where $Z_i$ is a scalar for $i = 1, 2$. These excluded regressors denote player-specific factors that enter the standalone utility additively and are excluded from the interaction term. That is, we let

$$u_i(X) = u_i(W) + Z_i \text{ and } v_i(X) = v_i(W) \text{ for } i = 1, 2.$$ 

As in the bundles model, these exclusion restrictions are key to identifying the joint distribution of unobservables separately from the interaction effects. In an entry game, $Z_i$ might be the distance between chain $i$’s headquarters and the geographic market in question.\footnote{In an entry game, $Z_i$ is typically said to shift the fixed costs of entry and not marginal costs.}

The first set of identification conditions is as follows.

**G1:** $Z_i \mid W, Z_{-i}$ has support on all $\mathbb{R}$ for $i = 1, 2$.

**G2:** (i) $F_{\varepsilon, \eta|X} = F_{\varepsilon, \eta|W}$; (ii) $E(\varepsilon \mid W) = (0, 0)$; and (iii) $E(\eta \mid w) = (1, 1)$.

**G3:** $\varepsilon \mid W$ has an everywhere positive Lebesgue density on its support.
**G4:** For each $W = w$, $\text{sign}(v_1(w)) = \text{sign}(v_2(w))$ and the econometrician knows the signs.

Assumptions G1–3 are similar to B1–3 for the bundles model. Thus, we omit their discussion.

Assumption G4 has no precedent in the bundles model. It requires symmetry and knowledge of the signs of the interaction terms. The symmetry of signs means the game is either a game of strategic complements or a game of strategic substitutes. This condition guarantees existence of at least one equilibrium in pure strategies; see Yildiz (2011) for identification in cases where only a mixed strategy equilibrium may exist.

The symmetry condition also ensures that certain profiles of actions are unique equilibria whenever they are part of the equilibrium set. A unique equilibrium $a$ is such that $\mathcal{D}(x, \varepsilon, \eta) = \{a\}$ whenever $a \in \mathcal{D}(x, \varepsilon, \eta)$. As in the case of bundles, our identification strategy exploits information on the signs of the interaction effects to recover the model structure. An important difference between the two models is that for the case of games, we conjecture that the sign of the interaction effects cannot be recovered from the data without imposing an equilibrium selection rule. For this reason, we assume the sign of the interaction effects is known by the econometrician; see also Bresnahan and Reiss (1991a, 1991b), Berry (1992), and Tamer (2003).

In many applications, such as entry games in industrial organization, the sign of the interaction effects can be motivated by economic theory.

Under G2(1) and G3, the conditional probability of an action profile $a$ in the data is given by

$$
\Pr(a \mid x) = \int 1[a \in \mathcal{D}(x, \varepsilon, \eta)] \Pr[a \text{ selected } \mid x, \varepsilon, \eta] dF_{\varepsilon, \eta \mid W}. \tag{6}
$$

The conditional probability of an action profile $a$ in the data is the joint probability that $a$ is part of the equilibrium set $\mathcal{D}(x, \varepsilon, \eta)$ and that $a$ is the selected equilibrium. $\Pr(a \mid x)$ is thereby a mixture distribution over equilibria in the regions of non-uniqueness. The mixture distribution potentially depends on the realizations of the observables and unobservables: $x$, $\varepsilon$ and $\eta$. The identification result is captured by the next theorem.

**Theorem 3.** Under G1–4, $((u_i)_{i \leq 2}, (v_i)_{i \leq 2})$ is identified. In addition:

- If $v_i(w) < 0$ for $i = 1, 2$, then $F_{\varepsilon \mid W}$ is identified. Moreover, if $\eta$ and $\varepsilon$ are independent conditional on $W$, then $F_{\eta \mid W}$ is also identified.

- If $v_i(w) \geq 0$ for $i = 1, 2$ and $\eta = (1, 1)$, then $F_{\varepsilon \mid W}$ is identified.

\(^{17}\)Kline (2013a) identifies the sign of interaction effects in a two-player binary game, but uses restrictions on equilibrium selection to do so.
The proof of this result is quite similar to the proof for the bundles model as the current proof also depends on the sign of the interaction effects. In the context of games, the reason is that for each sign of the interaction effects there are two profiles of actions (whose identity is sign dependent) that are unique equilibria, as defined above, whenever they are part of the equilibrium set. These profiles of actions share the following feature: **whenver unilateral changes of actions are non-profitable for the players, then the action profile is a unique equilibrium.**

When the game is submodular, then the two action profiles that share this feature are (0, 0) and (1, 1). The probability of observing action profile (0, 0), conditional on the covariates, is given by

\[
\Pr ((0, 0) \mid x) = \Pr (\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq -z_2 \mid x). \quad (7)
\]

We do not need to consider the deviation from (0, 0) to (1, 1) because, being unique, this equilibrium only requires that unilateral deviations be unprofitable (see also Tamer 2003). We therefore use the probabilities \( \Pr ((0, 0) \mid x) \) to identify the CDF of \((\varepsilon_1 + u_1(w), \varepsilon_2 + u_2(w))\) at all points in the support of \(-Z\). With full support on the excluded regressors and the mean independence assumption on the unobservables, we identify \(u_1, u_2\) and the bivariate distribution \(F_{\varepsilon|W}\).

To recover \(F_{\eta|W}, v_1\) and \(v_2\), we then use the observed probabilities of selecting the profile of actions (1, 1) in the data. In the context of strategic substitutes, the choice probability is given by

\[
\Pr ((1, 1) \mid x) = \Pr (\varepsilon_1 + u_1(w) + \eta_1 \cdot v_1(w) \geq -z_1, \varepsilon_2 + u_2(w) + \eta_2 \cdot v_2(w) \geq -z_2 \mid x). \quad (8)
\]

As before, we do not need to consider the deviation from (1, 1) to (0, 0) because, being unique, Nash equilibrium examines only unilateral deviations. We then recover \(v_1, v_2\) and \(F_{\eta|W}\) by using properties of characteristic functions under the independence of \(\varepsilon\) and \(\eta\). In games, \(\eta\) is a vector and there is one interaction term \(v_i(W)\) for each player; otherwise the algebraic relationship between outcome probabilities and CDFs is nearly identical to the case of bundles.

If the game is of strategic complements, then the two action profiles that are unique equilibria are (0, 1) and (1, 0). To recover the homogeneous functions \((u_i, v_i)_{i \leq 2}\), we use the observed probabilities of choosing these action profiles. As with complements in the bundles model, our argument does not identify the distribution of \(\eta\) and we therefore assume \(\eta = (1, 1)\) if we wish.
to identify $F_{ε|W}$.

Theorem 3 identifies the game under the key restrictions of symmetric signs of interaction effects and knowledge of them. Berry and Tamer (2006, Result 4), following Tamer (2003), study identification in two-player, submodular games where the key excluded regressors enter multiplicatively instead of additively (without heterogeneous interaction effects). The advantage of our identification strategy with respect to their provided proof is that ours does not rely on identification at infinity to recover the payoff terms. We discussed the simultaneous work of Kline (2013a) and Dunker et al. (2013) in the introduction.

As we did for the bundles model, we next show that we can recover part of the model structure without the large support restriction on the excluded regressors, if we also change the location restrictions on the unobservables. This result closely but not exactly reproduces a result originally found in Kline (2013b).

As with bundles, point identification with bounded excluded regressors relies on ex ante bounds specified by the researcher on the unknown homogeneous functions, $u_i(W)$ and $v_i(W)$, for $i = 1, 2$. For each realization $W = w$, let the scalars $u_1(w)$, $u_2(w)$, $u_1(w) + v_1(w)$ and $u_2(w) + v_2(w)$ all lie in a bounded interval $Θ ⊆ \mathbb{R}$. If the interval varies with $w$, let $Θ$ be the union of all such $w$-specific intervals.

**Theorem 4.** Under the next restrictions, $(u_i, v_i)_{i \leq 2}$ is identified.

\[ G1': \ Z_i \mid W, Z_{-i} \text{ has support on a superset of } Θ \text{ for } i = 1, 2. \]

\[ G2': (i) \ F_{ε|X} = F_{ε|W}; (ii) \ ε \mid W \text{ has mode } (0,0); \text{ and } (iii) \ F_{ε|W} \text{ is } C^2. \]

\[ G3': \ ε \mid W \text{ has an everywhere positive Lebesgue density on its support.} \]

\[ G4': \ η = (1,1). \]

\[ G5': \text{ For each } W = w, \text{ sign}(v_1(w)) = \text{ sign}(v_2(w)) \text{ and the econometrician knows the signs.} \]

The results of Zhou (2013) on estimation at the parametric rate mentioned before are stated for games.

### 4 Two-Player Potential Games

The identification strategies we followed for the bundles model and the binary game are quite similar, in that almost identical key algebraic equations involving outcome probabilities and
CDFs appeared in the identification arguments for both models. In this section, we show that the two models are indeed mathematically equivalent if we restrict our attention to potential games and impose a specific equilibrium selection rule in the data generating process. This result is important because it means that any progress regarding identification of one of these models can be used with the other.

