Identification of Average Demand Models*

Roy Allen
Department of Economics
University of California, San Diego
rhallen@ucsd.edu

John Rehbeck
Department of Economics
University of California, San Diego
jrehbeck@ucsd.edu

Click here for the latest version

November 15, 2016

Abstract

This paper studies the nonparametric identification of a model of average demand with 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.

*We thank Brendan Beare, Keri Hu, Ivana Komunjer, Mark Machina, Andres Santos, Jeffrey Shrader, Yixiao Sun, Kaspar Wüthrich, and seminar participants at UC San Diego for helpful comments. Any remaining errors are our own.

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 satisfies a first order condition. 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 their desirability. Each index function depends on the characteristics of that good. Formally, assume the demand vector Y satisfies

$$Y = \tilde{M}(v_1(X_1), \ldots, v_K(X_K), \varepsilon), \tag{1}$$

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 $\vec{v} = (v_1, \ldots, v_K)'$ 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 gradient of the individual-specific indirect utility function. Welfare is then calculated as the average indirect utility.

We are motivated to study (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 (2015), 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) 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 $\tilde{M}$. 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

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

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 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) aggregates. Specifically, when latent variables are independent of characteristics, (1) implies

$$\mathbb{E}[Y \mid X = x] = M(v_1(x_1), \ldots, v_K(x_K)), \tag{2}$$

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

[2] Such as a special regressor structure.

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

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 2 contains a literature review. Section 3 formalizes our setup, provides examples and further literature review, and illustrates restrictions of the model under smoothness conditions. Section 4 establishes conditions for nonparametric identification of $\vec{v}$. Section 5 constructively identifies changes in D and average welfare. Section 6 characterizes out-of-sample bounds using the model. Section 7 contains an illustrative application. Section 8 concludes.

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 (2015b) 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 (2015a). 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 that $\tilde{M}$ (or the aggregate counterpart M) is the gradient 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).

To our knowledge, exploiting Slutsky symmetry constitutes a new identification 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, $v_k(X_k) = g(Z_k) + \tilde{v}_k(W_k)$, 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. By defining $\tilde{Z}_k = g(Z_k)$, 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]

A structure similar to (1) or (2) has been widely used. 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 gradient 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 (2012) study a representation related to (1) focusing on budget variation. Fudenberg, Iijima, and Strzalecki (2015) study a stochastic choice model that implies (1). Chiong, Galichon, and Shum (2016) study a structure similar to (1) in dynamic discrete choice. Roughly, their results can be used to show that if M is known then $\vec{v}$ is identified. Fosgerau

[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 alternatives for our nonparametric results, so we do not contribute to the special regressor technique when there is only a single good.

and de Palma (2015) use a special case of (1) to model and estimate demand for differentiated products

This paper is part of a broader study of (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.

3 Model and Examples

We study models in which conditional means satisfy

$$\mathbb{E}[Y \mid X = x] \in \underset{y \in B}{\operatorname{argmax}} \sum_{k=1}^{K} y_k v_k(x_k) + D(y). \tag{3}$$

We call this the *perturbed utility model* (PUM). This is an average demand version of the structure discussed in the Introduction.[7] Heuristically, in the Introduction we presented the indirect utility version of the model (the choice vector is the gradient of a convex function), whereas here we present the direct utility version of the model.

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

[6]A related result appears in Gentzkow (2007), but assumes there are no income effects.
[7]This is a slight generalization to allow nonunique maximizers.

encodes substitutability/complementarity patterns between the goods. Importantly, characteristics do not enter D.

We state the following maintained assumptions.

Assumption 1. *(i)* $\mathbb{E}[Y \mid X = x]$ *satisfies (3) for every* $x \in \operatorname{supp}(X)$.[8]

(ii) B *is a nonempty 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.

Assumption 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 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 (3) and then provide some examples that fit into our framework.

