Title
Essays on Identification

Author
Allen, Roy Henry

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays on Identification

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Economics

by

Roy Henry Allen

Committee in charge:

Professor Ivana Komunjer, Co-Chair
Professor Andres Santos, Co-Chair
Professor Ery Arias-Castro
Professor Jelena Bradic
Professor Yixiao Sun

2017

The dissertation of Roy Henry Allen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Co-Chair

University of California, San Diego

2017

DEDICATION

To my parents.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my dissertation committee for their support. I would especially like to thank my co-chairs Andres and Ivana for their guidance and patience.

Chapter 1 is being prepared for submission for publication. Allen, Roy; Rehbeck, John. The dissertation author was a primary author and investigator of this paper.

Chapter 2 is being prepared for submission for publication. Allen, Roy; Rehbeck, John. The dissertation author was a primary author and investigator of this paper.

Chapter 3 is being prepared for submission for publication. Allen, Roy. The dissertation author was the sole author.

VITA

2010	B. A. in Economics, *high honors*, Swarthmore College
2017	Ph. D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Essays on Identification

by

Roy Henry Allen

Doctor of Philosophy in Economics

University of California, San Diego, 2017

Professor Ivana Komunjer, Co-Chair
Professor Andres Santos, Co-Chair

The first two chapters of this dissertation study identification of a new class of demand models termed *perturbed utility models*. The first chapter provides sufficient conditions under which structural functions in these models can be uniquely determined from knowledge of conditional means. The second chapter proposes a definition of complementarity/substitutability for these models and shows how to recover this measure

from data.

The third chapter of this dissertation studies inference in a class of partially identified models. Specifically, this chapter provides a finite-sample power comparison between two existing tests of moment inequalities.

Chapter 1

Identification of Average Demand Models

Abstract: This paper studies the nonparametric identification of a model of average demand for multiple goods, once unobservable heterogeneity has been integrated out. The model can be used for bundles, decisions under uncertainty, stochastic choice, and other examples. Optimizing behavior implies an analogue of Slutsky symmetry, which we exploit to show nonparametric identification of the model. Our main results do not rely on special regressors or identification at infinity. As a special case we provide new conditions for identification of additive random utility models (ARUM). These conditions also apply to a stochastic choice model allowing bounded rationality. In an illustrative application, we refute ARUM in favor of this more general model.

1.1 Introduction

Individuals often choose quantities of multiple goods at once. The most familiar example is the choice of quantity bundles in a standard consumer problem. In decisions under uncertainty, an individual chooses a lottery (= "quantities") over outcomes (= "goods"). While there are many tractable econometric approaches to modeling choice of a *single* good, the literature on demand of multiple goods is less well-developed. This paper studies identification of an "average demand" framework for multiple goods.

This model can be used to understand how individuals value different characteristics of goods, make counterfactual statements, and quantify welfare changes. This paper shows that even when there are multiple goods, conditional means can answer many questions without needing a full specification of the underlying behavioral model. For example, in a model of decisions under uncertainty, fairly general forms of heterogeneity or departures from expected utility are allowed, and we can still identify risk aversion.

This paper shows that optimization implies certain *equalities* that provide simple identification results for a large class of models. The equalities we exploit are analogues of Slutsky symmetry in the standard consumer problem. Such equalities arise in a class of models where the choice vector is the derivative of the indirect utility function. This is a common consequence of the envelope theorem, and a number of existing latent utility models fit into the class we study, including additive random utility models (McFadden (1973)).

We assume each good has an unknown index function controlling its desirability. Each index function depends on the characteristics of that good. Formally, assume the demand vector Y satisfies

$$Y = \tilde{M}(v_1(X_1), \dots, v_K(X_K), \varepsilon), \tag{1.1}$$

where X_k includes characteristics of good k as well as demographic variables, v_k controls the desirability of good k, and ε represents unobservable heterogeneity. The structure we impose is that for fixed ε, $\tilde{M}(\cdot, \varepsilon)$ is the derivative of a convex function.

This paper studies what we can learn about features of preferences given knowledge of $\mathbb{E}[Y \mid X = x]$. The main contribution of this paper is that $\vec{v} = (v_1, \dots, v_K)'$ is identified under mild conditions. Knowledge of $\vec{v}$ allows us to understand how individuals value the characteristics of goods. For example, the curvature of v_k may describe an individual's risk aversion. While it is fairly straightforward to identify the ordinal ranking of v_k, this paper shows that $\vec{v}$ is identified up to an affine transformations. Once $\vec{v}$ is identified, we show welfare differences are identified using only conditional means. This is possible because we interpret $\tilde{M}(\cdot, \varepsilon)$ as the derivative of the individual-specific indirect utility function. Welfare is then calculated as the average indirect utility.

We are motivated to study (1.1), rather than a more general model, for several reasons. First, this structure is implied by many latent utility models. Examples include discrete choice additive random utility models (McFadden (1973), McFadden (1978)), the bundles model of Gentzkow (2007) and Fox and Lazzati (forthcoming), and an expected utility model considered in Agarwal and Somaini (2014). A byproduct of our analysis is to provide weaker conditions for identification of these models. Second, conditional means are sufficient to identify differences in average welfare for this model. It is unclear when this is possible for more general models. Third, since our results require only knowledge of conditional means (and not full conditional distributions), we can handle less than ideal data. For example, in decisions under uncertainty we may not observe choices of lottery, but instead only the outcomes of the lotteries.[1] Finally, this is a *cardinal* model in the sense that the structural functions $\vec{v}$ are identified up to affine transformations. Relaxing (1.1) to a more general index model may lose this feature. Models that are not cardinal can be hard to interpret or take to data because they require a careful handling of normalizations.

We identify $\vec{v}$ without specifying $\tilde{M}$ or the distribution of ε. Assuming independence between observable characteristics and unobservable heterogeneity, we show $\vec{v}$ is identified if: (i) each good has a continuous regressor specific to it that affects its desirability, (ii) there is sufficient complementarity and/or substitutability among goods, and (iii) some smoothness and support conditions are satisfied. Identification is up to a location and multiplicative scale normalization for $\vec{v}$, which in general cannot be weakened.

Specializing the analysis, we obtain a new set of conditions for identification of the additive random utility model (ARUM). This model includes logit, nested logit, and probit as special cases. We establish nonparametric identification of $\vec{v}$ without specifying a distribution of the latent variables. In contrast with the nonparametric identification results of Matzkin (1993), we do not require *a priori* knowledge of how a regressor enters v_k [2] and do not need to reduce the problem to a binary choice identification problem.[3] Instead, we impose smoothness conditions to obtain constructive identification results. Smoothness conditions are not imposed by Matzkin (1993), so our results are not strictly more general.

In ARUM, we identify $\vec{v}$ without using the full structure of the model. The

[1] From such data one can estimate the conditional probability of each outcome, which in our setup is the "average demand."

[2] Such as a special regressor structure.

[3] The latter technique is sometimes termed identification at infinity.

results apply to a strictly more general model of stochastic choice studied in Allen and Rehbeck (2016b). A motivation for this model is that a common violation of random utility models is that adding an alternative to a menu can *increase* the probability of choosing an existing alternative (e.g. Huber, Payne, and Puto (1982)). The model of Allen and Rehbeck (2016b) accommodates this behavior and also allows the closely related possibility that a good could become more attractive and an *existing* good could be chosen with higher probability. Such behavior is ruled out by ARUM. As an illustrative application, we use the "no complementarity" implication to provide a parametric test of ARUM against the more general model. Using data from Louviere et al. (2013) on preference for types of pizza, we refute ARUM because there is not enough stochastic substitution among alternatives.

Our identification results use an analogue of Slutsky symmetry, and the core argument is simple. First, the latent utility model of (1.1) aggregates. Specifically, when latent variables are independent of characteristics, (1.1) implies

$$\mathbb{E}[Y \mid X = x] = M(v_1(x_1), \ldots, v_K(x_K)), \tag{1.2}$$

where M is the gradient of a convex function. Symmetry states that the effect of an increase in the index v_k on the mean of Y_ℓ is *equal* to that of an increase of v_ℓ on the mean of Y_k:

$$\frac{\partial M_k(\overline{v})}{\partial \overline{v}_\ell} = \frac{\partial M_\ell(\overline{v})}{\partial \overline{v}_k}.$$

If for example each x_k is scalar and only enters the index for good k, then using symmetry (and the chain rule) we show that

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_k} = \frac{\partial v_\ell(x_\ell)}{\partial x_\ell} \Bigg/ \frac{\partial v_k(x_k)}{\partial x_k}.$$

This provides constructive identification of partial derivatives of $\vec{v}$ at certain points x_ℓ, x_k. By repeated use of this equality, the fundamental theorem of calculus, and a location and scale normalization, we obtain constructive identification of $\vec{v}$.

This paper proceeds as follows. Section 1.2 contains a literature review. Section 1.3 formalizes our setup, provides examples and further literature review, and illustrates restrictions of the model under smoothness conditions. Section 1.4 establishes conditions for nonparametric identification of $\vec{v}$. Section 1.5 constructively identifies changes in D and average welfare. Section 1.6 characterizes out-of-sample bounds using the model. Section 1.7 contains an illustrative application. Section 1.8 concludes.

1.2 Literature Review

The literature on identification is large, and so we discuss only some of the most closely related papers. See Matzkin (2007), Matzkin (2013), and Berry and Haile (2015a) for recent surveys.

Identification of general simultaneous equations models with multiple endogenous variables has been studied in Matzkin (2008), Matzkin (2015), and Berry and Haile (2015b). The present paper differs from this line of work in focusing on specific features of preferences. Indeed, our identification results do *not* cover $\tilde{M}$ or the distribution of ε. In general, we expect the distribution of ε is not identified because we do not restriction its dimension. We instead identify a feature of its distribution in the form of average welfare changes. A second difference is we work with conditional means and allow for discrete Y, whereas the referenced line of work uses conditional densities and thus requires continuous Y.

The techniques of this paper use the property that the choice vector is the derivative of a convex function. This is a shape restriction coming from optimization. The identifying power of shape restrictions, such as monotonicity, convexity, and homotheticity, has been demonstrated in Matzkin (1992), Matzkin (1993), and Matzkin (1994). Using tools from convex analysis and related fields, a wave of recent papers has leveraged the hypothesis of optimization to establish identification. Examples include Galichon and Salanié (2015) and Chernozhukov et al. (2014).

While analogues of symmetry are known in a number of settings, to our knowledge, exploiting Slutsky symmetry for identification of index functions constitutes a new technique. We discuss its relationship to the *special regressor* (Lewbel (1998)) approach, which has been a powerful and influential tool in the identification literature. This approach assumes additional separability in $\vec{v}$, such as that Z_k enters v_k in a known way $v_k(X_k) = Z_k + \tilde{v}_k(W_k)$, where $X_k = (Z_k, W_k)$. We show in examples that when latent variables are independent of characteristics, this structure is unnecessary.[4] The functions $\vec{v}$ are identified not because Z_k enters v_k in a known way, but because Z_k is a relevant, good-specific regressor. When latent variables are independent conditional on certain characteristics (Lewbel (1998), Lewbel (2000)), we show in several examples that the partial linearity of the special regressor approach can be relaxed to additive separability, $v_k(X_k) = g_k(Z_k) + \tilde{v}_k(W_k)$, where $X_k = (Z_k, W_k)$. This means that Z_k does not need to enter v_k monotonically. Our results can be used to identify g_k. By defining $\tilde{Z}_k = g_k(Z_k)$, the insights of the special regressor approach can then be readily applied to the constructed special regressor $\tilde{Z}_k$. We thus place some identification results that use special regressors on firmer foundations.[5]

[4]Use of the term "special regressor" has commonly been applied any setup such as $v_k(X_k) = Z_k + \tilde{v}_k(W_k)$, though Lewbel (1998) is motivated by *failures* of independence between observable characteristics and latent variables.

[5]We discuss this in greater detail in examples. Recall we require multiple goods for our nonparametric results, so we do not contribute to the special regressor technique when there is only a single good.

A structure similar to (1.1) or (1.2) has been widely used. However, previous work typically begins with a model of utility maximization that implies one of these representations. We mention a few of the most closely related papers. For ARUM, the celebrated Williams-Daly-Zachary theorem (McFadden (1981)) states that conditional choice probabilities are the derivative of a convex function. In a panel setting, Shi, Shum, and Song (2016) use this feature to identify $\vec{v}$ when it is assumed to be linear. McFadden and Fosgerau (2012a) study a representation related to (1.1) focusing on budget variation. Fudenberg, Iijima, and Strzalecki (2015) study a stochastic choice model that implies (1.1). Chiong, Galichon, and Shum (2016) study a structure similar to (1.1) in dynamic discrete choice. Roughly, their results can be used to show that if M is known then $\vec{v}$ is identified. Fosgerau and Palma (2015) use a special case of (1.1) to model and estimate demand for differentiated products

This paper is part of a broader study of (1.2) and slightly more general *perturbed utility models* (PUM). In Allen and Rehbeck (2016a), we show that PUM provides a setting in which to define complementarity when prices are not available. We show in the consumer problem that if prices *were* available, our definition of complementarity would agree with the Hicksian definition stated in terms of cross-price elasticties.[6] We also show an aggregation property of a class of latent utility models. We use this result in the present paper to show that several existing latent utility models have conditional means consistent with PUM. In Allen and Rehbeck (2016b) we study the model specialized to stochastic choice. We show it is testable, but do not study identification.

[6]A related result appears in Gentzkow (2007), but assumes there are no income effects.

1.3 Model and Examples

We study models in which conditional means satisfy

$$\mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k) + D(y). \tag{1.3}$$

We call this the *perturbed utility model* (PUM). This is a generalization of the structure discussed in the Introduction. The vector $\mathbb{E}[Y \mid X = x]$ can be interpreted as the vector of average demands for K goods, conditional on characteristics. We can interpret $B \subseteq \mathbb{R}^K$ as a fixed budget. Each good k has a d_k-dimensional vector of characteristics X_k. These include good-specific regressors as well as demographic and other individual-specific regressors. We collect these in $X = (X_1', \ldots, X_K')'$, which we treat as a random variable. The functions $\vec{v} = (v_1, \ldots, v_K)'$ shift the marginal utility of each good. The function D encodes substitutability/complementarity patterns between the goods. Importantly, characteristics do not enter D.

We state the following maintained assumptions.

Assumption 1.1. *(i)* $\mathbb{E}[Y \mid X = x]$ *satisfies (1.3) for every* $x \in \operatorname{supp}(X)$.[7]

(ii) B *is a non-empty set.*

(iii) $v_k : \mathbb{R}^{d_k} \to \mathbb{R}$ *for* $k = 1, \ldots, K$.

(iv) $D : \mathbb{R}^K \to \mathbb{R} \cup \{-\infty\}$ *is an extended real-valued function that is finite at some* $y \in B$.

For some of our results, we require an additional assumption.

[7] The support of a random variable Z, which is denoted $\operatorname{supp}(Z)$, is the smallest closed set K such that $P(Z \in K) = 1$.

Assumption 1.2. *(Unique Maximizer) For each* $x \in \operatorname{supp}(X)$*,*

$$M(\vec{v}(x)) = \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k) + D(y)$$

is a singleton.

A sufficient condition for Assumption 1.2 is that D is strictly concave and B is convex and nonempty. Under this assumption, we can write

$$\mathbb{E}[Y \mid X = x] = M(\vec{v}(x)),$$

where $M : \mathbb{R}^K \to \mathbb{R}^K$.

We first describe a class of latent utility models that imply (1.1) and then provide some examples that fit into our framework.

1.3.1 Examples

One possible interpretation of (1.3) is that the function being maximized is a utility function for a representative agent and $\mathbb{E}[Y \mid X = x]$ is the demand for that agent. The restrictions of this model also arise from many latent utility models. Suppose that the quantity vector Y satisfies

$$Y \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(X_k) + \tilde{D}(y, \varepsilon). \tag{1.4}$$

This is utility maximization subject to a budget constraint. The econometrician observes characteristics X_k, but does not observe individual heterogeneity, ε, which is possibly infinite-dimensional. We can interpret $\tilde{D}(y, \varepsilon)$ as some (possibly unknown) function allowing individual heterogeneity in preferences over the goods.

Allen and Rehbeck (2016a) show that if X and ε are independent and other mild regularity conditions hold, conditional means satisfy (1.3) for some $D \in \mathcal{D}$. The set $\mathcal{D}$ denotes extended real-valued functions that are finite at some $y \in B$ and never attain ∞. We will use this aggregation theorem to show several examples that fit in our framework.

The fact that $v_k(X_k)$ does not contain latent variables is a key homogeneity assumption, but this assumption is weaker than may initially appear. An equivalent restatement of the model is given by,

$$Y \in \tilde{M}(\vec{v}(X), \varepsilon), \tag{1.5}$$

where for fixed ε, $\tilde{M}$ is the subgradient of a convex function and is derived from $\tilde{D}$ and B. (See Appendix 1.10 for details.) When there is a unique maximizer to (1.4), the subgradient is just the derivative and we have the equality $Y = \tilde{M}(\vec{v}(X), \varepsilon)$, as presented in the Introduction. The response of Y to changes in X is flexible, and can vary widely as ε varies. The core shape restriction that must hold for each ε is a multivariate version of monotonicity.

We now turn to specific examples. Previous work has needed to specify the form of $\tilde{D}$. Our results apply *regardless* of the form, provided certain expectations exist. Recall that we differ from previous work in that we are interested in $\vec{v}$, whereas previous work is often interested in the distribution of ε. To identify this distribution, more structure on $\tilde{D}$ is typically needed.

Bundles

We present a version of the model studied in Gentzkow (2007), Fox and Lazzati (forthcoming), and Allen and Rehbeck (2016a). Assume for simplicity that there are two goods, and an individual can buy either 0 or 1 unit of each good. Let $u_{j,k}$ denote utility obtained from quantity j of good 1 and quantity k of good 2. We assume utilities are given by

$$\begin{aligned} u_{0,0} &= 0 \\ u_{1,0} &= v_1(X_1) + \varepsilon_{1,0} \\ u_{0,1} &= v_2(X_2) + \varepsilon_{0,1} \\ u_{1,1} &= u_{1,0} + u_{0,1} + \varepsilon_{1,1}, \end{aligned} \tag{1.6}$$

where X_k is a vector of characteristics for good k. The vector $\varepsilon = (\varepsilon_{1,0}, \varepsilon_{0,1}, \varepsilon_{1,1})$ consists of latent random variables known to the individual but not the econometrician. The variable $\varepsilon_{1,1}$ parametrizes whether the goods are complements or substitutes.

Under the following conditions, this model fits into our setup.

Lemma 1.1. *Let $Y \in \{0,1\}^2$ denote the utility maximizing quantities of (1.6). Assuming X and ε are independent, Y is measurable, and ε has finite mean, then there is some $D \in \mathcal{D}$ such that*

$$\mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in [0,1]^2} \sum_{k=1}^{2} y_k v_k(x_k) + D(y).$$

In particular,

$$D(y) = \sup_{\tilde{Y} \in \mathcal{Y} : \mathbb{E}[\tilde{Y}(\varepsilon)] = y} \mathbb{E}\left[\tilde{Y}_1(\varepsilon)\varepsilon_{1,0} + \tilde{Y}_2(\varepsilon)\varepsilon_{0,1} + 1\{\tilde{Y}_1(\varepsilon) = \tilde{Y}_2(\varepsilon) = 1\}\varepsilon_{1,1}\right],$$

where $\mathcal{Y}$ is the set of measurable functions from the support of ε to $\{0,1\}^2$.

This result uses the additive separability in (1.6). If similar separability is maintained, this example can generalize to multiple goods, regardless of whether the quantities are discrete.

Fox and Lazzati (forthcoming) formally study identification of this model, focusing on identification of the distribution of ε (and other structural features not present in our simplified model). They rely on special regressors Z_k that enter $v_k(X_k) = Z_k + \tilde{v}(W_k)$ in a known way, where $X_k = (Z_k, W_k)$. Our results show that $\vec{v}$ can be identified without this structure, and so we show that their insights apply to a more general setup.

Additive Random Utility Models

Additive random utility models have been widely used since the seminal work of McFadden (1973). These models include familiar examples like logit, nested logit, and probit. Suppose an individual chooses from a set of K alternatives (called goods previously). Alternative k has utility given by

$$u_k = v_k(X_k) + \varepsilon_k.$$

In an *additive random utility model* (ARUM), an individual's choice satisfies

$$Y \in \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1}^{K} y_k v_k(X_k) + \sum_{k=1}^{K} y_k \varepsilon_k, \tag{1.7}$$

where $\Delta^{K-1} = \{y \in \mathbb{R}^K \mid \sum_{k=1}^{K} y_k = 1,\ y_k \geqslant 0\ \forall k\}$. We let the individual choose from the probability simplex Δ^{K-1} to handle utility ties. Typically the individual is just picking the alternative with the highest latent utility.

We are interested in identification of $\vec{v}$ *without* specifying a distribution for the latent variable $\varepsilon = (\varepsilon_1, \dots \varepsilon_K)'$. Instead, we assume X and ε are independent. The following result, with slightly stronger assumptions, is due to Hofbauer and Sandholm (2002).

Lemma 1.2. *Assume Y is consistent with (1.7). If ε and X are independent, Y is measurable, and ε has finite mean, then there is some $D \in \mathcal{D}$ such that*

$$\mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1}^{K} y_k v_k(x_k) + D(y).$$

In particular, $D(y) = \sup_{\tilde{Y} \in \mathcal{Y} : \mathbb{E}[\tilde{Y}(\varepsilon)] = y} \mathbb{E}\left[\sum_{k=1}^{K} \tilde{Y}_k(\varepsilon)\varepsilon_k\right]$, where $\mathcal{Y}$ is the set of measurable functions from the support of ε to Δ^{K-1}.

Note we do not have to assume that ε_j and ε_k are independent for $j \neq k$, and moreover we do not have to specify a distribution for ε. Identification of the distribution of ε is straightforward once $\vec{v}$ is identified.

We provide a new set of conditions for identification of ARUM. Under the assumption that ε and X are independent, Matzkin (1993) provides two sets of conditions to identify $\vec{v}$. The first requires that one function v_k be *known* at a large set of points $\mathcal{S}^k$ such that $v_k(\mathcal{S}^k) = \mathbb{R}$. This is typically interpreted as assuming $v_k(X_k) = Z_k + \tilde{v}_k(W_k)$ for a special regressor Z_k. The second set of conditions ensures that the identification problem can be reduced to that of a binary choice problem by an assumption such as $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$. Our results show that neither of these assumptions is necessary.

Our results do not handle general failures of independence between ε and X. We show in Remark 1.3 that we can relax independence if we impose additive separability, $v_k(X_k) = g(Z_k) + \tilde{v}_k(W_k)$. Suppose that conditional on $(W_1, \dots, W_K)$, the vector $Z = (Z_1, \dots, Z_K)$ is independent of ε. Our results show g_k is identified. By defining $\tilde{z}_k = g(z_k)$, we can use $\tilde{z}_k$ as a special regressor (Lewbel (2000)) to identify $\tilde{v}_k$. Lemma 1.2 also rules out random coefficients models. Matzkin (2007) and Berry and Haile (2009) use special regressors to identify models with nonseparable heterogeneity. Because we show how to relax assumptions of the special regressor approach, we widen the applicability of these papers.[8]

Our results do not apply if we replace independence with conditional median restrictions as in the maximum score approach (Manski (1975), Manski (1985), Matzkin (1993), Fox (2007)). Heuristically, the maximum score approach has only been shown to deliver *ordinal* identification of v_k whereas additive, independent errors ensures *cardinal* identification of v_k.

Expected Utility / Moral Hazard

Suppose an individual chooses a lottery Y that maximizes expected utility plus a heterogenous term $\tilde{D}$:

$$Y \in \underset{y \in \tilde{B}}{\operatorname{argmax}} \sum_{k=1}^{K} y_k v_k(X_k) + \tilde{D}(y, \varepsilon). \tag{1.8}$$

There are K outcomes. Outcome k is a "good" treated in our general setup. Individual heterogeneity is represented by ε.

Agarwal and Somaini (2014) study a special case of (1.8) in their study of school match. An agent chooses a report, but in equilibrium the agent knows the allocation probabilities associated with this report. Thus, choice of report is equivalent to choice from a set of lotteries. They assume expected utility with additive errors so that $\tilde{D}(y, \varepsilon) = \sum_{k=1}^{K} y_k \varepsilon_k$. We can interpret $v_k(X_k) + \varepsilon_k$ as the von Neumann-Morgenstern utility index. This index is assumed known to the individual but not the econometrician.

[8]Berry and Haile (2009) show that this can be done using the Matzkin (1993) conditions for nonparametric identification in ARUM. Recall we provide an alternative set of conditions to Matzkin (1993).

The budget $\tilde{B}$ in the latent utility specification (1.8) is a finite set of lotteries.[9]

With this structure, Agarwal and Somaini (2014), Section 6.2 establishes identification of the distribution of ε using knowledge of $\tilde{B}$ and a special regressor. We strengthen their results by showing that $\vec{v}$ is identified without needing a special regressor. It is possible to identify $\vec{v}$ without knowledge of $\tilde{B}$, without observing the choice of lottery or report (many observations of ex-post assignment is sufficient), and with deviations from expected utility.[10] Together with the results of Agarwal and Somaini (2014), this establishes identification of $\vec{v}$ and the latent distribution of ε.[11]

The setup of (1.8) also admits a moral hazard interpretation. Suppose an individual can exert costly effort to affect the probability a particular outcome is realized. Formally, the individual chooses an effort level $e \in \mathcal{E}$ to maximize

$$\sum_{k=1}^{K} p_k(e)v_k(X_k) - c(\vec{p}(e), \varepsilon).$$

The vector $\vec{p}(e) = (p_1(e), \ldots, p_K(e))$ represents the probability of each outcome occurring given the effort level. If effort is unobserved, this can be written in the form of (1.8). The budget is given by $\tilde{B} = \{y \mid y = p(e) \text{ for some } e \in \mathcal{E}\}$ and $\tilde{D}$ is defined by $\tilde{D}(y, \varepsilon) = -c(\vec{p}(e), \varepsilon)$ whenever $y = \vec{p}(e)$. If the econometrician does not observe effort, then the individual's choice of effort is analogous to choice of a probability vector.

The following result shows that (1.8) fits into our setup.

Lemma 1.3. *Assume Y is consistent with (1.8). If ε is independent of X, Y is measurable,*

[9] This set is fixed in one of their identification approaches (Agarwal and Somaini (2014), Section 6.2). This budget is convexified in the average demand version of the model.

[10] These deviations could be due to costly effort or hedging Fudenberg, Iijima, and Strzalecki (2015), Section 5.1.

[11] While $\vec{v}$ is identified without needing to know the budget $\tilde{B}$, the results of Agarwal and Somaini (2014) use knowledge of the budget to identify the distribution of ε. It is unknown whether knowledge of $\tilde{B}$ is needed to identify the distribution of ε.

and $\mathbb{E}[\tilde{D}(Y,\varepsilon) \mid X = x]$ *is finite for each* $x \in \operatorname{supp}(X)$*, then there is some* $D \in \mathcal{D}$ *such that*

$$\mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1}^{K} y_k v_k(x_k) + D(y).$$

Stochastic Choice

In discrete choice models, a growing literature studies choice that is stochastic at the *individual* level. One way to represent this is by

$$Y \in \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1}^{K} y_k v_k(X_k) + \tilde{D}(y,\varepsilon), \tag{1.9}$$

where Y is the choice of lottery for a specific individual (ε) given characteristics X. The interpretation of stochastic choice as deliberate choice of lottery is due to Machina (1985) and has received renewed theoretical and empirical interest.[12] We study this model in Allen and Rehbeck (2016b). Lemma 1.3 shows (1.9) this model fits into our setup once we integrate out ε.

Choice may be stochastic because it is costly to make or implement a decision (Mattsson and Weibull (2002)). For a "trembling hand" example (Selten (1975)), suppose that an individual chooses the alternative with highest index $v_k(X_k)$ a fraction $(1-\varepsilon)$ of the time, and otherwise uniformly randomizes over the remaining alternatives. The variable ε is specific to the individual and can be thought of as a probabilistic chance of

[12]Recent work includes Cerreia-Vioglio et al. (2015) and Agranov and Ortoleva (2017).

making a mistake. This is a special case of (1.9) if we set

$$\tilde{D}(y,\varepsilon) = \begin{cases} 0 & \text{if } \exists k \text{ s.t. } y_k = 1-\varepsilon \text{ and } y_j = \frac{\varepsilon}{K-1} \text{ for } j \neq k \\ -\infty & \text{otherwise} \end{cases}.$$ [13]

The representation (1.9) can capture other forms of costly optimization, preference for variety, and ambiguity aversion arising from uncertainty over the true utility of a good (Fudenberg, Iijima, and Strzalecki (2015)). A related setup has been used to model rational inattention (Matejka and McKay (2014), Caplin and Dean (2015)).

One qualitative feature allowed in (1.9) is a form of complementarity. For example, v_k can increase due to a change in x_k, and the conditional probability of choosing some *other* alternative can increase. This behavior may be natural in a model of mistakes or preference for variety. Complementarity is formally ruled out in ARUM. See Allen and Rehbeck (2016a) for a further discussion of complementarity.

In practice, we may not observe the choice of lottery Y, only the realizations of the lottery (i.e. the actual choices). Observing only realizations is sufficient for our identification results, since we only need the vector of conditional choice probabilities, $\mathbb{E}[Y \mid X = x]$.

Remark 1.1. Shi, Shum, and Song (2016) study identification of ARUM in a panel setting when $\vec{v}$ is linear. Theorem 1.8 shows that their identified set uses (only) the implications of the more general perturbed stochastic choice model discussed in this section. Thus, our results can be seen as a nonparametric counterpart to their identification results. We note that in addition to linearity, they impose v_k does not depend on k. This imposes an *a priori* common scale of v_k with respect to characteristics. While this is reasonable in a

[13] Note that Lemma 1.8 requires that $\mathbb{E}[\tilde{D}(Y,\varepsilon) \mid X = x]$ be finite. This allows $\tilde{D}$ to take on value $-\infty$ since Y is a choice. We include this function for motivation, but note that it is not formally covered by our identification results. If we smooth it slightly, our identification results apply.

discrete choice setting, one message of the present paper is that this is not necessary.