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 gain in the potential function. A game that admits such a function is called a potential game. This concept was first used in economics as a way to prove the existence of Nash equilibria in pure strategies. The reason is that the global maximum of the potential function always corresponds to a pure strategy Nash equilibrium of the related game. In a finite game, the potential function is defined on a finite set of values. Thus, it always has a global maximizer. It follows that the equilibrium set (in pure strategies) of any finite potential game is guaranteed to be non-empty.

Monderer and Shapley (1996) show that when the game admits a potential function, this function is uniquely defined up to an additive constant. Thus, from a technical perspective, the potential function offers an equilibrium refinement. Important work has been done to address whether the selection rule based on potential maximizers is economically meaningful. Lab experiments studying the so-called minimum effort games have shown that observed choice data are consistent with the maximization of objects close to the potential function associated to 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 displays a very large number of

18Monderer and Shapley (1996) also define ordinal and weighted potential games. We restrict attention to exact potential games.

19According to Monderer and Shapley (1996), this concept appeared for the first time in Rosenthal (1973) to prove equilibrium existence in congestion games.

20Ui (2001) shows that if a unique Nash equilibrium maximizes the potential function, then that equilibrium is generically robust in the sense of Kajii and Morris (1997b). Roughly speaking, a Nash equilibrium of a complete information game is said to be robust if every incomplete information game with payoffs almost always given by the complete information game has an equilibrium that generates behavior close to the original equilibrium. (See Ui (2001) for a formal definition.) All robust equilibria do not necessarily maximize a potential function in a potential game, but, at a minimum, the maximizer of the potential selects one such robust equilibrium. Ui writes that “It is an open question when robust equilibria are unique, if they exist.” In a certain two-player, exact potential game, Blume (1993) argues that a log-linear strategy revision process selects potential maximizers. Morris and Ui (2005) discuss the so-called generalized potential functions and robust sets of equilibria. Weinstein and Yildiz (2007) present a deep critique of all refinements of rationalizability, including Nash equilibrium. They write about the robustness definition in Kajii and Morris (1997a): “Then the key difference between our notions of perturbation is that they focus on small changes to prior beliefs without regard to the size of changes to interim beliefs, while our focus is the reverse. Their approach is appropriate when there is an ex ante stage along with well understood inference rules and we know the prior to some degree.”
We next elaborate on the mathematical equivalence between potential games and bundle models under the additional restriction of an equilibrium selection rule based on potential maximizers. To achieve our goal, we first define the concept of a potential function for a two-player game and then provide a necessary and sufficient condition for $\Gamma (x, \varepsilon, \eta)$ to admit a potential function.

**Potential:** A function $U : \{0, 1\}^2 \times X \times \mathbb{R}^2 \times \mathbb{R}^2_+ \rightarrow \mathbb{R}$ is a potential function for $\Gamma (x, \varepsilon, \eta)$ if, for all $i \leq 2$, and all $a_{-i} \in \{0, 1\}$,

$$U (a_i = 1, a_{-i}, x, \varepsilon, \eta) - U (a_i = 0, a_{-i}, x, \varepsilon, \eta) = U_{1,i} (a_{-i}, x, \varepsilon_i, \eta_i).$$

$\Gamma (x, \varepsilon, \eta)$ is called a potential game if it admits a potential function.

Monderer and Shapley (1996) show that $\Gamma (x, \varepsilon, \eta)$ is a potential game if and only if the interaction effects on players’ payoffs are symmetric.

**Proposition 1.** $\Gamma (x, \varepsilon, (\eta_1, \eta_2))$ is a potential game if and only if $\eta_1 \cdot v(x) = \eta_2 \cdot v(x)$. Letting a scalar $\eta = \eta_1 = \eta_2$ and a scalar-valued function $v = v_1 = v_2$, the potential function can be written as

$$U (a, x, \varepsilon, \eta) = \sum_{i \leq 2} (u_i (x) + \varepsilon_i) 1 (a_i = 1) + \eta \cdot v(x) 1 (a_1 = 1, a_2 = 1).$$

The proof of this proposition follows directly from Ui (2000, Theorem 3) and is therefore omitted. It states that when the interaction effects are the same across the players, meaning $v_1 (W) = v_2 (W)$ and $\eta_1 = \eta_2$, then the game admits a potential function. Moreover, the potential is identical to the overall utility in the bundle model — see (1). We simply need to interpret the players as different alternatives. Proposition 1 emphasizes that the potential function is uniquely characterized — up to an additive constant — by the structure of the game and is not a construction of the modeler, like a social welfare function. The assumption of equal interaction effects is restrictive. In the empirical literature, it is often maintained when players are anonymous, meaning that player indices have no particular meaning.

We mentioned earlier that when the game admits a potential function, then the potential maximizer is always a pure strategy Nash equilibrium in the underlying game. Thus, if we

---

21See Monderer and Shapley (1996, Section 5) for the original discussion of how Van Huyck et al.’s experimental evidence relates to potential games.
assume that players coordinate according to the potential maximizer rule, identification of the game is mathematically equivalent to identification of the bundles model. In turn, this guarantees that the sign of the interaction term in the game can be recovered from the data.

**Proposition 2.** If in the binary game we let \( \eta_1 = \eta_2 = \eta \) and \( v_1 = v_2 = v \) for scalar \( \eta \) and scalar-valued \( v(x) \) and assume that players coordinate on the potential maximizer, then identification of the game is mathematically equivalent to identification of the bundle model.

We end this section by elaborating on the meaning of an equilibrium selection rule based on potential maximizers. When the equilibrium of the potential game is unique, the unique equilibrium is always the potential maximizer. Under multiple equilibria, the potential function selects an equilibrium depending on the sign of the interaction effect. When \( \eta \cdot v(x) \leq 0 \) (e.g., an entry game) and the game has multiple equilibria, the equilibrium set is \( D(x, \varepsilon, \eta) = \{(0, 1), (1, 0)\} \). In this case, \((0, 1)\) maximizes the potential if

\[
u_2(x) + \varepsilon_2 > u_1(x) + \varepsilon_1,
\]

while \((1, 0)\) is the maximizer otherwise. Thus, the equilibrium selection rule induced by the potential function 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 \( \eta \cdot v(x) \geq 0 \) (e.g., a game of peer effects) and the game has multiple equilibria, the equilibrium set is \( D(x, \varepsilon, \eta) = \{(0, 0), (1, 1)\} \). It can be easily checked that \((1, 1)\) maximizes the potential if

\[
(-u_1(x) - \varepsilon_1 - \eta \cdot v(x)) (-u_2(x) - \varepsilon_2 - \eta \cdot v(x)) > (u_1(x) + \varepsilon_1) (u_2(x) + \varepsilon_2),
\]

while \((0, 0)\) is the maximizer otherwise. That is, players coordinate on \((1, 1)\) if the product of deviation losses from selecting action 0 as compared to action 1 while the other player selects action 1 are lower than the product of deviation loses from selecting action 1 instead of action 0 when the other player selects action 0. In this case, the potential maximizer corresponds to the less risky equilibrium of Harsanyi and Selten (1988). Many laboratory experiments have shown that the less risky equilibrium is selected even when another equilibrium gives higher payoffs to both players (e.g., Van Huyck, Battalio and Beil 1990, Goeree and Holt 2005, and Chen and Chen 2011).
5 Three or More Goods and Players

We extend our identification results to three or more goods and three or more players.

5.1 Discrete Choice Model for Bundles

5.1.1 Bundles Model with $n$ Goods

Consider an agent that faces $N = \{1, 2, \ldots, n\}$ binary alternatives that are not mutually exclusive. Thus, her choice set is $\{0, 1\}^n$, where as before we use 1 to indicate that the corresponding alternative is selected. We define $a = (a_i)_{i \leq n}$.

If the agent selects only alternative $i$ then her payoff is $u_i (X) + \varepsilon_i$ for $i = 1, 2, \ldots, n$, where $X \in \mathcal{X} \subseteq \mathbb{R}^k$ is a vector of observable state variables and $\varepsilon = (\varepsilon_i)_{i \leq n} \in \mathbb{R}^n$ indicates a vector of random terms that are observed by the agent but not by the econometrician. The random vector $\varepsilon$ is distributed according to $F_{\varepsilon|X}$. We write $S(a) = \{i \in N : a_i = 1\}$ for the set of alternatives that are actually selected if bundle $a$ is chosen. The agent selects bundle $a$ to maximize her utility

$$U(a, X, \varepsilon) = \sum_{i \leq n} (u_i (X) + \varepsilon_i) 1 (a_i = 1) + v (S(a), X),$$

where $v (S(a), X)$ is the interaction effect between the selected alternatives. This specification allows the interaction term to vary with the alternatives selected. It nests an example where the joint selection of alternatives 1 and 2 increases the overall payoff as compared to the standalone utilities, while the joint selection of 1 and 3 reduces it. We normalize $v (S(a), x)$ so that it is 0 when $|S(a)| < 2$ and let the overall utility be 0 if bundle $(0,0,\ldots,0)$ is chosen.\footnote{Gentzkow (2007b, Proposition 2) explores comparative statics of choice probabilities in the excluded regressors $Z$; they depend on the interaction effects in a somewhat complex manner.}

The purpose of our analysis is to recover the structure of the discrete choice for bundles model, $((u_i)_{i \leq n}, v, F_{\varepsilon|X})$, from available data on choices and covariates $(a, x)$.