3.1 Examples

One possible interpretation of (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

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

that the quantity vector Y satisfies

$$Y \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(X_k) + \tilde{D}(y, \varepsilon). \tag{4}$$

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 (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 does not contain latent variables is a key homogeneity assumption, but this assumption is weaker than may initially appear. An equivalent restatement of the latent utility model (4) is given by

$$Y \in \tilde{M}(\vec{v}(X), \varepsilon), \tag{5}$$

where for fixed ε, $\tilde{M}$ is the subgradient of a convex function and is derived from $\tilde{D}$ and B. (See Appendix A for details.) When there is a unique maximizer to (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. For example, for fixed ε, Y_k weakly increases if v_k changes, all else equal.

We now turn to specific examples. Previous work needs to specify a functional form for $\tilde{D}$. 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.

3.1.1 Bundles

We present a version of the model studied in Gentzkow (2007), Fox and Lazzati (2015), 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

$$\begin{aligned} u_{0,0} &= 0 \\ u_{1,0} &= v_1(X_1) + \varepsilon_{1,0} \\ u_{0,1} &= v_2(X_2) + \varepsilon_{0,1} \\ u_{1,1} &= u_{1,0} + u_{0,1} + \varepsilon_{1,1}, \end{aligned} \tag{6}$$

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 3.1. *Let $Y \in \{0,1\}^2$ denote a utility maximizing quantity vector for (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 (6). If similar separability is maintained, this example can generalize to multiple goods, regardless of whether the quantities are discrete.

Fox and Lazzati (2015) 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. Dunker, Hoderlein, and Kaido (2015) study identification of this and related models with linear random coefficients and with endogeneity (so ε is not independent of X).

3.1.2 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

$$Y \in \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1}^{K} y_k v_k(X_k) + \sum_{k=1}^{K} y_k \varepsilon_k, \tag{7}$$

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 $\varepsilon = (\varepsilon_1, \ldots \varepsilon_K)'$. Instead, we assume X and ε are independent. The following result, with slightly stronger assumptions than we impose, is due to Hofbauer and Sandholm (2002).

Lemma 3.2. *Assume Y is consistent with (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. This is possible if there are values of $\vec{v}$ such that only two alternatives are chosen with positive probability. 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 2 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 $(W_1, \ldots, W_K)$, the vector $Z = (Z_1, \ldots, Z_K)$ 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 3.2 also rules out random coefficients models, but our results may be combined with others to tackle such models. Matzkin (2007), Berry and Haile (2009), and Briesch, Chintagunta, and Matzkin (2012) use special regressors to identify discrete choice models with random coefficients or other forms of nonseparable heterogeneity. Specifically, they consider latent utilities of the form $u_k = Z_k + \tilde{v}_k(W_k, \varepsilon)$. We show how to identify g in the more general latent utility model $u_k = g(Z_k) + \tilde{v}_k(W_k, \varepsilon)$, provided Z is independent of ε, conditional on W. Our results thus widen the applicability of these papers.[9]

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

[9]Berry and Haile (2009) mention 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).

to deliver *ordinal* identification of v_k whereas additive, independent errors ensures *cardinal* identification of v_k.

3.1.3 Expected Utility / Moral Hazard

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

$$Y \in \underset{y \in \tilde{B}}{\operatorname{argmax}} \sum_{k=1}^{K} y_k v_k(X_k) + \tilde{D}(y, \varepsilon). \tag{8}$$

There are K outcomes. Outcome k is a "good" treated in our general setup. In the utility function, y_k may be interpreted as the probability of outcome k. Individual heterogeneity is represented by ε.

Agarwal and Somaini (2014) study a special case of (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. The budget $\tilde{B}$ in the latent utility specification (8) is a finite set of lotteries.[10]

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 ($v_k(X_k) = X_k$ in their setup). 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.[11] Together with the results of Agarwal and Somaini (2014), this establishes identification of $\vec{v}$ and the latent distribution of ε.[12]

[10] This set is fixed in one of their identification approaches (Agarwal and Somaini (2014), Section 6.2).