1.3.2 Model Structure

We now show how PUM implies analogous of the Slutsky restrictions. This helps illustrate the structure of the model and provides a foundation for our identification results. Recall that under Assumption 1.2, $\mathbb{E}[Y \mid X = x] = M(\vec{v}(x))$ for some M. Let $J(\overline{v})$ denote the Jacobian matrix of $M(\overline{v})$, which has j,k element equal to $\frac{\partial M_j(\overline{v})}{\partial \overline{v}_k}$.

Proposition 1.1 ("Slutsky Conditions")**.** *Let Assumption 1.1 and 1.2 hold. Assume M is continuously differentiable in a neighborhood of $\overline{v}$ with Jacobian matrix $J(\overline{v})$. Then*

(i) $u'J(\overline{v})u \geqslant 0, \forall u \in \mathbb{R}^K$.

(ii) $J_{k,\ell}(\overline{v}) = J_{\ell,k}(\overline{v})$ *for* $k, \ell = 1, \ldots, K$.

We refer to (i) as positive semi-definiteness and (ii) as symmetry. Symmetry provides a cross-equation *equality* that will be the foundation for our identification results.

Symmetry follows from the fact that when $J(\overline{v})$ exists, it is the Hessian of a convex function. We provide a binary choice example further illustrating why symmetry appears. This is a special case of ARUM (Section 1.3.1).

Example 1 (Binary Choice)**.** *Suppose that there are only two goods. In addition, assume* $(\varepsilon_1, \varepsilon_2)$ *is independent of X and has a continuous density. Then conditional on characteristics, the probability of choosing good* 1 *is given by*

$$\begin{aligned}\mathbb{E}[Y_1 \mid X = x] &= P(v_1(x_1) + \varepsilon_1 > v_2(x_2) + \varepsilon_2 \mid X = x) \\ &= F(v_1(x_1) - v_2(x_2)),\end{aligned}$$

where F denotes the cumulative distribution function of $\varepsilon_2 - \varepsilon_1$. Similarly,

$$\mathbb{E}[Y_2 \mid X = x] = 1 - F(v_1(x_1) - v_2(x_2)).$$

If we write,

$$\mathbb{E}[Y \mid X = x] = M(\vec{v}(x)) = (F(v_1(x_1) - v_2(x_2)), 1 - F(v_1(x_1) - v_2(x_2)))',$$

we can easily see that

$$\frac{\partial M_2(\overline{v})}{\partial \overline{v}_1} = \frac{\partial M_1(\overline{v})}{\partial \overline{v}_2}.$$

If in addition $(\varepsilon_1, \varepsilon_2)$ has a strictly positive density, these partial derivatives are strictly negative.

1.4 Identification of $\vec{v}$

This section provides conditions under which $\vec{v}$ is nonparametrically identified. We treat D as an unknown function. We assume knowledge of $\mathbb{E}[Y \mid X = x]$ over $x \in \text{supp}(X)$. This is because we aim to understand whether $\vec{v}$ can be uniquely determined with *ideal*, population-level information. Thus, we abstract from sampling error. One can think of knowledge of $\mathbb{E}[Y \mid X = x]$ as obtained from an "infinite" number of independent and identically distributed draws of (Y, X).

Our results require that each good has a continuous regressor that is excluded from the other equations. To formalize this, recall X_k is a d_k-dimensional vector of characteristics for good k. Partition each such vector into $X_k = (Z_k', W')'$. The d_z^k-dimensional vector Z_k contains regressors that are *excluded* from the function v_ℓ for

$\ell \neq k$. We require $d_z^k > 0$. The d_w-dimensional vector W contains characteristics that are common across goods, such as socioeconomic characteristics.

For a discrete choice example, suppose the K goods are modes of transportation. Interpret $\mathbb{E}[Y \mid X = x]$ as the probability distribution of choosing the goods, conditional on observable characteristics. Let $k = 1$ denote "bus." The vector Z_1 may include bus fare and the number of bus lines. The vector W may include regressors that could affect the desirability of several (or all) goods, such as an individual's income.

In order to identify $\vec{v}$ we require a normalization. Let $c \in \mathbb{R}^k$ and let $\lambda > 0$ be a scalar. Then we have the equality

$$\operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k) + D(y) = \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k(\lambda v_k(x_k) + c_k) + \left(\lambda D(y) - \sum_{k=1}^{K} y_k c_k\right). \tag{1.10}$$

Because we do not specify D, identification requires two normalizations, to handle both λ and c. This is because if $\vec{v}$ is consistent with the model, then $\lambda \vec{v} + c$ is as well. We provide conditions under which $\vec{v}$ is identified up to a multiplicative scale normalization and additive location normalization.

Identification of $\vec{v}$ is established as follows. First, we assume all characteristics are continuous and that there are no common characteristics, i.e. $d_w = 0$. Using the analogue of Slutsky symmetry, we constructively identify $\vec{v}$ up to a location/scale normalization. We then leverage this analysis to identify $\vec{v}$ when $d_w > 0$.

1.4.1 Identification for Good-Specific Regressors

We first provide identification results if each regressor shows up in exactly one index v_k. Specifically, we treat the case $d_w = 0$ in this section. These results hold as well if we condition on a fixed value $\overline{w}$ of common characteristics. We will use this fact to identify $\vec{v}$ when there are common characteristics.

We provide an informal sketch how we use symmetry for identification. For simplicity suppose that x_k is a scalar for $k = 1, \ldots, K$. Under smoothness conditions, for arbitrary k, ℓ,

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} = \left.\frac{\partial M_k(\overline{v})}{\partial \overline{v}_\ell}\right|_{\overline{v}=\vec{v}(x)} \frac{\partial v_\ell(x_\ell)}{\partial x_\ell}, \tag{1.11}$$

from the chain rule. This relies on the fact that x_ℓ is excluded from $v_j(\cdot)$ for $j \neq \ell$. Secondly, x_ℓ must be continuous so we can take a derivative. An analogous equation holds for k and ℓ interchanged. Assuming all involved derivatives are nonzero, combining (1.11) with symmetry of cross-partials of M (Proposition 1.1(ii)), we obtain

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_k} = \frac{\partial v_\ell(x_\ell)}{\partial x_\ell} \Bigg/ \frac{\partial v_k(x_k)}{\partial x_k}. \tag{1.12}$$

Thus, we identify the ratio of two specific partial derivatives of $\vec{v}$ at points x_ℓ and x_k.

We now consider the general case where x_k is not a scalar and formalize the arguments leading to (1.12). We only need (1.12) to hold at certain *points*. We thus make explicit the points of evaluation of partial derivatives. We maintain Assumption 1.2 to state the following definition.

Definition 1.1 (Pairs). *Say that the points* $\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell = x_\ell^*}$ *and* $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k = x_k^*}$ *are paired if they exist and the following conditions hold.*

(i) There exists a known value $x^* = (x_1^{*\prime}, \ldots, x_K^{*\prime})' \in \operatorname{supp}(X)$ *that has* ℓ*-th component* x_ℓ^* *and* k*-th component* x_k^**.*

(ii) $\left.\frac{\partial \mathbb{E}[Y_\ell \mid X=x]}{\partial x_{k,q}}\right|_{x=x^*}$ *and* $\left.\frac{\partial \mathbb{E}[Y_k \mid X=x]}{\partial x_{\ell,p}}\right|_{x=x^*}$ *exist.*

(iii) M *is continuously differentiable in a neighborhood of* $\vec{v}(x^*)$*.*

(iv) $\left.\frac{\partial M_k(\overline{v})]}{\partial \overline{v}_\ell}\right|_{\overline{v}=\vec{v}(x^*)} \neq 0.$

If in addition $\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^*}$ *and* $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$ *are nonzero, we say that they are strictly paired.*

One important fact is that if $\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^*}$ and $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$ are paired, then either $\ell \neq k$ or the values are equal, $x_\ell^* = x_k^*$. This is because of part (i). In particular, if $K = 1$, then in order for partial derivatives to be paired, they must be evaluated at the same arguments.

Part (i) is stated in terms of x^* being known so that (in principle) we know precisely where to evaluate derivatives of conditional means. Part (ii) is a support condition. In order for these derivatives to exist, we need to be able to continuously vary $x_{k,q}$ and $x_{\ell,p}$ separately from the other characteristics. Part (iii) ensures symmetry,

$$\left.\frac{\partial M_k(\overline{v})}{\partial \overline{v}_\ell}\right|_{\overline{v}=\vec{v}(x^*)} = \left.\frac{\partial M_\ell(\overline{v})}{\partial \overline{v}_k}\right|_{\overline{v}=\vec{v}(x^*)}$$

Part (iv) is a behavioral restriction. It requires that if good ℓ becomes more attractive, there need to be "spillovers" to good k. These occur precisely when goods satisfy a local form of substitutability or complementarity.[14]

Proposition 1.2. *Let Assumptions 1.1 and 1.2 hold and assume $x_{\ell,p}$ and $x_{k,q}$ are regressors specific to ℓ and k, respectively. If the points* $\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^*}$ *and* $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$ *are*

[14]See Allen and Rehbeck (2016a) for more discussion of complementarity and substitutability in these models.

paired and $\frac{\partial v_k(x_k)}{\partial x_{k,q}}\Big|_{x_k=x_k^*} \neq 0$*, there is some known* $x^* \in \text{supp}(X)$ *such that*

$$\frac{\partial \mathbb{E}[Y_k \mid X=x]}{\partial x_{\ell,p}}\bigg|_{x=x^*} \bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X=x]}{\partial x_{k,q}}\bigg|_{x=x^*} = \frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\bigg|_{x_\ell=x_\ell^*} \bigg/ \frac{\partial v_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=x_k^*}, \tag{1.13}$$

where x_ℓ^* *and* x_k^* *are components of* x^* *specific to goods* ℓ *and* k*, respectively, and* $\partial \mathbb{E}[Y_\ell \mid X=x]/\partial x_{k,q}\big|_{x=x^*} \neq 0$*. In particular, the right hand side of (1.13) is identified.*

It is noteworthy that the left hand side of (1.13) involves the characteristics for all goods x^*, whereas the right hand side only involves the characteristics for alternatives ℓ and k. This suggests the equality could potentially be used to perform a specification test.[15]

This result tells us that a derivative ratio can be identified by comparing two goods at a time. Identification is established by varying good-specific regressors. We will leverage this result to identify $\vec{v}$ by using the ideas of the following lemma, which is a consequence of the fundamental theorem of calculus.

Lemma 1.4. *Let* $f:\mathbb{R}\to\mathbb{R}$ *and* $g:\mathbb{R}\to\mathbb{R}$ *be differentiable and define*

$$h(a,b) = \frac{\partial f(a)}{\partial a} \bigg/ \frac{\partial g(b)}{\partial b}$$

whenever the denominator is nonzero. If $h(\cdot,b^*)$ *exists and is known for a fixed* b^**,* $\frac{\partial g(b)}{\partial b}\Big|_{b=b^*} = 1$*, and* $f(0)=0$*, then* f *is identified. In particular,*

$$f(a^*) = \int_0^{a^*} h(a,b^*)da.$$

[15]For stochastic choice, when X is discrete and takes on finitely many values, Allen and Rehbeck (2016b) characterize the complete testable implications of the model. An interesting question is whether an alternative characterization of the testable implications can be given when X is continuous and smoothness conditions hold.

This lemma can be applied as well to identify g once f is identified. The multivariate extension is immediate. In our setting, $f(a)$ and $g(b)$ are replaced by $v_\ell(x_\ell)$ and $v_k(x_k)$. The function h is replaced by derivative ratios of conditional means.[16]

The following theorem is our main result. Its assumptions ensure that the points $\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\Big|_{x_\ell=x_\ell^*}$ and $\frac{\partial v_k(x_k)}{\partial x_{k,q}}\Big|_{x_k=x_k^*}$ are paired whenever $j \neq \ell$ and these derivatives are nonzero. Identification of $\vec{v}$ follows using Proposition 1.2 and Lemma 1.4.

Theorem 1.1. *Let Assumptions 1.1 and 1.2 hold, assume all regressors are good-specific and assume $K \geqslant 2$. Assume X has full support,*[17] *$\vec{v}$ is differentiable, $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$, and M is continuously differentiable with nonzero cross-partial derivatives. Then $\vec{v} = (v_1, \ldots, v_K)$ is identified under the following normalization:*

i. (Scale) $\frac{\partial v_k(x_k)}{\partial x_{k,q}}\Big|_{x_k=x_k^*} \in \{-1, 1\}$ *for a tuple (k, q, x_k^*).*

ii. (Location) $v_\ell(0_{d_\ell}) = 0$ for each $\ell = 1, \ldots, K$, where 0_{d_k} denotes a d_k-dimensional vector of zeros.

Remark 1.2. We assume X has full support to identify $\vec{v}$ over its entire domain. We assume $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$ for simplicity.[18]

In light of (1.10), the scale and location normalization cannot be weakened without further restrictions on D.[19] To obtain identification with this normalization, it is necessary that the model involve *multiple* goods in a non-trivial way. If $K = 1$ or there is no complementarity/substitutability across goods, the best we can hope for is identification of v_k up to a monotonic transformation. See Remark 1.10.

[16] We contrast this result with what could be obtained if we knew $f = g$ but only observed h at points $a = b$. In this case, we could only identify f up to a *monotonic transformation*.

[17] That is, $\text{supp}(X) = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_K}$.

[18] This assumption allows us to rule out boundary issues. We expect it can be relaxed to the assumption that $\vec{v}(\text{supp}(X))$ is convex with nonempty interior but do not provide details. See Appendix 1.16 for a result that significantly weakens the assumptions of Theorem 1.1.

[19] Though other normalizations could ensure identification. An example would be replacing the scale normalization with $|v_k(x_k^*)| = 1$, provided this value is nonzero and $x_k^* \neq 0_{d_k}$.

As a corollary of Theorem 1.1, we provide new conditions for identification of additive random utility models. The theorem readily applies because in ARUM, all cross-partials of M are nonpositive. Under mild conditions, the cross-partials are strictly negative.

Corollary 1.1. *Let the assumptions of Lemma 1.2 hold and assume $K \geqslant 2$. Suppose X has full support, $\vec{v}$ is differentiable, $\vec{v}(\operatorname{supp}(X)) = \mathbb{R}^K$, and ε has a density that is everywhere positive and continuously differentiable. We maintain the assumption that all regressors are good-specific. Then the conditions of Theorem 1.1 are satisfied. In particular, $\vec{v}$ is nonparametricially identified up to a location and scale normalization.*

The assumptions of Theorem 1.1 are overly strong in some contexts. In particular, it is not innocuous to assume that all cross-partials of M are *everywhere* nonzero. In ARUM, having everywhere nonzero cross-partials implies choice probabilities are always on the interior of the simplex. More generally, having nonzero cross-partials implies goods can never switch from being complements to being substitutes, since this would imply a cross-partial derivative is zero somewhere by continuity. In Appendix 1.16, we show identification of $\vec{v}$ under a weaker set of assumptions that can accomodate these cases. The basic idea is that we can use Proposition 1.2 to identify derivative ratios of components of $\vec{v}$. We can multiply these derivative ratios to identify new derivative ratios. If all derivative ratios of $\vec{v}$ can be identified, we then identify $\vec{v}$ up to location and scale by the mean value theorem.

Remark 1.3 (Special Regressors). Our results can be combined with insights of the special regressor (Lewbel (1998)) approach to relax independence conditions in some of our motivating examples. For concreteness, consider the discrete choice setting with latent utility for alternative k given by

$$u_k = g_k(X_{k,1}) + \tilde{v}_k(X_{k,-1}) + \varepsilon_k, \tag{1.14}$$

where $X_{k,1}$ is a scalar and $X_{k,-1}$ collects components of X_k other than $X_{k,1}$. We assume (1.14) holds for each $k = 1, \ldots, K$. We also assume each $X_{k,1}$ is continuous. The canonical version of the special regressor approach imposes the additional assumption that g_k be the identity mapping. This implies that $X_{k,1}$ enters v_k monotonically and that $X_{k,1}$ and ε_k are in the same units.

We can relax independence between characteristics and unobservables to a conditional independence condition. Specifically, suppose that $X_1 = (X_{1,1}, \ldots, X_{K,1})$ and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_K)$ are independent, conditional on $X_{-1} = (X_{1,-1}, \ldots, X_{K,-1})$. Then conditioning on $X_{-1} = x_{-1}$, Lemma 1.2 applies to yield

$$\mathbb{E}\left[Y \mid X_1 = x_1, X_{-1} = x_{-1}\right] \in \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1}^{K} y_k g_k(x_{k,1}) + D(y, x_{-1})$$

for some D. Note that x_{-1} enters D. By fixing x_{-1} and varying x_1, Theorem 1.1 provides conditions under which $g_k(\cdot)$ is identified for each k.[20] Conditions from Lewbel (2000) can then be used to identify $\tilde{v}_k$.

Remark 1.4 (Invertibility). While the Jacobian of M is assumed to exist at certain points, we do not place invertibility restrictions on it. Thus, we are not using the inverse function theorem for any of the results in this section.

1.4.2 Identification for Common Characteristics

In this section we provide identification results when there are discrete characteristics or characteristics that are not good-specific. We require that each good have at least

[20]To apply our results, a location and scale assumption is needed for g_k. It is important to note that a scale assumption such as $g_k'(0) \in \{-1, 1\}$ must hold for *arbitrary* conditioning values x_{-1}. Thus, while imposing $g_k'(0) \in \{-1, 1\}$ is weaker than assuming $g_k(X_{k,1}) = X_{k,1}$ as in Lewbel (2000), it may not be an "innocous" normalization.

one good-specific, continuous regressor. All other characteristics can be discrete.

Recall we assume $X_k = (Z_k', W')'$, where W consists of characteristics that are common across alternatives, and Z_k consists of regressors that may be specific to alternative k. To state the following assumption, partition $Z_k = \left(Z_k^{(1)'}, Z_k^{(2)'}\right)'$. Think of $Z_k^{(1)}$ as a subvector of continuous regressors specific to good k. Let $Z^{(m)} = \left(Z_1^{(m)'}, \ldots, Z_K^{(m)'}\right)'$ for $m = 1, 2$.

Assumption 1.3. *There exists a known* $\left(\tilde{z}^{(2)'}, \tilde{w}'\right)' \in \operatorname{supp}\left(Z^{(2)}, W\right)$ *such that*

i. $\vec{v}\left(\operatorname{supp}\left(Z^{(1)} \mid Z^{(2)} = \tilde{z}^{(2)}, W = \tilde{w}\right), \tilde{z}^{(2)}, \tilde{w}\right) = \vec{v}(\operatorname{supp}(Z, W))$.

ii. $\vec{v}\left(z^{(1)}, \tilde{z}^{(2)}, \tilde{w}\right)$ *is identified for each* $z^{(1)} \in \operatorname{supp}\left(Z^{(1)} \mid Z^{(2)} = \tilde{z}^{(2)}, W = \tilde{w}\right)$.

Part (i) states that conditional on a value $\left(\tilde{z}^{(2)'}, \tilde{w}'\right)'$, $Z^{(1)}$ can move sufficiently to trace out the variation in $\vec{v}$. This is a relevance and support condition. Under Assumption 1.2, (i) is equivalent to assuming that for each $(z', w')' \in \operatorname{supp}(Z, W)$, there exists $\tilde{z}^{(1)} \in \operatorname{supp}(Z^{(1)} \mid Z^{(2)} = \tilde{z}^{(2)}, W = \tilde{w})$ such that

$$\mathbb{E}[Y \mid Z = z, W = w] = \mathbb{E}[Y \mid Z = \tilde{z}, W = \tilde{w}],$$

where $\tilde{z} = (\tilde{z}^{(1)'}, \tilde{z}^{(2)'})'$. Sufficient conditions for (ii) are given in Theorem 1.1. Recall that the results of the theorem go through *conditional* on $Z^{(2)} = \tilde{z}^{(2)}, W = \tilde{w}$.

For the following theorem, define

$$D_B(y) = \begin{cases} D(y) & \text{if } y \in B \\ -\infty & \text{otherwise} \end{cases}.$$

Theorem 1.2. *Let Assumptions 1.1, 1.2, and 1.3 hold. Assume D is concave and B is convex. If the derivative of* D_B *exists at* $\mathbb{E}[Y \mid Z = z, W = w]$ *and* $(z, w) \in \operatorname{supp}(Z, W)$*,* *then $\vec{v}(z,w)$ is identified.*

The assumptions on D_B are used to establish that

$$\mathbb{E}[Y \mid Z = z, W = w] = \mathbb{E}[Y \mid Z = \tilde{z}, W = \tilde{w}] \implies \vec{v}(z,w) = \vec{v}(\tilde{z},\tilde{w}). \tag{1.15}$$

From this implication we identify $\vec{v}(z,w)$ by "matching" it with a value $\vec{v}(\tilde{z},\tilde{w})$ that is already identified. With the maintained assumptions that D is concave and B is convex, differentiability is actually *necessary* for this implication without further restrictions on the parameter space for $\vec{v}$. This is formalized in Proposition 1.5.

Remark 1.5. Concavity of D is innocuous on its own (see Theorem 1.8), but places restrictions when combined with differentiability. Differentiability of D_B at $\mathbb{E}[Y \mid Z = z, W = w]$ implies this value cannot be on the boundary of B. In particular, differentiability requires that B have a non-empty interior when viewed as a subset of $\mathbb{R}^K$. This rules out the probability simplex. However, the theorem can be extended to handle the probability simplex after a change of variables. This can be done with the normalization $v_1(\cdot) = 0$, which is sensible if the first good is an outside option such as "buy nothing." With this normalization, the problem can be reparametrized to eliminate the first good. See Appendix 1.18.1 for more details.

1.5 Identification of D and Welfare Changes

We now study identification of D, assuming $\vec{v}$ is identified from our previous results. When these are both identified, many counterfactual questions can be answered. Secondly, D provides information on the complementarity/substitutability in the model, and so is of interest in its own right.

To study welfare, define the *social surplus function* (McFadden (1978))

$$V(x) = \sup_{y \in B} \left\{ \sum_{k=1}^{K} y_k v_k(x_k) + D(y) \right\}.$$

This function has been widely use to quantify welfare changes in a discrete choice environment (Small and Rosen (1981)).[21] In ARUM, $V(x) = \mathbb{E}[\max_k \{v_k(X_k) + \varepsilon_k\} \mid X = x]$. Allen and Rehbeck (2016a) show that for a class of latent utility models (which includes our examples), V may interpreted as the average indirect utility function for optimizing agents, once latent variables have been integrated out. Alternative objects may be more natural measures of welfare. In discrete choice, if D arises purely due to costly optimization for an individual, then differences in $\sum_{k=1}^{K} \mathbb{E}[Y_k \mid X = x] v_k(x_k)$ for different choices of x may be welfare relevant. We will refer to V as a welfare measure, with the caveat that its interpretation depends on the context.

The units of V are "utils," but utils can be converted to characteristics because $\vec{v}$ is assumed known. For example, if price enters as a characteristic (as is common in discrete choice models that assume away income effects), then differences in V can be converted to dollars.

The functions V and D are linked by the identity

$$V(x) = \sum_{k=1}^{K} \mathbb{E}[Y_k \mid X = x] v_k(x_k) + D(\mathbb{E}[Y \mid X = x]), \tag{1.16}$$

which follows from the fact that $\mathbb{E}[Y \mid X = x]$ is a maximizer. We will state results in terms of differences in V and D, depending on which is more convenient.

In order to characterize what can be learned about D, we will use the following

---

[21] See also Bhattacharya (2015). The results of Bhattacharya (2015) do not apply to the general setting we consider.

inequality, which follows from writing the necessary conditions for optimality:

$$D(\mathbb{E}[Y \mid X=\tilde{x}]) - D(\mathbb{E}[Y \mid X=x]) \leqslant (\mathbb{E}[Y \mid X=x] - \mathbb{E}[Y \mid X=\tilde{x}])' v(x). \tag{1.17}$$

Remark 1.6 (Normalizations). Nonparametric identification of $\vec{v}$ requires a location and scale normalization, as shown in (1.10). The results in this section are specific to a particular normalization. The scale normalization affects the bounds on both D and V, and the effect of the scale is easy to see. A less transparent fact is that the location normalization only affects the bounds on D. For example if we normalize $\vec{v}(x) = 0_K$, where 0_K is vector of zeros of length K, then from (1.17) we have

$$D(\mathbb{E}[Y \mid X=\tilde{x}]) - D(\mathbb{E}[Y \mid X=x]) \leqslant 0 \tag{1.18}$$

for any $\tilde{x} \in \text{supp}(X)$. This tells us that the highest value of D is obtained at the point where we normalize $\vec{v}(\cdot) = 0_K$. Fortunately, the particular location normalization will not affect identification of differences in V. This is shown in the integral representation in Theorem 1.4.

1.5.1 Partial Identification of D and Welfare Changes

We first provide bounds on the differences of certain values of D. Recall these bounds immediately give bounds on differences in V by (1.16). The bounds are sharp if we only know that $D \in \mathcal{D}$, i.e. D is finite at some $y \in B$ and never attains $-\infty$.

Let $x^0, \ldots, x^S$ be a sequence of points in $\text{supp}(X)$. By summing up (1.17), we obtain

$$
\begin{aligned}
D(\mathbb{E}[Y \mid X = x^S] - D(\mathbb{E}[Y \mid X = x^0]) \leqslant \\
\sum_{s=0}^{S-1} (\mathbb{E}[Y \mid X = x^s] - \mathbb{E}[Y \mid X = x^{s+1}])' \vec{v}(x^s).
\end{aligned} \tag{1.19}
$$

We need a bit more notation to present a strengthening of this bound. For $x, \tilde{x} \in \operatorname{supp}(X)$, let $C(x, \tilde{x})$ be the set of finite sequences in $\operatorname{supp}(X)$ that begin at x and end at $\tilde{x}$. Define

$$
\begin{aligned}
\overline{\Delta} D(x, \tilde{x}) &= \inf_S \inf_{\{x^s\}_{s=0}^S \subseteq C(x,\tilde{x})} \left\{ \sum_{s=0}^{S-1} (\mathbb{E}[Y \mid X = x^s] - \mathbb{E}[Y \mid X = x^{s+1}])' \vec{v}(x^s) \right\} \\
\underline{\Delta} D(x, \tilde{x}) &= \sup_S \sup_{\{x^s\}_{s=0}^S \subseteq C(x,\tilde{x})} \left\{ \sum_{s=0}^{S-1} (\mathbb{E}[Y \mid X = x^s] - \mathbb{E}[Y \mid X = x^{s+1}])' \vec{v}(x^{s+1}) \right\}.
\end{aligned}
$$

By (1.19), we obtain that for arbitrary $x, \tilde{x} \in \operatorname{supp}(X)$,

$$
\underline{\Delta} D(x, \tilde{x}) \leqslant D(\mathbb{E}[Y \mid X = \tilde{x}]) - D(\mathbb{E}[Y \mid X = x]) \leqslant \overline{\Delta} D(x, \tilde{x}). \tag{1.20}
$$

We formalize below that these bounds are sharp if D is only restricted to satisfy $D \in \mathcal{D}$. Note we do not assume a unique maximizer. This is for technical reasons.[22]

Theorem 1.3. *Suppose Assumption 1.1 holds, $\vec{v}$ is known, and $x, \tilde{x} \in \operatorname{supp}(X)$. Then there exists a function $\tilde{D} \in \mathcal{D}$ such that*

$$
\tilde{D}(\mathbb{E}[Y \mid X = \tilde{x}]) - \tilde{D}(\mathbb{E}[Y \mid X = x]) = \overline{\Delta} D(x, \tilde{x})
$$

and

$$
\forall x \in \operatorname{supp}(X), \mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k) + \tilde{D}(y).
$$

[22] The bounds still hold with the additional restriction of a unique maximizer. Studying sharpness is technically challenging because one must work with strict inequalities.

An analogous statement holds for $\underline{\Delta} D(x, \tilde{x})$.

To bound differences in V, we can use (1.16) to obtain,

$$V(\tilde{x}) - V(x) = \left(\sum_{k=1}^{K} \mathbb{E}[Y_k \mid X = \tilde{x}] v_k(\tilde{x}_k) - \sum_{k=1}^{K} \mathbb{E}[Y_k \mid X = x] v_k(x_k) \right) + \left(D(\mathbb{E}[Y \mid X = \tilde{x}]) - D(\mathbb{E}[Y \mid X = x]) \right).$$

Theorem 1.3 shows differences in D are bounded by $\underline{D}$ and $\overline{D}$, which then bound differences in V.

1.5.2 Point Identification of D and Welfare Changes

In this section we establish point identification of differences in D and V. We first define the $\vec{v}$-subdifferential of V at x:

$$\partial_v V(x) = \{u \in \mathbb{R}^k \mid \forall \tilde{x}, V(\tilde{x}) - V(x) \geqslant u \cdot (\vec{v}(\tilde{x}) - \vec{v}(x))\}.$$

It can be shown that

$$\mathbb{E}[Y \mid X = x] \in \partial_v V(x). \tag{1.21}$$

If $\partial_v V(x)$ is a singleton, then

$$\mathbb{E}[Y \mid X = x] = M(\vec{v}(x)),$$

as used previously. The following result identifies differences in V and D by "integrating" (1.21).

Theorem 1.4. *Let Assumption 1.1 hold. Assume* $\vec{v}$ *is known,* V *is everywhere finite, and let* $x^0, x^1 \in \operatorname{supp}(X)$. *Suppose there is a function* $x(t)$ *such that* $\vec{v}(x(t)) = t\vec{v}(x^1) + (1-t)\vec{v}(x^0)$

and $x(t) \in \operatorname{supp}(X)$ *for* $t \in [0,1]$. *Then*

$$V(x^1) - V(x^0)$$

and

$$D(\mathbb{E}[Y \mid X = x^1]) - D(\mathbb{E}[Y \mid X = x^0])$$

are identified. In particular,

$$V(x^1) - V(x^0) = \int_0^1 \mathbb{E}[Y \mid X = x(t)] \cdot (\vec{v}(x^1) - \vec{v}(x^0))dt.$$

This result shows that conditional means are sufficient for identification of average welfare. (Recall V may be interpreted as the average indirect utility function in many examples.) The integral representation of V is analogous to changes in consumer surplus in the standard consumer problem. It is also closely related to the revenue equivalence theorem (Riley and Samuelson (1981), Myerson (1981)).