5.1.2 Identification of the Model

The identification strategy we follow again relies on excluded regressors. Specifically, we assume $X = (W, Z)$, where $W$ is a sub-vector of standard covariates and $Z$ represents the excluded regressors. We define $Z = (Z_{i \leq n})$, where $Z_i$ is a scalar for $i = 1, 2, \ldots, n$. These regressors denote alternative-specific factors that enter the standalone utility additively and are excluded from the interaction terms. That is, we let

$$S(a) = \{i \in N : a_i = 1\}.$$
\[ u_i(X) = u_i(W) + Z_i \text{ for } i = 1, 2, \ldots, n \text{ and } v(S(a), X) = v(S(a), W) \text{ for all } S(a) \subseteq \mathcal{N}. \]

We next provide the set of identifying restrictions. We add “M” in front of each assumption to differentiate this case from the previous, two-alternatives, model. Let \( Z_{\sim i} \) be all other excluded regressors than the one for choice \( i \).

**MB1:** \( Z_i | W, Z_{\sim i} \) has support on all \( \mathbb{R} \) for \( i = 1, 2, \ldots, n \).

**MB2:** (i) \( F_{\varepsilon|X} = F_{\varepsilon|W} \) and (ii) \( E(\varepsilon | W) = (0, 0, \ldots, 0) \).

**MB3:** \( \varepsilon | W \) has an everywhere positive Lebesgue density on its support.

**MB4:** For each \( W = w \), there exists a known vector \( \hat{a}_w \in \{0, 1\}^n \) with the following property.

\[
\text{For all } z, \text{ with } x = (w, z), \text{ and for all } \varepsilon, U(a, x, \varepsilon) \text{ is maximized at } \hat{a}_w \text{ if } U(a, \hat{a}_{\sim i}, x, \varepsilon) \geq U(a_i, \hat{a}_{\sim i}, x, \varepsilon) \text{ for all } i \in \mathcal{N}, \text{ where } a_i = 1 - \hat{a}_i^w. \]

We do not allow for heterogeneity in the interaction terms in the multiple alternatives model.

In addition, we need to add MB4; as we show below, this condition is always fulfilled when there are only two alternatives. This assumption requires the existence of a vector of choices (known by the econometrician) such that if the vector is a local maximizer, then it is a global maximizer as well. Assumption MB4 can be thought of as a notion of local concavity for discrete domains.

We use this assumption to trace \( F_{\varepsilon|W} \) using variation in \( Z \). We provide sufficient conditions for assumption MB4 after describing the identification result.

The identification result is as follows.

**Theorem 5.** Under MB1–4, \( ((u_i)_{i \leq n}, v, F_{\varepsilon|W}) \) is identified.

The proof works by first using condition MB4 to trace \( F_{\varepsilon|W} \) using variation in the excluded regressors \( Z \). We then use the known \( F_{\varepsilon|W} \) to show that there exist realizations of \( Z \) where different values of \( ((u_i)_{i \leq n}, v) \) lead to different bundle probabilities. This latter step is formalized by a contradiction argument and involves a subtle detail not present in proofs of identification for multinomial choice models without bundles: some inequalities within choice probabilities are always implied by other inequalities. This was illustrated earlier for the two good case: in the case of substitutes, if it does not maximize utility to buy either good alone, then the bundle

---

23The number of such interaction terms grows quickly with the number of goods. Indeed, there is one interaction term for each bundle with more than two elements. Our strategy, for the case of two goods, of identifying one distribution of unobservables for each outcome probability \( \Pr(a | x) \) used in identification will not allow the identification of the joint distribution of heterogeneous interaction terms.
of both goods will also not maximize utility. In a discrete choice model without bundles, it is always the case that increasing a parameter entering the payoff for alternative \( i \) will strictly increase the probability of purchasing \( i \). However, for bundles, changing unknown parameter values locally may not affect choice probabilities if those parameter values enter only inequalities that are implied by other inequalities. This detail makes the proof harder.

The next three conditions are sufficient for assumption MB4 to hold.

**Two Goods**

MB4 always holds if there are two goods, so that the assumption in this case is not stronger than the previous analysis. If MB1–3 are satisfied and \( \mathcal{N} = \{1, 2\} \), then by Lemma 1 the sign of \( v(w) = v((1, 1), w) \) is identified for all \( w \). It can be easily shown that MB4 holds with \( \tilde{a}^w = (1, 0) \) or \( \tilde{a}^w = (0, 1) \) when \( v(w) \geq 0 \) and \( \tilde{a}^w = (0, 0) \) or \( \tilde{a}^w = (1, 1) \) when \( v(w) \leq 0 \).

**Negative Interaction Effects**

The second sufficient condition relies on the goods being substitutes. The next lemma shows that MB4 holds if \( U(a, x, \varepsilon) \) has the negative single-crossing property in \((a_i; a_{-i})\) for all \( i \in \mathcal{N} \).

**Lemma 2.** Assume \( U(a, x, \varepsilon) \) has the negative single-crossing property in \((a_i; a_{-i})\) for all \( i \in \mathcal{N} \); meaning that for all \( a'_i - a_i \geq a_{-i} \) (in the coordinatewise order) and for all \( x, \varepsilon \) we have

\[
U(a_i = 1, a_{-i}, x, \varepsilon) - U(a_i = 0, a_{-i}, x, \varepsilon) \leq (\leq) 0 \iff U(a_i = 1, a'_{-i}, x, \varepsilon) - U(a_i = 0, a'_{-i}, x, \varepsilon) \leq (\leq) 0.
\]

Then assumption MB4 holds with \( \tilde{a}^w = (0, 0, \ldots, 0) \) for all \( w \).

In this lemma, \( a'_i > a_i \) means that (leaving aside alternative \( i \)), bundle \( a' \) includes more selected alternatives as compared to \( a \). In this case, identification of \( F_{(x)} \) uses data on the fraction of consumers who do not purchase any of the goods, \( \Pr(\{(0, 0, \ldots, 0) | x\}) \). The single-crossing condition in Lemma 2 captures the idea that alternatives are substitutes.

**Global Concavity for Discrete Domains**

MB4 is a local notion of concavity for discrete domains. The last sufficient condition for the assumption is a more global analog of discrete concavity.

**Lemma 3.** Assume that, for all \( a, a' \in \{0, 1\}^n \) with \( \|a - a'\| = 2 \),

\[
\max_{a'': \|a - a''\| = \|a' - a''\| = 1} U(a'', x, \varepsilon) \begin{cases} > \min \{U(a, x, \varepsilon), U(a', x, \varepsilon)\} & \text{if } U(a, x, \varepsilon) \neq U(a', x, \varepsilon) \\ \geq U(a, x, \varepsilon) = U(a', x, \varepsilon) & \text{otherwise} \end{cases}
\]
Then assumption MB4 holds with any \( \hat{\alpha}^w \in \{0, 1\}^n \) for all \( w \).

Global concavity imposes non-trivial restrictions on the cross effects of multivariate functions and, in our model, the support of unobservables and regressors. In particular, if the unobservables and regressors can take values on the entire real line, then the interaction effects must be identically zero for concavity to hold globally. Thus, this requires MB1 to be relaxed. For this reason, we cannot recommend basing identification on global concavity explicitly. However, assumption MB4 itself can be thought as a weaker (local) version of discrete concavity that is compatible with our other restrictions.

5.2 \( n \)-Player Game

5.2.1 Game Structure

Consider an extension of the game in Section 3 to three or more players. The set of players is \( \mathcal{N} = \{1, ..., n\} \). Each player \( i \in \mathcal{N} \) chooses an action \( a_i \in \{0, 1\} \). We denote by \( X \in \mathcal{X} \subseteq \mathbb{R}^k \) a vector of observable state variables and by \( \varepsilon = (\varepsilon_i)_{i \leq n} \in \mathbb{R}^n \) a vector of random terms that are observed by the players but not by the econometrician. The random vector \( \varepsilon \) is distributed according to \( F_{\varepsilon|X} \). The payoff of player \( i \) from choosing action 1 is

\[
U_{1,i} (a_{-i}, X, \varepsilon_i) = u_i (X) + v_i (a_{-i}, X) + \varepsilon_i,
\]

while the return from action 0, \( U_{0,i} (a_{-i}, X, \varepsilon_i) \), is normalized to 0. In addition, we normalize \( v_i (a_{-i}, X) \) to 0 when the actions of all players but \( i \) are 0. We denote this game by \( \Gamma (x, \varepsilon) \).