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

[12] While $\vec{v}$ is identified without needing to know the budget $\tilde{B}$, the results of Agarwal and Somaini

The setup of (8) also admits a moral hazard interpretation.[13] 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 (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 (8) fits into our setup.

Lemma 3.3. *Assume Y is consistent with (8). If ε is independent of X, Y is measurable, and $\mathbb{E}[\tilde{D}(Y, \varepsilon) \mid X = x]$ is finite for each $x \in \operatorname{supp}(X)$, then there is some $D \in \mathcal{D}$ such that*

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

We include this example to demonstrate the versatility of our setup, but it has limitations. Additive separability in (8) may be particularly unappealing for aggregate data, since it implies that individuals with the same observable characteristics (which may be part of X) have the same risk aversion $\vec{v}$. The model may be more appropriate for repeated observations from the same individual. To our knowledge, the search for an appropriate model for such data is ongoing (see Apesteguia and Ballester (2016) and the references therein).

(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.

[13] We call this moral hazard because of the presence of costly effort, but see for example Escanciano, Salanié, and Yıldız (2016) for a more canonical version of moral hazard.

3.1.4 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

$$Y \in \operatorname*{argmax}_{y \in \Delta^{K-1}} \sum_{k=1}^{K} y_k v_k(X_k) + \tilde{D}(y, \varepsilon), \tag{9}$$

where Y is the choice of lottery (conditional probabilities) 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.[14] We study this model in Allen and Rehbeck (2016b). Lemma 3.3 shows (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 making a mistake. This is a special case of (9) if we set

$$\tilde{D}(y, \varepsilon) = \begin{cases} 0 & \text{if } \exists k \text{ s.t. } y_k = 1 - \varepsilon \text{ and } y_j = \frac{\varepsilon}{K-1} \text{ for } j \neq k \\ -\infty & \text{otherwise} \end{cases}.$$ [15]

The representation (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 (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

[14] Recent work includes Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (2015) and Agranov and Ortoleva (2015).

[15] Note that Lemma 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.

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. Shi et al. (2016) study identification of ARUM in a panel setting when $\vec{v}$ is linear. Theorem H.1 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.[16] 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 a common assumption in discrete choice models, one message of the present paper is that this is not necessary.

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 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 k, ℓ element equal to $\frac{\partial M_k(\overline{v})}{\partial \overline{v}_\ell}$.

Proposition 3.1 ("Slutsky Conditions")**.** *Let Assumption 1 and 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

[16] In light of their work, it is likely that our results may be combined with theirs to deliver nonparametric identification for certain grouped (e.g. panel) data.

appears. This is a special case of ARUM (Section 3.1.2).

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.*

4 Identification of $\vec{v}$

This section provides conditions under which $\vec{v}$ is nonparametrically identified. We treat D (and M) 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

$$\operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k) + D(y) = \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k(\lambda v_k(x_k) + c_k) + \left(\lambda D(y) - \sum_{k=1}^{K} y_k c_k\right). \tag{10}$$

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 $d_w > 0$.

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,

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} = \left.\frac{\partial M_k(\overline{v})}{\partial \overline{v}_\ell}\right|_{\overline{v}=\vec{v}(x)} \frac{\partial v_\ell(x_\ell)}{\partial x_\ell}, \tag{11}$$

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 (11) with symmetry of cross-partials of M (Proposition 3.1(ii)), we obtain

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_k} = \frac{\partial v_\ell(x_\ell)}{\partial x_\ell} \Bigg/ \frac{\partial v_k(x_k)}{\partial x_k}. \tag{12}$$

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 (12). We only need (12) to hold at certain *points*. We thus make explicit the points of evaluation of partial derivatives. We maintain Assumption 2 to state the following definition.