Corollary 1.2. *Let Assumption 1.1 hold. Suppose* $\vec{v}$ *is identified and the set* $\vec{v}(\operatorname{supp}(X))$ *is convex. Then for every* $x^0, x^1 \in \operatorname{supp}(X)$,

$$V(x^1) - V(x^0)$$

and

$$D(\mathbb{E}[Y \mid X = x^1]) - D(\mathbb{E}[Y \mid X = x^0])$$

are identified.

Remark 1.7. These results only identify differences in D for points in $\{\mathbb{E}[Y \mid X = x]\}_{x \in \operatorname{supp}(X)}$. The set $\{\mathbb{E}[Y \mid X = x]\}_{x \in \operatorname{supp}(X)}$ need not be convex even under the assumptions of Corollary 1.2.

1.5.3 Direct Identification of Welfare Changes

The previous results in this section are of a two-step nature. They rely on the fact that $\vec{v}$ has first been identified, and then use this to identify or bound differences in D and welfare changes. In this section we provide an explicit one-step mapping from conditional means to welfare. For brevity, we provide an informal outline how to leverage our previous results.

For notational simplicitly we assume each good has one characteristic, i.e. $d_k = 1$. The following lemma follows from the integral representation in Theorem 1.4.

Proposition 1.3. *Assume $d_k = 1$ for each k. Let Assumptions 1.1 and 1.2 hold and assume V and $\vec{v}$ are differentiable. Let x be in the interior of* $\operatorname{supp}(X)$*. Then for each ℓ,*

$$\frac{\partial V(x)}{\partial x_\ell} = \mathbb{E}[Y_\ell \mid X = x] \frac{\partial v_\ell(x_\ell)}{\partial x_\ell}. \tag{1.22}$$

Proposition 1.2 provides conditions under which

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_k} = \frac{\partial v_\ell(x_\ell)}{\partial x_\ell} \Bigg/ \frac{\partial v_k(x_k)}{\partial x_k}.$$

Combining this with (1.22) we have for each ℓ that,

$$\frac{\partial V(x)}{\partial x_\ell} = \mathbb{E}[Y_\ell \mid X = x] \frac{\frac{\partial \mathbb{E}[Y_k | X = x]}{\partial x_\ell}}{\frac{\partial \mathbb{E}[Y_\ell | X = x]}{\partial x_k}} \frac{\partial v_k(x_k)}{\partial x_k}.$$

By integrating over x_ℓ (other characteristics fixed), this equality identifies differences in V up to the scale term $\frac{\partial v_k(x_k)}{\partial x_k}$. This scale term has the interpretation as a conversion rate between utils and characteristic x_k.

1.6 Counterfactual Bounds

While complete knowledge of $\vec{v}$ and D can answer many counterfactual questions, it may be unreasonable to assume these are both identified. We now use the structure of the model to provide counterfactual or out-of-sample bounds without the assumption D is identified. We are formally interested in the set

$$\operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k^0) + D(y), \tag{1.23}$$

where $x^0 \notin \text{supp}(X)$.[23] We provide bounds on a generic element of the set (1.23), denoted $\tilde{M}(x^0)$. We do not assume (1.23) is a singleton.

We formalize the sharp bounds on $\tilde{M}(x^0)$, assuming $\vec{v}$ is known at out-of-sample points. In practice, such knowledge typically requires extrapolation such as assuming that $\vec{v}$ belongs to a parametric class of functions, and must confront estimation error. The results in this section assume that (somehow) $\vec{v}$ is known and defined everywhere, and shows how to provide sharp bounds on $\tilde{M}(x^0)$. Importantly, we do *not* place assumptions on D other than $D \in \mathcal{D}$. This allows us to highlight that the shape restrictions of the model are capable of delivering non-trivial out-of-sample restrictions.

For a simple version of the bounds on $\tilde{M}(x^0)$, suppose $K = 1$. Bounds on $\tilde{M}(x^0)$ are given by

$$\begin{aligned} \sup_{x^\ell \in \text{supp}(X): \vec{v}(x^\ell) < \vec{v}(x^0)} \mathbb{E}[Y \mid X = x^\ell] &\leqslant \tilde{M}(x^0) \\ &\leqslant \inf_{x^u \in \text{supp}(X): \vec{v}(x^u) > \vec{v}(x^0)} \mathbb{E}[Y \mid X = x^u]. \end{aligned} \tag{1.24}$$

The following proposition characterizes the sharp bounds in the general setting ($K \geqslant 1$),

[23] We implicitly assume the argmax set is non-empty.

provided the only thing we know about D is $D \in \mathcal{D}$.

Theorem 1.5 (Sharp Out-of-Sample Bounds)**.** *Let Assumption 1.1 hold, assume $\vec{v}$ is known, and assume $D \in \mathcal{D}$. Let $x^0 \notin \operatorname{supp}(X)$ and assume*

$$\tilde{M}(x^0) \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k^0) + D(y).$$

Then the most that can be said about $\tilde{M}(x^0)$ is the following: $\tilde{M}(x^0) \in B$ and for every integer S, and sequence $x^1 \ldots, x^{S-1}$ of points in $\operatorname{supp}(X)$*,*

$$\begin{aligned}\tilde{M}(x^0)'(\vec{v}(x^0) - \vec{v}(x^{S-1})) \geqslant \mathbb{E}[Y \mid X = x^1]'\vec{v}(x^0) - \mathbb{E}[Y \mid X = x^{S-1}]\vec{v}(x^{S-1}) \\ - \sum_{s=1}^{S-2} (\mathbb{E}[Y \mid X = x^s] - \mathbb{E}[Y \mid X = x^{s+1}])'\vec{v}(x^s).\end{aligned} \tag{1.25}$$

When $\operatorname{supp}(X)$ is finite, the counterfactual restrictions on $\tilde{M}(x^0)$ are given by a finite set of linear inequalities. This is because each cycle constitutes a linear restriction and there are finitely many such cycles. The number of cycles grows quickly, however, and so directly operationalizing these inequalities can be computationally expensive.

Remark 1.8. While the focus of our analysis is on identification, this theorem has some practical implications. Suppose we have an estimate of $\vec{v}$ over its entire domain. Suppose we also have estimates of $\mathbb{E}[Y \mid X = x]$ for values of x in a user-chosen set. To apply Theorem 1.5, the estimates of $\vec{v}$ and $\mathbb{E}[Y \mid X = x]$ must be consistent with the restrictions of the model. Otherwise, restrictions like (1.25) would rule out *any* possible conjectured value of $\tilde{M}(x^0)$. This highlights a potential advantage of using shape-restricted estimation methods: if estimates of $\vec{v}$ and $\mathbb{E}[Y \mid X = x]$ are constrained to satisfy the restrictions of the model, then Theorem 1.5 can directly be used to provide counterfactual bounds.

1.7 Application

Nonparametric estimation enforcing all of the constraints of the model is computationally challenging and is left for future work. The fundamental problem is enforcing the constraint that the Jacobian of M is everywhere positive semidefinite, which is a continuum of constraints. We are pursuing a computationally feasible estimation approach in ongoing work.

One convenient feature for discrete choice is that a semiparametric model can be used for maximum-likelihood-based inference. Specifically, set

$$\mathbb{E}[Y \mid X = x] = P(\vec{v}(x), \lambda),$$

where Y is a vector of indicators, and P now replaces the notation M to highlight that conditional means are conditional probabilities. The index functions $\vec{v}$ are nonparametric, and λ is a finite-dimensional parameter that determines the mapping from the index to the conditional mean. By an appropriate choice of P, enforcing all of the constraints is straightforward. We will illustrate this with the paired combinatorial model (PCL) (Chu (1981), Chu (1989), Koppelman and Wen (2000)). We choose this model because it allows us to provide a parametric test of ARUM (Section 1.3.1) against the strictly more general perturbed stochastic choice model (Section 1.3.1). Violations of ARUM occur precisely when stochastic complementarity can occur.

1.7.1 Paired Combinatorial Logit

First we describe PCL. This model generalizes logit and allows pairs of alternatives to be in the same nest, with a parameter controlling substitution patterns *within* the nest. This model consists of an individual first choosing a pair according to a logit-like formula, and then second choosing an alternative within this pair according to another logit-like formula. Specifically, the model can be written

$$P_k(\bar{v},\lambda) = \sum_{j\neq k} P_{k|jk}(\bar{v},\lambda) P_{jk}(\bar{v},\lambda),$$

where

$$P_{k|jk}(\bar{v},\lambda) = \frac{e^{\bar{v}_k/\lambda_{jk}}}{e^{\bar{v}_j/\lambda_{jk}} + e^{\bar{v}_k/\lambda_{jk}}}$$

denotes the probability of choosing k conditional on choosing the nest jk, and

$$P_{jk}(\bar{v},\lambda) = \frac{\left(e^{\bar{v}_j/\lambda_{jk}} + e^{\bar{v}_k/\lambda_{jk}}\right)^{\lambda_{jk}}}{\sum_{\ell=1}^{K-1}\sum_{m=\ell+1}^{K}\left(e^{\bar{v}_\ell/\lambda_{\ell m}} + e^{\bar{v}_\ell/\lambda_{\ell m}}\right)^{\lambda_{\ell m}}}$$

denotes the probability of choosing nest jk. Combining these, choice probabilities are given by,

$$P_k(\bar{v},\lambda) = \frac{\sum_{j\neq k} e^{\bar{v}_k/\lambda_{jk}}\left(e^{\bar{v}_j/\lambda_{jk}} + e^{\bar{v}_k/\lambda_{jk}}\right)^{\lambda_{jk}-1}}{\sum_{\ell=1}^{K-1}\sum_{m=\ell+1}^{K}\left(e^{\bar{v}_\ell/\lambda_{\ell m}} + e^{\bar{v}_\ell/\lambda_{\ell m}}\right)^{\lambda_{\ell m}}}. \tag{1.26}$$

We see that this model reduces to logit when each of the nesting parameters λ_{jk} equals 1.

The model is perhaps best understood by the substitution patterns it allows. Higher values of λ_{jk} denote less substitutability between j and k. If $\lambda_{jk} > 1$ and there are at least 3 alternatives, this model allows complementarity

$$\frac{\partial P_k(\bar{v},\lambda)}{\partial \bar{v}_j} > 0$$

for certain values of $\bar{v}$. This is ruled out in ARUM, but is allowed in the perturbed utility model. The following result shows that this model is a strict generalization of ARUM if the parameter space is enlarged to allow $\lambda_{jk} > 1$. We show a similar result for nested logit in Allen and Rehbeck (2016a).

Proposition 1.4. *Let* $\vec{v}(\text{supp}(X)) = \mathbb{R}^k$*. Then PCL is consistent with ARUM if and only if* $0 < \lambda_{jk} \leqslant 1$ *for each pair* jk*. PCL is consistent with PUM if* $0 < \lambda_{jk}$*.*

The proposition is proven by showing that for fixed λ, $P(\bar{v}, \lambda)$ is the gradient of a convex function

The threshold $\lambda_{jk} \leqslant 1$ allows us to provide a simple parametric test of ARUM versus PUM. Formally, we can test

$$H_0 : 0 < \lambda_{jk} \leqslant 1 \text{ for each pair} \qquad H_a : \lambda_{jk} > 1 \text{ for some pair.}$$

1.7.2 Data Description

To test H_0 we use data from Louviere et al. (2013) on pizza choice. This is a stated-preference, panel dataset collected from an opt-in web survey. Each individual is randomized to a design, and faces either 16 or 32 decisions in that design.

An individual faces 5 alternatives for each decision. The alternatives are ordered left to right. Each alternatives specifies the values of the different characteristics (price, brand, number of toppings, etc.). We treat alternative k (or good k in our general setup) as the alternative in the k-th position.

In the survey, different individuals are shown different characteristics. For example, some individuals never see the delivery time. We include the characteristics that are available for all individuals: price ($12-$18 Australian dollars), number of toppings (up to 4), dummies for brand, whether the pizza comes with a free drink, and whether it comes with free dessert. These characteristics will shift the values of $v_k(x_k)$, which we will use to identify the nesting parameters λ.

1.7.3 Identification and Normalizations

Recall our nonparametric identification results require several ingredients. (i) First, each good needs a continuous regressor affecting its desirability. Here, all regressors we use are good-specific. (ii) Second, there needs to be sufficient complementarity/substitutability among the alternatives. Except in pathological cases, discrete choice models imply a rich amount of complementarity/substitution. We may thus invoke results in Appendix 1.16 for identification. We reiterate that the conditions in that appendix, while difficult to state, are weak. (iii) Finally, we need a location and scale normalization.

The location and scale normalization of (iii) can be weakened for the PCL relative to the normalizations used for our nonparametric results. This is because of our particular specificaton for $P(\bar{v},\lambda)$. We need to normalize the intercept of one alternative to be 0, and need some restriction on the nesting parameters λ_{jk} (Koppelman and Wen (2000)). We impose that these are 1 for all pairs *except* adjacent alternatives. This gives us the nesting parameters $\lambda_{12},\lambda_{23},\lambda_{34},\lambda_{45},\lambda_{51}$. Note we treat the first and last alternative as adjacent.

1.7.4 Test

We wish to test

$$H_0 : 0 < \lambda_{jk} \leqslant 1 \text{ for each adjacent pair} \qquad H_a : \lambda_{jk} > 1 \text{ for some pair.}$$

This is a parametric specification test of ARUM against PUM. For this test, $\vec{v}(x)$ is an unknown nuisance function. To get an intuition for how to construct a feasible test of H_0, suppose that we have

$$\sqrt{n}\left(\hat{\lambda}-\lambda\right) \xrightarrow{d} N(0,V).$$

Let $\hat{V}$ denote a consistent estimate for V and let $\hat{\sigma}_{jk}$ denote the diagonal component of $\hat{V}$ associated with $\hat{\lambda}_{jk}$. Using the normal approximation, we can construct a simple but conservative test by rejecting H_0 when

$$\max_{jk} \sqrt{n}\frac{\hat{\lambda}_{jk}-1}{\hat{\sigma}_{jk}} \geqslant \Phi^{-1}(1-\alpha/5), \tag{1.27}$$

where Φ denotes the standard normal cumulative distribution function. The 5 comes from the fact that H_0 involves 5 inequality restrictions for the adjacent pairs $\lambda_{12}, \lambda_{23}, \lambda_{34}, \lambda_{45}, \lambda_{51}$.

Approximate size control of this test follows from the following, which holds under H_0 and the normal approximation to $\hat{\lambda}$,

$$\begin{aligned}
P\left(\max_{jk} \sqrt{n}\frac{\hat{\lambda}_{jk}-1}{\hat{\sigma}_{jk}} \geqslant \Phi^{-1}(1-\alpha/5)\right) &\leqslant P\left(\max_{jk} \sqrt{n}\frac{\hat{\lambda}_{jk}-\lambda_{jk}}{\hat{\sigma}_{jk}} \geqslant \Phi^{-1}(1-\alpha/5)\right) \\
&\leqslant 5\max_{jk} P\left(\sqrt{n}\frac{\hat{\lambda}_{jk}-\lambda_{jk}}{\hat{\sigma}_{jk}} \geqslant \Phi^{-1}(1-\alpha/5)\right) \\
&\approx 5(\alpha/5) = \alpha.
\end{aligned}$$

The naive test of (1.27) is potentially conservative whenever these inequalities are strict. To address potential conservativeness of the test, an alternative test of H_0 can be constructed by drawing on ideas from the literature on moment inequalities. See for example Romano, Shaikh, and Wolf (2014). We omit the details because for our application the conservative test given by (1.27) handily rejects H_0.

1.7.5 Results

Table 1.1: Logit vs. Paired Combinatorial Logit

	Multinomial Logit	PCL
PRICE	-0.119	-0.140
	(0.0137)	(0.0163)
NUMTOP	0.347	0.414
	(0.0320)	(0.0417)
FREEDRK	0.348	0.442
	(0.0391)	(0.0543)
FREEDSRT	0.206	0.243
	(0.0360)	(0.0447)
PIZZAHUT	0.278	0.323
	(0.0866)	(0.107)
DOMINOS	0.284	0.321
	(0.0921)	(0.109)
EAGLEBOYS	0.00747	0.00756
	(0.0764)	(0.0917)
λ_{12}	1	2.47
		(0.300)
λ_{23}	1	1.20
		(0.370)
λ_{34}	1	1.85
		(0.283)
λ_{45}	1	.545
		(0.362)
λ_{51}	1	2.047
		(0.260)
Observations	4,928	4,928

Notes: Robust standard errors in parentheses, clustered at the individual level. Omitted brand is Pizza Haven.

We use (1.26) to estimate the model with maximum likelihood, treating observations as independent. To facilitate comparison with existing work, we specify linear index functions $v_k(x_k) = \beta' x_k$ with the same β for each alternative.[24] The vector x_k is a

[24] Given that we take a semiparametric approach, one could instead take $v_k(x_k)$ to be a linear combination of a tensor product of polynomials of the characteristics.

vector of characteristics for good k. It includes a constant, though the intercept on the k^{th} good is normalized to 0. Recall that our primary interest is on testing H_0, which only involves λ.

Table 1.1 reports the coefficient estimates obtained from PCL. We include multinomial logit estimates for comparability. The β coefficients are similar for the two models, especially when they are rescaled by dividing by β_{PRICE} to compute a measure of willingness to pay. We are most interested in the estimates of λ_{jk} for PCL. Recall our null hypothesis is that $\lambda_{jk} \leqslant 1$ for each pair. We can reject this if a single t statistic is high enough. Calculating the statistic $t_{12} = (2.47 - 1)/.3 = 4.9$, we reject the null hypothesis at conventional significance levels. (For $\alpha = .05$, the critical value from (1.27) is $\Phi^{-1}(1 - .05/5) = 2.33$.) All nesting parameters except $\hat{\lambda}_{45}$ are above 1, overall suggesting there is "not enough substitution" in the data to be consistent with ARUM.

The PCL is a member of the large family of generalized nested logit models (Wen and Koppelman (2001)). In Appendix 1.19, we show that this entire family of models is *sometimes* consistent with ARUM, but is always consistent with the perturbed utility model. This provides a large class of likelihood models that can readily be used. Indeed, our reading of the literature is that these models are often estimated *without* imposing parameter restrictions necessary for ARUM. The results of this paper – together with Allen and Rehbeck (2016b) and Allen and Rehbeck (2016a) – provide a theoretical foundation for use of these and other models when ARUM may not hold.[25]

[25] See also Fosgerau and Palma (2015) and Shi, Shum, and Song (2016) for parametric estimation of perturbed utility models. These papers do not use the full structure of ARUM.

1.8 Conclusion

This paper shows that perturbed utility models are identified under mild conditions using conditional means. We show this by leveraging the fact that agents in the model are optimizers. Optimization implies equality restrictions that allow us to obtain cardinal identification of the model.

Using our results, we show in several examples that some existing work using special regressors applies to more general setups. We do this by either showing $\vec{v}$ is nonparametricially identified without actually needing a special regressor structure,[26] or that an additively separable structure is sufficient for identification.[27] Relaxing the latter assumption means that the special regressor for good k does not need to enter the index v_k monotonically.

Our identification results apply to a model that strictly generalizes additive random utility models (ARUM) and allows a form of stochastic complementarity. As an illustrative example, we show that the paired combinatorial logit model can be used for a parametric test of ARUM against the perturbed utility model. We test this using data from Louviere et al. (2013) and find evidence against ARUM.

A detailed study of estimation in the general case is left for future work. There are both computational and econometric challenges to fully nonparametric estimation. Recall that under mild assumptions we have

$$\mathbb{E}[Y \mid X = x] = M(\vec{v}(x)),$$

where M has a Jacobian that is symmetric and positive semi-definite. We essentially

[26] As in ARUM with errors independent of characteristics.
[27] As in ARUM with conditional independence conditions as in Lewbel (2000).

only use symmetry for our identification results,[28] but to estimate conditional means that are consistent with the model, it is necessary to impose full semi-definiteness. For independent and identically distributed data $\{(Y^i, X^i)\}_{i=1}^n$, an intuitive approach is to construct a constrained least-squares estimate by solving the problem,

$$\min_{\hat{M}\in\mathcal{M}_n, \hat{\vec{v}}\in\mathcal{V}_n} \sum_{i=1}^{n} (Y^i - \hat{M}(\hat{\vec{v}}(X^i)))'(Y^i - \hat{M}(\hat{\vec{v}}(X^i)))$$
$$s.t. \qquad \nabla\hat{M}(u) \geq 0, \forall u \in \mathbb{R}^K,$$

where $\mathcal{M}_n$ and $\mathcal{V}_n$ are sets that grow to be dense in the parameter spaces for M and $\vec{v}$, respectively. The Jacobian of M is denoted ∇M and $\geq$ denotes the positive semi-definite order. The primary computational challenge is that the semi-definiteness constraint is actually a continuum of constraints. In ongoing work, we are pursuing a computationally feasible approach to enforcing this constraint. The econometric challenge to studying a constrained estimator is to develop a theory designed to reflect the finite sample impact of imposing a constraint.

[28] We also use that the diagonals of the Jacobian of M are weakly positive.

1.9 Acknowledgements

Chapter 1 is being prepared for submission for publication. Allen, Roy; Rehbeck, John. The dissertation author was a primary author and investigator of this paper.

1.10 Preliminaries in Convex Analysis

Definition 1.2 (Subdifferential)**.** *Let* $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ *be a convex function. The subdifferential of* f *at* z *is the set*

$$\partial f(z) = \{w \in \mathbb{R}^K : \forall \tilde{z} \in \mathbb{R}^K, f(\tilde{z}) - f(z) \geqslant w'(\tilde{z} - z)\}.$$

An element of $\partial f(z)$ is called a subgradient at z. ∂f is a multi-valued mapping called the subdifferential of f.

Definition 1.3 (Convex Conjugate)**.** *Let* f *be a function from* $\mathbb{R}^K$ *to* $[-\infty, \infty]$*. Then the convex conjugate of* f *is denoted*

$$f^*(w) = \sup_{z \in \mathbb{R}^K} \{z'w - f(z)\}.$$

The function f^* is convex (regardless of whether f is convex) as discussed in Rockafellar (1970), p. 104.

Lemma 1.5 (Rockafellar (1970), Theorem 23.5)**.** *Let* $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ *be a convex function such that* $f(z) < \infty$ *for some* z*. Then the following are equivalent,*

(i) $w^* \in \partial f(z^*)$.

(ii) $z'w^* - f(z)$ *attains its supremum in* z *at* $z = z^*$.

If in addition f is lower semi-continuous, then the following conditions are also equivalent to the ones above,

(i) $z^* \in \partial f^*(w^*)$.

(ii) $w'z^* - f^*(w)$ *attains its supremum in* w *at* $w = w^*$.

If we rewrite the first part of the lemma as $0 \in \partial f(z^*) - w^*$, we may recognize it as a sort of generalized first order condition.

Note that we do not assume f is convex in the following result.

Lemma 1.6. *Let* $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ *be a function such that* $f(z) < \infty$ *for some* z*. If*

$$z'w^* - f(z) \text{ attains its supremum in } z \text{ at } z = z^*,$$

then

$$z^* \in \partial f^*(w^*).$$

Proof. By assumption the supremum is attained, so

$$f^*(w^*) = z^{*'} w^* - f(z^*).$$

Because f^* is defined as a supremum, for arbitrary $\tilde{w}$,

$$f^*(\tilde{w}) \geqslant z^{*'} \tilde{w} - f(z^*).$$

Thus,

$$f^*(\tilde{w}) - f^*(w^*) \geqslant z^{*'}(\tilde{w} - w^*),$$

so $z^* \in \partial f^*(w^*)$. □

Lemma 1.7 (Rockafellar (1970), Theorem 25.1). *Let $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ be convex and assume $f(z^*)$ is finite. Then f is differentiable at z^* if and only if $\partial f(z^*)$ is a singleton.*

1.11 Appendix: Proofs for Section 1.3

We make use of the following result, which is proven in Allen and Rehbeck (2016a). The lemmas for our examples are immediate corollaries.

Theorem 1.6 (Allen and Rehbeck (2016a)). *Let*

$$Y \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(X_k) + \tilde{D}(y, \varepsilon),$$

where ε is possibly infinite-dimensional. Suppose ε is independent of X, Y is measurable, and $\mathbb{E}[\tilde{D}(Y,\varepsilon) \mid X = x]$ is finite for each $x \in \operatorname{supp}(X)$.[29] *Then*

$$\mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in conv(B)} \sum_{k=1}^{K} y_k v_k(x_k) + D(y)$$

for $D(y) = \sup_{\tilde{Y} \in \mathcal{Y} : \mathbb{E}[\tilde{Y}(\varepsilon)] = y} \mathbb{E}\left[\tilde{D}(\tilde{Y}(\varepsilon), \varepsilon)\right]$,[30] *where $\mathcal{Y}$ is the set of measurable functions from the support of ε to B, and $conv(B)$ denotes the convex hull of B.*[31]

This result may be useful for identification of some other applications so we discuss its requirements. The key requirement is that unobservables enter the latent utility

[29] Allen and Rehbeck (2016a) provide sufficient conditions for existence of a measurable selector using results from Stinchcombe and White (1992). These results are mild and allow $\tilde{D}$ to take on value ∞, which allows "random budgets" that are proper subsets of B.

[30] D is taken to be $-\infty$ if there is no $Y \in \mathcal{Y}$ such that $\mathbb{E}[Y(\varepsilon)] = y$. D is always finite when evaluated at $\mathbb{E}[Y \mid X = x]$.

[31] This is a trivial extension of results in Allen and Rehbeck (2016a). There, we assume B is convex to ensure that a measurable selector exists. Here, we take measurability of Y as a high-level condition. The theorem does not restrict $\tilde{D}$ to be finite, so the budget B can be absorbed into $\tilde{D}$; thus, setting $B = \mathbb{R}^K$ is without loss of generality. We explicitly maintain the budget for ease of understanding, noting that when we do this we have to convexify B in the statement of the theorem.

function *separably* from the characteristics. This rules out random coefficients models. Moreover, ε must be independent of characteristics. Because $\tilde{D}$ can be $-\infty$, the theorem allows random budget sets. Specifically,

$$B(\varepsilon) = \{y \in B \mid \tilde{D}(y, \varepsilon) > -\infty\}$$

can depend on ε. These budget sets just need to be independent of characteristics.

Our results establish identification of $\vec{v}$ and the aggregated function D. Our results do *not* cover identification of $\tilde{D}$ or the distribution of ε. It may be possible to use the characterization $D(y) = \sup_{\tilde{Y} \in \mathcal{Y}:\mathbb{E}[\tilde{Y}(\varepsilon)]=y} \mathbb{E}\left[\tilde{D}(\tilde{Y}(\varepsilon), \varepsilon)\right]$ to identify these latent objects, though we do not pursue this.

Proof of Proposition 1.1. Let

$$D_B(y) = \begin{cases} D(y) \text{ if } y \in B \\ -\infty \text{ otherwise} \end{cases}.$$

By Lemma 1.6, $\mathbb{E}[Y \mid X = x] \in \partial(-D_B)^*(\vec{v}(x))$, where $(-D_B)^*$ is the convex conjugate of $-D_B$. By Assumption 1.2 and Lemma 1.7, $\mathbb{E}[Y \mid X = x] \in \partial(-D_B)^*(\vec{v}(x))$, is a singleton. Thus when $J(\overline{v})$ exists, it is the Hessian of a twice continuously differentiable convex function, and so the result follows from Rockafellar (1970), Theorem 4.5. □

1.12 Appendix: Proofs for Section 1.4

1.12.1 Proofs for Section 1.4.1

Proof of Proposition 1.2. The proof is a direct extension of the arguments in the text. Under the assumptions of Definition 1.1, there is some $x \in \text{supp}(X)$ such that

$$\begin{aligned} \left.\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_{\ell,p}}\right|_{x=x^*} &= \left.\frac{\partial M_k(\overline{v})}{\partial \overline{v}_\ell}\right|_{\overline{v}=\vec{v}(x^*)} \left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^*} \\ \left.\frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_{k,q}}\right|_{x=x^*} &= \left.\frac{\partial M_\ell(\overline{v})}{\partial \overline{v}_k}\right|_{\overline{v}=\vec{v}(x^*)} \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}. \end{aligned} \tag{1.28}$$

Recall these equalities hold at the specified *points*. Under condition (iv) of Definition 1.1, we may take the ratio of the equations in (1.28) to yield the result. □

Proof of Theorem 1.1. This is implied by Corollary 1.3 so we provide only a brief discussion.

Using Proposition 1.2, the conditions of the theorem provide constructive identification of

$$\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=\tilde{x}_\ell} \Bigg/ \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=\tilde{x}_k} \tag{1.29}$$

whenever $\ell \neq k$. Thus we only need to deal with $\ell = k$. Because $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$, we can always find a "path" between equations as in the discussion leading to (1.33). Thus, (1.29) is identified without restrictions.

Corollary 1.3 shows the sign of $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$ is identified. Using the scale and location normalization, $\vec{v}$ is identified by the mean value theorem. Constructive identification can also be established by the fundamental theorem of calculus (recall Lemma 1.4).

We can relax the assumption $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$. If we assume $\vec{v}(\text{supp}(X))$ is convex with nonempty interior,[32] then the preceeding arguments identify $\vec{v}$ at characteristic values in the interior of $\vec{v}(\text{supp}(X))$. The basic idea is that derivative ratios such as (1.29) can be identified for a sufficiently rich set of inputs. Since $\vec{v}(\text{supp}(X))$ is convex, we can use the mean value theorem to identify $\vec{v}(x)$ if it is in the interior of $\vec{v}(\text{supp}(X))$. Identification of arbitrary $\vec{v}(x)$ then follows by continuity. □

Proof of Corollary 1.1. Because ε has a density that is continuously differentiable, M is continuously differentiable. Because ε has a positive density, M has strictly negative cross-partial derivatives. All of the other conditions of Theorem 1.1 hold. □

1.12.2 Proofs for Section 1.4.2

Proof of Theorem 1.2. By Assumption 1.3(i) there exists $(\tilde{z}, \tilde{w}) \in \text{supp}(Z, W)$ such that

$$\mathbb{E}[Y \mid Z = z, W = w] = \mathbb{E}[Y \mid Z = \tilde{z}, W = \tilde{w}].$$

Since D_B is differentiable at $\mathbb{E}[Y \mid Z = z, W = w]$, Assumption 1.2 and Proposition 1.5 establish that

$$\mathbb{E}[Y \mid Z = z, W = w] = \mathbb{E}[Y \mid Z = \tilde{z}, W = \tilde{w}] \iff \vec{v}(z, w) = \vec{v}(\tilde{z}, \tilde{w}).$$

Assumption 1.3(ii) completes the proof since $\vec{v}(\tilde{z}, \tilde{w})$ is identified. □

[32]Our assumptions imply that $\vec{v}(\text{supp}(X))$ is a Cartesian product of connected sets, each in $\mathbb{R}$. This implies that $\vec{v}(\text{supp}(X))$ is convex.