The definition of a pure strategy Nash equilibrium \( a^* = (a^*_i)_{i \leq n} \) naturally extends from the two-player case. We write \( \mathcal{D} (x, \varepsilon) \) for the equilibrium set (in pure strategies) of \( \Gamma (x, \varepsilon) \). The same conditions that facilitate identification of the game guarantee that \( \mathcal{D} (x, \varepsilon) \) is non-empty.

The purpose of our analysis is to recover the structure of the game, \( ((u_i, v_i)_{i \leq n}, F_{\varepsilon|x}) \), from available data on equilibrium choices and covariates, \( (a, x) \), for each play of the game.

5.2.2 Previous Identification Strategy with No Selection Rule

We first explore our identification strategy without imposing an equilibrium selection rule, which for the two player case was done in Section 3. Recall that the identification strategy for two players used two unique equilibria. For a submodular (example, entry) game where \( \eta = (1, 1) \), we used the probabilities \( \Pr (0, 0 \mid x) \) to identify \( F_{\varepsilon|W} \) and each \( u_i (W) \). We then used the probabilities \( \Pr (1, 1 \mid x) \) to identify the interaction effects, which here we call \( v_i (1, W) \).
In a submodular game of three players, the equilibria $a = (0, 0, 0)$ and $a = (1, 1, 1)$ are also unique equilibria, in the sense that $\mathcal{D}(x, \varepsilon) = \{a\}$ whenever $a \in \mathcal{D}(x, \varepsilon)$. However, the number of interaction effects is now quite large. There are three interaction effect functions for each of the three players. For example, for player 1 the interaction effects are $v_1((1, 0), W)$, $v_1((0, 1), W)$, and $v_1((1, 1), W)$, yielding $3 \cdot 3 = 9$ interaction effect values to identify for each $W = w$. While the probabilities $\Pr(0, 0, 0 \mid W)$ are enough to identify $F_{\varepsilon|W}$ and each $u_i(W)$, the probabilities $\Pr(1, 1, 1 \mid x)$ and the proof strategy used in Section 3 will not identify nine interaction terms.

The number of interaction effects increases quickly with the number of players. For submodular games of four players, there will be $6 \cdot 4 = 24$ interaction effects to identify from the probabilities $\Pr(1, 1, 1, 1 \mid x)$. For these combinatorial reasons, the strategy of identifying all the model parameters by relying on the unique equilibria $(0, \ldots, 0)$ and $(1, \ldots, 1)$ for submodular games likely applies only to two-player games. In addition, supermodular games with three or more players may not even have unique equilibria. The entire identification strategy in Section 3 relies on unique equilibria. As a comment on the identification literature for games, we note that many papers rigorously establish identification for games of two players and make comments about generalizations to three or more players being natural. Our arguments here suggest that generalizations to games of three or more players are not simple.

### 5.2.3 Representation of the Game via Potential Functions

Because the identification strategy relying on unique equilibria does not appear to extend to the case of three or more players, we turn our attention to potential games.

We first specify the definition of a potential function for the $n$-player case and then provide a necessary and sufficient condition for $\Gamma(x, \varepsilon)$ to be a potential game.

**Potential:** A function $U : \{0, 1\}^n \times \mathcal{X} \times \mathbb{R}^n \to \mathbb{R}$ is a potential function for $\Gamma(x, \varepsilon)$ if, for all $i \leq n$, for all $a_{-i} \in \{0, 1\}^{n-1}$,

$$U(a_i = 1, a_{-i}, x, \varepsilon) - U(a_i = 0, a_{-i}, x, \varepsilon) = U_{1,i}(a_{-i}, x, \varepsilon).$$

$\Gamma(x, \varepsilon)$ is a potential game if it admits a potential function.

We write

$$\mathcal{S}(a) = \{S \subseteq S(a) \mid |S| \geq 2\} \text{ and } \mathcal{S}(a, i) = \{S \subseteq S(a) \mid |S| \geq 2, i \in S\}.$$
Each player’s payoff from action 0 is normalized to 0; the notation $S \in S(a,i)$ returns the empty set when $a_i = 0$. The next proposition derives from Ui (2000, Theorem 3).

**Proposition 3.** $\Gamma (x, \varepsilon)$ is a potential game if and only if there exists a function

$$\{ \tilde{v}(S, x) \mid \tilde{v}(S, x) : \mathcal{N} \times \mathcal{X} \rightarrow \mathbb{R}, |S| \geq 2 \}$$

such that, for all $a \in \{0,1\}^n$ and all $i \in \mathcal{N}$,

$$U_{1,i}(a, x, \varepsilon_i) = u_i(x) + \varepsilon_i + \sum_{S \in S(a,i)} \tilde{v}(S, x).$$

A potential function is given by

$$U(a, x, \varepsilon) = \sum_{i \in \mathcal{N}} (u_i(x) + \varepsilon_i) 1(a_i = 1) + v(S(a), x) \text{ with } v(S(a), x) = \sum_{S \in S(a)} \tilde{v}(S, x).$$

Briefly, an $n$-player game admits a potential representation if the interaction terms are group-wise symmetric. That is, if players $i$ and $j$ are in $S(a)$, then the corresponding interaction effects for players $i$ and $j$ are the same. The restriction does not imply that the effect of $i$’s entry, say, on player $j$ is the same as $i$’s effect on player $k$. Rather, $i$’s effect on $j$ must be the same as $j$’s effect on $i$.

The purpose of our analysis is to combine data with initial assumptions to learn about the structure of the game $(U_i, v, F_{\varepsilon|x})$.

### 5.2.4 Identification for Potential Games

We first state that identification of a potential game under an equilibrium selection rule based on potential maximizers is the same as identification of the bundle model. We then interpret the required conditions for identification of the bundle model in the context of a game.

**Proposition 4.** If a binary game admits a potential function and players coordinate on the potential maximizer, then identification of the game is mathematically equivalent to identification of the bundle model.

By Theorem 5, $(U_i)_{i \leq n}, v, F_{\varepsilon|x})$ is identified in the bundle model and it is readily verified that we can recover $\tilde{v}$ from $v$ by a one-to-one change of notation. The three sufficient conditions provided above for the notion of local concavity in condition MB4 have counterparts for the analysis of the game. That is, the analog to MB4 holds in the game context if one or more of the following conditions are satisfied: there are only two players; the game is of strategic substitutes.
– this restriction is often assumed in the entry games estimated in industrial organization; and
the potential function is discrete globally concave.

It is likely our results will lead to nonparametric identification in other classes of potential
games. For example, Monderer and Shapley (1996) and Qin (1996) discuss the connection
between noncooperative potential games and cooperative games. It follows from their results
that many cooperative solution concepts can be expressed as a potential game, and therefore
our previous results can likely be applied to these types of interactions.

6 Conclusion

We explore identification of discrete choice models for bundles and binary choice games of
complete information. We show that identification uses similar algebraic relationships between
outcome probabilities and unknown CDFs when these models include only two alternatives and
players, respectively. Moreover, there is an exact equivalence for identification between bundle
models and binary games of any number of goods and players when attention is restricted to
the class of potential games under an equilibrium selection rule based on potential maximizers.

We show how our models are identified. Specifically, we recover from data the standalone
utility function of each alternative or each player, the interaction effects among each bundle
or set of players, the joint distribution of potentially correlated, alternative- or player-specific
unobservables and, under additional assumptions, the distribution of heterogeneous interaction
effects.

A Proofs

A.1 Proof of Lemma 1

The proof of Lemma 1 involves three steps. The first two steps show that, for each \( W = w \),
the sign of \( \eta \cdot v(w) \) has different observable implications. (Appendix B reviews some of
the tools we use in this proof.) The last step uses the first two steps to show that the sign of the
interaction effect can be recovered from the data.

**Step 1** If \( \eta \cdot v(w) \geq 0 \), then \( U(a, w, z, \varepsilon, \eta) \) 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 \((a_2, z_1)\) — in the last case, the cross effect between
\( a_2 \) and \( z_1 \) is 0. By B3, the maximizer is unique with probability 1. Let us define the optimal
decision vector as

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

Then, by Topkis’s theorem, \( a^* (w, z, \varepsilon, \eta) \) increases (in the coordinatewise order) in \( z_1 \) with probability 1. By B2(i), the unobservables \( \varepsilon \) are independent of \( z_1 \). Therefore, for all \( z'_1 > z_1 \) and every upper set \( U \in \{0, 1\}^2 \), we have

\[ \Pr (a^* (w, z'_1, z_2, \varepsilon, \eta) \in U \mid w, z_1, z_2) \geq \Pr (a^* (w, z_1, z_2, \varepsilon, \eta) \in U \mid w, z_1, z_2). \]