Definition 1 (Pairs)**.** *Say that* $\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell = x_\ell^*}$ *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) $\left.\frac{\partial \mathbb{E}[Y_\ell | X=x]}{\partial x_{k,q}}\right|_{x=x^*}$ *and* $\left.\frac{\partial \mathbb{E}[Y_k | X=x]}{\partial x_{\ell,p}}\right|_{x=x^*}$ *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.[17]

Proposition 4.1. *Let Assumptions 1 and 2 hold and assume $x_{\ell,p}$ and $x_{k,q}$ are regressors specific to ℓ and k, respectively. 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 and $\left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*} \neq 0$, there is some known $x^* \in \operatorname{supp}(X)$ such that*

$$\left.\frac{\partial \mathbb{E}[Y_k \mid X=x]}{\partial x_{\ell,p}}\right|_{x=x^*} \Bigg/ \left.\frac{\partial \mathbb{E}[Y_\ell \mid X=x]}{\partial x_{k,q}}\right|_{x=x^*} = \left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^*} \Bigg/ \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}, \tag{13}$$

where x_ℓ^* and x_k^* are components of x^* specific to goods ℓ and k, respectively, and $\partial\mathbb{E}[Y_\ell \mid X=x]/\partial x_{k,q}|_{x=x^*} \neq 0$. In particular, the right hand side of (13) is identified.

It is noteworthy that the left hand side of (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.[18]

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.

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

[18] 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.

Lemma 4.1. *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,

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

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.[19]

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 4.1 and Lemma 4.1.

Theorem 4.1. *Let Assumptions 1 and 2 hold, assume all regressors are good-specific, and assume $K \geqslant 2$. Assume X has full support,[20] $\vec{v}$ is differentiable, $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$,[21] 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_ℓ} denotes a d_ℓ-dimensional vector of zeros.*

In light of (10), the scale and location normalization cannot be weakened without further restrictions on D.[22] To obtain identification with this normalization, the model

[19] 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*.

[20] That is, $\text{supp}(X) = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_K}$. We assume X has full support to identify $\vec{v}$ over its entire domain.

[21] 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 G for a result that significantly weakens the assumptions of Theorem 4.1.

[22] 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 $x_k^* \neq 0_{d_k}$.

needs to 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 H.1.

As a corollary of Theorem 4.1, 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 4.1. *Let the assumptions of Lemma 3.2 hold and assume $K \geqslant 2$. Suppose X has full support, $\vec{v}$ is differentiable, $\vec{v}(\text{supp}(X)) = \mathbb{R}^K$, 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 4.1 are satisfied. In particular, $\vec{v}$ is nonparametricially identified up to a location and scale normalization.*

The assumptions of Theorem 4.1 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 G, 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 4.1 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 2 (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

$$u_k = g_k(X_{k,1}) + \tilde{v}_k(X_{k,-1}) + \varepsilon_k, \tag{14}$$

where $X_{k,1}$ is a scalar and $X_{k,-1}$ collects components of X_k other than $X_{k,1}$. We assume (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 is 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 3.2 applies to yield

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

for some D. (The maximizer is unique under mild conditions.) Note that x_{-1} enters D. By fixing x_{-1} and varying x_1, Theorem 4.1 provides conditions under which $g_k(\cdot)$ is identified for each k.[23] Conditions from Lewbel (2000) or the subsequent literature can then be used to identify $\tilde{v}_k$.

Remark 3 (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.

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 has at least 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 $Z_k = \left(Z_k^{(1)\prime}, Z_k^{(2)\prime}\right)'$. Think of $Z_k^{(1)}$ as a subvector of continuous regressors specific to good k. Let $Z^{(m)} = \left(Z_1^{(m)\prime}, \ldots, Z_K^{(m)\prime}\right)'$ for $m = 1, 2$.

Assumption 3. *There exists a known* $\left(\tilde{z}^{(2)\prime}, \tilde{w}'\right)' \in \operatorname{supp}\left(Z^{(2)}, W\right)$ *such that*

[23] 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.