1.13 Appendix: Proofs for Section 1.5

Proof of Theorem 1.3. Kos and Messner (2013) prove a result related to this in their study of incentive compatible transfers in a mechanism design setting.

Fix $x^* \in \text{supp}(X)$. Define $\tilde{D}$ over $y \in \{\mathbb{E}[Y \mid X = x\}_{x \in \text{supp}(X)}$ by

$$\tilde{D}(y) = \inf\{(y - \mathbb{E}[Y \mid X = x^S])'\vec{v}(x^S) + \cdots + (\mathbb{E}[Y \mid X = x^1] - \mathbb{E}[Y \mid X = x^*])'\vec{v}(x^*)\},$$

where the infimum is taken over finite sequences such that $x^s \in \text{supp}(X)$ for each s. For $y \notin \{\mathbb{E}[Y \mid X = x]\}_{x \in \text{supp}(X)}$, set $\tilde{D}(y) = -\infty$. This function satisfies

$$\tilde{D}(\mathbb{E}[Y \mid X = x^*]) \leqslant 0$$

because we can take a sequence with all terms equal to $\mathbb{E}[Y \mid X = x^*]$. Theorem 1.8(iv) establishes the opposite inequality, so we have

$$\tilde{D}(\mathbb{E}[Y \mid X = x^*]) = 0.$$

By the construction of $\overline{D}$, we have

$$\tilde{D}(\mathbb{E}[Y \mid X = \tilde{x}]) = \overline{D}(x^*, \tilde{x})$$

for every $\tilde{x} \in \text{supp}(X)$ (recall x^* is fixed). Thus,

$$\tilde{D}(\mathbb{E}[Y \mid X = \tilde{x}]) - \tilde{D}(\mathbb{E}[Y \mid X = x^*]) = \overline{\Delta}D(x^*, \tilde{x})$$

The proof of Rockafellar (1970), Theorem 24.8 establishes that $\tilde{D}$ satisfies

$$\forall x \in \operatorname{supp}(X), \mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k) + \tilde{D}(y).$$

The analogous statement for $\underline{\Delta} D(x^*, \tilde{x})$ follows from similar arguments. □

Proof of Theorem 1.4. Step 1: Change of variables.

Define

$$\tilde{V}(\bar{v}) = \sup_{y \in B} \left\{ \sum_{k=1}^{K} y_k \bar{v}_k + D(y) \right\},$$

where we recognize $\tilde{V}(\bar{v}) = V(x)$ when $\bar{v} = \vec{v}(x)$. To state a version of the envelope theorem with this change of variables, define

$$\partial \tilde{V}(\bar{v}) = \left\{ u \in \mathbb{R}^k \mid \forall \tilde{v}, \tilde{V}(\tilde{v}) - \tilde{V}(\bar{v}) \geqslant u \cdot (\tilde{v} - \bar{v}) \right\}.$$

For each $x \in \operatorname{supp}(X)$ such that $\vec{v}(x) = \bar{v}$, we have

$$\mathbb{E}[Y \mid X = x] \in \partial \tilde{V}(\bar{v}). \tag{1.30}$$

This follows from Lemma 1.5 and the fact that $\mathbb{E}[Y \mid X = x]$ is a maximizer.

Step 2: Convert to single variable problem and invoke Rockafellar (1970), Corollary 24.2.1.

For $t \in [0, 1]$, let

$$h(t) = \tilde{V}\left(t\vec{v}(x^1) + (1-t)\vec{v}(x^0)\right).$$

The function $\tilde{V}$ is convex and $s(t) = t\vec{v}(x^1) + (1-t)\vec{v}(x^0)$ is affine so h is convex. The function h is thus directionally differentiable. Let its left derivative be denoted $h'_-(t)$ and let $h'_+(t)$ denote the right derivative of h. The directional derivative of $\tilde{V}$ at y in direction z is denoted $\tilde{V}'(y;z)$; see Rockafellar (1970) for the formal definition. From (1.30) we have,

$$\mathbb{E}[Y \mid X = x(t)] \in \partial\tilde{V}(t\vec{v}(x^1) + (1-t)\vec{v}(x^0)).$$

Combining this with Rockafellar (1970), Theorem 23.2 we have

$$h'_-(t) = \tilde{V}'(t\vec{v}(x^1) + (1-t)\vec{v}(x^0); -(\vec{v}(x^1) - \vec{v}(x^0))) \leqslant \mathbb{E}[Y \mid X = x(t)] \cdot (\vec{v}(x^1) - \vec{v}(x^0))$$
$$h'_+(t) = \tilde{V}'(t\vec{v}(x^1) + (1-t)\vec{v}(x^0); \vec{v}(x^1) - \vec{v}(x^0)) \geqslant \mathbb{E}[Y \mid X = x(t)] \cdot (\vec{v}(x^1) - \vec{v}(x^0)).$$

From Rockafellar (1970), Corollary 24.2.1 we obtain

$$\tilde{V}(\vec{v}(x^1)) - \tilde{V}(\vec{v}(x^0)) = \int_0^1 \mathbb{E}[Y \mid X = x(t)] \cdot (\vec{v}(x^1) - \vec{v}(x^0))dt.$$ [33]

Since $\vec{v}$ is known,

$$V(x^1) - V(x^0) = \tilde{V}(\vec{v}(x^1)) - \tilde{V}(\vec{v}(x^0))$$

is identified.

□

1.14 Appendix: Proofs for Section 1.6

Proof of Theorem 1.5. Obviously, $\tilde{M}(x^0) \in B$ by use of the *a priori* knowledge of the budget.

[33] Rockafellar (1970), Corollary 24.2.1 establishes that the Riemann integrals of h'_- and h'_+ from 0 to 1 exist and are equivalent. Riemann integrability of $\mathbb{E}[Y \mid X = x(t)] \cdot (\vec{v}(x^1) - \vec{v}(x^0))$ from 0 to 1 then follows from a sandwiching argument.

To prove the rest, let $\tilde{M}(x^0)$ be a conjectured value and let $\{\mathbb{E}[Y \mid X = x]\}_{x \in \text{supp}(X)} \cup \tilde{M}(x^0)$ be the original values of the conditional mean augmented with this conjectured value. The conjectured value is consistent with the model if and only if the restrictions of Theorem 1.8(iv) hold. By rearranging the cyclic monotonicity inequalities, we obtain (1.25). □

1.15 Appendix: Proofs for Section 1.7

Proof of Proposition 1.4. The first part of the lemma is known. It remains to show PCL is consistent with the perturbed utility model.

The PCL probabilities satisfy

$$P(\overline{v}, \lambda) = \nabla \ln \left(\sum_{j \neq k} (e^{\overline{v}_j/\lambda_{jk}} + e^{\overline{v}_k/\lambda_{jk}})^{\lambda_{jk}} \right).^{34}$$

By the arguments in Appendix 1.19, the function

$$\ln \left(\sum_{j \neq k} (e^{\overline{v}_j/\lambda_{jk}} + e^{\overline{v}_k/\lambda_{jk}})^{\lambda_{jk}} \right)$$

is convex in $\overline{v}$, completing the proof. □

To see that PCL allows complementarity, we use the fact that the sign of

$$\frac{\partial P_k(\overline{v}, \lambda)}{\partial \overline{v}_j}$$

[34] See e.g. Koppelman and Wen (2000).

equals that of

$$-\left[P_k + \frac{\left(\frac{1}{\lambda_{jk}-1}\right)(P_j+P_k)(P_{j|jk})(P_{k|jk})}{P_j}\right].^{35} \tag{1.31}$$

We suppress dependence of each P term on $\bar{v}$ and λ for simplicitly. We show that (1.31) can be positive, indicating complementarity, when $\lambda_{jk} > 1$ and there are at least 3 alternatives. To that end, let $v_\ell = 0$ for ℓ except j and k. Then set $v_j = v_k$. This implies $P_{j|jk} = .5$, so the sign of (1.31) is determined by the sign of

$$-P_kP_j + \frac{1}{4}\left(\frac{1}{\lambda_{jk}-1}\right)(P_j+P_k). \tag{1.32}$$

By setting $v_j = v_k$ to sufficiently small values, (1.32) becomes positive.

1.16 Appendix: Identification of $\vec{v}$ for "Nonstandard" Cases

As discussed previously, Theorem 1.1 rules out some examples of interest. We now provide weaker conditions under which $\vec{v}$ is identified. Instead of assuming cross-partials of M are everywhere nonzero, we assume cross-partials are nonzero at a "rich" set of points.

The basic idea is that if we identify many ratios of partial derivatives of $\vec{v}$, then we can identify $\vec{v}$ itself by two different approaches. The first, which is feasible given the assumptions of Theorem 1.1, allows us to integrate these derivatives and obtain constructive identification; recall Lemma 1.4. The second approach, taken in this section, is to use the mean value theorem to obtain non-constructive results. Recall

[35] See Koppelman and Wen (2000).

that by the mean value theorem, a differentiable function is uniquely determined by its partial derivatives. Thus, we only need to identify ratios of all partial derivatives of $\vec{v}$. We describe how to do this by multiplying ratios that are directly identified by Proposition 1.2.

For an example, suppose

$$\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^1} \Bigg/ \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^1} \quad \text{and} \quad \left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j=x_j^2} \Bigg/ \left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^2}$$

are identified from data (via Proposition 1.2). If $x_\ell^1 = x_\ell^2$, we can multiply these derivative ratios to identify

$$\left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j=x_j^2} \Bigg/ \left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^1}. \tag{1.33}$$

Note that this derivative ratio can be identified *even if* $j = k$ and $x_j^2 \neq x_\ell^2$. This shows that if ratios of derivatives are identified and overlap in a specific sense, then we can multiply these ratios to identify new ratios. We need to handle sequences of derivative ratios of arbitrary finite length, so we introduce some more notation.

Definition 1.4 (Paths)**.** *There is a path from the point* $a := \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$ *to* $b := \left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j=\tilde{x}_j}$ *if a is nonzero and there is a sequence of partial derivatives beginning at a and ending at b such that each adjacent element is paired, and these pairs are strict except possibly between the final two elements of the sequence.*

In order for there to be a path between partial derivatives, several conditions must hold. We must have $K \geqslant 2$ or $x_k = \tilde{x}_j$. The function M must be continuously differentiable over (at least) a finite set of points. Importantly, it is *not* necessary that all cross-partials be nonzero or that M be continuously differentiable everywhere

Theorem 1.7. *Let Assumptions 1.1 and 1.2 hold and assume* $x_{k,q}$ *and* $x_{j,r}$ *are regres-*

sors specific to k and j, respectively. If there is a path from the point $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$ *to* $\left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j=\tilde{x}_j}$*, then*

$$\left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j=\tilde{x}_j} \Bigg/ \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*} \tag{1.34}$$

is identified.

Proof of Theorem 1.7. Let $\left.\frac{\partial v_{\ell_1}(x_{\ell_1})}{\partial x_{\ell_1,p_1}}\right|_{x_{\ell_1}=x_{\ell_1}^1}, \ldots, \left.\frac{\partial v_{\ell_M}(x_{\ell_M})}{\partial x_{\ell_M,p_M}}\right|_{x_{\ell_M}=x_{\ell_M}^M}$ be a finite sequence as in Definition 1.4. For each $m = 2, \ldots, M$, let

$$S_{m-1,m} = \left.\frac{\partial v_{\ell_m}(x_{\ell_m})}{\partial x_{\ell_m,p_m}}\right|_{x_{\ell_m}=x_{\ell_m}^m} \Bigg/ \left.\frac{\partial v_{\ell_{m-1}}(x_{\ell_{m-1}})}{\partial x_{\ell_{m-1},p_{m-1}}}\right|_{x_{\ell_{m-1}}=x_{\ell_{m-1}}^{m-1}}. \tag{1.35}$$

This ratio is identified due to Proposition 1.2. This follows because for $m < M$, the numerator and denominator are strictly paired. For $m = M$, the numerator and denominator are paired and the denominator is nonzero.

By construction,

$$\prod_{m=1}^{M} S_{m-1,m} = \left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j=\tilde{x}_j} \Bigg/ \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$$

since all intermediate terms cancel out. This is valid because there is never any division by zero. □

The following corollary relaxes assumptions in Theorem 1.1.

Corollary 1.3. *Let Assumptions 1.1 and 1.2 hold and assume all regressors are good-specific. Assume there is a tuple* (k, q, x_k^*) *such that* $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}$ *has a path to* $\left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j=\tilde{x}_j}$ *for any j, r, and* $\tilde{x}_j \in \mathbb{R}^{d_j}$*. Then* $\vec{v}$ *is identified under the following normalization:*

i. (Scale) $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*} \in \{-1, 1\}$.

ii. (Location) $v_\ell(0_{d_k}) = 0$ *for each* $\ell = 1, \ldots, K$*, where* 0_{d_k} *denotes a* d_k*-dimensional vector of zeros.*

Proof of Corollary 1.3. First we identify the sign of $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k = x_k^*}$. Under Assumptions 1.1 and 1.2, it can be shown that

$$\mathbb{E}[Y \mid X = x] \neq \mathbb{E}[Y \mid X = \tilde{x}] \iff (\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'(\vec{v}(x) - \vec{v}(\tilde{x})) > 0. \tag{1.36}$$

(This is a straightforward extension of Lemma 1.8.) From the assumptions of the corollary, there is some $x^* \in \text{supp}(X)$ that has x_k^* as its k-th row. Moreover, there must be some ℓ such that

$$\left.\frac{\partial M_\ell(\bar{v})}{\partial \bar{v}_k}\right|_{\bar{v} = \vec{v}(x^*)} \neq 0$$

This follows from the definition of a path. Since $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k = x_k^*} \neq 0$, we see that for sufficiently small changes in $x_{k,q}$ there must be a change in $\mathbb{E}[Y_\ell \mid X = x]$. From (1.36) this implies that there must be a change in $\mathbb{E}[Y_k \mid X = x]$ as well. Again using (1.36), we determine the sign of $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k = x_k^*}$ depending on whether $\mathbb{E}[Y_k \mid X = x]$ is locally increasing or decreasing with respect to $x_{k,q}$.

Normalizing $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k = x_k^*}$ to 1 or -1 depending on its sign, we identify all partial derivatives using Theorem 1.7. Recall that from the mean value theorem, two functions that share partial derivatives can differ by at most an additive constant. Given the location normalization, $\vec{v}$ is identified. □

We provide an example where the path condition holds even though M may lie on the boundary of B and may not even be continuous. This illustrates how identification in Corollary 1.3 is established using restrictions on pairs of partial derivatives. This is in contrast with Theorem 1.1, which places global restrictions that rule out boundary behavior and discontinuities as in the following example.

Remark 1.9 (Boundaries and Discontinuities). For illustration of how we can handle boundary issues, let B be the probability simplex, $K > 2$, and let D be given by,

$$D(y) = \begin{cases} -\sum_{k=1}^{K} p_k \ln p_k & \text{if } p_k \neq 0 \text{ for at most 2 distinct } k \\ -\infty & \text{otherwise.} \end{cases}$$

We set $0 \ln 0$ to 0. For simplicity suppose x_k is scalar. This choice of D ensures exactly 2 goods will be chosen with positive probability, and requires that they be the ones with the highest values of the indices $v_k(x_k)$ (assuming the two highest values are unique). Suppose the second highest value of $v_k(x_k)$ is unique and let $k(1)$ and $k(2)$ attain the highest values of v_k.[36] Then

$$\mathbb{E}[Y_{k(1)} \mid X = x] = \frac{e^{v_{k(1)}(x_{k(1)})}}{e^{v_{k(1)}(x_{k(1)})} + e^{v_k(2)(x_{k(2)})}}, \mathbb{E}[Y_{k(2)} \mid X = x] = \frac{e^{v_{k(2)}(x_{k(2)})}}{e^{v_{k(1)}(x_{k(1)})} + e^{v_{k(2)}(x_{k(2)})}}.$$

If $\vec{v}$ is differentiable, sufficient conditions for Corollary 1.3 are fairly mild. One sufficient condition is that X has full support, all partial derivatives of $\vec{v}$ are everywhere nonzero, and $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$. Note that while we may write $\mathbb{E}[Y \mid X = x] = M(\vec{x})$, M is not differentiable everywhere. In fact, it is not even continuous everywhere.

1.17 Appendix: Partial Identification of $\vec{v}$

Our sufficient conditions for identification of $\vec{v}$ may fail. Our conditions do not apply if all covariates are discrete, sufficient substitution/complementarity does not exist, or if $\mathbb{E}[Y \mid X = x]$ is not suitably differentiable. We provide a complete characterization

[36]These implicitly depend on x.

of the identifying power of the model for $\vec{v}$, assuming only that $D \in \mathcal{D}$. The results in this appendix are the discrete analogues of positive semi-definiteness and symmetry (Proposition 1.1).

Characterizing the identifying power of the model is beneficial even when $\vec{v}$ is identified. Section 1.4 attempts to answer the condition "Under what conditions is $\vec{v}$ identified?" To answer this we used symmetry. The results in this appendix use conditions more closely related to monotonicity. These conditions characterize the identifying power of the model and may provide a more transparent answer to the question "What variation in the data identifies $\vec{v}$?"

Now, we allow the possibility that there are multiple values of $\vec{v}$ that are consistent with the restrictions of the model. The set of such values is called the *identified set* for $\vec{v}$ and is denoted

$$\mathcal{V}_{ID} = \left\{ \vec{v} \in \mathcal{V} \mid \exists D \in \mathcal{D} \ s.t. \ \forall x \in \operatorname{supp}(X), \mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{B} \sum_{k=1}^{K} y_k v_k(x_k) + D(y) \right\}.$$

The set $\mathcal{V}$ is the parameter space for $\vec{v}$. We assume $\mathcal{V}$ consists of real-valued functions. It could be further restricted. For example, it could be a parametric class of functions. We are again agnostic about the function D. As previously, $\mathcal{D}$ is the set of functions $D : \mathbb{R}^K \to \mathbb{R} \cup \{-\infty\}$ that are finite at some point $y \in B$.

The following lemma is a convenient restatement of restrictions of the optimizing model.

Lemma 1.8. *If $\vec{v} \in \mathcal{V}_{ID}$, then there is some $D \in \mathcal{D}$ such that for every $x, \tilde{x} \in \operatorname{supp}(X)$,*

$$\begin{aligned}(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'\vec{v}(x) &\geqslant D(\mathbb{E}[Y \mid X = \tilde{x}]) - D(\mathbb{E}[Y \mid X = x]) \\ &\geqslant (\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'\vec{v}(\tilde{x}).\end{aligned}$$

Moreover, $D(\mathbb{E}[Y \mid X = \tilde{x}])$ *and* $D(\mathbb{E}[Y \mid X = x])$ *are finite.*

Proof. We use necessary conditions for optimality. If $\vec{v} \in \mathcal{V}_{ID}$, then for some $D \in \mathcal{D}$ we must have,

$$\mathbb{E}[Y \mid X = x]'\vec{v}(x) + D(\mathbb{E}[Y \mid X = x]) \geqslant \mathbb{E}[Y \mid X = \tilde{x}]'\vec{v}(x) + D(\mathbb{E}[Y \mid X = \tilde{x}])$$
$$\mathbb{E}[Y \mid X = \tilde{x}]'\vec{v}(\tilde{x}) + D(\mathbb{E}[Y \mid X = \tilde{x}]) \geqslant \mathbb{E}[Y \mid X = x]'\vec{v}(\tilde{x}) + D(\mathbb{E}[Y \mid X = x]).$$

Since $D \in \mathcal{D}$, it is finite at the referenced points because of optimality. The inequalities of the lemma follow from rearranging these inequalities. □

One feature captured in Lemma 1.8 is the *monotonicity* condition

$$(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'(\vec{v}(x) - \vec{v}(\tilde{x})) \geqslant 0. \quad (1.37)$$

This resembles the law of compensated demand if we relate $\mathbb{E}[Y \mid X = x]$ to Hicksian demand and $-\vec{v}(x)$ to the price vector. For further illustration of (1.37), suppose that $\vec{v}(x)$ and $\vec{v}(\tilde{x})$ only differ with respect to their first component. Then (1.37) becomes

$$(\mathbb{E}[Y_1 \mid X = x] - \mathbb{E}[Y_1 \mid X = \tilde{x}])(v_1(x_1) - v_1(\tilde{x}_1)) \geqslant 0,$$

which states that the conditional expectation of Y_1 is weakly increasing in v_1.

We now use Lemma 1.8 to remove the nuisance function D. To that end let $x^0, \ldots, x^{M-1}, x^M = x^0$ be a cycle of points in $\text{supp}(X)$. By repeated application of

Lemma 1.8 we obtain,

$$\begin{aligned}\sum_{m=0}^{M-1} (\mathbb{E}[Y \mid X = x^m] - \mathbb{E}[Y \mid X = x^{m+1}])'\vec{v}(x^m) \\ \geqslant \sum_{m=0}^{M-1} D(\mathbb{E}[Y \mid X = x^{m+1}]) - D(\mathbb{E}[Y \mid X = x^m]) \\ = 0.\end{aligned} \tag{1.38}$$

By summing up over a cycle, we "sum out" the unknown function D. An alternative way to state the inequalities obtained in this way is as follows. Suppose that $\{x^m\}_{m=0}^{M-1} \subseteq \text{supp}(X)$. Then for every permutation π we have

$$\sum_{m=0}^{M-1} \mathbb{E}[Y \mid X = x^m]'\vec{v}(x^m) \geqslant \sum_{m=0}^{M-1} \mathbb{E}[Y \mid X = x^{\pi(m)}]'\vec{v}(x^{\pi(m)}).$$

This inequality highlights the connection to optimizing behavior. Intuitively, no permutation can improve the "match" between choices (= conditional expectations) and payoffs (= marginal utility shifters). We now show that inequalities such as (1.38) capture the complete restrictions of the model for $\vec{v}$.

Theorem 1.8 (Sharp Characterization of $\mathcal{V}_{ID}$)**.** *Let $\vec{v} \in \mathcal{V}$. The following are equivalent:*

i. $\vec{v} \in \mathcal{V}_{ID}$, i.e. there is a function $D \in \mathcal{D}$ such that

$$\forall x \in \text{supp}(X), \mathbb{E}[Y \mid X = x] \in \underset{y \in B}{\text{argmax}} \sum_{k=1}^{K} y_k v_k(x_k) + D(y).$$

ii. There is a concave function $D \in \mathcal{D}$ such that

$$\forall x \in \text{supp}(X), \mathbb{E}[Y \mid X = x] \in \underset{y \in B}{\text{argmax}} \sum_{k=1}^{K} y_k v_k(x_k) + D(y).$$

iii. There exist finite numbers $\{D_x\}_{x\in\text{supp}(X)}$ *such that for every* $x,\tilde{x}\in\text{supp}(X)$*,*

$$(\mathbb{E}[Y\mid X=x]-\mathbb{E}[Y\mid X=\tilde{x}])'\vec{v}(x)\geqslant D_{\tilde{x}}-D_x.$$

iv. For every integer M and cycle of points $x^0,\ldots,x^{M-1},x^M=x^0$ *each in* $\text{supp}(X)$*,*

$$\sum_{m=0}^{M-1}(\mathbb{E}[Y\mid X=x^m]-\mathbb{E}[Y\mid X=x^{m+1}])'\vec{v}(x^m)\geqslant 0.$$

Proof. We shall show (i) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (ii) $\Longrightarrow$ (i).

By relating D_x with $D(\mathbb{E}[Y\mid X=x])$, the previous discussion shows (i) $\Longrightarrow$ (iii) $\Longrightarrow$ (iv). Note that while D may take on value $-\infty$ over some points, $D(\mathbb{E}[Y\mid X=x])>-\infty$ for each $x\in\text{supp}(X)$. This is because $D\in\mathcal{D}$, $\vec{v}(x)$ is finite, and $\mathbb{E}[Y\mid X=x]$ a maximizer. This is why the numbers in (iii) are finite.

The implication (iv) $\Longrightarrow$ (ii) follows from Rockafellar (1970), Theorem 24.8, so we provide only a sketch of this implication. Let $\Gamma=\mathbb{R}^K\times\mathbb{R}^K$. Let $S=\{(\mathbb{E}[Y\mid X=x],\vec{v}(x))\}_{x\in\text{supp}(X)}$, so we have $S\subseteq\Gamma$. The set S is contained in the graph of a cyclically monotone multi-valued mapping (see Rockafellar (1970), which generalizes Definition 1.5). By the constructive extension result of Rockafellar (1970), Theorem 24.8, we have $\vec{v}(x)\in\partial f(\mathbb{E}[Y\mid X=x])$, where f is a convex function that never attains $-\infty$ and that is finite at some point. By Lemma 1.5 and the fact that $\mathbb{E}[Y\mid X=x]\in B$ for $x\in\text{supp}(X)$, we have

$$\mathbb{E}[Y\mid X=x]'\vec{v}(x)-f(\mathbb{E}[Y\mid X=x])=\sup_{y\in B}\left\{y'\vec{v}(x)-f(y)\right\}.$$

By letting $D=-f$, we have (ii).

Obviously, (ii) $\implies$ (i). $\square$

This result is closely related to results in Brown and Calsamiglia (2007) and Chambers and Echenique (2009a).[37] Related results that simultaneously vary budgets are established in McFadden and Fosgerau (2012a).

The fact that (i) and (ii) are equivalent means that if we assume D is concave, we obtain no additional identifying power for $\vec{v}$. Moreover, it is not possible to separately test whether D is concave aside from testing the entire model. This insight is fairly well-known in other settings (Afriat (1967), Varian (1982)).

Part (iii) is helpful for computational reasons such as checking whether a particular point is in the identified set. Note that we need not worry about forcing D_x and $D_{\tilde{x}}$ to agree whenever $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y \mid X = \tilde{x}]$, since (iii) implies $D_x = D_{\tilde{x}}$ by double inequalities.

If the parameter space $\mathcal{V}$ contains constant functions, these functions will always be in $\mathcal{V}_{ID}$. This can easily be seen from (iv). We refer to (iv) as the *cyclic monotonicity* inequalities in light of the following definition.

Definition 1.5 (Cyclic Monotonicity)**.** *$f : \mathbb{R}^\ell \to \mathbb{R}^\ell$ is said to be cyclically monotone if for every integer M and cycle $x^0, x^1, \ldots, x^{M-1}, x^M = x^0$ of points each in $\mathbb{R}^\ell$,*

$$\sum_{m=0}^{M-1} (f(x^m) - f(x^{m+1}))' x^m \geqslant 0.$$

To the best of our knowledge, cyclic monotonicity was introduced in the econometrics literature by Shi, Shum, and Song (2016). Shi, Shum, and Song (2016) have previously demonstrated that additive random utility models (Section 1.3.1) imply the

[37] A similar result is shown in Allen and Rehbeck (2016b), which focuses on formal testability of a version of this model. That paper uses different techniques because strict inequalities are required.

restrictions of (iv).

Remark 1.10 (Single Dimensional Case). When $K = 1$, it can be shown that Theorem 1.8(iv) is equivalent to the condition that for every $x, \tilde{x} \in \text{supp}(X)$,

$$(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])(\vec{v}(x) - \vec{v}(\tilde{x})) \geqslant 0.$$[38]

This shows that when $K = 1$, the most we can say about $\vec{v}$ is that it must be consistent with the monotonicity statements:

$$\mathbb{E}[Y \mid X = x] > \mathbb{E}[Y \mid X = \tilde{x}] \implies \vec{v}(x) \geqslant \vec{v}(\tilde{x})$$
$$\mathbb{E}[Y \mid X = x] < \mathbb{E}[Y \mid X = \tilde{x}] \implies \vec{v}(x) \leqslant \vec{v}(\tilde{x}).$$[39]

This is purely ordinal information, and so point identification is impossible for many choices of the parameter space. If $\mathcal{V}$ is unrestricted, then in the single dimensional case whenever $\tilde{v} \in \mathcal{V}_{ID}$, we also have $g(\tilde{v}) \in \mathcal{V}_{ID}$ for any strictly increasing function g. Even if $\mathcal{V}$ is restricted to a class of differentiable functions with a location/scale normalization, $\mathcal{V}_{ID}$ may not be a singleton.

1.18 Appendix: Injectivity

This section provides some injectivity results that are used in establishing identification of $\vec{v}$ in Section 1.4. We use the same mathematical setup as before, but change notation a bit. Formally, we are interested in when the following mapping is at most a

[38] See Rockafellar (1970), p. 240.

[39] When $\vec{v}$ is assumed linear, these restrictions are implied by but do not generally imply the restrictions of the generalized regression model of Han (1987).

singleton,

$$\rho^{-1}(y^*) = \left\{ u \in \mathbb{R}^K \mid y^* \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k u_k + D(y) \right\}.$$

This question is relevant for identification because when $\rho^{-1}(\mathbb{E}[Y \mid X = x])$ is a singleton, we have

$$\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y \mid X = \tilde{x}] \implies \vec{v}(x) = \vec{\tilde{x}}.$$

This implication is the key to our identification results for common characteristics. In this section, we consider a slightly different structure than previously. We use the abstract notation y^* because we do not restrict attention to $u \in \vec{v}(\operatorname{supp}(X))$.

Existence of an inverse function ρ^{-1} is relevant to handle certain forms of endogeneity. We provide a sketch and pointers to some relevant papers. Suppose for example that conditional on characteristics we have

$$u_{k,i} = v_k(x_{k,i}) + \xi_{k,i}.$$

The random vector $\xi_i = (\xi_{1,i}, \ldots, \xi_{k,i})'$ need not be independent of the characteristics $X_i = (X'_{1,i}, \ldots, X'_{k,i})'$. Let

$$y_i^* = \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k (v_k(X_{k,i}) + \xi_{k,i}) + D(y).$$

We assume this is a singleton. For concreteness, one may think of $y_{i,k}^*$ as the market share of good k in market i, as in Berry (1994). Note y_i^* is a random variable. Suppose an inverse ρ^{-1} exists for every value in the support of y_i^*, so that we have

$$\rho_k^{-1}(y_i^*) = v_k(X_{k,i}) + \xi_{k,i} \tag{1.39}$$

almost surely. Using instruments that satisfy a completeness or conditional mean restriction, several recent papers contain identification results for equations similar to (1.39). See Berry and Haile (2014), Chen (2013), and Dunker, Hoderlein, and Kaido (2015). We complement these papers by providing sufficient conditions for the inverse ρ^{-1} to exist.[40]

We now turn to our formal results.