That is, the random vector \( a^* (w, z, \varepsilon, \eta) \mid w, z_1, z_2 \) increases with respect to first order stochastic 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 \( \eta \cdot v (w) \leq 0 \), then \( U (a, w, z, \varepsilon, \eta) \) 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 \((-a_2, z_1)\) — in the last case, the cross effect between \(-a_2\) and \( z_1 \) is 0. By B3, the maximizer is unique with probability 1. Then, by Topkis’s theorem, \((a_1^* (w, z, \varepsilon, \eta), -a_2^* (w, z, \varepsilon, \eta))\) increases in \( z_1 \) with probability 1. By B2(i), the unobservables are independent of \( z_1 \). Therefore, for all \( z'_1 > z_1 \) and every upper set \( U \in \{0, 1\}^2 \), we have

\[ \Pr ((a_1^* (w, z'_1, z_2, \varepsilon, \eta), -a_2^* (w, z'_1, z_2, \varepsilon, \eta)) \in U \mid w, z_1, z_2) \geq \Pr ((a_1^* (w, z_1, z_2, \varepsilon, \eta), -a_2^* (w, z_1, z_2, \varepsilon, \eta)) \in U \mid w, z_1, z_2). \]

That is, the random vector \((a_1^* (w, z, \varepsilon, \eta), -a_2^* (w, z, \varepsilon, \eta)) \mid w, z_1, z_2 \) increases with respect to first order stochastic dominance in \( z_1 \). Because stochastic dominance is preserved under marginalization, in the data, \( \Pr (a_2 = 0 \mid w, z_1, z_2) \) increases in \( z_1 \) and thus \( \Pr (a_2 = 1 \mid w, z_1, z_2) \) decreases in \( z_1 \). Similarly, we can show that \( \Pr (a_1 = 1 \mid w, z_1, z_2) \) decreases in \( z_2 \).

**Step 3** By B3, 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. \( \square \)
A.2 Proof of Theorem 1

This proof relies on the sign of the interaction effect. By Lemma 1, this sign can be recovered from the data.

**Alternatives are Substitutes** \((\eta \cdot v(w) < 0)\)

Under B3, the probability of buying neither good is

\[
\Pr ((0, 0) \mid x) = \Pr (\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq \varepsilon_1 + u_1(w) + \varepsilon_2 + u_2(w) + \eta \cdot v(w) \leq -z_1 - z_2 \mid x).
\]

Because \(\eta \cdot v(w) \leq 0\), the third inequality above is implied by the first two. Thus,

\[
\Pr ((0, 0) \mid x) = \Pr (\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq -z_2 \mid x).
\]

Define the random vector

\[
\alpha = (\alpha_1, \alpha_2) \equiv (\varepsilon_1 + u_1(w), \varepsilon_2 + u_2(w)).
\]

By B2(1), the vector \((\varepsilon_1, \varepsilon_2, \eta)\) and hence \(\varepsilon\) are independent of \(Z\). Therefore, for an arbitrary point of evaluation \(\alpha^* = (\alpha_1^*, \alpha_2^*)\) of the CDF \(F_{\alpha | w}\),

\[
F_{\alpha | w} (\alpha^*) = \Pr (\alpha_1 \leq \alpha_1^*, \alpha_2 \leq \alpha_2^* \mid w) = \Pr (\alpha_1 \leq -z_1, \alpha_2 \leq -z_2 \mid x) = \Pr ((0, 0) \mid x),
\]

for excluded regressor choices \(z_1 = -\alpha_1^*\) and \(z_2 = -\alpha_2^*\). Therefore, the variation in \(Z\) from B1 and \(\Pr ((0, 0) \mid x)\) identifies the CDF of \(\alpha\), for each \(W = w\).

Under B3, the probability of buying both goods is

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

Because \(\eta \cdot v(w) \leq 0\), the third inequality above is implied by the first two. Thus,

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

Define the random vector

\[
\beta = (\beta_1, \beta_2) \equiv (-\varepsilon_1 - u_1(w) - \eta \cdot v(w), -\varepsilon_2 - u_2(w) - \eta \cdot v(w)).
\]
Given B2(i) and applying the same logic as before, we can use variation in $Z$ and the probability $Pr((1, 1) \mid x)$ to recover $F_{\beta \mid w}$. By B2(ii) and B2(III), $\varepsilon$ and $\eta$ have mean $(0, 0)$ and 1, respectively, conditional on $W = w$. Therefore,

$$
E[(\alpha_1, \alpha_2) \mid w] = (u_1(w), u_2(w))
$$

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

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

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

Once $((u_i)_{i \leq 2}, v)$ is identified, it is a simple, known shift of the location of a random vector to go from the identified distribution of $\alpha$ to the distribution of $(\varepsilon_1, \varepsilon_2)$. Therefore, we identify $F_{\varepsilon \mid W}$. 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 $\eta$ and $\varepsilon$ are independent conditional on $W$, the distribution of $F_{\eta \mid W=w}$ is identified from the distributions $F_{\varepsilon \mid W=w}$ and $F_{\beta=w}$ because the joint characteristic function of the sum of two independent random vectors, here the vector $\varepsilon$ and the scalar $\eta \cdot v(w)$, is equal to the products of the joint characteristic functions of $\eta \cdot v(w)$ and $\varepsilon$. Therefore, the characteristic function of $\eta \cdot v(w)$ is the ratio of the characteristic functions of $\beta$ and $\varepsilon$. Because there is a bijection between the spaces of distribution and characteristic functions, $F_{\eta \cdot v(w) \mid W=w}$ is identified. A known, multiplicative change of variables, given that $v(w)$ is known, identifies $F_{\eta \mid W=w}$.

**Alternatives are Complements ($\eta \cdot v(w) \geq 0$)**

Under B3, when $(1, 0)$ is selected, it is the only bundle that maximizes utility with positive probability. Thus,

$$
Pr((1, 0) \mid x) =
$$

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

Because $\eta \cdot v(w) \geq 0$, the third inequality above is implied by the first two. Thus,

$$
Pr((1, 0) \mid x) = Pr(-\varepsilon_1 - u_1(w) \leq z_1, \varepsilon_2 + u_2(w) + \eta \cdot v(w) \leq -z_2 \mid x).
$$

Define the random vector

$$
\alpha = (\alpha_1, \alpha_2) \equiv (-\varepsilon_1 - u_1(w), \varepsilon_2 + u_2(w) + \eta \cdot v(w)).
$$
By B2(i), \((\varepsilon_1, \varepsilon_2, \eta)\) is independent of \(Z\). Therefore, for an arbitrary point of evaluation \(\alpha^* = (\alpha_1^*, \alpha_2^*)\) of the CDF \(F_{\alpha|w}\),

\[
F_{\alpha|w}(\alpha^*) = \Pr(\alpha_1 \leq \alpha_1^*, \alpha_2 \leq \alpha_2^* \mid w) = \Pr(\alpha_1 \leq z_1, \alpha_2 \leq -z_2 \mid x) = \Pr((1, 0) \mid x)
\]

for choices \(z_1 = \alpha_1^*\) and \(z_2 = -\alpha_2^*\). Therefore, the variation in \(Z\) from B1 and \(\Pr((1, 0) \mid x)\) identifies the CDF of \(\alpha\), for each \(W = w\). Applying the same logic, we can use variation in \(Z\) and probability \(\Pr((0, 1) \mid x)\) to recover \(F_{\beta|w}\), where the random vector \(\beta\) is

\[
\beta = (\beta_1, \beta_2) = (\varepsilon_1 + u_1(w), \eta \cdot v(w), -\varepsilon_2 - u_2(w)).
\]

By B2(II) and B2(III), \(\varepsilon\) has \((0, 0)\) mean and \(\eta\) has a mean of 1 conditional on \(W = w\). Therefore,

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

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

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

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

Once \(((u_i)_{i \leq 2}, v)\) is identified and we impose the further restriction that \(\eta = 1\), it is a known shift of the location of a random vector to go from the identified distribution of \(\alpha\) to the distribution of \((-\varepsilon_1, \varepsilon_2)\). Then, a known multiplicative change of variables is used to learn the distribution of \((\varepsilon_1, \varepsilon_2)\) from the distribution of \((-\varepsilon_1, \varepsilon_2)\). Therefore, we identify \(F_{\varepsilon|w}\).

If we had included heterogeneous interaction effects \(\eta \neq 1\) in the last step, we could not have learned either \(F_{\varepsilon|w}\) or \(F_{\eta|w}\) from the distribution of \(F_{\alpha|w}\) alone. \(\square\)

**A.3 Proof of Theorem 2**

This proof relies on the sign of the interaction effect. By Lemma 1, this sign can be recovered from the data.

**Alternatives are Substitutes** \((\eta \cdot v(w) \leq 0)\)

Under B3', given the proof of Theorem 1, the probability of buying neither good is

\[
\Pr((0, 0) \mid x) = \Pr(\varepsilon_1 + u_1(w) \leq -z_1, \varepsilon_2 + u_2(w) \leq -z_2 \mid x).
\]

Thus, by B2'(III), we can obtain the bivariate density \(f\) corresponding to the CDF \(F_{\varepsilon_1+u_1(w), \varepsilon_2+u_2(w)|w}\).
(in the support of $Z$) as follows

\[ f_{\varepsilon_1 + u_1 (w), \varepsilon_2 + u_2 (w)} | W (z_1, z_2) = \frac{\partial^2 \Pr ((0,0) \mid x)}{\partial z_1 \partial z_2} = \frac{\partial^2 \Pr (\varepsilon_1 + u_1 (w) \leq -z_1, \varepsilon_2 + u_2 (w) \leq -z_2 \mid x)}{\partial z_1 \partial z_2}. \]

By assumption B2'(ii), this expression has mode $(u_1 (w), u_2 (w))$. Thus, given B1', it 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 recovered from the data.