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

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

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

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

where $\tilde{z} = (\tilde{z}^{(1)\prime}, \tilde{z}^{(2)\prime})'$. Sufficient conditions for (ii) are given in Theorem 4.1. 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 4.2. *Let Assumptions 1, 2, and 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 $(z, w) \in \text{supp}(Z, W)$, then $\vec{v}(z, w)$ is identified.*

The assumptions on D_B are used to establish that

$$\mathbb{E}[Y \mid Z = z, W = w] = \mathbb{E}[Y \mid Z = \tilde{z}, W = \tilde{w}] \implies \vec{v}(z, w) = \vec{v}(\tilde{z}, \tilde{w}). \tag{15}$$

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 I.1.

Remark 4. Concavity of D is innocuous on its own (see Theorem H.1), 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 nonempty 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 I.1 for more details.

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)).[24] 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.

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

The functions V and D are linked by the identity

$$V(x) = \sum_{k=1}^{K} \mathbb{E}[Y_k \mid X = x]v_k(x_k) + D(\mathbb{E}[Y \mid X = x]), \tag{16}$$

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

$$D(\mathbb{E}[Y \mid X = \tilde{x}]) - D(\mathbb{E}[Y \mid X = x]) \leqslant (\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'v(x). \tag{17}$$

Remark 5 (Normalizations). Nonparametric identification of $\vec{v}$ requires a location and scale normalization, as shown in (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, not V. For example if we normalize $\vec{v}(x^*) = 0_K$, where 0_K is vector of zeros of length K, then from (17) we have

$$D(\mathbb{E}[Y \mid X = \tilde{x}]) - D(\mathbb{E}[Y \mid X = x^*]) \leqslant 0 \tag{18}$$

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 5.2.

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 (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 $x^0, \dots, x^S$ be a sequence of points in $\text{supp}(X)$. By summing up (17), we obtain

$$D(\mathbb{E}[Y \mid X = x^S] - D(\mathbb{E}[Y \mid X = x^0]) \leqslant \sum_{s=0}^{S-1} (\mathbb{E}[Y \mid X = x^s] - \mathbb{E}[Y \mid X = x^{s+1}])' \vec{v}(x^s). \tag{19}$$

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

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

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

By (19), we obtain that for arbitrary $x, \tilde{x} \in \text{supp}(X)$,

$$\underline{\Delta} D(x, \tilde{x}) \leqslant D(\mathbb{E}[Y \mid X = \tilde{x}]) - D(\mathbb{E}[Y \mid X = x]) \leqslant \overline{\Delta} D(x, \tilde{x}). \tag{20}$$

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.[25]

Theorem 5.1. *Suppose Assumption 1 holds, $\vec{v}$ is known, and $x, \tilde{x} \in \text{supp}(X)$. 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

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

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

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

To bound differences in V, we can use (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 5.1 shows differences in D are bounded by $\underline{D}$ and $\overline{D}$, which then bound differences in V.

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

$$\mathbb{E}[Y \mid X = x] \in \partial_v V(x). \tag{21}$$

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" (21).

Theorem 5.2. *Let Assumption 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 5.1. *Let Assumption 1 hold. Assume $\vec{v}$ is identified and the set $\vec{v}(\mathrm{supp}(X))$ is convex. Then for every $x^0, x^1 \in \mathrm{supp}(X)$,*

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

and

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

are identified.

Remark 6. These results only identify differences in D for points in $\{\mathbb{E}[Y \mid X = x]\}_{x \in \mathrm{supp}(X)}$. The set $\{\mathbb{E}[Y \mid X = x]\}_{x \in \mathrm{supp}(X)}$ need not be convex even under the assumptions of Corollary 5.1.

5.3 Direct Identification of Welfare Changes

The previous results in this section use two steps. 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 proposition follows from the integral representation in Theorem 5.2.

Proposition 5.1. *Assume $d_k = 1$ for each k. Let Assumptions 1 and 2 hold and assume V and $\vec{v}$ are differentiable. Let x be in the interior of $\mathrm{supp}(X)$. Then for* *each ℓ,*

$$\frac{\partial V(x)}{\partial x_\ell} = \mathbb{E}[Y_\ell \mid X = x] \frac{\partial v_\ell(x_\ell)}{\partial x_\ell}. \tag{22}$$

Proposition 4.1 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 (22) we have for each ℓ that,

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

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.