Assumption 1.4. *$B \subseteq \mathbb{R}^K$ is convex and has nonempty interior.*

Assumption 1.4 rules out the probability simplex, which does not have a nonempty interior when viewed as a subset of $\mathbb{R}^K$. We consider the probability simplex separately in Section 1.18.1.

Assumption 1.5. *$D : B \to \mathbb{R} \cup \{-\infty\}$ is a concave function.*

Define

$$T = \left\{ y \in \operatorname{int}(B) \mid y \in \underset{y \in B}{\operatorname{argmax}} \sum_{k=1}^{K} y_k u_k + D(y) \text{ for some } u_k \in \mathbb{R}^K \right\}.$$

The following result is a consequence of Lemma 1.7.

Proposition 1.5. *Suppose $y^* \in T$. Let Assumptions 1.4 and 1.5 hold. Then the following are equivalent:*

(i) $\rho^{-1}(y^*)$ *is a singleton.*

(ii) *D is differentiable at y^*.*

[40]Relative to the invertibility results of Berry, Gandhi, and Haile (2013), we work with a specific model. In return we can handle complementarity without a reparametrization. Our results are thus distinct from theirs.

Proof. Let

$$D_B(y) = \begin{cases} D(y) \text{ if } y \in B \\ -\infty \text{ otherwise} \end{cases}.$$

This allows the budget B to remain implicit. Recall that B is convex, so D_B is concave. Note that $D_B(y^*)$ must be finite.

We can now prove the result using the preceding lemmas once we identify $-D_B$ with f, y^* with z^*, and $\overline{v}$ with w.

By Lemma 1.5,

$$\rho^{-1}(y^*) = \partial(-D_B(y^*)).$$

From Lemma 1.7, we conclude that $\rho^{-1}(y^*)$ is a singleton if and only if $(-D_B)$ is differentiable at y^*. Since y^* is in the interior of B, $(-D_B)$ is differentiable at y^* if and only if D is differentiable at y^*. □

This result explains why differentiability is used in Theorem 1.2.

We can state a global version of the proposition.

Corollary 1.4 (Global Injectivity)**.** *Let the conditions of Proposition 1.5 hold. Let $\tilde{T} \subseteq T$. Then the following are equivalent:*

(i) For each $y^* \in \tilde{T}$, $\rho^{-1}(y^*)$ is a singleton.

(ii) D is differentiable at each $y^* \in \tilde{T}$.

1.18.1 Injectivity on the Simplex

We now assume that B is the probability simplex,

$$B = \left\{ y \in \mathbb{R}^K \mid \sum_{k=1}^{K} y_k = 1, y_k \geqslant 0 \text{ for } k = 1, \ldots, K \right\}.$$

In order to obtain an injectivity result, we need to restrict the set of possible values of the vector u. This is because for fixed D,

$$\operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k u_k + D(y)$$

is the same set with u replaced by $u + c$, where c is a constant vector. We restrict the parameter space u lives in with the following normalization for its first component:

$$\mathcal{U} = \{u \in \mathbb{R}^K \mid u_1 = 0\}.$$

We are now interested in when

$$\rho_{\mathcal{U}}^{-1}(y^*) = \left\{ u \in \mathcal{U} \mid y^* \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k u_k + D(y) \right\}$$

is a singleton.

Assumption 1.6. *$D : B \to \mathbb{R} \cup \{-\infty\}$ is a concave function. Moreover, $\{y \in B \mid D(y) > -\infty\}$ has nonempty interior when viewed as a subset of B.*

A nonempty interior will be needed because we invoke (Fréchet) differentiability of D. Because B is the probability simplex, we need to replace T with a set defined over the relative interior of B:[41]

$$T_{\mathcal{U}} = \left\{ y \in \mathrm{ri}(B) \mid y \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k u_k + D(y) \text{ for some } u \in \mathcal{U} \right\}.$$

Proposition 1.6. *Suppose $y^* \in T_{\mathcal{U}}$. Assume $K \geqslant 2$. Let B be the probability simplex and let Assumption 1.6 hold. Then the following are equivalent:*

(i) $\rho_{\mathcal{U}}^{-1}(y^*)$ *is a singleton.*

(ii) *D is Fréchet differentiable at y^*.*

Proof. We can prove this from Proposition 1.5 with a change of variables. The basic idea will be that over the probability simplex, y_1 is uniquely determined by $(y_2, \ldots, y_K)$. Using the normalization defining $\mathcal{U}$, we may convert the problem from a K-dimensional problem to a $K-1$-dimensional problem and then invoke Proposition 1.5.

We define a new function,

$$\tilde{D}(y_2, \ldots, y_k) = \begin{cases} D\left(\left(1 - \sum_{k=2}^{K} y_k\right), y_2, \ldots, y_K\right) & \text{if } \sum_{k=2}^{K} y_k \leqslant 1, y_k \geqslant 0 \\ -\infty & \text{otherwise.} \end{cases}$$

This function removes y_1 by using the budget constraint. The function $\tilde{D}$ is concave since D is concave and B is convex.

Now define the multi-valued mapping $\rho_{\mathcal{U}}$ that maps points in $\mathbb{R}^{K-1}$ to subsets of

---

[41] The relative interior of B is the set

$$\mathrm{ri}(B) = \left\{ y \in \mathbb{R}^K \mid \sum_{k=1}^{K} y_k = 1, y_k > 0 \text{ for } k = 1, \ldots, K \right\}.$$

$\mathbb{R}^K$ by

$$\rho_{\mathcal{U}}(u) = \operatorname*{argmax}_{y \in B} \sum_{k=2}^{K} y_k u_k + D\left(\left(1 - \sum_{k=2}^{K} y_k\right), y_2, \ldots, y_K\right). \tag{1.40}$$

Note that the choice of the first component (y_1) now enters trivially. Similarly define ρ from $\mathbb{R}^{K-1}$ to subsets of $\mathbb{R}^{K-1}$ by

$$\rho(u) = \operatorname*{argmax}_{y_2,\ldots,y_K \in \mathbb{R}^{K-1}} \sum_{k=2}^{K} y_k u_k + \tilde{D}(y_2, \ldots, y_K). \tag{1.41}$$

Over the probability simplex, we may put $\rho_{\mathcal{U}}(u)$ and $\rho(u)$ in one-to-one correspondence by the mapping $\pi(y_1, \ldots, y_K) = (y_2, \ldots, y_K)$.

Finally, note Fréchet differentiability of D at y^* is equivalent to differentiability of $\tilde{D} : \mathbb{R}^{K-1} \to \mathbb{R} \cup \{+\infty\}$ at $\pi(y^*)$. Invoking Proposition 1.5, $\rho_{\mathcal{V}}^{-1}(y^*)$ is a singleton if and only if D is Fréchet differentiable at y^*. □

1.19 Appendix: Perturbed Generalized Nested Logit

We first describe the generalized nested logit (GNL) (Wen and Koppelman (2001)). See Train (2009) for a textbook treatment. Let there be a finite set of nests N_m, indexed by m. Each alternative k has a weight $0 \leqslant \alpha_{km} \leqslant 1$ reflecting membership in nest m. For each k, these weights sum to 1. Each nest has a nesting parameter λ_m.

The GNL specifies that conditional on utility index $\bar{v}$, the choice probability vector satisfies

$$P(\bar{v}) = \nabla \ln G_{GNL}(\bar{v}),$$

where

$$G_{GNL}(\bar{v}) = \sum_m \left(\sum_{k \in N_m} \alpha_{km} e^{\bar{v}_k/\lambda_m} \right)^{\lambda_m}.$$

We suppress dependence of G and P on λ for notational convenience.

For example, if each alternative is in its own nest, this reduces to

$$G_{Logit}(\bar{v}) = \sum_{k=1}^{K} e^{\bar{v}_k},$$

from which we arrive at the multinomial logit probabilities,

$$P_k(\bar{v}) = \nabla_k \ln \left(\sum_{k=1}^{K} e^{\bar{v}_k} \right) = \frac{e^{\bar{v}_k}}{\sum_{j=1}^{K} e^{\bar{v}_j}}.$$

We will show that GNL is consistent with PUM even when it is consistent with ARUM by using the following result.

Proposition 1.7. *Let* $P(\bar{v}) = \nabla \ln \left(\sum_m \left(\sum_{k \in N_m} \alpha_{km} e^{\bar{v}_k/\lambda_m} \right)^{\lambda_m} \right)$. *Suppose* $0 \leqslant \alpha_{km} \leqslant 1$ *for each km,* $\sum_m \alpha_{km} = 1$ *for each k, and* $\lambda_m > 0$ *for each m. Then there is some* $D \in \mathcal{D}$ *such that*

$$P(\bar{v}) = \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1} y_k \bar{v}_k + D(y).$$

By setting $\bar{v} = \vec{v}(x)$ in Proposition 1.7, we see that $\mathbb{E}[Y \mid X = x] := P(\vec{v}(x))$ is a perturbed utility model. It is known that when $0 < \lambda_m \leqslant 1$, this model is consistent with ARUM. Thus, by enlarging the parameter space to requiring only $0 < \lambda_m$, GNL provides a strict generalization of ARUM. We will prove this proposition using several lemmas.

Lemma 1.9. *Let* $f : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ *be a lower semi-continuous, convex function such* *that $f(z) < \infty$ for some z. Let $P(\bar{v}) = \nabla f(\bar{v})$ be a probability vector. Then*

$$P(\bar{v}) = \underset{y \in \Delta^{K-1}}{\operatorname{argmax}} \sum_{k=1} y_k \bar{v}_k - f^*(y),$$

where f^* denotes the convex conjugate of f.

Proof. This follows immediately from Lemma 1.5. □

It remains to show that $\ln G_{GNL}(\cdot)$ is convex, or equivalently that $G_{GNL}(\cdot)$ is log-convex.

The following result is well-known (see e.g. Boyd and Vandenberghe (2004)).

Lemma 1.10. *Let $f(\vec{u}) = \sum_{k=1}^{J} e^{u_k}$. Then f is log-convex, i.e.*

$$\ln f(\alpha\vec{u} + (1-\alpha)\vec{v}) \leqslant \alpha \ln f(\vec{u}) + (1-\alpha)\ln f(\vec{v}).$$

Lemma 1.11. *Let $g_p(\vec{v}) = (\sum_{k=1}^{J} \beta_k e^{v_k/p})^p$, where $\beta_k \geqslant 0$ for each $k = 1, \ldots, J$. Then for every $p > 0$, g_p is log-convex:*

$$\ln g_p(\alpha\vec{u} + (1-\alpha)\vec{v}) \leqslant \alpha \ln g_p(\vec{u}) + (1-\alpha)\ln g_p(\vec{v}).$$

Proof. This is straightforward.

$$\ln g_p(\alpha\vec{u} + (1-\alpha)\vec{v}) = p \ln\left(\sum_{k=1}^{J} e^{\ln\beta_k(\alpha v_k + (1-\alpha)u_k))/p}\right).$$

The proof follows by the previous lemma. □

The following result is well-known (see e.g. Boyd and Vandenberghe (2004)).

Lemma 1.12. *The sum of log-convex functions is log-convex.*

Proof of Proposition 1.7. By Lemma 1.9, we only need to show that $\ln G_{GNL}(\cdot)$ is log-convex.

By Lemma 1.11, the component

$$\left(\sum_{k\in N_m} \alpha_{km} e^{\bar{v}_k/\lambda_m}\right)^{\lambda_m}$$

is log-convex in $\bar{v}$ due to the parameter restrictions $\alpha_{km} \geqslant 0$ and $\lambda_m > 0$. Since $G_{GNL}(\cdot)$ is the sum of such components, Lemma 1.12 completes the proof. □

Chapter 2

Complementarity in Perturbed Utility Models

Abstract. This paper extends the Hicksian definition of complementarity to settings outside of the standard consumer problem. Using the structure of *perturbed utility models*, we show how to identify complementarity using variation in observable characteristics. This does not require price variation. We propose a derivative ratio as a measure of the degree of complementarity, which can be estimated with standard instrumental variables techniques. Turning to specific settings we show (i) the assumption of no income effects is not needed to identify complementarity in the analysis of Gentzkow (2007) and (ii) complementarity is ruled out by many, but not all, discrete choice models.

2.1 Introduction

The classical definition of complementarity labels two goods complements if the cross-price elasticity of compensated demand is negative (Hicks and Allen (1934)). Despite the importance of complementarity, there are many examples where this definition does not apply. This could occur either because prices do not vary (Gentzkow (2007)) or the setting is outside of the standard consumer problem. For example, "Are exercise and sleep complements or substitutes?" is not a precise question. We extend the Hicksian definition to answer this question using non-price characteristic variation. Our definition is based on comparative statics and can be used for prediction. For example, if exercise and sleep are identified as complements for an individual, this suggests building bicycle paths will induce that person to sleep more.

We propose a model-based definition of complementarity that preserves key features of the Hicksian definition of complementarity. First, our definition uses alternative-specific characteristics to identify complementarity. In the Hicksian case these must be prices. Second, our definition is symmetric. *Perturbed utility models* (McFadden and Fosgerau 2012b) are a class of models with just enough structure to ensure these properties. These models generate conditional means (or "average quantities") that satisfy

$$\mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in B} \sum_{j=1}^{k} y_j v_j(x_j) + D(y), \tag{2.1}$$

where k is the number of alternatives and y is a k-dimensional vector of quantities of each alternative. An "alternative" is general and may constitute a specific time activity such as exercise. Each alternative has a vector of characteristics associated with it, and these are collected in the observable random variable $X = (X_1, \ldots, X_k) \in \mathbb{R}^d \times \ldots \times \mathbb{R}^d$. For example, Gentzkow (2007) uses accessibility to the internet as a characteristic of online news. The collection of functions $v = (v_1, \ldots, v_k)$ encodes how characteristics affect the desirability of each alternative. Let D be an extended real-valued function and $B \subseteq \mathbb{R}^k$ be a compact, convex set that is interpreted as a budget.

Instead of variation in prices, we use variation in v (due to changes in x) to define and identify complementarity. This requires alternative-specific characteristics that shift the value of one alternative relative to the others. Our approach thus follows Gentzkow (2007). Our primitive definition of complementarity is that if v_j increases (due to a change in x_j) and the ℓ^{th} component of $\mathbb{E}[Y \mid X = x]$ increases, then goods j and ℓ are local complements. The definition is formally stated in terms of partial derivatives. Thus, our definition is *local* and preserves the comparative statics features of the Hicksian definiton.

We show how to identify complementarity even when v is not known *a priori*. This is possible by checking the sign of derivative ratios such as

$$\frac{\partial \mathbb{E}[Y_j \mid X = x]}{\partial x_{\ell,q}} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_{\ell,q}} \tag{2.2}$$

and their discrete counterparts. This uses variation in $x_{\ell,q}$, which is the q-th characteristic specific to alternative ℓ. The derivative ratio is positive exactly when goods j and ℓ are local complements according to our primitive definition. When conditional means are linear, derivative ratios can be estimated using standard linear instrumental variables methods.

We apply and analyze our definition of complementarity in two settings. First, we study the bundles model of Gentzkow (2007), in which $\mathbb{E}[Y \mid X = x]$ is the average quantity of k different goods purchased. We show that the assumption of no income effects is unnecessary to identify whether goods are complements or substitutes. For perturbed utility models, a derivative ratio like (2.2) identifies not only the sign of Hicksian substitution effects, but also the ratio of cross- and own-price substitution effects. Thus, our definition of complementarity formally extends the Hicksian definition.

Next, we turn to the setting of discrete choice. Now we interpret $\mathbb{E}[Y \mid X = x]$ as a vector of conditional choice probabilities. While perturbed utility models allow complementarity, there are special cases that rule it out. Additive random utility models are one important class that rules out complementarity. Complementarity is also ruled out in the additive perturbed utility models proposed by Fudenberg, Iijima, and Strzalecki (2015). More broadly, violations of regularity from the random utility hypothesis are a form of complementarity (Block and Marschak (1960)).[1] Therefore, complementarity is a qualitative feature that can be used to sort among models of discrete choice.

While we focus attention on the bundles and discrete choice settings, perturbed utility models can be used in other settings. For example, $\mathbb{E}[Y \mid X = x]$ can represent conditional probabilities of choosing an action in a game theoretic context, average time spent on k different activities, and so forth. The wide range of applicability of our analysis is highlighted by Theorem 2.1, which shows that the structure of the perturbed utility model is implied by a large class of latent utility models. This allows us to accommodate certain forms of unobserved heterogeneity or measurement error. This result may be of independent interest since it shows that the the structure of the perturbed utility model aggregates.

The literature on complementarity is large. Samuelson (1974) provides a comprehensive survey of alternative definitions of complementarity in the standard consumer problem. Gentzkow (2007) pioneers the use of non-price characteristics to identify

[1] Regularity states that adding an item to a menu of alternatives cannot increase the probability of choosing any specific alternative that was previously available.

complementarity. Complementing the focus of this paper, Manzini, Mariotti, and Ülkü (2015) study complementarity when there is *no* characteristic variation in a model free setting.

Focusing on settings that admit a lattice structure such as certain firm or game-theoretic problems, a large literature relates complementarity with shape restrictions such as supermodularity (Topkis (1978), Vives (1990), Milgrom and Shannon (1994), Athey and Stern (1998)). We depart from this literature in two directions: we focus on *local* definitions of complementarity; and we *define* complements based on comparative statics of the model. In the context of the standard consumer problem, the relationship between supermodularity of the utility function and a comparative statics definition of complementarity has been discussed in Samuelson (1974) and Chambers and Echenique (2009b).

The rest of this paper is organized as follows. Section 2.2.1 presents a preview of our analysis by showing how (non-price) characteristic variation can identify complementarity in the standard consumer problem. Section 2.2.2 shows how measurement error and unobservable heterogeneity can be accommodated in our setup. Section 2.3 presents our primitive definition of complementarity and shows how to identify whether goods are complements or substitutes using characteristic variation. Section 2.4 studies the bundles model of Gentzkow (2007). Section 2.5 studies complementarity in a discrete choice setting. Section 2.6 summarizes the analysis.

2.2 Perturbed Utility Models

Perturbed utility models provide a framework to maintain important features of Hicksian complementarity. We illustrate this in an example. This example motivates our analysis of complementarity and provides a direct link to Hicksian complementarity.

2.2.1 Relation to Hicksian Complementarity

In this section, we specialize the perturbed utility model to a variant of the standard consumer problem. Suppose a consumer chooses y to maximize

$$U(y;x) = \sum_{j=1}^{k} y_j v_j(x_j) + D(y) \qquad \text{subject to} \qquad \sum_{j=1}^{k} p_j y_j \leqslant I. \tag{2.3}$$

This functional form allows alternative-specific characteristics to enter the utility function. Characteristics x are fixed and the agent chooses quantities to maximize utility. Let y_j^* denote Marshallian demand for good j and suppose $y_j^* > 0$ for all j. We write y_j^* instead of $\mathbb{E}[Y_j \mid X = x]$ for notational convenience. Let h_j^* be the associated Hicksian demand function for good j.[2] Appendix 2.8 shows that the derivative of Marshallian demand with respect to characteristics is just a scaled version of the substitution effect:

$$\frac{\partial y_j^*}{\partial x_{\ell,q}} = -\frac{\partial v_\ell / \partial x_{\ell,q}}{\lambda^*} \frac{\partial h_j^*}{\partial p_\ell}, \tag{2.4}$$

where λ^* is the marginal utility of income. Note that the scale term involves only alternative ℓ. We can cancel out the unobservable scale to obtain the following result.

Proposition 2.1. *Suppose $\partial h_\ell^* / \partial p_\ell \neq 0$ and $\partial y_\ell^* / \partial x_{\ell,q} \neq 0$. Then assuming the derivatives exist,*

$$\frac{\partial y_j^*}{\partial x_{\ell,q}} \Big/ \frac{\partial y_\ell^*}{\partial x_{\ell,q}} = \frac{\partial h_j^*}{\partial p_\ell} \Big/ \frac{\partial h_\ell^*}{\partial p_\ell}. \tag{2.5}$$

[2] All dependence on parameters is left implicit. See Appendix 2.8 for more details.

In particular, by the law of compensated demand,

$$\operatorname{sign}\left(\frac{\partial y_j^*}{\partial x_{\ell,q}}\Big/\frac{\partial y_\ell^*}{\partial x_{\ell,q}}\right) = -\operatorname{sign}(\partial h_j^*/\partial p_\ell).^3 \tag{2.6}$$

Even without price or income variation, we can still recover the sign of the Hicksian measure of complementarity.[4] This is possible due to additive separability of $U(y;x)$ and the fact that changes in characteristics do not affect the budget set. In our subsequent analysis, a derivative ratio of conditional expectations will similarly be used to identify complementarity.

For some intuition behind (2.5) and Proposition 2.1, note that the expenditure minimization problem simply finds a supporting hyperplane to the upper contour set of a utility function. The price derivatives of Hicksian demand are characterized by pivoting this supporting hyperplane, whose slope is determined by prices, around the upper contour set. Even without price or income variation, we can "reproduce" this same pivoting with characteristic variation *provided* the characteristics enter as in (2.3).

2.2.2 A Latent Utility Foundation

We now return to the general setting. We show that the restrictions of the perturbed utility models arise from a class of latent utility models. This illustrates that the structure of perturbed utility models is preserved even under certain forms of heterogeneity or measurement error. Allowing unobservable heterogeneity is essential for the practical measurement of complementarity, as emphasized in Athey and Stern

[3]While *DR* is related to the Hicks-Allen definition of complementarity, this is a good opportunity to note that to our knowledge, neither of the authors of the present paper is related to Roy G. D. Allen.

[4]Note that using the Slutsky Equation, (2.5) allows for *partial* recoverability of income effects using characteristic variation. This is not assumption-free since it relies on the form of the utility function.

(1998) and Gentzkow (2007).

Consider the utility function

$$U(y;X,\varepsilon) = \sum_{j=1}^{k} y_j v_j(X_j) + \tilde{D}(y,\varepsilon), \tag{2.7}$$

where ε is a random variable of possibly infinite-dimension and captures unobservable heterogeneity. Here $\tilde{D}$ is an unknown, extended real-valued function. The set B is a budget that does not depend on unobservables. Note this assumption is mild since $\tilde{D}$ can take on value $-\infty$.

We comment on some key aspects of this utility function. Importantly, observable characteristics X do not enter $\tilde{D}$. To relax this assumption, we can allow a subvector of X to enter $\tilde{D}$ and condition analysis on a fixed value of those observables. It is necessary for our anlaysis that at least one characteristic does not enter $\tilde{D}$. A second key assumption is that ε enters the utility specification separably from X. This assumption is needed for our aggregation result (Theorem 2.1).

To fix ideas, we present an example.

Example 2 (Bundles)**.** *This example considers a bundles model studied in Gentzkow (2007) and Fox and Lazzati (forthcoming). For simplicitly, we consider a model with two goods in which a consumer can buy between* 0 *and* 1 *unit of each good. The consumption vector* $y = (y_1, y_2)$ *obtains utility*

$$U(y;X,\varepsilon,\eta) = \sum_{j \leqslant 2} y_j v_j(X_j) + \tilde{D}(y,\varepsilon,\eta),$$

where

$$\tilde{D}(y,\varepsilon,\eta) = y_1\varepsilon_1 + y_2\varepsilon_2 + \eta \cdot \Delta 1(y_1 = 1, y_2 = 1)$$

and $B = [0,1]^2$. *Here, both* ε *and* η *are latent random variables.* Δ *is a parameter.*

Let $Y^*(X,\varepsilon)$ be a measurable maximizer of $U(y;X,\varepsilon)$ subject to the budget constraint $y \in B$.[5] We do not observe $Y^*(X,\varepsilon)$ directly because ε is unobservable. Instead, we assume knowledge of

$$\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y^*(X,\varepsilon) \mid X = x].$$

Knowledge of conditional means can be thought of as "limiting information" if we observe an arbitrarily large independent and identically distributed sample $\{(Y^i, X^i)\}_{i=1}^n$. The focus of this paper is on what is theoretically identifiable given this knowledge, not the finite sample reality that must deal with estimation error.

Under regularity conditions ensuring existence of a measurable $Y^*(X,\varepsilon)$ as well as finiteness of certain expectations, we now show that $\mathbb{E}[Y^*(X,\varepsilon) \mid X = x]$ is consistent with the perturbed utility model.

Theorem 2.1. *Let*

$$Y^*(X,\varepsilon) \in \operatorname{argmax} \sum_{j=1}^{k} y_j v_j(X_j) + \tilde{D}(y,\varepsilon).$$

Under Assumptions 2.5 and 2.6 (stated in Appendix 2.9), if $\varepsilon \perp X$*, then*

i. $\mathbb{E}[Y^*(X,\varepsilon) \mid X = x]$ *is consistent with a perturbed utility model with the same* $v = (v_1, \ldots, v_k)$ *and*

$$D(y) = \sup_{Y \in \mathcal{Y} : \mathbb{E}[Y(\varepsilon)] = y} \mathbb{E}\left[\tilde{D}(Y(\varepsilon),\varepsilon)\right],^{6}$$

where $\mathcal{Y}$ *is the set of measurable functions from* $\operatorname{supp}(\varepsilon)$ *to B.*[7]

---

[5] Sufficient conditions for existence of a measurable $Y^*(X,\varepsilon)$ are given in Appendix 2.9.

[6] D is taken to be $-\infty$ if there is no $Y \in \mathcal{Y}$ such that $\mathbb{E}[Y(\varepsilon)] = y$. D is always finite when evaluated at $\mathbb{E}[Y^*(X,\varepsilon) \mid X = x]$.

[7] The support of a random variable Z is the smallest closed set K such that $P(Z \in K) = 1$.

ii. Define the Social Surplus Function (McFadden 1978)

$$\mathcal{G}(x) = \mathbb{E}\left[\max_{y \in B} U(y; X, \varepsilon) \mid X = x\right]. \tag{2.8}$$

Then

$$\mathcal{G}(x) = \max_{y \in B} \left\{ \sum_{j=1}^{k} y_j v_j(x) + D(y) \right\}. \tag{2.9}$$

Theorem 2.1 demonstrates the link between the latent utility and representative agent formulations.[8] It shows how restrictions on conditional means arise *without* specifying the distribution of latent variables. Specifically, this aggregation result occurs regardless of correlation or other types of dependence of ε. While we explicitly characterize how $\tilde{D}$ and the distribution of ε induce a function D for a representative agent, obtaining analytical results for D appears challenging in general.

We now require an additional assumption. This will be used to establish that our definition of complementarity is symmetric.

Assumption 2.1 (Index Sufficiency)**.** *If* $v(x) = v(\tilde{x})$*, then* $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y \mid X = \tilde{x}]$.

Assumption 2.1 states that $\mathbb{E}[Y \mid X = x]$ depends on x only through $v(x)$. Assumption 2.1 is automatically satisfied if the argmax of (2.1) is always single-valued. With minor abuse of notation we may write, $\mathbb{E}[Y \mid \bar{v}] := \mathbb{E}[Y \mid v(X) = \bar{v}]$.

Corollary 2.1. *Define the v-subdifferential*

$$\partial_v \mathcal{G}(x) = \{u \in \mathbb{R}^k \mid \forall \tilde{x}, \mathcal{G}(\tilde{x}) - \mathcal{G}(x) \geqslant u \cdot (v(\tilde{x}) - v(x))\}.$$

Under Assumption 2.1 and the conditions of Theorem 2.1, $\partial_v \mathcal{G}(x)$ *is nonempty and*

[8] Theorem 2.1 appears new at this level of generality and these results may be of independent interest. In the setting of discrete choice, related results appear in McFadden (1981), Hofbauer and Sandholm (2002), and Fosgerau and Palma (2015).

satisfies

(i) $\mathbb{E}[Y \mid X = x] \in \partial_v \mathcal{G}(x)$.

(ii) When $\mathbb{E}[Y \mid \overline{v}]$ *is differentiable,*

$$\frac{\partial \mathbb{E}[Y_j \mid \overline{v}]}{\partial \overline{v}_\ell} = \frac{\partial \mathbb{E}[Y_\ell \mid \overline{v}]}{\partial \overline{v}_j} \qquad \forall j, \ell = 1, \ldots, k. \tag{2.10}$$

Proof. This Corollary follows from McFadden and Fosgerau (2012b) Theorem 1. □

Note that we are not assuming we know $\mathbb{E}[Y \mid \overline{v}]$, just stating properties of it. Part (i) is a form of Roy's Identity, and may be seen as a generalization of the Williams-Daly-Zachary Theorem in the discrete choice literature (see e.g. McFadden (1981)). When v is the identity mapping and $\partial_v \mathcal{G}(x)$ is a singleton, part (i) states that the conditional mean vector is the Jacobian of $\mathcal{G}$ at x. In this case, part (ii) follows for the same reason the Slutsky matrix is symmetric: the Jacobian of $\mathbb{E}[Y \mid \overline{v}]$ is the Hessian of a convex function. Note that Corollary 2.1 holds for general compact budget sets as emphasized in McFadden and Fosgerau (2012b). Our measurement of complementarity will not require (or use) knowledge of B. When this budget set is known *a priori*, additional shape restrictions on D can potentially be used to try to separate alternative notions of "true complementarity" due to preferences from those due to the shape of the constraint set.[9] This important distinction is left for future work.

[9] Suppose $B = \{y : y = \alpha(1, \ldots, 1) + (1 - \alpha)(0, \ldots, 0)$ for some $\alpha \in [0, 1]\}$. Then changing v moves the optimal solution along this manifold in ways that are only consistent with our definition of complementarity.

2.3 Measuring Complementarity

We now propose formal definitions of complementarity measures. Each measure is rooted in the standard Hicksian definition. Only some of these measures can be recovered from data without knowledge of the functions v. We first present differential measures and then discrete measures.

2.3.1 Differential Measures of Complementarity

We maintain Assumption 2.1 so that we can write $\mathbb{E}[Y \mid \bar{v}] = \mathbb{E}[Y \mid v(X) = \bar{v}]$ for simplicity. In addition we assume differentiability.