Under B3' and B4', given the proof of Theorem 1, the probability of buying both goods is

\[ \Pr ((1,1) \mid x) = \Pr (\varepsilon_1 + u_1 (w) + v (w) \geq -z_1, \varepsilon_2 + u_2 (w) + v (w) \geq -z_2 \mid x). \]

Thus, by B2'(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 — recall that the density of a bivariate distribution function can be obtained as the cross partial derivative of the survival function —

\[ f_{\varepsilon_1 + u_1 (w) + v (w), \varepsilon_2 + u_2 (w) + v (w)} | W (z_1, z_2) = \frac{\partial^2 \Pr ((1,1) \mid x)}{\partial z_1 \partial z_2} = \frac{\partial^2 \Pr (\varepsilon_1 + u_1 (w) + v (w) \geq -z_1, \varepsilon_2 + u_2 (w) + v (w) \geq -z_2 \mid x)}{\partial z_1 \partial z_2}. \]

By assumption B2'(ii), this expression has mode $(u_1 (w) + v (w), u_2 (w) + v (w))$. Thus, given B1' it is maximized at $-z_1'$ and $-z_2'$ when $u_1 (w) + v (w) = -z_1'$ and $u_2 (w) + v (w) = -z_2'$. Thus, $v$ can be recovered from the data.

**Alternatives are Complements ($\eta \cdot v (w) \geq 0$)**

Under B3' and B4', from the proof of Theorem 1 we know that

\[ \Pr ((1,0) \mid x) = \Pr (\varepsilon_1 - u_1 (w) \leq z_1, \varepsilon_2 + u_2 (w) + v (w) \leq -z_2 \mid x). \]

Thus, by B2'(iii), we can obtain the density of $F_{\varepsilon_1 - u_1 (w), \varepsilon_2 + u_2 (w) + v (w)} | W$ (in the support of $Z$) as follows

\[ f_{\varepsilon_1 - u_1 (w), \varepsilon_2 + u_2 (w) + v (w)} | W (z_1, z_2) = -\frac{\partial^2 \Pr ((-1,0) \mid x)}{\partial z_1 \partial z_2} = -\frac{\partial^2 \Pr (\varepsilon_1 - u_1 (w) \leq z_1, \varepsilon_2 + u_2 (w) + v (w) \leq -z_2 \mid x)}{\partial z_1 \partial z_2}. \]

By assumptions B1' and B2'(ii), this expression has mode $(-u_1 (w), u_2 (w) + v (w))$. Thus, given B1' it 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 recovered from the data.

We can then use $\Pr ((0,1) \mid x)$ to recover $u_2$ and $u_1 + v$. Thus, $((u_i)_{i \leq 2}, v)$ is identified. □
A.4 Proof of Theorem 3

By G5, for each \( W = w \), the game is either supermodular or submodular.

**Submodular Game** \((\eta \cdot v_1 (w) < 0 \text{ and } \eta_2 \cdot v_2 (w) < 0)\)

We first consider the case where \( \Gamma (x, \varepsilon, \eta) \) is a submodular game, meaning that \( \eta_i \cdot v_i (w) \leq 0 \) for \( i = 1, 2 \). Under G3, when \((0, 0)\) is an equilibrium, it is a unique equilibrium (e.g., Tamer 2003). Using expression (6), the probability of both agents taking action 0 is

\[
\Pr ((0, 0) \mid x) = \Pr (\varepsilon_1 + (w) \leq -z_1, \varepsilon_2 + u_2 (w) \leq -z_2 \mid x).
\]

Define the random vector

\[
\alpha = (\alpha_1, \alpha_2) \equiv (\varepsilon_1 + u_1 (w), \varepsilon_2 + u_2 (w)).
\]

By G2(i), the vector \((\varepsilon_1, \varepsilon_2, \eta_1, \eta_2)\) and hence \( \varepsilon \) are independent of \( Z \). Therefore, for an arbitrary point of evaluation \( \alpha^* = (\alpha_1^*, \alpha_2^*) \) of the CDF \( F_{\alpha \mid w} \),

\[
F_{\alpha \mid w} (\alpha^*) = \Pr (\alpha_1 \leq \alpha_1^*, \alpha_2 \leq \alpha_2^* \mid w) = \Pr (\alpha_1 \leq -z_1, \alpha_2 \leq -z_2 \mid x) = \Pr ((0, 0) \mid x),
\]

for excluded regressor choices \( z_1 = -\alpha_1^* \) and \( z_2 = -\alpha_2^* \). Therefore, the variation in \( Z \) from G1 and \( \Pr ((0, 0) \mid x) \) identifies the CDF of \( \alpha \), for each \( w \). Applying the same logic and condition G2(i), we can use variation in \( Z \) and the probability \( \Pr ((1, 1) \mid x) \) of the other unique equilibrium, where both agents take action 1, to recover \( F_{\beta \mid w} \), where the random vector \( \beta \) is

\[
\beta = (\beta_1, \beta_2) \equiv (-\varepsilon_1 - u_1 (w) - \eta_1 \cdot v_1 (w), -\varepsilon_2 - u_2 (w) - \eta_2 \cdot v_2 (w)).
\]

By G2(ii) and G2(iii), \( \varepsilon \) and \( \eta \) have mean \((0, 0)\) and \((1, 1)\), respectively, conditional on \( W \). Therefore,

\[
E [(\alpha_1, \alpha_2) \mid w] = (u_1 (w), u_2 (w))
\]

\[
E [(\beta_1, \beta_2) \mid w] = (-u_1 (w) - v_1 (w), -u_2 (w) - v_2 (w))
\]

\[
E [(\alpha_1, \alpha_2) \mid w] + E [(\beta_1, \beta_2) \mid w] = (-v_1 (w), -v_2 (w)).
\]

One can see that \((u_i, v_i)_{i \leq 2}\) are identified for this \( W = w \).

The rest of the proof is similar to the one of Theorem 1 for the case of substitutes, thus we
Supermodular Game \((\eta_1 \cdot v_1(w) \geq 0 \text{ and } \eta_2 \cdot v_2(w) \geq 0)\)

We next consider the case where \(\Gamma(x, \varepsilon, \eta)\) is a supermodular game, i.e., \(\eta_i \cdot v_i(w) \geq 0\) for \(i = 1, 2\). Recall that \(\eta = (1, 1)\) by G2(III). Under G4, when \((1, 0)\) is an equilibrium, it is a unique equilibrium (e.g., Tamer 2003). Using expression (6) and G3, the probability of the first agent taking action 1 and the second agent taking action 0 is

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

Define the random vector

\[
\alpha = (\alpha_1, \alpha_2) \equiv (-\varepsilon_1 - u_1(w), \varepsilon_2 + u_2(w) + \eta_2 \cdot v_2(w)).
\]

By G2(1), \(\varepsilon\) is independent of \(Z\). Therefore, for an arbitrary point of evaluation \(\alpha^* = (\alpha_1^*, \alpha_2^*)\) of the CDF \(F_{\alpha \mid w}\),

\[
F_{\alpha \mid w}(\alpha^*) = \Pr(\alpha_1 \leq \alpha_1^*, \alpha_2 \leq \alpha_2^* \mid w) = \Pr(\alpha_1 \leq z_1, \alpha_2 \leq -z_2 \mid x) = \Pr((1, 0) \mid x)
\]

for choices \(z_1 = \alpha_1^*\) and \(z_2 = -\alpha_2^*\). Therefore, the variation in \(Z\) from G1 and \(\Pr((1, 0) \mid x)\) identifies the CDF of \(\alpha\), for each \(W = w\). Applying the same logic, we can use variation in \(Z\) and the probability \(\Pr((0, 1) \mid x)\) of the other unique equilibrium to recover \(F_{\beta \mid w}\), where the random vector \(\beta\) is

\[
\beta = (\beta_1, \beta_2) \equiv (\varepsilon_1 + u_1(w) + \eta_1 \cdot v_1(w), -\varepsilon_2 - u_2(w)).
\]

By G2(II), \(\varepsilon\) has zero mean conditional on \(W\). Therefore,

\[
\begin{align*}
E[\alpha_1, \alpha_2 \mid w] &= (-u_1(w), u_2(w) + v_2(w)) \\
E[\beta_1, \beta_2 \mid w] &= (u_1(w) + v_1(w), -u_2(w)) \\
E[\alpha_1, \alpha_2 \mid w] + E[\beta_1, \beta_2 \mid w] &= (v_1(w), v_2(w)).
\end{align*}
\]

One can see that \((u_i, v_i)_{i \leq 2}\) are identified for this \(W = w\).