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. Instead, in this section we only assume $\vec{v}$ is identified and characterize the out-of-sample restrictions of the model.[26] This exercise is in the spirit of Blundell, Browning, and Crawford (2003). Nontrivial out-of-sample restrictions are possible because the model contains a nonparametric shape restriction akin to monotonicity.

We are formally interested in the set

$$\operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k v_k(x_k^0) + D(y), \tag{23}$$

where $x^0 \notin \text{supp}(X)$.[27] We denote the restrictions on this set by providing bounds on

[26] In practice, knowledge of $\vec{v}$ typically requires extrapolation such as assuming $\vec{v}$ belongs to a parametric class of functions.

[27] We implicitly assume the argmax set is nonempty.

a generic element of it, denoted $\tilde{m}(x^0)$. We do not assume (23) is a singleton.

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

$$\begin{aligned}\sup_{x^\ell \in \text{supp}(X): \vec{v}(x^\ell) < \vec{v}(x^0)} \mathbb{E}[Y \mid X = x^\ell] &\leqslant \tilde{m}(x^0) \\ &\leqslant \inf_{x^u \in \text{supp}(X): \vec{v}(x^u) > \vec{v}(x^0)} \mathbb{E}[Y \mid X = x^u].\end{aligned} \tag{24}$$

Thus in the univariate case, the restrictions of the model are monotonicity restrictions. The following proposition characterizes the sharp bounds in the general setting ($K \geqslant 1$), provided the only thing we know about D is $D \in \mathcal{D}$.

Theorem 6.1 (Sharp Out-of-Sample Bounds)**.** *Let Assumption 1 hold, assume $\vec{v}$ is known, and assume $D \in \mathcal{D}$. Let $x^0 \notin \text{supp}(X)$ 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 supp(X)*,*

$$\begin{aligned}\tilde{m}(x^0)'(\vec{v}(x^0) - \vec{v}(x^{S-1})) \geqslant \mathbb{E}[Y \mid X = x^1]'\vec{v}(x^0) - \mathbb{E}[Y \mid X = x^{S-1}]\vec{v}(x^{S-1}) \\ - \sum_{s=1}^{S-2}(\mathbb{E}[Y \mid X = x^s] - \mathbb{E}[Y \mid X = x^{s+1}])'\vec{v}(x^s).\end{aligned} \tag{25}$$

When 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 7. 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 6.1, the estimates of $\vec{v}$ and $\mathbb{E}[Y \mid X = x]$ must be consistent with the restrictions of the model. Otherwise, restrictions like (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 6.1 can directly be used to provide counterfactual bounds.

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 3.1.2) against the strictly more general perturbed stochastic choice model (Section 3.1.4). Violations of ARUM occur precisely when stochastic complementarity can occur.

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

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

where

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

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

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

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

$$P_k(\overline{v}, \lambda) = \frac{\sum_{j \neq k} e^{\overline{v}_k/\lambda_{jk}} \left(e^{\overline{v}_j/\lambda_{jk}} + e^{\overline{v}_k/\lambda_{jk}}\right)^{\lambda_{jk}-1}}{\sum_{\ell=1}^{K-1} \sum_{m=\ell+1}^{K} \left(e^{\overline{v}_\ell/\lambda_{\ell m}} + e^{\overline{v}_\ell/\lambda_{\ell m}}\right)^{\lambda_{\ell m}}}. \tag{26}$$

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

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

for certain values of $\overline{v}$ (with $j \neq k$). 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 7.1. *Let* $\vec{v}(\mathrm{supp}(X)) = \mathbb{R}^k$. *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, $P(\overline{v}, \lambda)$ is the gradient of a convex function

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

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

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.