Assumption 2.2. $\mathbb{E}[Y \mid \bar{v}]$ *is differentiable in* $\bar{v}$*.*

Definition 2.1. *Goods* j *and* ℓ *are local complements at* $\bar{v}$ *if*

$$CD_{j,\ell}(\bar{v}) = \frac{\partial \mathbb{E}[Y_j \mid \bar{v}]}{\partial \bar{v}_\ell} \geqslant 0, \tag{2.11}$$

and substitutes if

$$CD_{j,\ell}(\bar{v}) \leqslant 0.$$

We will show how to identify whether goods are local complements using data, but first we consider some alternative measures of complementarity that are stated only in terms of $\mathbb{E}[Y \mid \bar{v}]$. We call *CD* the *cross-derivative* measure due to the similarity with cross-price elasticities. Note that by Corollary 2.1, $CD_{j,\ell}(\bar{v}) = CD_{\ell,j}(\bar{v})$.

When the function v is not known *a priori*, we cannot directly identify $CD_{j,\ell}(\bar{v})$ from data. We thus propose two alternative measures of complementarity. Our preferred measure is the derivative ratio (*DR*) measure

$$DR_{j,\ell}(\overline{v}) = \frac{\partial \mathbb{E}[Y_j \mid \overline{v}]}{\partial \overline{v}_\ell} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid \overline{v}]}{\partial \overline{v}_\ell}, \tag{2.12}$$

which can heuristically be thought of as a ratio of indirect effect to direct effect. We also introduce

$$CM_{j,\ell}(\overline{v}) = \frac{\partial \mathbb{E}[Y_j \mid \overline{v}]}{\partial \overline{v}_\ell} \frac{\partial \mathbb{E}[Y_\ell \mid \overline{v}]}{\partial \overline{v}_\ell}, \tag{2.13}$$

which we call the *CM* measure to emphasize this as a *comovement* measure. This measure is superficially related to measures such as $\mathrm{Cov}(Y_j, Y_\ell \mid \overline{v})$, but the interpretation is fundamentally different. $\mathrm{Cov}(Y_j, Y_\ell \mid \overline{v})$ can be positive in a way that is unrelated to the comparative statics measure *CM*. We explore this further in Section 2.4.

It can be shown that under Assumption 2.2, $\frac{\partial \mathbb{E}[Y_\ell \mid \overline{v}]}{\partial \overline{v}_\ell} \geqslant 0$. This ensures that the denominator of *DR* has a known sign.[10]

Assumption 2.3 (Index Relevance)**.** $\frac{\partial \mathbb{E}[Y_\ell \mid \overline{v}]}{\partial \overline{v}_\ell} \neq 0$.

Assumption 2.3 formalizes that there must be a direct effect due to a change in $\overline{v}_\ell$.[11]

Proposition 2.2. *Under Assumptions 2.1-2.3,*

$$\mathrm{sign}\left(CD_{j,\ell}(\overline{v})\right) = \mathrm{sign}(CM_{j,\ell}(\overline{v})) = \mathrm{sign}(DR_{j,\ell}(\overline{v})). \tag{2.14}$$

The proof of this proposition is simple and is omitted. Proposition 2.2 demonstrates that once we can link the measures to observables we can test the sign of any of the three measures. While *CM* and *DR* are not necessarily symmetric in magnitude with

[10] See Lemma 2.2 in the Appendix.

[11] Of course it is possible that $\mathbb{E}[Y_\ell \mid \overline{v}]$ is differentiable and strictly increasing in $\overline{v}_\ell$ and yet the derivative is 0.

respect to j and ℓ, they are symmetric in sign.

We are interested not only in defining and testing complementarity but also a measure of complementarity with cardinal meaning. For this purpose, *DR* conveys the most useful information when v is not known *a priori*. This is because *DR* is scale free, while *CD* and *CM* are not invariant to monotonic transformations of v.[12] In fact, $DR_{j,\ell}(\overline{v})$ can be identified directly from the data under an additional assumption.

Assumption 2.4 (Characteristic Relevance)**.** *There exists a characteristic q for alternative ℓ such that $\partial v(x)/\partial x_{\ell,q} \neq 0$.*

Together with Assumption 2.3, this is a relevance assumption for $x_{\ell,q}$. Relying on Assumptions 2.1-2.4, we have

$$\frac{\partial \mathbb{E}[Y_j \mid X = x]}{\partial x_{\ell,q}} = \frac{\partial \mathbb{E}[Y_j \mid \overline{v}]}{\partial \overline{v}_\ell}\bigg|_{\overline{v}=v(x)} \frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,q}}. \tag{2.15}$$

This equation holds with $j = \ell$ and $j \neq \ell$. By dividing two such equations we obtain

$$\frac{\partial \mathbb{E}[Y_j \mid X = x]}{\partial x_{\ell,q}} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_{\ell,q}} = DR_{j,\ell}(v(x)). \tag{2.16}$$

Recall that we used the same derivative ratio to identify Hicksian complementarity in Section 2.2.1.

Because the characteristics of alternative ℓ enter the perturbed utility models only through the index function v_ℓ, (2.16) is the same value for any q such that $\partial v(x)/\partial x_{\ell,q} \neq 0$. In other words, derivatives taken with respect to different characteristics will identify the same measure of complementarity. Thus, if one learns the derivative ratio by using

[12]It is clear from the definition of perturbed utility models that v is not identified up to scale when D is not known, since we can multiply the right hand side of (2.1) without affecting the argmax set. Moreover, when the only variation in the data is in characteristics for a single good, say, ℓ, then it can be shown that v_ℓ can at best only be identified up to a monotonic transformation without further restrictions.

characteristic variation in $x_{\ell,q}$, then we obtain the same ratio for $x_{\ell,r}$ for any other characteristic r.

Proposition 2.3. *Under Assumptions 2.1-2.4, $DR_{j,\ell}(\overline{v})$ is identified from data.*

When $\mathbb{E}[Y \mid X = x]$ is linear in x, $DR_{j,\ell}(v(x))$ can be estimated by linear instrumental variables. See Appendix 2.14 for details. The sign of CM is also directly recoverable from data by

$$\frac{\partial \mathbb{E}[Y_j \mid \overline{v}]}{\partial x_{\ell,q}} \frac{\partial \mathbb{E}[Y_\ell \mid \overline{v}]}{\partial x_{\ell,q}} = \frac{\partial \mathbb{E}[Y_j \mid \overline{v}]}{\partial \overline{v}_\ell} \frac{\partial \mathbb{E}[Y_\ell \mid \overline{v}]}{\partial \overline{v}_\ell} \left(\frac{\partial \overline{v}_\ell}{\partial x_{\ell,q}} \right)^2 = CM_{j,\ell}(\overline{v}) \left(\frac{\partial \overline{v}_\ell}{\partial x_{\ell,q}} \right)^2. \quad (2.17)$$

We will use a related comovement measure for our discrete measures of complementarity.

2.3.2 Discrete Measures of Complementarity

We extend our analysis to a discrete setting for two reasons. First, the support of characteristics, denoted $\text{supp}(X)$, may be discrete. Second, we may wish to test a hypothesis involving non-local variation such as that goods are everywhere complements (over the support of X).

We focus on complementarity between goods j and ℓ. We make use of the following definition, which implicitly depends on j and ℓ.

Definition 2.2 (Comparability)**.** *Let $x, \tilde{x} \in \text{supp}(X)$. Then $x = (x_1, \ldots, x_k)$ and $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_k)$ are comparable if they differ only with respect to components corresponding to x_j or x_ℓ but not both.*

We introduce a discrete version of the CM measure,

$$DCM_{j,\ell}(x, \tilde{x}) = (\mathbb{E}[Y_j \mid X = x] - \mathbb{E}[Y_j \mid X = \tilde{x}])(\mathbb{E}[Y_\ell \mid X = x] - \mathbb{E}[Y_\ell \mid X = \tilde{x}]). \quad (2.18)$$

We use this measure rather than a discrete analogue of the derivative ratio to avoid division by 0. We propose the following definition of complementarity (relative to a point $\bar{x}$) for discrete changes.

Definition 2.3. *Goods j and ℓ are complements at $\bar{x} \in \operatorname{supp}(X)$ if for every vector x that is comparable to $\bar{x}$,*

$$DCM_{j,\ell}(x,\bar{x}) \geqslant 0.$$[13]

We say they are substitutes at $\bar{x}$ if the opposite inequality holds.

We show this definition is consistent with the differentiable analysis.

Proposition 2.4. *Let $\bar{v}^1 = v(x), \bar{v}^0 = v(\bar{x})$, where x and $\bar{x}$ are comparable. Then under the assumptions of the previous section, if j and ℓ are local complements for every $\bar{v}$ on the line segment between $\bar{v}^1$ and $\bar{v}^0$, then*

$$DCM_{j,\ell}(x,\bar{x}) \geqslant 0.$$

Our definition of complementarity is relative to a single vector $\bar{x}$.[14] There are of course alternative definitions. One approach would be to label the goods complements at $\bar{x}$ if the above inequality holds for every x that is comparable to $\bar{x}$ in a specific neighborhood. This would provide a local discrete definition of complementarity.

We now compare our definition with some previous notions involving covariances and conditional covariances.

Proposition 2.5. *Let $S \subseteq \operatorname{supp}(X)$ and suppose any two elements of S are comparable.*[15] *Let $\bar{x} \in S$ be arbitrary.*

---

[13] This paper studies complementarity and substitutability in terms of population-level data. See Kaplan (2015) for tests of the hypothesis $H_0 : DCM_{j,\ell}(x,\bar{x}) \geqslant 0$.

[14] Two goods can be neither complements nor substitutes at $\bar{x}$ according to our definition.

[15] The conditioning set S is similar to sets used in Proposition 2 of De Paula and Tang (2012).

i. If goods j and ℓ are complements at $\bar{x}$, then

$$\mathrm{Cov}(\mathbb{E}[Y_j \mid X], \mathbb{E}[Y_\ell \mid X] \mid X \in S) \geqslant 0.$$

ii. If goods j and ℓ are substitutes at $\bar{x}$, then

$$\mathrm{Cov}(\mathbb{E}[Y_j \mid X], \mathbb{E}[Y_\ell \mid X] \mid X \in S) \leqslant 0.$$

Proposition 2.5 clarifies how (an implication of) our definition of complementarity differs from alternative approaches that associate complementarity with the sign of $\mathrm{Cov}(Y_j, Y_\ell)$ or $\mathrm{Cov}(Y_j, Y_\ell \mid X)$.

2.4 Bundles Model

We now revisit Example 2. Using Theorem 2.1, we may integrate out the latent variables and study complementarity in terms of $\mathbb{E}[Y \mid X = x]$. This is just a vector of average quantitites of goods 1 and 2. Our definition of complementarity agrees with that of Gentzkow (2007) at this level of aggregation. Proposition 2.1 shows that if we embed this example in the standard consumer problem, we can identify complementarity without assuming no income effects.[16]

We now relate supermodularity of $\tilde{D}$ in the latent utility model with our definition of complementarity. Supermodularity has been used in this setting by Fox and Lazzati (forthcoming).

[16]This embedding requires minor changes including addition of a third (outside) good and inclusion of a budget reflecting both the quantity constraints of Example 2 and the standard income budget.

Definition 2.4. $\tilde{D}(y,\varepsilon,\eta)$ *is supermodular in y at* (ε,η) *if*

$$\tilde{D}((1,1),\varepsilon,\eta)+\tilde{D}((0,0),\varepsilon,\eta)\geqslant\tilde{D}((1,0),\varepsilon,\eta)+\tilde{D}((0,1),\varepsilon,\eta). \tag{2.19}$$

$\tilde{D}(y,\varepsilon,\eta)$ *is submodular in y at* (ε,η) *if the opposite inequality holds.*

We say $\tilde{D}$ is supermodular (resp. submodular) if it is supermodular (resp. submodular) for almost every (ε,η). It is clear that $\tilde{D}$ is supermodular if and only if $\eta\cdot\Delta\geqslant 0$ almost surely.

In the following result, $DR_{1,2}$ denotes the derivative ratio of the conditional means.

Proposition 2.6. *Suppose* (ε,η) *and X are independent,* $\varepsilon\mid\eta$ *has a density with respect to Lebesgue measure, and* $\mathbb{E}[\eta]=1$.

i. If $\tilde{D}$ *is supermodular, then* $DR_{1,2}(v(x))\geqslant 0$ *for every x at which it is defined, and for any comparable vectors* $x,\tilde{x}$, $DCM_{1,2}(x,\tilde{x})\geqslant 0$.

ii. If $\tilde{D}$ *is submodular, then* $DR_{1,2}(v(x))\leqslant 0$ *for every x at which it is defined, and for any comparable vectors* $x,\tilde{x}$, $DCM_{1,2}(x,\tilde{x})\leqslant 0$.

Fox and Lazzati (forthcoming) show how to identify whether goods are substitutes or complements under an additional assumption. They partition $X_j=(Z_j,W_j)$ and assume $v_j(X_j)=Z_j+\tilde{v}_j(W_j)$ for some $\tilde{v}_j$. They show that variation in Z_j identifies the sign of Δ. In addition to separability, this assumes *a priori* knowledge that Z_j enters monotonically in $v_j(X_j)$. We highlight that we do not need *a priori* knowledge of how a characteristic enters v_j to identify the sign of Δ. This complements the work of Fox and Lazzati (forthcoming), whose focus is instead on identification of all aspects of the model.

We close this example by contrasting our measures of complementarity with

$\mathrm{Cov}(Y_1, Y_2 \mid X = x)$. Our conclusion essentially agrees with results in Athey and Stern (1998) that conditional covariance may contain little or no information about complementarity. In particular, conditional covariance of choices may capture only the covariance of unobservables.

Proposition 2.7. *Assume $\Delta = 0$, $v(x)$ is the zero vector in $\mathbb{R}^k$, and that $(\varepsilon_1, \varepsilon_2)|X = x$ is normally distributed with mean 0. Then*

$$\mathrm{sign}(\mathrm{Cov}(Y_1, Y_2 \mid X = x)) = \mathrm{sign}(\mathrm{Cov}(\varepsilon_1, \varepsilon_2 \mid X = x)).$$

2.5 Discrete Choice

When making a single discrete choice, it is natural to label all goods substitutes since only one good is chosen. We will instead focus on a higher level of aggregation than a single discrete choice. Specifically, we focus on complementarity in terms of conditional choice probabilities. One reason to consider complementarity at this level of aggregation is that complementarity is ruled out by many existing discrete choice models.

We make a slight change of notation and write

$$\mathbb{E}[Y \mid X = x] \in \operatorname*{argmax}_{y \in \Delta^k} \sum_{j=1}^{k} y_j v_j(x_j) - C(y), \tag{2.20}$$

where Y_j is an indicator for whether good j is purchased. The vector of conditional choice probabilities is denoted $\mathbb{E}[Y \mid X = x]$. The budget is now the probability simplex $\Delta^k = \{y \in \mathbb{R}^k \mid \sum_{j=1}^{k} y_k = 1\ ,\ y_j \geqslant 0\ \forall j\}$. Following the literature, we place a negative in front of C so that it may be interpreted as a cost to perfect optimization.

An important special case of this model is the class of additive random utility models, which includes familiar examples such as logit and probit. It is easy to see that these models do not allow complementarity.

Example 3 (Additive Random Utility Models)**.** *Suppose alternative j has conditional indirect utility*

$$v_j(x_j) + \varepsilon_j,$$

and an individual chooses an alternative that maximizes utility. Letting $\tilde{D}(y,\varepsilon) = -\sum_{j=1}^{k} y_k \varepsilon_k$, Theorem 2.1 shows these models fit into our framework when $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_k)$ is independent of $X = (X_1,\ldots,X_k)$.[17]

Perturbed utility models have been used to model costly optimization,[18] mistakes, rational inattention,[19] and ambiguity aversion arising from not knowing the true utility (or characteristics) of an alternative. Many existing models impose that C is additively separable. For example, Fudenberg, Iijima, and Strzalecki (2015) introduce the model with general additively separable C functions such as $C(y) = \sum_{j=1}^{k} c(y_j)$.[20] Additively separable C do not allow alternatives to be (strict) complements according to our definition (Proposition 2.8). Thus, one can use the "no complementarity" implication to test the restrictions of the entire class of additively separable cost functions.

Proposition 2.8. *Suppose C is differentiable, strictly convex,*[21] *and $\mathbb{E}[Y \mid X = x]$ is always on the relative interior of Δ^k. Moreover, assume C is additively separable, i.e. $C(y) = \sum_{j=1}^{k} c_j(y_j)$. Then for each pair of alternatives j,ℓ and comparable vectors $x,\tilde{x}$,*

$$DCM_{j,\ell}(x,\tilde{x}) < 0.$$

---

[17]Under additional differentiability assumptions, Hofbauer and Sandholm (2002) have previously shown that the perturbed utility model is *strictly* more general than the additive random utility model.

[18]Mattsson and Weibull (2002).

[19]Matejka and McKay (2014).

[20]Fudenberg, Iijima, and Strzalecki (2015) also study alternative C functions that are not covered here.

[21]Strict convexity is mild and is "almost" equivalent to assuming a unique maximizer as shown in (Allen and Rehbeck (2016b)).

A related result is shown in Feng, Li, and Wang 2015.

We now relate the shape of C with the DR measure of complementarity. Shape restrictions such as supermodularity have been related with a global notion of complementarity in other settings (Topkis (1978), Milgrom and Shannon (1994)). Without a change of variables, we find that the relationship between the shape of C and complementarity involves a number of cross derivatives of C.

For simplicity, suppose there are 3 alternatives. We write y_j^* instead of $\mathbb{E}[Y_j \mid X = x]$ for notational convenience. Under regularity conditions, Appendix 2.13 shows that $DR_{2,1}(v(x))$ is given by

$$DR_{2,1}(v(x)) = \frac{-C_{3,3}(y^*) + C_{2,3}(y^*) - C_{1,2}(y^*) + C_{1,3}(y^*)}{C_{2,2}(y^*) + C_{3,3}(y^*) - 2C_{2,3}(y^*)}.$$

Subscripts on C denote mixed partial derivatives. The denominator of $DR_{2,1}(v(x))$ is strictly positive under our regularity conditions, so the sign of $DR_{2,1}(v(x))$ is determined by

$$-C_{3,3}(y^*) + C_{2,3}(y^*) - C_{1,2}(y^*) + C_{1,3}(y^*).$$

The sign of this sum is not determined by supermodularity of C.[22]

2.5.1 Perturbed Nested Logit

We consider an example that generalizes the standard nested logit model. We show that the nested logit probability formula is *sometimes* an additive random utility model, but is *always* a perturbed utility model. The parameters for which the model is not an additive random utility model are precisely the parameters that admit complementarity.

[22]Under differentiability, supermodularity of the utility function requires $C_{j,\ell}(y) \geqslant 0$ for $j \neq \ell$. Note that $C_{j,\ell}(y) \geqslant 0$ is neither necessary nor sufficient for $DR_{j,\ell}(v(x)) > 0$.

We also show that a logarithmic version of *DR* identifies nesting parameters in the model.

To define this model, we partition the alternatives $\{1,\ldots,k\}$ into nests $A_1, A_2, \ldots, A_M$. Now assign to each nest A_m a weight $\eta_m > 0$. Let the cost function be given by

$$C(y) = \sum_{m=1}^{M}\left[\eta_m \sum_{r\in A_m} y_r \ln y_r + (1-\eta_m)\left(\sum_{r\in A_m} y_r\right)\ln\left(\sum_{j\in A_m} y_r\right)\right]. \tag{2.21}$$

Let j be in nest A_m. The nested logit probabilities (see e.g. Train (2009)) are given by

$$y_j^*(x) = \frac{e^{v_j(x_j)/\eta_m}}{\sum_{r\in A_m} e^{v_r(x_r)/\eta_m}} \frac{\left(\sum_{r\in A_m} e^{v_r(x_r)/\eta_m}\right)^{\eta_m}}{\sum_{s=1}^{M}\left(\sum_{r\in A_s} e^{v_r(x_r)/\eta_s}\right)^{\eta_s}}. \tag{2.22}$$

The first fraction is the probability of choosing j conditional on choosing the nest A_m. The second fraction is the probability of choosing the nest A_m. We see that $\eta_m = 1$ for $m = 1,\ldots,M$ corresponds to the standard logit formula. The following result builds on results in Fosgerau and Palma 2015.[23]

Proposition 2.9. *If $\eta_m > 0$ for $m = 1,\ldots,M$, then C defined in (2.21) generates the probabilities defined in (2.22). In particular, the nested logit formula is a perturbed utility model even when it is not an additive random utility model.*

When $\eta_m > 1$ for some nest m, it is known that (2.22) can be inconsistent with random utility maximization (see Train (2009) for discussion and references). We show that the inconsistency with random utility maximization occurs precisely when complementarity is allowed. To that end, let j and ℓ belong to the same nest A_m. We then

[23] Fosgerau and Palma 2015 consider the same cost function in the context of market shares and show the equivalence with nested logit. They formally require nesting parameters to be in the interval $(0,1]$. See their paper for a comprehensive analysis of "flexible entropy" models, which are a subclass of perturbed utility models.

obtain

$$\ln(y_j^*(x)) = \frac{v_j(x_j)}{\eta_m} - \frac{v_\ell(x_\ell)}{\eta_m} + \ln(y_\ell^*(x)).$$

If v_ℓ were a known function, η_m could be recovered by varying x_ℓ in this expression, holding x_j fixed. We must instead deal with the case of an *unknown* function v_ℓ. Let $x_{\ell,q}$ be a characteristic of alternative ℓ that is excluded from the characteristics of good j. Assuming differentiability and that $\partial \ln(y_\ell^*(x))/\partial x_{\ell,q} \neq 0$,

$$\begin{aligned}\frac{\partial \ln(y_j^*(x))/\partial x_{\ell,q}}{\partial \ln(y_\ell^*(x))/\partial x_{\ell,q}} &= \frac{\partial \ln(y_j^*(x))/\partial v_\ell}{\partial \ln(y_\ell^*(x))/\partial v_\ell} \\ &= -\left(1 + (\eta_m - 1)\frac{y_\ell^*(x)}{\sum_{r\in A_m} y_r^*(x)} - \eta_m y_\ell^*(x)\right)^{-1} + 1.\end{aligned} \tag{2.23}$$

This equation[24] shows that a derivative ratio directly identifies η_m since all other terms on the right hand side of (2.23) are identified. Moreover, only local information is needed to determine these nesting parameters, not full identification of v.

Equation (2.23) is strictly positive, indicating complementarity, whenever

$$\frac{\eta_m - 1}{\eta_m} > \sum_{r\in A_m} y_r^*(x). \tag{2.24}$$

Note that this *requires* $\eta_m > 1$.[25] We provide a possible interpretation of complementarity. If a nest is not typically noticed by the individual (think: $\sum_{r\in A_m} y_r^*(x)$ small), then making an alternative in the nest more attractive could make it more likely the individual notices *other* alternatives in the nest. This could result in the individual choosing those items more often due to this complementarity.

[24]See Appendix 2.13.1 for derivation. The sign of this derivative ratio (with logs) is the same as the sign of $DR_{j,\ell}(v(x))$.

[25]Whenever $\eta_m > 2$ for $m = 1,\ldots,k$ and there are two distinct nests with at least two alternatives, the weak inequality version of (2.24) must be satisfied for at least one nest m. Thus for *any* value of x, there are always two goods that are local complements.

If $\eta_m > 1$, then complementarity occurs when $\sum_{r \in A_m} y_r^*(x)$ is close to 0. When v can take on arbitrarily large and small values, we can always ensure that the sum of these probabilities is sufficiently small. This yields the following result.

Proposition 2.10. *Assume $v(\text{supp}(X)) = \mathbb{R}^k$ and for every x there is an alternative ℓ and a characteristic q such that $\partial \ln(y_\ell^*(x))/\partial x_{\ell,q} \neq 0$, and $\partial \ln(y_j^*(x))/\partial x_{\ell,q}$ exists. If $1 < \eta_m$, $j, \ell \in A_m$, and there exists $s \notin A_m$, then there exists a characteristic vector x such that*

$$\frac{\partial \ln(y_j^*(x))/\partial x_{\ell,q}}{\partial \ln(y_\ell^*(x))/\partial x_{\ell,q}} > 0.$$

2.6 Conclusion

This paper extends the definition of Hicksian complementarity to a wide variety of settings. Using the perturbed utility model, we provide a primitive definition of complementarity and show how to identify complementarity using characteristic variation. This definition agrees with the Hicksian definition and we show how to identify the sign of Hicksian substitution effects without price variation. The study of complementarity is important for reasons beyond just documenting whether two goods are substitutes or complements. We show that "no complementarity" is a testable implication of a large class of discrete choice models. Measures of complementarity can also be used for identification of structural parameters. For example, a logarithmic version of the derivative ratio (*DR*) measure of complementarity identifies nesting parameters in nested logit. We also relate our definition of complementarity with the concept of supermodularity. Complementarity and supermodularity are consistent in a bundles setting, but do not agree in the discrete choice setting.

2.7 Acknowledgements

Chapter 2 is being prepared for submission for publication. Allen, Roy; Rehbeck, John. The dissertation author was a primary author and investigator of this paper.

2.8 Appendix: Proofs for Section 2.2.1

In this section, we examine comparative statics with respect to characteristic variation and prove Proposition 2.1. We assume that we are evaluating demand functions at an interior point. Throughout the remainder of this section, assumptions on differentiability are left implicit in the analysis. Consider the Lagrangian given by the standard consumer problem

$$\mathcal{L}(y,\lambda,\mu) = \sum_{j=1}^{k} y_j v_j(x_j) + D(y) + \lambda \left(I - \sum_{j=1}^{k} p_j y_j \right) + \sum_{j=1}^{k} \mu_j y_j$$

with $\lambda \geqslant 0$ and $\mu_j \geqslant 0$. The first order conditions at an interior point yield

$$v_j(x_j) + D_j(y) - \lambda p_j = 0 \quad \text{for } j = 1,\ldots,k$$
$$\sum_{j=1}^{k} p_j y_j = I.$$

Let the function for Marshallian Demand of good j be denoted $y_j^* = y_j^*(p,x,I)$ and the Lagrange multiplier at the maximum be denoted $\lambda^* = \lambda^*(p,x,I)$. Assume y^* is a regular maximizer.[26] We suppress dependence on parameters throughout the analysis for

[26] See Varian (1992), p. 498. This ensures the second order conditions are satisfied with a strict inequality.

convenience. Let $p = (p_1, \ldots, p_k)' \in \mathbb{R}^k_+$ and

$$H = \begin{bmatrix} \nabla^2 D(y^*) & -p \\ -p' & 0 \end{bmatrix}.$$

Examining the total derivative with respect to p and applying Cramer's rule, we can write the Slutsky Equation

$$\begin{aligned}\frac{\partial y_j^*}{\partial p_\ell} &= \lambda^* \frac{(-1)^{\ell+j}\det(H_{\ell,j})}{\det(H)} + \frac{(-1)^{k+1+j}\det(H_{k+1,j})}{\det(H)} y_\ell^* \\ &= \frac{\partial h_j^*}{\partial p_\ell} - \frac{\partial y_j^*}{\partial I} y_\ell^*\end{aligned}$$

where $H_{r,c}$ is the submatrix of H which removes row r and column c. Moreover, let $h_j^* = h_j^*(p,x,u)$ be the Hicksian demand function for good j. Similarly, we can examine the change in Marshallian demand when there is a change in characteristic q of alternative ℓ ($x_{\ell,q}$). Differentiating the first order conditions and applying Cramer's rule yields

$$\frac{\partial y_j^*}{\partial x_{\ell,q}} = -\frac{\partial v_\ell}{\partial x_{\ell,q}} \frac{(-1)^{\ell+j}\det(H_{\ell,j})}{\det(H)}.$$

We can rewrite this as a scaled version of the Hicksian substitution term,

$$\frac{\partial y_j^*}{\partial x_{\ell,q}} = -\frac{\partial v_\ell / \partial x_{\ell,q}}{\lambda^*} \frac{\partial h_j^*}{\partial p_\ell}. \tag{2.25}$$

Thus,

$$\frac{\partial y_j^*}{\partial x_{\ell,q}} \Big/ \frac{\partial y_\ell^*}{\partial x_{\ell,q}} = \frac{\partial h_j^*}{\partial p_\ell} \Big/ \frac{\partial h_\ell^*}{\partial p_\ell}.$$

To obtain an additional result, one can substitute equation (2.25) into the Slutsky

Equation and rearrange to obtain

$$\frac{\partial y_j^*}{\partial x_{\ell,q}} = (-\mathrm{BS}_\ell \varepsilon_{j,I} - \varepsilon_{j,p_\ell}) \left(\frac{y_j^*}{y_\ell^* p_\ell} \right) \frac{(\partial v_\ell / \partial x_{\ell,q}) y_\ell^*}{\lambda^*}.$$

$\varepsilon_{j,p_\ell} = \frac{\partial y_j^*}{\partial p_\ell} \frac{p_\ell}{y_j^*}$ is the cross-price elasticity of demand for alternative j with respect to p_ℓ, $\varepsilon_{j,I} = \frac{\partial y_j^*}{\partial I} \frac{I}{y_j^*}$ is the income elasticity of demand for alternative j, and $\mathrm{BS}_\ell = \frac{p_\ell y_\ell^*}{I}$ is the budget share of alternative ℓ. We examine the units of the last term in the above equation and find that $\left(\frac{\partial v_\ell}{\partial x_{\ell,q}} y_\ell^* / \lambda^* \right)$ is in units $\left(\frac{\Delta\$}{\Delta\text{characteristic } q \text{ of good } l} \text{ quantity } \ell \right)$. One could interpret this term as the willingness to pay for a marginal increase in a characteristic at current demand.

2.9 Appendix: Proof of Theorem 2.1

We first provide a set of sufficient conditions on $\tilde{D}$, marginal utility shifters v_j, and budgets B such that $\operatorname{argmax}_{y \in B} U(y; X, \varepsilon)$ admits a measurable selection, where U is defined in (2.7). Recall that $y \in \mathbb{R}^k$ and let $\mathcal{B}_{\mathbb{R}^k}$ be the Borel σ-algebra induced by the Euclidean metric. Let $\varepsilon : \Omega_\varepsilon \to E$ be a random variable defined from the probability space $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, P_\varepsilon)$ to the measure space $(E, \mathcal{E})$. Let $X : \Omega_X \to (\mathbb{R}^d)^k$ be a random variable from the probability space $(\Omega_X, \mathcal{F}_X, P_X)$ to the measurable space $((\mathbb{R}^d)^k, \mathcal{B}_{(\mathbb{R}^d)^k})$ where $\mathcal{B}_{(\mathbb{R}^d)^k}$ is the Borel σ-algebra of $(\mathbb{R}^d)^k$. Note that the projection of X in the j-th component is a random variable to $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ where $\mathcal{B}_{\mathbb{R}^d}$ is the Borel σ-algebra of $\mathbb{R}^d$.