The rest of the proof is similar to the one of Theorem 1 for the case of complements, thus we omit it. \(\square\)
A.5 Proof of Theorem 4

The proof of this result is similar to the one of Theorem 2, thus we omit it. □

A.6 Proof of Proposition 1

The proof of this result follows directly from Ui (2000, Theorem 3). □

A.7 Proof of Proposition 2

Under the assumed conditions, the result follows directly from Proposition 1.

A.8 Proof of Theorem 5

The proof of this result is divided into two steps. We first show that if MB1–4 are satisfied, then \( F_{\varepsilon|W} \) is identified. We then show that if MB1–3 are satisfied and \( F_{\varepsilon|W} \) is identified, then the standalone utility functions and interaction terms \((u_i)_{i \leq n}, v)\) are also identified.

A.8.1 Identification of \( F_{\varepsilon|W} \)

If MB1–4 hold, we show that \( F_{\varepsilon|W} \) is identified. By MB4, for each \( W = w \), there exists a known vector \( \hat{a}^w \in \{0, 1\}^n \) such that, for any \( x = (w, \varepsilon) \) and \( \varepsilon, U(a, x, \varepsilon) \) is maximized at \( \hat{a}^w \) if \( U(\hat{a}_i^w, \hat{a}_{-i}^w, x, \varepsilon) \geq U(\hat{a}_i', \hat{a}_{-i}^w, x, \varepsilon) \) for all \( i \in \mathcal{N} \) and \( a_i' = 1 - \hat{a}_i^w \). This condition holds if, for all \( i \in \mathcal{N} \),

\[
\begin{align*}
(1 (\hat{a}_i^w = 1) - 1 (\hat{a}_i^w = 0)) \varepsilon_i \geq & \\
v(S(a'), w) - v(S(\hat{a}^w), w) - (1 (\hat{a}_i^w = 1) - 1 (\hat{a}_i^w = 0)) (u_i(w) + z_i),
\end{align*}
\]

(11)

where \( a' \) is obtained from \( \hat{a}^w \) by changing only \( \hat{a}_i^w \).

We next show that we can recover \( F_{\varepsilon|W} \) from variation in \( z \) using \( \Pr(\hat{a} \mid x) = \Pr(\hat{a} \mid w, z) \). From MB4 we get that

\[
\Pr(\hat{a} \mid w, z) = \Pr((11) \text{ holds for all } i \in \mathcal{N} \mid w, z).
\]
Define the random variable $\mu_i$ for each $i \in \mathcal{N}$ to be

$$
\mu_i = (1 (\hat{a}_i^w = 1) - 1 (\hat{a}_i^w = 0)) \varepsilon_i \\
- (v(S(a'), W) - v(S(\hat{a}_i^w), W)) + (1 (\hat{a}_i^w = 1) - 1 (\hat{a}_i^w = 0)) (u_i (W)).
$$

Let $\mu \equiv (\mu_1, \ldots, \mu_n)$, which is independent of $Z$ conditional on $W$. Therefore,

$$
\Pr (\hat{a}_i^w \mid w, z) = \Pr (\mu_i \geq - (1 (\hat{a}_i^w = 1) - 1 (\hat{a}_i^w = 0)) z_i \text{ for all } i \in \mathcal{N} \mid w, z).
$$

We identify the upper probabilities of the vector $\mu$, conditional on $w$, at all points

$$
\tilde{z} = (- (1 (\hat{a}_1^w = 1) - 1 (\hat{a}_1^w = 0)) z_1, \ldots, - (1 (\hat{a}_n^w = 1) - 1 (\hat{a}_n^w = 0)) z_n).
$$

By the large support in MB1 and the fact that $\tilde{z}$ is at most a sign change from $z$, the random vector $\tilde{Z}$, defined in the obvious way, has support on all of $\mathbb{R}^n$. Therefore, we learn the upper tail probabilities of $\mu$ conditional on $W = w$ for all points of evaluation $\mu^*$. Upper tail probabilities completely determine a random vector’s distribution, so we also identify the lower tail probabilities of $\mu$ conditional on $w$, also known as the joint CDF of $\mu$ conditional on $w$. Note that $\varepsilon_i$ is the only random variable in $\mu_i$, conditional on $W = w$. By MB2(ii), $E(\varepsilon \mid W) = 0$. Therefore, up to the possible sign change in $(1 (\hat{a}_i^w = 1) - 1 (\hat{a}_i^w = 0))$, the distribution of $\varepsilon$ conditional on $w$ is obtained from the distribution of $\mu - E(\mu \mid w)$ conditional on $w$.

**A.8.2 Identification of $((u_i)_{i \leq n}, v)$**

The remaining argument conditions on $W = w'$. Recall the utility, equation (9). The non-$\varepsilon$ portion of utility plays a key role in the identification argument. Therefore, let

$$
Q(a, X) \equiv \sum_{i \in \mathcal{N}} (u_i (W) + Z_i) 1 (a_i = 1) + v(S(a), W).
$$

Our location normalization is that $Q(a, X) = 0$ for $a = (0, 0, \ldots, 0)$.

For expositional ease, we order the elements of $\{0, 1\}^n$ in terms of the lexicographic order so that $a^1 = (0, 0, \ldots, 0)$, $a^2 = (1, 0, \ldots, 0)$, ..., and $a^{2^n} = (1, 1, \ldots, 1)$. In addition, we define

$$
\Delta^j \varepsilon (a') \equiv \sum_{i \in \mathcal{N}} \varepsilon_i 1 (a_i^j = 1) - \sum_{i \in \mathcal{N}} \varepsilon_i 1 (a'_i = 1) \text{ and } \Delta^j Q(a', X) \equiv Q(a', X) - Q(a^j, X).
$$
We indicate by $\triangle Q (a', X)$ and $\triangle \varepsilon (a')$ the $(2^n - 1)$-dimensional vectors

$$(Q (a', X) - Q (a^j, X))_{j \leq 2^n, a^j \neq a'} \quad \text{and} \quad \left( \sum_{i \in \mathcal{N}} \varepsilon_i 1 (a^j_i = 1) - \sum_{i \in \mathcal{N}} \varepsilon_i 1 (a'_i = 1) \right)_{j \leq 2^n, a^j \neq a'}.$$  

Given this notation, for each $a'$,

$$P (\triangle Q (a', x) \mid x; F_{\varepsilon|x}) \equiv \Pr (\triangle \varepsilon (a') \leq \triangle Q (a', x) \mid x)$$  

captures the probability of observing the choice vector $a'$ conditional on $X = x$, or in simpler notation, $\Pr (a' \mid x)$. More formally, let $F_{\triangle \varepsilon(a')|x'}$ be the distribution of $\triangle \varepsilon (a')$. Then,

$$P (\triangle Q (a', x) \mid x; F_{\varepsilon|x}) \equiv \int \ldots \int 1 (\triangle_1 \varepsilon (a') \leq \triangle Q (a', x)) \ldots 1 (\triangle_2 \varepsilon (a') \leq \triangle Q (a', x)) dF_{\triangle \varepsilon(a')|x}.$$

The researcher can identify $\Pr (a' \mid x)$ directly from the data.

Let $\tilde{Q} (a, X) \neq Q (a, X)$. As $Z$ enters both $\tilde{Q}$ and $Q$ in the same way, this means that one or more of $(u_i (w))_{i \in \mathcal{N}}$ and $v (S (a), w)$ differ across $\tilde{Q}$ and $Q$ for $W = w$. We next show that, without loss of generality, $P (\triangle Q (a', x) \mid x; F_{\varepsilon|x}) > P (\triangle \tilde{Q} (a', x) \mid x; F_{\varepsilon|x})$, so that $Q$ is identified and we can then recover $((u_i)_{i \leq n}, v)$.

Let

$$C (x) \equiv \arg \max_a \left\{ (Q (a, x) - \tilde{Q} (a, x)) \mid a \in \{0, 1\}^n \right\}.$$  

Note that by the formulas for $\left( Q (a, x) - \tilde{Q} (a, x) \right)$, $C (x)$ is a constant set, as only $z$ varies once we condition on $w$. Also suppose $\max_a (Q (a, x) - \tilde{Q} (a, x)) > 0$; the other case follows by a similar argument. Define $D (x) \equiv \{ a \notin C (x) \mid a \in \{0, 1\}^n \}$. We know $C (x) \neq \emptyset$. The fact that $D (x) \neq \emptyset$ follows as

$$Q (a = (0, 0, ..., 0), x) = \tilde{Q} (a = (0, 0, ..., 0), x) = 0$$

and we supposed that $\max_a (Q (a, x) - \tilde{Q} (a, x)) > 0$.

Fix some $a' \in C (x)$. We know that $Q (a', x) - \tilde{Q} (a', x) = Q (a, x) - \tilde{Q} (a, x)$ for all $a \in C (x)$, and $Q (a', x) - \tilde{Q} (a', x) > Q (a, x) - \tilde{Q} (a, x)$ for all $a \in D (x)$. Rearranging terms,

$$Q (a', x) - Q (a, x) = \tilde{Q} (a', x) - \tilde{Q} (a, x) \quad \text{for all} \ a \in C (x) , \text{and}$$

$$Q (a', x) - Q (a, x) > \tilde{Q} (a', x) - \tilde{Q} (a, x) \quad \text{for all} \ a \in D (x).$$

Then, by MB2 and MB3, the argument in the following paragraphs ensures that we can find $z$
such that, for \( x = (w, z) \),

\[
\int \ldots \int 1 \left( \triangle^1 \varepsilon (a') \leq \triangle^1 Q (a', x) \right) \ldots 1 \left( \triangle^{2n} \varepsilon (a') \leq \triangle^{2n} Q (a', x) \right) dF_{\triangle \varepsilon (a') | w} > \int \ldots \int 1 \left( \triangle^1 \varepsilon (a') \leq \triangle^1 \tilde{Q} (a', x) \right) \ldots 1 \left( \triangle^{2n} \varepsilon (a') \leq \triangle^{2n} \tilde{Q} (a', x) \right) dF_{\triangle \varepsilon (a') | w},
\]

i.e., \( P \left( \Delta Q (a', x) \mid x; F_{\varepsilon | w} \right) > P \left( \Delta \tilde{Q} (a', x) \mid x; F_{\varepsilon | w} \right) \). Therefore, \( Q \) is non-constructively identified at \( w \), and hence \(( (u_i (w))_{i \leq n}, v (w)) \) is identified as well.

As mentioned previously, we need to find an appropriate value for \( z \). This choice of \( z \) involves an additional detail that we address next. The inequalities that allow us to show that \( P \left( \Delta W (a', x) \mid x; F_{\varepsilon | w} \right) > P \left( \Delta \tilde{W} (a', x) \mid x; F_{\varepsilon | w} \right) \) are the ones that involve \( a \in D (x) \). Some of the inequalities involving \( a \in D (x) \) may be implied by other inequalities for some choices of \( z \). To see this, suppose there are two, substitute goods and let \( a' = (0, 0) \), \( C (x) = \{(0, 0), (1, 0), (0, 1)\} \) and \( D (x) = \{(1, 1)\} \). Under substitutes, \( Q ((0, 0), x) \geq Q ((1, 0), x) \) and \( Q ((0, 0), x) \geq Q ((1, 0), x) \) together imply \( Q ((0, 0), x) > Q ((1, 1), x) \). In this example, the fact that two inequalities imply a third means that marginal changes in the interaction term, \( v \), will not affect the probability of the outcome \((0, 0) \). Therefore, \( a' = (0, 0) \) does not allow us to effectively distinguish \( Q \) from \( \tilde{Q} \) as we just claimed. Notice that we assumed \((0, 0) \in C (x) \), which is not possible as we are covering the case \( \max_a (Q (a, x) - \tilde{Q} (a, x)) > 0 \) and (12) holds.

The next argument extends this idea.

We now show that there always exists some \( a'' \in D (x) \) for which the inequality \( Q (a', x) \geq Q (a'', x) \) is not directly implied from \( Q (a', x) \geq Q (a, x) \) for all \( a \neq a', a'' \). By contradiction, as we illustrated above, we will show that if this were not true, then \((0, 0, \ldots, 0) \in C (x) \) which is not possible as \( \max_a (Q (a, x) - \tilde{Q} (a, x)) > 0 \) and (12) holds.

For each \( a' \in C (x) \), there are (at least) \( n \) inequalities that are not implied by the others. These inequalities correspond to vectors of actions that differ from \( a' \) regarding one single good. To see this, let \( a'' \) be equal to \( a' \) except for some good \( i \) that is chosen at \( a' \) but not at \( a'' \). Then \( Q (a', x) \geq Q (a'', x) \) if and only if

\[
u_i (w) + z_i + v (S (a'), w) \geq v (S (a''), w).
\]

All other inequalities, \( W (a', x) \geq W (a, x) \) with \( a \neq a', a'' \), will involve at least one other excluded regressor \( z_j \) with \((j \neq i) \). Thus, we can always find a vector \((z_i)_{i \leq n} \) such that \( Q (a', x) \geq Q (a, x) \) with \( a \neq a', a'' \) and yet \( Q (a', x) < Q (a'', x) \). Therefore assume \( a'' \in C (x) \). By repeating this process \( \|a'\| \) times (the number of goods being consumed in the bundle represented by \( a' \)),
we need to assume $(0,0,...,0) \in C(x)$. But this is not possible, as we explained before.

**A.9 Proof of Lemma 2**

Assume $V(a,x,\varepsilon)$ has the negative single-crossing property on $(a_i;a_{-i})$ for all $i \in \mathcal{N}$. That is, for all $a'_i > a_i$ and all $a'_{-i} > a_{-i}$ we have

$$V(a'_i, a_{-i}, x, \varepsilon) - V(a_i, a_{-i}, x, \varepsilon) \leq (\prec) 0 \implies V(a'_i, a'_{-i}, x, \varepsilon) - V(a_i, a'_{-i}, x, \varepsilon) \leq (\prec) 0.$$  (13)

We next show that Assumption MB4 holds with $\tilde{a}^w = (0,0,...,0)$ for all $w$.

Assume $V(\tilde{a}_i^w = 0, \tilde{a}_{-i}^w = (0,0,...,0), x, \varepsilon) \geq V(a_i = 1, \tilde{a}_{-i}^w = (0,0,...,0), x, \varepsilon)$ for all $i \in \mathcal{N}$. Then, by (13), for all $a_{-i} \in \{0,1\}^{n-1}$, and all $i \in \mathcal{N}$,

$$V(\tilde{a}_i^w = 0, a_{-i}, x, \varepsilon) \geq V(a_i = 1, a_{-i}, x, \varepsilon)$$

Thus, $V((0,0,...,0), x, \varepsilon) \geq V(a, x, \varepsilon)$ for all $a \in \{0,1\}^n$.

**A.10 Proof of Lemma 3**

Assume the required conditions of the lemma hold. Then, by Ui (2008, Proposition 1),

$$V(\tilde{a}^w, X, \varepsilon) \geq V(a, X, \varepsilon)$$

for all $a \in \{0,1\}^n$ with $\|\tilde{a}^w - a\| \leq 1$ is satisfied if and only if $V(\tilde{a}^w, X, \varepsilon) \geq V(a, X, \varepsilon)$ for all $a \in \{0,1\}^n$. Thus, Assumption MB4 holds at any $\tilde{a}^w \in \{0,1\}^n$.

**A.11 Proof of Proposition 3**

The proof of this result follows directly from Ui (2000, Theorem 3).

**B Monotone Comparative Statics and Stochastic Dominance**

The proof of Lemma 1 relies on Topkis’ theorem (Topkis (1998)) and the concept of stochastic dominance. We outline a simple version of these concepts.

**Proposition 5. Topkis’ theorem:** Let $f(a_1,a_2,x) : \mathcal{A}_1 \times \mathcal{A}_2 \times \mathbb{R} \to \mathbb{R}$, where $\mathcal{A}_1$ and $\mathcal{A}_2$ are finite ordered sets. Assume that $f(a_1,a_2,x)$ (i) is supermodular in $(a_1,a_2)$; and (ii) has increasing differences in $(a_1,x)$ and $(a_2,x)$. Then, arg max $\{f(a_1,a_2,x) | (a_1,a_2) \in \mathcal{A}_1 \times \mathcal{A}_2\}$
increases in $x$ with respect to the strong set order.\footnote{For any two elements $a, a' \in A_1 \times A_2$ we write $a \vee a'$ ($a \wedge a'$) for the least upper bound (greatest lower bound). We say $f(a_1, a_2, x)$ is supermodular in $(a_1, a_2)$ if, for all $a, a' \in A_1 \times A_2$,
\[
  f(a \vee a', x) + f(a \wedge a', x) \geq f(a, x) + f(a', x).
\]
We say $f(a_1, a_2, x)$ has increasing differences in $(a_1, x)$ if, for all $a_1' > a_1$ and $x' > x$,
\[
  f(a_1', a_2, x') - f(a_1, a_2, x') \geq f(a_1', a_2, x) - f(a_1, a_2, x).
\]}

\begin{footnotesize}
(According to this order, we write $A \succeq B$ if for every $a \in A$ and $b \in B$, we have that $a \vee b \in A$ and $a \wedge b \in B$.)
\end{footnotesize}

The concept of first order (or standard) stochastic dominance (FOSD), is based on upper sets. Let us consider $(\Omega, \succeq)$, where $\Omega$ is a set and $\succeq$ defines a partial order on it. A subset $U \subset \Omega$ is an upper set if and only if $x \in U$ and $x' \succeq x$ imply $x' \in U$.

**First Order Stochastic Dominance:** Let $X', X \in \mathbb{R}^n$ be two random vectors. We say $X'$ is higher than $X$ with respect to first order stochastic dominance if

\[
  \Pr(X' \in U) \geq \Pr(X \in U)
\]

for all upper set $U \subset \mathbb{R}^n$.

**References**