Assumption 2.5. *We assume $\tilde{D}$, marginal utility shifters v, and budget B satisfy the following:*

i. $\tilde{D} : \mathbb{R}^k \times E \to \bar{\mathbb{R}}$ is $\mathcal{E}$-measurable and continuous in $\mathbb{R}^k$.[27]

[27] $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ denotes the extended reals.

ii. For all $j = 1,\dots,k$, $v_j : \mathbb{R}^d \to \bar{\mathbb{R}}$ is $\mathcal{B}_{R^d}$-measurable

iii. Let $B \subseteq \mathbb{R}^n$ be a nonempty, compact, and convex budget

Lemma 2.1. *Under Assumption 2.5, there exists a measurable selector*

$$Y^*(X,\varepsilon) \in \operatorname*{argmax}_{y \in B} U(y;X,\varepsilon).$$

Proof. From (i) in Assumption 2.5 and Stinchcombe and White 1992 Lemma 2.15, $\tilde{D} : \mathbb{R}^k \times E \to \bar{\mathbb{R}}$ is $\mathcal{B}_{\mathbb{R}^k} \otimes \mathcal{E}$-measurable where $\mathcal{B}_{R^k}$ is the Borel σ-field of $\mathbb{R}^k$. From (ii) of Assumption 2.5, we have that $v_j(X_j)$ is $\mathcal{B}_{(\mathbb{R}^d)^k}$ measurable since it is a composition of a measurable function and the continuous projection of X to X_j. Therefore, each $g_j(y,X) = y_j v_j(X_j)$ is continuous in $\mathbb{R}^k$ for each X and $\mathcal{B}_{(R^d)^k}$-measurable for each $y \in \mathbb{R}^k$. By Stinchcombe and White 1992 Lemma 2.15, $g_j : \mathbb{R}^n \times (\mathbb{R}^d)^k$ is $\mathcal{B}_{\mathbb{R}^k} \otimes \mathcal{B}_{(\mathbb{R}^d)^k}$-measurable.

Since $\tilde{D}(y,\varepsilon)$ does not depend on X and $X \perp \varepsilon$, we may extend $\tilde{D}$ to be $\mathcal{B}_{\mathbb{R}^k} \otimes (\mathcal{B}_{(\mathbb{R}^d)^k} \otimes \mathcal{E})$-measurable.[28] Similarly, each $g_j(y,X)$ does not depend on ε and $X \perp \varepsilon$, so we may extend g_j to be $\mathcal{B}_{\mathbb{R}^k} \otimes (\mathcal{B}_{(\mathbb{R}^d)^k} \otimes \mathcal{E})$-measurable. Therefore, $U(y;X,\varepsilon) = \sum_{j=1}^k y_j v_j(X_j) + \tilde{D}(y,\varepsilon)$ is $\mathcal{B}_{\mathbb{R}^k} \otimes (\mathcal{B}_{(\mathbb{R}^d)^k} \otimes \mathcal{E})$-measurable since it is a sum of measurable functions. Moreover, for all $(X,\varepsilon) \in (\mathbb{R}^d)^k$, $U(\cdot;X,\varepsilon) : \mathbb{R}^k \to \bar{\mathbb{R}}$ is continuous.

Lastly from (iii) of Assumption 2.5, the assumptions of Corollary 2.21 and Lemma 2.22 in Stinchcombe and White 1992 are satisfied so for any probability measure P on $\left(((\mathbb{R}^d)^k, E), (\mathcal{B}_{(\mathbb{R}^d)^k} \otimes \mathcal{E})\right)$, there is a $(\mathcal{B}_{(\mathbb{R}^d)^k} \otimes \mathcal{E})^P$-measurable everywhere selection[29] Y^* that satisfies $Y^*(X,\varepsilon) \in \operatorname{argmax}_{y \in B} U(y;X,\varepsilon)$ for all $(X,\varepsilon) \in ((\mathbb{R}^d)^k, E)$. Note that $\operatorname{argmax}_{y \in B} U(y;X,\varepsilon)$ is non-empty and compact for all $(X,\varepsilon) \in ((\mathbb{R}^d)^k, E)$. □

[28] We suppress dependence of $\tilde{D}$ on X for convenience.

[29] $\mathcal{F}^P$ is the completion of a σ-field $\mathcal{F}$ with respect to the probability measure P.

The assumption that B does not depend on ε can be relaxed using conditions from Stinchcombe and White 1992.

Assumption 2.6. $\mathbb{E}[\tilde{D}(Y^*(X,\varepsilon),\varepsilon) \mid X = x]$ *exists and is finite.*

Assumption 2.6 is a high level condition that is easily checked in many examples. In additive random utility models, a sufficient condition is that ε has finite mean and is independent of X. Because we maintain the assumption that B is compact and convex, Assumption 2.6 implies that $\mathcal{G}(x)$ is finite.

Proof of Theorem 1. Fix $x \in \text{supp}(X)$. Let $\mathcal{Y}_x$ denote the set of measurable functions from $\text{supp}(\varepsilon)$ to B. We include the subscript x only for ease of exposition, but note that $\mathcal{Y}_x$ is defined with no reference to x. Under Assumption 2.6, $\mathcal{G}(x)$ is finite. Using the definition of $\mathcal{G}(x)$, we see that

$$\mathcal{G}(x) = \mathbb{E}\left[\max_{y\in B} U(y;X,\varepsilon) \mid X = x\right] \tag{2.26}$$

$$= \mathbb{E}\left[Y^*(X,\varepsilon)'v(x) \mid X = x\right] + \mathbb{E}[\tilde{D}(Y^*(X,\varepsilon),\varepsilon) \mid X = x] \tag{2.27}$$

$$= \mathbb{E}\left[Y^*(x,\varepsilon)\right]' v(x) + \mathbb{E}[\tilde{D}(Y^*(x,\varepsilon),\varepsilon) \mid X = x] \tag{2.28}$$

$$\leqslant \mathbb{E}\left[Y^*(x,\varepsilon)\right]' v(x) + \sup_{Y_x\in\mathcal{Y}_x:\mathbb{E}[Y_x(\varepsilon)\mid X=x]=\mathbb{E}[Y^*(x,\varepsilon)]} \mathbb{E}[\tilde{D}(Y_x(\varepsilon),\varepsilon) \mid X = x] \tag{2.29}$$

$$\leqslant \sup_{y\in B}\left\{\sum_{j=1}^{k} y_j v_j(x_j) + \sup_{Y_x\in\mathcal{Y}_x:\mathbb{E}[Y_x(\varepsilon)\mid X=x]=y} \mathbb{E}[\tilde{D}(Y_x(\varepsilon),\varepsilon) \mid X = x]\right\}^{30} \tag{2.30}$$

$$= \sup_{y\in B}\left\{\sum_{j=1}^{k} y_j v_j(x_j) + \sup_{Y_x\in\mathcal{Y}_x:\mathbb{E}[Y_x(\varepsilon)]=y} \mathbb{E}[\tilde{D}(Y_x(\varepsilon),\varepsilon)]\right\} \tag{2.31}$$

$$= \sup_{y\in B} \sup_{Y_x\in\mathcal{Y}_x:\mathbb{E}[Y_x(\varepsilon)]=y} \mathbb{E}\left[\sum_{j=1}^{k} y_j v_j(x_j) + \tilde{D}(Y_x(\varepsilon),\varepsilon)\right] \tag{2.32}$$

$$= \sup_{Y_x\in\mathcal{Y}_x} \mathbb{E}\left[\sum_{j=1}^{k} Y_{x,j}(\varepsilon) v_j(x_j) + \tilde{D}(Y_x(\varepsilon),\varepsilon)\right] \tag{2.33}$$

$$\leqslant \mathbb{E}\left[\sup_{y\in B} U(y;x,\varepsilon)\right] \tag{2.34}$$

$$= \mathcal{G}(x) \tag{2.35}$$

In the above derivation, $Y_{x,j}(\varepsilon)$ denotes the j-th component of $Y_x(\varepsilon)$. The final inequality follows from Jensen's inequality. We obtain that these inequalities are all equalities. This further implies existence of all expectations because $\mathcal{G}(x)$ is finite. Finally, we have that

$$\mathbb{E}[Y^*(X,\varepsilon) \mid X = x] \in \operatorname*{argmax}_{y\in B}\left\{\sum_{j=1}^{k} y_j v_j(x) + D(y)\right\},$$

where $D(y) = \sup_{Y_x\in\mathcal{Y}_x:\mathbb{E}[Y_x(\varepsilon)]=y} \mathbb{E}[\tilde{D}(Y_x(\varepsilon),\varepsilon)]$. Here we use that B is convex to ensure

[30] The supremum is taken to be $-\infty$ if there is no $Y_x \in \mathcal{Y}_x$ such that $\mathbb{E}[Y_x(\varepsilon) \mid X = x] = y$

that $\mathbb{E}[Y^*(X,\varepsilon) \mid X = x]$ is in the argmax set. It is clear that D does not depend on x; recall we included the subscript in $\mathcal{Y}_x$ purely for exposition. Finally, note that 2.6 ensures that $D(y)$ is finite when evaluated at $\mathbb{E}[Y^*(X,\varepsilon) \mid X = x]$ for any $x \in \operatorname{supp}(x)$. □

2.10 Appendix: A Monotonicity Lemma

We will use the following lemma in a few places. It is a version of the law of (compensated) demand.

Lemma 2.2. *Let $x, \tilde{x} \in \operatorname{supp}(X)$. Then*

i.

$$(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'(v(x) - v(\tilde{x})) \geqslant 0.$$

In particular, if $x = (x_1, \ldots, x_k)$ and $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_k)$ only differ with respect to the j-th component (i.e. $x_\ell = \tilde{x}_\ell$ when $\ell \neq j$), then

$$v_j(x_j) > v_j(\tilde{x}_j) \implies \mathbb{E}[Y_j \mid X = x] \geqslant \mathbb{E}[Y_j \mid X = \tilde{x}].$$

ii. Assume for x and $\tilde{x}$ that there is a unique maximizer to (2.1), or alternatively D is strictly concave and B is convex. Then whenever $\mathbb{E}[Y \mid X = x] \neq \mathbb{E}[Y \mid X = \tilde{x}]$,

$$(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'(v(x) - v(\tilde{x})) > 0.$$

In particular, if $x = (x_1, \ldots, x_k)$ and $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_k)$ only differ with respect to the j-th component, then

$$\mathbb{E}[Y_j \mid X = x] > \mathbb{E}[Y_j \mid X = \tilde{x}] \implies v_j(x_j) > v_j(\tilde{x}_j).$$

(i) and (ii) are shown in (Allen and Rehbeck (2016b)). (i) is straightforward and follows from writing out the necessary conditions for maximization. (ii) essentially follows from the definition of a unique maximizer.

2.11 Appendix: Proofs for Section 2.3

Proof of Proposition 2.4. This follows from Lemma 2.2 and the Fundamental Theorem of Calculus. □

Proof of Proposition 2.5. This result is closely related to Chebyshev's association inequality. We have by definition that,

$$(\mathbb{E}[Y_j \mid X = x] - \mathbb{E}[Y_j \mid X = \tilde{x}])(\mathbb{E}[Y_\ell \mid X = x] - \mathbb{E}[Y_\ell \mid X = \tilde{x}]) \geqslant 0, \qquad \forall x, \tilde{x} \in S.$$

Let X and $\tilde{X}$ be independent. Then,

$$\mathbb{E}\left[(\mathbb{E}[Y_j \mid X] - \mathbb{E}[Y_j \mid \tilde{X}])(\mathbb{E}[Y_\ell \mid X] - \mathbb{E}[Y_\ell \mid \tilde{X}]) \mid X, \tilde{X} \in S\right] \geqslant 0. \tag{2.36}$$

Using independence of X and $\tilde{X}$, we see the left hand side of (2.36) is equal to

$$\begin{aligned}
\mathbb{E}\Big[\mathbb{E}[Y_j \mid X]\mathbb{E}[Y_\ell \mid X] + \mathbb{E}[Y_j \mid \tilde{X}]\mathbb{E}[Y_\ell \mid \tilde{X}] \mid X, \tilde{X} \in S\Big] \\
&- \mathbb{E}\left[\mathbb{E}[Y_j \mid \tilde{X}]\mathbb{E}[Y_\ell \mid X] + \mathbb{E}[Y_j \mid \tilde{X}]\mathbb{E}[Y_\ell \mid X] \mid X, \tilde{X} \in S\right] \\
&= 2\mathbb{E}\left[\mathbb{E}[Y_j \mid X]\mathbb{E}[Y_\ell \mid X] \mid X \in S\right] - 2\mathbb{E}[Y_j \mid X \in S]\mathbb{E}[Y_\ell \mid X \in S] \\
&= 2\,\mathrm{Cov}(\mathbb{E}[Y_j \mid X], \mathbb{E}[Y_\ell \mid X] \mid X \in S).
\end{aligned}$$

Note that we used that

$$\mathbb{E}[\mathbb{E}[Y_j \mid \tilde{X}]\mathbb{E}[Y_\ell \mid X] \mid X, \tilde{X} \in S] = \mathbb{E}[Y_j \mid X \in S]\mathbb{E}[Y_\ell \mid X \in S],$$

which follows from the fact that X and $\tilde{X}$ are independent and identically distributed. This proves the first statement of the proposition. The second statement follows from analogous arguments. □

2.12 Appendix: Proofs for Section 2.4

Proof of Proposition 2.6. Recall that supermodularity is equivalent to assuming $\eta \cdot \Delta \geqslant 0$ almost surely. Suppose $x = (x_1, x_2)$ and $\tilde{x} = (\tilde{x}_1, x_2)$ differ only with respect to the characteristics of good 1. Suppose $v_1(x_1) > v_1(\tilde{x}_1)$. The proof of Lemma 1 of Fox and Lazzati (forthcoming) shows that $\mathbb{E}[Y_2 \mid X = x]$ is weakly increasing in v_1, and hence we conclude

$$\mathbb{E}[Y_2 \mid X = \tilde{x}] \geqslant \mathbb{E}[Y_2 \mid X = x].$$

Since $v_1(x_1) > v_1(\tilde{x}_1)$, from Lemma 2.2 we obtain

$$\mathbb{E}[Y_1 \mid X = \tilde{x}] \geqslant \mathbb{E}[Y_1 \mid X = x].$$

If instead $v_1(x_1) < v_1(\tilde{x}_1)$, then the opposite inequalities hold. Finally, if $v_1(x_1) = v_1(\tilde{x}_1)$, then $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y \mid X = \tilde{x}]$ because $\varepsilon \mid \eta$ has a density with respect to Lebesgue measure (so that utility ties occur with probability 0 and the argmax set of the perturbed utility model is a singleton). Thus we establish that regardless of the sign of

$v_1(x_1) - v_1(\tilde{x}_1)$,

$$C(x,\tilde{x}) =$$
$$(\mathbb{E}[Y_2 \mid X = \tilde{x}] - \mathbb{E}[Y_2 \mid X = x])\,(\mathbb{E}[Y_1 \mid X = \tilde{x}] - \mathbb{E}[Y_1 \mid X = x])$$
$$\geqslant 0.$$

From this we obtain that $DR_{2,1}(x) \geqslant 0$ when it exists.

The case $\eta \cdot \Delta \leqslant 0$ almost surely is proven by similar arguments. □

Proof of Proposition 2.7.

$$\begin{aligned}
\mathrm{Cov}(Y_1, Y_2 \mid X = x) &= P((Y_1, Y_2) = (1,1) \mid X = x) - P(Y_1 = 1 \mid X = x)P(Y_2 = 1 \mid X = x) \\
&= P(\varepsilon_1 \geqslant 0, \varepsilon_2 \geqslant 0) - P(\varepsilon_1 \geqslant 0)P(\varepsilon_2 \geqslant 0) \\
&= P(\varepsilon_1 \geqslant 0, \varepsilon_2 \geqslant 0) - 1/4.
\end{aligned}$$

Finally, note $P(\varepsilon_1 \geqslant 0, \varepsilon_2 \geqslant 0) > 1/4$ if and only if $\mathrm{Cov}(\varepsilon_1, \varepsilon_2 \mid X = x) > 0$. □

2.13 Appendix: Proofs for Section 2.5

Proof of Proposition 2.8. Since C is additively separable, $C(y) = \sum_{j=1}^{k} c_j(y_j)$. Let $v_\ell(x_\ell) > v_\ell(\tilde{x}_\ell)$, and let $x = (x_1, \ldots, x_k)$ and $\tilde{x} = (x_1, \ldots, \tilde{x}_\ell, \ldots x_k)$ only differ with respect to the ℓ-th component. By differencing the first order conditions of the respective

Lagrangians (see (2.39)) we obtain

$$v_\ell(x_\ell) - v_j(x_j) = c'_\ell(y^*_\ell(x)) - c'_j(y^*_j(x))$$
$$v_\ell(\tilde{x}_\ell) - v_j(x_j) = c'_\ell(y^*_\ell(\tilde{x})) - c'_j(y^*_j(\tilde{x})).$$

Differencing these again yields,

$$v_\ell(x_\ell) - v_\ell(\tilde{x}_\ell) - \left[c'_\ell(y^*_\ell(x)) - c'_\ell(y^*_\ell(\tilde{x}))\right] = c'_j(y^*_j(\tilde{x})) - c'_j(y^*_j(x)). \tag{2.37}$$

Suppose for the purpose of contradiction that $y^*_j(x) \geqslant y^*_j(\tilde{x})$ for some $j \neq \ell$. Then by strict monotonicity of the derivatives of strictly convex functions and (2.37) we have that this must hold for every $j \neq \ell$. Moreover, Lemma 2.2 implies $y^*_\ell(x) \geqslant y^*_\ell(\tilde{x})$. Note we cannot have $y^*_j(x) = y^*_j(\tilde{x})$ for every j, since then we would violate (2.37) (recall $v_\ell(x_\ell) > v_\ell(\tilde{x}_\ell)$). Thus we must have $y^*_j(x) \geqslant y^*_j(\tilde{x})$ for all j, with at least one strict inequality. This violates that all probabilities must sum to 1. Hence we obtain, by contradiction, that $y^*_j(x) < y^*_j(\tilde{x})$. This implies that $DCM_{j,\ell}(x) < 0$, as desired. □

We now characterize complementarity for an arbitrary number of goods. The Lagrangian is given by

$$\mathcal{L}(y, \lambda, \mu) = \sum_{j=1}^{k} y_j v_j(x_j) - C(y) + \lambda \left(1 - \sum_{j=1}^{k} y_j\right) + \sum_{j=1}^{k} \mu_j y_j \tag{2.38}$$

for $\lambda \in \mathbb{R}$ and $\mu_j \geqslant 0$.

We assume that $C(y) = C(y_1, \ldots, y_k)$ is a twice continuously differentiable strictly convex function and $v_j(\cdot)$ are continuously differentiable functions in x_j. We define $C_j(\cdot)$ as the partial derivative with respect to the j-th component. Similarly, $C_{j,\ell}(\cdot) = \frac{\partial^2}{\partial y_j \partial y_\ell} C(\cdot)$.

Lastly, assume that the solution y^* is at a point on the relative interior of the simplex. We consider changes in choice probabilities when there is a change in the q-th characteristic of good ℓ (denoted $x_{\ell,q}$).

Consider the optimization of (2.38) at an interior point. In this case, $\mu_j = 0$ for every j so that first order conditions are given by

$$v_j(x_j) - C_j(y) - \lambda = 0 \quad \text{for } j = 1, \ldots, k \tag{2.39}$$

$$\sum_{j=1}^{k} y_j = 1. \tag{2.40}$$

Define the optimal probability of choosing object j by $y_j^* = y_j^*(x_1, \ldots, x_n)$ and the Lagrange multiplier as $\lambda^* = \lambda^*(x_1, \ldots, x_k)$. Assume y^* is a regular maximizer. We suppress dependence on the characteristics for convenience. We focus on a change in $x_{1,q}$. Similar to the standard consumer problem, we examine the total differential with respect to the characteristic $x_{1,q}$. This gives the system of equations,

$$\begin{aligned}
\frac{\partial v_1}{\partial x_{1,q}} - \sum_{j=1}^{k} C_{1,j}(y^*) \frac{\partial y_j^*}{\partial x_{1,q}} - \frac{\partial \lambda^*}{\partial x_{1,q}} &= 0 \\
- \sum_{j=1}^{k} C_{j,\ell}(y^*) \frac{\partial y_j^*}{\partial x_{1,q}} - \frac{\partial \lambda^*}{\partial x_{1,q}} &= 0 \quad \text{for all } \ell \neq 1 \\
\sum_{j=1}^{k} \frac{\partial y_j^*}{\partial x_{1,q}} &= 0.
\end{aligned}$$

Let $\mathbf{1} = (1, \ldots, 1)' \in \mathbb{R}^k$ and

$$H = \begin{bmatrix} -\nabla^2 C(y^*) & -\mathbf{1} \\ -\mathbf{1}' & 0 \end{bmatrix}.$$

Combining H with the above system of equations gives

$$H \begin{bmatrix} \frac{\partial y_1^*}{\partial x_{1,q}} \\ \frac{\partial y_2^*}{\partial x_{1,q}} \\ \vdots \\ \frac{\partial y_k^*}{\partial x_{1,q}} \\ \frac{\partial \lambda^*}{\partial x_{1,q}} \end{bmatrix} = \begin{bmatrix} -\frac{\partial v_1}{\partial x_{1,q}} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Let $H_{r,c}$ be the submatrix of H which removes row r and column c. Now, use Cramer's rule and the definition of the determinant to obtain

$$\frac{\partial y_\ell^*}{\partial x_{1,q}} = -\frac{\partial v_1}{\partial x_{1,q}} \left(\frac{(-1)^{1+\ell} \det(H_{1,\ell})}{\det(H)} \right).$$

We assume these derivatives are non-zero. For $\ell = 1$, by second order conditions of a maximizer $\left(\frac{(-1)^{1+\ell} \det(H_{1,\ell})}{\det(H)} \right) < 0$ (recall we assume this is a regular maximizer; see Varian 1992, Howard 1977.). Thus,

$$\operatorname{sign}\left(\frac{\partial y_1^*}{\partial x_{1,q}} \right) = \operatorname{sign}\left(\frac{\partial v_1}{\partial x_{1,q}} \right).$$

Using the *DR* definition of complementary, we examine if

$$DR_{\ell,1}(x) = \frac{\partial y_\ell^*}{\partial x_{1,q}} \Big/ \frac{\partial y_1^*}{\partial x_{1,q}} = \frac{(-1)^{1+\ell} \det(H_{1,\ell})}{\det(H_{1,1})} \geqslant 0$$

for $\ell \neq 1$. Note the sign of $DR_{\ell,1}(x)$ depends only on C and may be positive or negative when $k \geqslant 3$ depending on the magnitude of the second derivatives. The equation in the main text is obtained from expanding the determinants above when $k = 3$.

2.13.1 Nested Logit Derivations

Proof of Proposition 2.9. Note that $0 < \eta_m$ for $m = 1, \ldots, M$ ensures we have an interior solution. To see this, recall $v_j(x_j)$ is finite for each j. Suppose for the purpose of contradiction that $y_j^* = 0$, $y_k^* \neq 0$. A differential shift in probability from y_k^* to y_j^* decreases $C(y^*)$ at rate ∞, so this cannot be a maximizer.

Let $m(j)$ denote the nest of alternative j. For simplicity, write $v_j := v_j(x_j)$ for $j = 1, \ldots, k$. The first order conditions of the Lagrangian for arbitrary y_j^* are given by,

$$v_j - \eta_{m(j)} \ln(y_j^*) - (1 - \eta_{m(j)}) \ln \left(\sum_{q \in A_{m(j)}} y_q^* \right) = \lambda + 1,$$

where λ is the Lagrange multiplier on the probability simplex constraint. For j, q in the same nest we conclude

$$y_q^* = y_j^* \frac{e^{v_q/\eta_{m(j)}}}{e^{v_j/\eta_{m(j)}}}.$$

Hence we may write

$$\sum_{q \in A_{m(j)}} y_q^* = \frac{y_j^*}{e^{v_j/\eta_{m(j)}}} \left(\sum_{q \in A_{m(j)}} e^{v_q/\eta_{m(j)}} \right),$$

and so our first order conditions can be rewritten after some rearrangement as

$$\frac{v_j}{\eta_{m(j)}} - \ln(y_j^*) - (1 - \eta_{m(j)}) \ln \left(\sum_{q \in A_{m(j)}} e^{v_q/\eta_{m(j)}} \right) = \lambda + 1.$$

Now for arbitrary j, ℓ not necessarily in the same nest, by differencing the first order conditions for j and ℓ and rearranging we obtain,

$$\ln(y_j^*) - \ln(y_\ell^*) =$$

$$\left(\frac{v_j}{\eta_{m(j)}} + \ln\left(\sum_{q\in A_{m(j)}} e^{v_q/\eta_{m(j)}}\right)^{\eta_{m(j)}-1}\right) - \left(\frac{v_\ell}{\eta_{m(\ell)}} + \ln\left(\sum_{q\in A_{m(j)}} e^{v_q/\eta_{m(j)}}\right)^{\eta_{m(\ell)}-1}\right).$$

By exponentiating, we see that the ratio of y_j^*/y_ℓ^* coincides exactly with that of $y_j^*(x)/y_\ell^*(x)$ generated according to (2.22). Since the solution to our constrained maximization problem is determined uniquely by ratios of probabilities, this completes the proof. □

Derivation of Equation (2.23). Let j and ℓ belong to nest A_m. From (2.22), we obtain

$$\frac{y_j^*(x)}{y_\ell^*(x)} = \frac{e^{v_j(x_j)/\eta_m}}{e^{v_\ell(x_\ell)/\eta_m}},$$

and hence

$$\ln(y_j^*(x)) = \frac{v_j(x_j)}{\eta_m} - \frac{v_\ell(x_\ell)}{\eta_m} + \ln(y_\ell^*(x)). \tag{2.41}$$

Let $x_{\ell,q}$ be a characteristic of alternative ℓ that is excluded from the characteristics of good j. Assuming differentiability, taking the derivative of (2.41) we obtain,

$$\frac{\partial \ln(y_j^*(x))}{\partial v_\ell}\frac{\partial v_\ell}{\partial x_{\ell,q}} = -\frac{1}{\eta_m}\frac{\partial v_\ell}{\partial x_{\ell,q}} + \frac{\partial \ln(y_\ell^*(x))}{\partial v_\ell}\frac{\partial v_\ell}{\partial x_{\ell,q}}. \tag{2.42}$$

By direct computation, we see,

$$\begin{aligned}\frac{\partial \ln y_\ell^*(x)}{\partial v_\ell} &= \frac{1}{\eta_m} + \left(\frac{\eta_m - 1}{\eta_m}\right) \frac{e^{v_\ell/\eta_m}}{\sum_{r\in A_m} e^{v_r(x_r)/\eta_m}} - \frac{\left(\sum_{r\in A_m} e^{v_r(x_r)/\eta_m}\right)^{\eta_m - 1} e^{v_\ell/\eta_m}}{\sum_{q=1}^{M} \left(\sum_{r\in A_q} e^{v_r(x_r)/\eta_q}\right)^{\eta_q}} \\ &= \frac{1}{\eta_m} - \left(\frac{\eta_m - 1}{\eta_m}\right) \frac{y_\ell^*(x)}{\sum_{r\in A_m} y_r^*(x)} - y_\ell^*(x).\end{aligned}$$

The replacement by terms with probabilities can be made by inspecting (2.22). Combining this equation with (2.42), we obtain (2.23). □

Proof of Proposition 2.10. This follows from (2.23). □

2.14 Appendix: Estimation with Instrumental Variables

Suppose that there are 2 alternatives and X_1 is a characteristic of alternative 1. Consider the reduced form,

$$\begin{aligned}Y_2 &= \alpha + \beta\pi X_1 + \varepsilon_2 \\ Y_1 &= \gamma + \pi X_1 + \varepsilon_1.\end{aligned}$$

If $\mathbb{E}[\varepsilon_j \mid X_1] = 0$ for $j = 1, 2$, it can be shown that this reduced form is a perturbed utility model with $v_1(X_1) = \pi X_1$, $v_2(X_2) = 0$, and a particular choice of D. We obtain (for $\pi \neq 0$),

$$DR_{2,1}(v(x_1)) = \frac{\frac{\partial \mathbb{E}[Y_2 | X_1 = x_1]}{\partial x_1}}{\frac{\partial \mathbb{E}[Y_1 | X_1 = x_1]}{\partial x_1}} = \frac{\beta\pi}{\pi} = \beta.$$

Treating X_1 as a instrument for Y_1, β can be estimated by standard instrumental variables methods. When there are 3 or more alternatives or more than 1 instrument, this analysis is more complicated. One reason is that the perturbed utility model allows conditional means that are linear in covariates, *but* it places certain restrictions on the parameters. Thus, a model-constent estimation approach would require enforcing these parameter restrictions.

For a non-linear generalization of the preceeding analysis, now consider the reduced form.

$$Y_2 = g(f(X_1)) + \varepsilon_1$$
$$Y_1 = f(X_1) + \varepsilon_2,$$

where we assume $\mathbb{E}[\varepsilon_j \mid X_1] = 0$ for $j = 1,2$ and again interpret X_1 as a characteristic specific to alternative 1.[31] Then provided the derivatives exist and $f' > 0$, we have that

$$\frac{\frac{\partial \mathbb{E}[Y_2|X_1=x_1]}{\partial x_1}}{\frac{\partial \mathbb{E}[Y_1|X_1=x_1]}{\partial x_1}} = \frac{g'(f(X_1))f'(X_1)}{f'(X_1)} = g'(f(X_1)),$$

so that complementarity is determined by the sign of g'.

[31]This reduced form is consistent with a perturbed model with $v_1(X_1) = f(X_1)$, $v_j(\cdot) = 0$ for $j \neq 1$, and some D.

Chapter 3

Testing Moment Inequalities: Selection Versus Recentering

Abstract. This paper provides finite-sample power comparisons and improvements for some existing tests of moment inequalities. We show that the "two-step" recentering approach of Romano, Shaikh, and Wolf (2014) rejects whenever a moment selection approach of Chernozukov, Chetverikov, and Kato (2014) rejects. Motivated by the computational appeal of moment selection, we show how to improve the power of the latter procedure by adding a correction factor to the moment selection threshold.

3.1 Introduction

Moment inequalities have shown up in many forms, from testing parameter restrictions in regressions (Wolak (1987)) to inference in partially identified models (Imbens and Manski (2004), Stoye (2009)).[1] Motivated by possible power improvements, many proposed tests of moment inequalities have incorporated some form of pre-test (Andrews and Soares (2010), Andrews and Barwick (2012)). The most common form of pre-testing amounts to selecting a data-dependent point in the compound null hypothesis and constructing critical values at that point. This can raise the rejection probability against many alternatives while lowering it (hopefully just slightly) against others. This paper makes two contributions to this literature.

First, we show that the recentering approach of Romano, Shaikh, and Wolf (2014) rejects whenever a moment selection approach of Chernozukov, Chetverikov, and Kato (2014) rejects. (Henceforth RSW and CCK, respectively.) We intentionally make this observation *only* in the setting of a "small" number of moment inequalities: the approach of CCK is motivated by testing a large number of moment inequalities relative to the sample size and the approach of RSW has not been shown to asymptotically control size in this setting.[2]

Next, we propose an improvement to the approach of CCK that retains its computational appeal. This *improved moment selection* (IMS) adds a correction factor to the moment selection threshold of CCK, which weakly increases power. Despite the power improvement, IMS is still (weakly) less powerful than RSW.

[1] See Ho and Rosen (2015) and Canay and Shaikh (2016) for reviews of theoretical and empirical work.

[2] With the aim of testing hypotheses on parameters in partially identified models, CCK also introduce a novel three-step test that is not analyzed here.

3.2 Setup and Preliminaries

We now formalize our hypothesis testing problem and introduce most of our notation. We are interested in testing the hypothesis that a finite vector of means is weakly negative,

$$H_0 : \mu_j \leqslant 0, \text{ for } j = 1, \ldots, p, \qquad H_a : \mu_j > 0 \text{ for some } j. \tag{3.1}$$

Suppose we have an independent and identically distributed (i.i.d.) sample of random vectors $X_1, \ldots, X_n$, each in $\mathbb{R}^p$, where $X_i = (X_{i1}, \ldots, X_{ip})^T$ and $\mu_j = \mathbb{E}[X_{1j}]$. We focus on tests that reject for large values of the statistic

$$T(X) = \max_{1 \leqslant j \leqslant p} \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j}, \tag{3.2}$$

where $X = (X_1, \ldots, X_n)$, $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^n X_{ij}$, and $\hat{\sigma}_j^2 = \frac{1}{n}\sum_{i=1}^n (X_{ij} - \hat{\mu}_j)^2$.

Now define $J = \{1, \ldots, p\}$ and for a set $H \subseteq J$ let $q_\gamma(H)$ satisfy

$$P^* \left(\max_{j \in H} \frac{\sqrt{n}\hat{\mu}_j^*}{\hat{\sigma}_j} \geqslant q_\gamma(H) \right) = \gamma,^3 \tag{3.3}$$

where P^* is a probability measure over $\hat{\mu}^* = (\hat{\mu}_1^*, \ldots, \hat{\mu}_p^*)$. Let $q_\gamma(\varnothing) = \infty$. The measure P^* may be interpreted as a bootstrap distribution. We treat this measure abstractly because the power comparisons in this paper are numerical results that hold for *any* probability measure, including the empirical bootstrap and other types of exchangeable bootstraps (Præstgaard and Wellner (1993)).

[3] For technical convenience we assume γ is chosen such that this probability can be set equal to γ.

Now for a vector $m \in \{\mathbb{R} \cup \{-\infty\}\}^p$, define $\tilde{q}_\gamma(m)$ by

$$P^*\left(\max_{j \in J} \frac{\sqrt{n}\hat{\mu}_j^* + \sqrt{n}m_j}{\hat{\sigma}_j} \geqslant \tilde{q}_\gamma(m)\right) = \gamma.^4 \quad (3.4)$$

Importantly, P^* is the same probability measure over $\hat{\mu}^*$ as that used to define $\tilde{q}_\gamma(H)$. Note that letting m be a vector with j^{th} component given by $m_j = 0$ if $j \in H$ and $-\infty$ otherwise, we have the relationship $q_\gamma(H) = \tilde{q}_\gamma(m)$.

3.3 Power Ranking

In this section we show that the *moment recentering* approach of RSW is weakly more powerful than a version of *moment selection* studied in CCK.

We first describe the moment recentering approach of RSW. First choose β such that $0 \leqslant \beta < \alpha$. Positive β helps direct power to alternatives in which $\mu_j < 0$ for some j. RSW choose the recentering vector $\tilde{\mu}$ defined by

$$\tilde{\mu}_j = \min\left\{\hat{\mu}_j + \frac{\hat{\sigma}_j}{\sqrt{n}} q_\beta(J), 0\right\}. \quad (3.5)$$

The moment recentering test of H_0 is given by

$$\phi^{RSW} = 1\left\{T(X) \geqslant \tilde{q}_{\alpha-\beta}(\tilde{\mu}) \text{ and } T(X) \geqslant 0\right\}, \quad (3.6)$$

where $\phi^{RSW} = 1$ denotes rejection of H_0. We include the requirement that $T(X) \geqslant 0$ (i.e. at least one sample mean needs to violate the null) since we deal with abstract P^*.

We now describe a version of the *moment selection* approach (see in particular

[4] Again, for simplicity we assume $\tilde{q}_\gamma(m)$ can be chosen to satisfy this probability equality.

CCK, Section 4.2.2). In order for the test to be well-defined we assume $\beta < \alpha/2$. We obtain a preliminary estimate of the set of "relevant" moments,

$$\hat{J} = \left\{ j \in J : \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j} \geqslant -2q_\beta(J) \right\}. \tag{3.7}$$

The moment selection approach can be thought of as guessing $\mu_j = 0$ if $j \in \hat{J}$, and $\mu_j = -\infty$ otherwise. The moment selection test of H_0 is given by

$$\phi^{MS} = 1\left\{T(X) \geqslant q_{\alpha-2\beta}(\hat{J}) \text{ and } T(X) \geqslant 0\right\},[5] \tag{3.8}$$

where $\phi^{MS} = 1$ denotes rejection of H_0.

We formalize our assumptions.

Assumption 3.1. *$q_\beta(J), q_{\alpha-2\beta}(\hat{J})$, and $\tilde{q}_{\alpha-\beta}(\tilde{\mu})$ exist and are defined by (3.3) and (3.4). They are constructed with the same measure P^*. $\beta < \alpha/2$.*

Assumption 3.2. *The same value $q_\beta(J)$ is used in (3.5) and (3.7).*

Existence in Assumption 3.1 is mild and is mainly for technical convenience.

Proposition 3.1. *Under Assumptions 3.1 and 3.2,*

$$\phi^{MS} \leqslant \phi^{RSW}. \tag{3.9}$$

In order to formally relate ϕ^{MS} and ϕ^{RSW}, it is useful to note that

$$\hat{J} = \left\{ j : \frac{\sqrt{n}\tilde{\mu}_j}{\hat{\sigma}_j} \geqslant -q_\beta(J) \right\}. \tag{3.10}$$

The approach of CCK may be seen as first constructing $\tilde{\mu}$, then replacing $\tilde{\mu}_j$ by 0 if $j \in \hat{J}$

[5] Recall $q_{\alpha-2\beta}(\varnothing) = \infty$.

and $-q_\beta(J)$ otherwise. Intuitively, this can lose some of the "heterogeneity" information in the moments. Finally, components satisfying $j \notin \hat{J}$ are deleted. This deletion comes at a cost of an additional β, which is why the final critical value is $q_{\alpha-2\beta}(\hat{J})$ instead of $q_{\alpha-\beta}(\hat{J})$.

Remark 3.1 (Other Statistics). With appropriate modifications, Proposition 3.1 extends to most statistics in the moment inequalities literature, including for example

$$\sum_{j=1}^{p} \max\left\{0, \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j}\right\}^2.$$

The key feature of this statistic is that it is flat in $\hat{\mu}_j$ whenever $\hat{\mu}_j < 0$. We omit the details for brevity since the next section contains results specific to T.

3.4 Improved Moment Selection

Motivated by the computational appeal of moment selection procedures, we propose an improved moment selection test. This test weakly dominates ϕ^{MS} in terms of power but in turn is weakly dominated by ϕ^{RSW}.

The improved moment selection test constructs the following estimate of "relevant" moments,

$$\hat{J}^I = \left\{ j \in J : \frac{\sqrt{n}\hat{\mu}_j}{\hat{\sigma}_j} \geqslant -2q_\beta(J) + \kappa \right\}, \tag{3.11}$$

where

$$\kappa = \max\left\{ \min_{j \in J} q_{\alpha-2\beta}(j), 0 \right\}.[6] \tag{3.12}$$

$\hat{J}^I$ differs from $\hat{J}$ (defined in (3.7)) due to the presence of κ, and we see $\hat{J}^I \subseteq \hat{J}$. κ can be

[6]In practice, for many choices of P^* a lower bound on κ can be approximated as $\Phi^{-1}(1-\alpha+2\beta)$.

introduced due to the *a priori* knowledge that rejection only occurs when (at least) one moment is characterized as binding.

The improved moment selection test is given by,

$$\phi^{IMS} = 1\left\{T(X) \geqslant q_{\alpha-2\beta}(\hat{J}^I) \text{ and } T(X) \geqslant 0\right\}. \tag{3.13}$$

To be explicit we state the following assumption.

Assumption 3.3. *The same value* $q_\beta(J)$ *is used in (3.5), (3.7) and (3.11).* $q_{\alpha-2\beta}(\hat{J}^I)$ *exists.* κ *provides an a priori bound on the critical value.* $\beta < \alpha/3$.

Proposition 3.2. *Under Assumptions 3.1 and 3.3,*

$$\phi^{MS} \leqslant \phi^{IMS} \leqslant \phi^{RSW}. \tag{3.14}$$

Remark 3.2 (Size Control). RSW show that when P^* is constructed by the empirical bootstrap,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} \mathbb{E}_P[\phi^{RSW}] \leqslant \alpha,$$

where $\mathcal{P}_0$ is a set of null distributions that satisfy a uniform integrability condition. This directly implies

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} \mathbb{E}_P[\phi^{IMS}] \leqslant \alpha.$$

3.5 Monte Carlo Study

We conduct a small Monte Carlo study to illustrate an example of the power gap between the procedures discussed in this paper. The rejection probabilities are illustrated in Table 3.1.

We set nominal size at $\alpha = .1$ and set $\beta = \alpha/10$ following the recommendation of RSW. Our power calculations are reported in Table 3.1 and are conducted with the following design:

$$n = 100, p = 50, \qquad X_i \sim N(\mu, I_{p\times p}),$$

where $I_{p\times p}$ is the $p \times p$ identity matrix. μ_1 varies from 0 to .4 to trace out a power curve. For the other means $j \geqslant 2$, we set $\mu_j = 5\left(-.1 - \frac{(j-2)}{49-2}\right)$.

Our simulations cover the 3 tests discussed in this paper: ϕ^{MS}, ϕ^{IMS}, and ϕ^{RSW}. We run 5,000 Monte Carlo simulations. Each quantile needed for each test is approximated with 999 bootstrap repetitions. P^* is given by the Gaussian multiplier bootstrap of CCK. To describe this bootstrap, let $\{w_i\}_{i=1}^n$ be i.i.d. standard normal, independent of $\{X_i\}_{i=1}^n$. Now let $\hat{\mu}_j$ have j^{th} component given by,

$$\hat{\mu}_j^* = \frac{1}{n}\sum_{i=1}^{n} w_i X_{ij}.$$

The Gaussian distribution over $\{w_i\}_{i=1}^n$ induces the distribution P^* over $\hat{\mu}^*$.

Table 3.1: Power Curve

	μ_1								
Test	0	.05	.1	0.15	.2	.25	0.3	.35	.4
MS	2.9	8.4	18.7	34.5	53.4	72.5	86.1	94.3	98.2
IMS	5.1	12.9	26.3	43.2	62.6	79.7	90.6	96.1	99.2
RSW	7.9	18.1	33.8	53.2	72.2	86.2	94.0	98.0	99.7

Notes: Entries in the table denote rejection probabilities. Nominal size is 10%. Each scenario is based on 5,000 Monte Carlo repetitions, $p = 50$ moment inequalities, and $n = 100$ observations.

As shown in Table 3.1, the power differences between the three procedures are moderately large. Recall we construct the means to be heterogeneous. RSW use additional heterogeneity in the sample moments relative to moment selection which helps understand why this design yields power differences.

3.6 Discussion

This paper shows a nesting of the rejection regions of an existing form of moment selection, an improved form of moment selection, and the recentering approach of RSW. This has computational implications beyond the power comparison, since it shows that implementing RSW can be done in two steps: first do a moment selection test (which need not involve a bootstrap); then run standard RSW only when the first step fails to reject. Further, the result provides impetus to extend the approach of RSW to more general settings beyond testing a finite number of moment inequalities.

3.7 Acknowledgements

Chapter 3 is being prepared for submission for publication. Allen, Roy. The dissertation author was the sole author.

3.8 Appendix

Proof of Proposition 3.1. We consider the case $q_{\alpha-2\beta}(\hat{J}) \geqslant 0$ since both tests reject only when $T(X) \geqslant 0$.

Recall that P^* is a measure only over $\hat{\mu}^*$, and so other terms (including $\hat{J}, \hat{\sigma}_j, \tilde{\mu}_j$)

are treated as constants in the following derivation:

$$\begin{aligned}
&P^*\left(\max_{j\in J}\frac{\sqrt{n}\hat{\mu}_j^* + \sqrt{n}\tilde{\mu}_j}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J})\right)\\
&\leqslant P^*\left(\max_{j\in \hat{J}}\frac{\sqrt{n}\hat{\mu}_j^*}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J})\right) + P^*\left(\max_{j\in \hat{J}^c}\frac{\sqrt{n}\hat{\mu}_j^* - \hat{\sigma}_j q_\beta(J))}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J})\right)\\
&\leqslant P^*\left(\max_{j\in \hat{J}}\frac{\sqrt{n}\hat{\mu}_j^*}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J})\right) + P^*\left(\max_{j\in \hat{J}^c}\frac{\sqrt{n}\hat{\mu}_j^* - \hat{\sigma}_j q_\beta(J))}{\hat{\sigma}_j} \geqslant 0\right)\\
&\leqslant (\alpha - 2\beta) + \beta.
\end{aligned} \tag{3.15}$$

The first inequality is due to a (weakly) conservative recentering, replacing $\sqrt{n}\tilde{\mu}_j$ with 0 if $j \in \hat{J}$ and with $-\hat{\sigma}_j q_\beta(J)$ otherwise; this is conservative by construction of $\hat{J}$ as easily seen in (3.10). The second inequality replaces $q_{\alpha-2\beta}(\hat{J})$ ($\geqslant 0$) with 0. The third inequality follows from the definitions of $q_{\alpha-2\beta}(\hat{J})$ and $q_\beta(J)$.

Recall that $q_{\alpha-\beta}(\tilde{\mu})$ is *defined* as the $(1-\alpha+\beta)$-quantile of the distribution of $\max_{j\in J}\frac{\sqrt{n}\hat{\mu}_j^* + \sqrt{n}\tilde{\mu}_j}{\hat{\sigma}_j}$ under P^*. From this we conclude that $q_{\alpha-2\beta}(\hat{J}) \geqslant \tilde{q}_{\alpha-\beta}(\tilde{\mu})$, and hence $\phi^{MS} \leqslant \phi^{RSW}$.

□

Proof of Proposition 3.2. The first inequality is straightforward. $\kappa \geqslant 0$ implies $\hat{J}^I \subseteq \hat{J}$ which in turn implies $q_{\alpha-2\beta}(\hat{J}^I) \leqslant q_{\alpha-2\beta}(\hat{J})$.

Now note that

$$\hat{J}^I = \left\{ j : \frac{\sqrt{n}\tilde{\mu}_j}{\hat{\sigma}_j} \geqslant -q_\beta(J) + \kappa \right\}.$$

The rest of the proof mirrors the proof of Proposition 3.1.

$$
\begin{aligned}
&P^*\left(\max_{j\in J}\frac{\sqrt{n}\hat{\mu}_j^* + \sqrt{n}\tilde{\mu}_j}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J}^I)\right)\\
&\leqslant P^*\left(\max_{j\in \hat{J}^I}\frac{\sqrt{n}\hat{\mu}_j^*}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J}^I)\right) + P^*\left(\max_{j\in (\hat{J}^I)^c}\frac{\sqrt{n}\hat{\mu}_j^* - \hat{\sigma}_j(q_\beta(J)-\kappa)}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J}^I)\right)\\
&\leqslant P^*\left(\max_{j\in \hat{J}^I}\frac{\sqrt{n}\hat{\mu}_j^*}{\hat{\sigma}_j} \geqslant q_{\alpha-2\beta}(\hat{J}^I)\right) + P^*\left(\max_{j\in (\hat{J}^I)^c}\frac{\sqrt{n}\hat{\mu}_j^* - \hat{\sigma}_j q_\beta(J))}{\hat{\sigma}_j} \geqslant 0\right)\\
&\leqslant (\alpha-2\beta)+\beta.
\end{aligned}
\tag{3.16}
$$

Recall that we may assume without loss of generality that $q_{\alpha-2\beta}(\hat{J}^I) \geqslant 0$. The first inequality follows from the previous arguments and the fact that $\kappa \leqslant q_{\alpha-2\beta}(J) \leqslant q_\beta(J)$ since $\beta \leqslant \alpha/3$. The second inequality follows from the fact that $\kappa \leqslant q_{\alpha-2\beta}(\hat{J}^I)$, which holds by construction. The rest of the proof follows from previous arguments.

□

Bibliography

Afriat, Sidney N. 1967. “The Construction of Utility Functions from Expenditure Data”. *International Economic Review* 8 (1): 67–77.

Agarwal, Nikhil, and Paulo Somaini. 2014. “Demand Analysis using Strategic Reports: An application to a school choice mechanism”. *NBER Working Paper 20775*.

Agranov, Marina, and Pietro Ortoleva. 2017. “Stochastic choice and preferences for randomization”. *Journal of Political Economy* 125 (1): 40–68.

Allen, Roy, and John Rehbeck. 2016a. “Complementarity in Perturbed Utility Models”. *Working Paper*.

– . 2016b. “Revealed preference analysis of characteristics in discrete choice”. *Working Paper*.

Andrews, Donald W.K., and Panle Jia Barwick. 2012. “Inference for parameters defined by moment inequalities: A recommended moment selection procedure”. *Econometrica* 80 (6): 2805–2826.

Andrews, Donald W.K., and Gustavo Soares. 2010. “Inference for parameters defined by moment inequalities using generalized moment selection”. *Econometrica* 78 (1): 119–157.

Athey, Susan, and Scott Stern. 1998. *An empirical framework for testing theories about complimentarity in organizational design*. Tech. rep. National Bureau of Economic Research.

Berry, Steven T. 1994. “Estimating discrete-choice models of product differentiation”. *The RAND Journal of Economics*: 242–262.

Berry, Steven T, and Philip A Haile. 2015a. “Identification in differentiated products markets”. *NBER Working Paper No. 21500*.

– . 2014. “Identification in differentiated products markets using market level data”. *Econometrica* 82 (5): 1749–1797.

– . 2009. “Nonparametric identification of multinomial choice demand models with heterogeneous consumers”. *NBER Working Paper No. 15276*.

Berry, Steven, Amit Gandhi, and Philip Haile. 2013. “Connected substitutes and invertibility of demand”. *Econometrica* 81 (5): 2087–2111.

Berry, Steven, and Philip A Haile. 2015b. “Identification of nonparametric simultaneous equations models with a residual index structure”. *Cowles Foundation Discussion Paper*.

Bhattacharya, Debopam. 2015. “Nonparametric welfare analysis for discrete choice”. *Econometrica* 83 (2): 617–649.

Block, Henry David, and Jacob Marschak. 1960. “Random orderings and stochastic theories of responses”. *Contributions to Probability and Statistics* 2:97–132.

Boyd, Stephen, and Lieven Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Brown, Donald J, and Caterina Calsamiglia. 2007. “The nonparametric approach to applied welfare analysis”. *Economic Theory* 31 (1): 183–188.

Canay, Ivan A., and Azeem M. Shaikh. 2016. “Practical and Theoretical Advances in Inference for Partially Identified Models”. *Working Paper*.

Caplin, Andrew, and Mark Dean. 2015. “Revealed Preference, Rational Inattention, and Costly Information Acquisition”. *American Economic Review* 105 (7): 2183–2203.

Cerreia-Vioglio, Simone, David Dillenberger, Pietro Ortoleva, and Gil Riella. 2015. “Deliberately stochastic”. *Manuscript Columbia University*.

Chambers, Christopher P, and Federico Echenique. 2009a. “Profit maximization and supermodular technology”. *Economic Theory* 40 (2): 173–183.

– . 2009b. “Supermodularity and preferences”. *Journal of Economic Theory* 144 (3): 1004–1014.

Chen, Linchun Lily. 2013. “Identification and Estimation of Linear Transformation Models with Endogeneity: an Instrumental Variables Method”. *Working Paper*.

Chernozhukov, Victor, Alfred Galichon, Marc Henry, and Brendan Pass. 2014. “Single Market Nonparametric Identification of Multi-Attribute Hedonic Equilibrium Models”. *Available at SSRN 2392549*.

Chernozukov, Victor, Denis Chetverikov, and Kengo Kato. 2014. “Testing many moment inequalities”. *arXiv preprint arXiv:1312.7614v4*.

Chiong, Khai Xiang, Alfred Galichon, and Matt Shum. 2016. “Duality in dynamic discrete-choice models”. *Quantitative Economics* 7 (1): 83–115.

Chu, Chaushie. 1989. “A paired combinatorial logit model for travel demand analysis”. In *Transport Policy, Management & Technology Towards 2001: Selected Proceedings of the Fifth World Conference on Transport Research*, vol. 4.

– . 1981. “Structural Issues and Sources of Bias in Residential Location and Travel Choice Models”. PhD thesis, Northwestern University.

De Paula, Aureo, and Xun Tang. 2012. “Inference of signs of interaction effects in simultaneous games with incomplete information”. *Econometrica* 80 (1): 143–172.

Dunker, Fabian, Stefan Hoderlein, and Hiroaki Kaido. 2015. “Nonparametric identification of endogenous and heterogeneous aggregate demand models: complements, bundles and the market level”. *Working Paper*.

Feng, Guiyun, Xiaobo Li, and Zizhuo Wang. 2015. “Analysis of discrete choice models: A welfare-based framework”. *arXiv preprint arXiv:1503.01854*.

Fosgerau, Mogens, and André de Palma. 2015. “Demand systems for market shares”. *Working Paper*.

Fox, Jeremy T. 2007. “Semiparametric estimation of multinomial discrete-choice models using a subset of choices”. *RAND Journal of Economics*: 1002–1019.

Fox, Jeremy T, and Natalia Lazzati. Forthcoming. “A Note on Identification of Discrete Choice Models for Bundles and Binary Games”. *Quantitative Economics*.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki. 2015. “Stochastic choice and revealed perturbed utility”. *Econometrica* 83 (6): 2371–2409.

Galichon, Alfred, and Bernard Salanié. 2015. “Cupid’s invisible hand: Social surplus and identification in matching models”. *Available at SSRN 1804623*.

Gentzkow, Matthew. 2007. “Valuing New Goods in a Model with Complementarity: Online Newspapers”. *The American Economic Review*: 713–744.

Han, Aaron K. 1987. “Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator”. *Journal of Econometrics* 35 (2): 303–316.

Hicks, John R, and Roy G D Allen. 1934. “A Reconsideration of the Theory of Value”. *Economica* 1 (1934): 52–176.

Ho, Kate, and Adam Rosen. 2015. “Partial identification in applied research: benefits and challenges”. *Working Paper*.

Hofbauer, Josef, and William H Sandholm. 2002. “On the global convergence of stochastic fictitious play”. *Econometrica* 70 (6): 2265–2294.

Howard, David H. 1977. “Rationing, quantity constraints, and consumption theory”. *Econometrica: Journal of the Econometric Society*: 399–412.

Huber, Joel, John W Payne, and Christopher Puto. 1982. “Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis”. *Journal of Consumer Research* 9 (1): 90–98.

Imbens, Guido W., and Charles F. Manski. 2004. “Confidence intervals for partially identified parameters”. *Econometrica* 72 (6): 1845–1857.

Kaplan, David M. 2015. “Bayesian and frequentist tests of sign equality and other non-convex, nonlinear inequalities”. Working Paper.

Koppelman, Frank S, and Chieh-Hua Wen. 2000. “The paired combinatorial logit model: properties, estimation and application”. *Transportation Research Part B: Methodological* 34 (2): 75–89.

Kos, Nenad, and Matthias Messner. 2013. “Extremal incentive compatible transfers”. *Journal of Economic Theory* 148 (1): 134–164.

Lewbel, Arthur. 1998. “Semiparametric latent variable model estimation with endogenous or mismeasured regressors”. *Econometrica*: 105–121.

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

Louviere, Jordan J, Richard T Carson, Leonie Burgess, Deborah Street, and AAJ Marley. 2013. “Sequential preference questions factors influencing completion rates and response times using an online panel”. *Journal of Choice Modelling* 8:19–31.

Machina, Mark J. 1985. “Stochastic choice functions generated from deterministic preferences over lotteries”. *The Economic Journal* 95 (379): 575–594.

Manski, Charles F. 1975. “Maximum score estimation of the stochastic utility model of choice”. *Journal of Econometrics* 3 (3): 205–228.

– . 1985. “Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator”. *Journal of Econometrics* 27 (3): 313–333.

Manzini, Paola, Marco Mariotti, and Levent Ülkü. 2015. *Stochastic Complementarity*. Tech. rep. Centro de Investigacion Economica, ITAM.

Matejka, Filip, and Alisdair McKay. 2014. "Rational inattention to discrete choices: A new foundation for the multinomial logit model". *The American Economic Review* 105 (1): 272–298.

Mattsson, Lars-Göran, and Jörgen W Weibull. 2002. "Probabilistic choice and procedurally bounded rationality". *Games and Economic Behavior* 41 (1): 61–78.

Matzkin, Rosa L. 2015. "Estimation of nonparametric models with simultaneity". *Econometrica* 83 (1): 1–66.

– . 2008. "Identification in nonparametric simultaneous equations models". *Econometrica* 76 (5): 945–978.

– . 1992. "Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models". *Econometrica: Journal of the Econometric Society*: 239–270.

– . 2007. "Nonparametric identification". *Handbook of Econometrics* 6:5307–5368.

– . 1993. "Nonparametric identification and estimation of polychotomous choice models". *Journal of Econometrics* 58 (1): 137–168.

– . 2013. "Nonparametric identification in structural economic models". *Annu. Rev. Econ.* 5 (1): 457–486.

– . 1994. "Restrictions of economic theory in nonparametric methods". *Handbook of Econometrics* 4:2523–2558.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior", ed. by P. Zarembka. New York: Academic Press.

– . 1981. "Econometric Models of Probabilistic Choice". In *Structural Analysis of Discrete Data with Econometric Applications*, ed. by Charles F. Manski and Daniel McFadden. Mit Press Cambridge, MA.

– . 1978. "Modelling the choice of residential location". In *Spatial Interaction Theory and Residential Location*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. North Holland Pub., Co.

McFadden, Daniel L, and Mogens Fosgerau. 2012a. *A theory of the perturbed consumer with general budgets*. Tech. rep. National Bureau of Economic Research.

– . 2012b. *A theory of the perturbed consumer with general budgets*. Tech. rep. National Bureau of Economic Research.

Milgrom, Paul, and Chris Shannon. 1994. “Monotone comparative statics”. *Econometrica: Journal of the Econometric Society*: 157–180.

Myerson, Roger B. 1981. “Optimal auction design”. *Mathematics of Operations Research* 6 (1): 58–73.

Præstgaard, Jens, and Jon A. Wellner. 1993. “Exchangeably weighted bootstraps of the general empirical process”. *The Annals of Probability*: 2053–2086.

Riley, John G, and William F Samuelson. 1981. “Optimal auctions”. *The American Economic Review* 71 (3): 381–392.

Rockafellar, R Tyrrell. 1970. *Convex Analysis*. Princeton University Press.

Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2014. “A Practical Two-Step Method for Testing Moment Inequalities”. *Econometrica* 82 (5): 1979–2002.

Samuelson, Paul A. 1974. “Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory”. *Journal of Economic literature*: 1255–1289.

Selten, Reinhard. 1975. “Reexamination of the perfectness concept for equilibrium points in extensive games”. *International Journal of Game Theory* 4 (1): 25–55.

Shi, Xiaoxia, Matthew Shum, and Wei Song. 2016. “Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity”. *Available at SSRN 2748647*.

Small, Kenneth A, and Harvey S Rosen. 1981. “Applied welfare economics with discrete choice models”. *Econometrica*: 105–130.

Stinchcombe, Maxwell B, and Halbert White. 1992. “Some measurability results for extrema of random functions over random sets”. *The Review of Economic Studies* 59 (3): 495–514.

Stoye, Jörg. 2009. “More on confidence intervals for partially identified parameters”. *Econometrica* 77 (4): 1299–1315.

Topkis, Donald M. 1978. “Minimizing a submodular function on a lattice”. *Operations research* 26 (2): 305–321.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Varian, Hal R. 1992. *Microeconomic analysis*. WW Norton.

– . 1982. "The nonparametric approach to demand analysis". *Econometrica*: 945–973.

Vives, Xavier. 1990. "Nash equilibrium with strategic complementarities". *Journal of Mathematical Economics* 19 (3): 305–321.

Wen, Chieh-Hua, and Frank S Koppelman. 2001. "The generalized nested logit model". *Transportation Research Part B: Methodological* 35 (7): 627–641.

Wolak, Frank A. 1987. "An exact test for multiple inequality and equality constraints in the linear regression model". *Journal of the American Statistical Association* 82 (399): 782–793.