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 G for identification. We reiterate that the conditions in that appendix 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 $P(\overline{v}, \lambda)$. 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. Our motivation for this is that an individual may first consider a pair of adjacent alternatives when making a decision, then choose among this pair. This choice behavior is consistent with the PCL, which can be written as a two-stage decision problem. Our free nesting parameters are then $\lambda_{12}, \lambda_{23}, \lambda_{34}, \lambda_{45}, \lambda_{51}$. Note we treat the first and last alternative as adjacent. This is to allow for the possibility of endpoint effects.

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 gain 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

$$\max_{jk} \sqrt{n} \frac{\hat{\lambda}_{jk} - 1}{\hat{\sigma}_{jk}} \geqslant \Phi^{-1}(1 - \alpha/5), \tag{27}$$

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 inequalities, which hold under H_0 and the normal approximation to $\hat{\lambda}$,

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

Because size control is established under these inequalities, the naive test of (27) is potentially conservative. 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 (potentially) conservative test given by (27) handily rejects H_0.

7.5 Results

We use (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.[28] The vector x_k is a 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 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 (27) is $\Phi^{-1}(1-.05/5) = 2.33$.) All nesting parameters except $\hat{\lambda}_{45}$ are above

[28]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.

Table 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.

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 J, 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.[29]

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,[30] or that an additively separable structure is sufficient for identification.[31] 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 to construct a parametric test of ARUM against the perturbed utility model. We test

[29] See also Fosgerau and de 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.

[30] As in ARUM with errors independent of characteristics.

[31] As in ARUM with conditional independence conditions as in Lewbel (2000).

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 only use symmetry for our identification results,[32] 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.

References

Afriat, S. N. 1967. The construction of utility functions from expenditure data. *International Economic Review 8,* 1, 67–77.

Agarwal, N. and Somaini, P. 2014. Demand analysis using strategic reports: An application to a school choice mechanism. *NBER Working Paper 20775*.

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

Agranov, M. and Ortoleva, P. 2015. Stochastic choice and preferences for randomization. *Available at SSRN 2644784*.

Allen, R. and Rehbeck, J. 2016a. Complementarity in perturbed utility models. *Working Paper*.

Allen, R. and Rehbeck, J. 2016b. Revealed preference analysis of characteristics in discrete choice. *Working Paper*.

Apesteguia, J. and Ballester, M. A. 2016. Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy, Forthcoming*.

Berry, S., Gandhi, A., and Haile, P. 2013. Connected substitutes and invertibility of demand. *Econometrica 81,* 5, 2087–2111.

Berry, S. and Haile, P. A. 2015a. Identification of nonparametric simultaneous equations models with a residual index structure. *Cowles Foundation Discussion Paper*.

Berry, S. T. 1994. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Berry, S. T. and Haile, P. A. 2009. Nonparametric identification of multinomial choice demand models with heterogeneous consumers. *NBER Working Paper No. 15276*.

Berry, S. T. and Haile, P. A. 2014. Identification in differentiated products markets using market level data. *Econometrica 82,* 5, 1749–1797.

Berry, S. T. and Haile, P. A. 2015b. Identification in differentiated products markets. *NBER Working Paper No. 21500*.

Bhattacharya, D. 2015. Nonparametric welfare analysis for discrete choice. *Econometrica 83,* 2, 617–649.

Blundell, R. W., Browning, M., and Crawford, I. A. 2003. Nonparametric engel curves and revealed preference. *Econometrica 71,* 1, 205–240.

Boyd, S. and Vandenberghe, L. 2004. *Convex Optimization.* Cambridge University Press.

Briesch, R. A., Chintagunta, P. K., and Matzkin, R. L. 2012. Nonparametric discrete choice models with unobserved heterogeneity. *Journal of Business & Economic Statistics*.

Brown, D. J. and Calsamiglia, C. 2007. The nonparametric approach to applied welfare analysis. *Economic Theory 31,* 1, 183–188.

Caplin, A. and Dean, M. 2015. Revealed preference, rational inattention, and costly information acquisition. *American Economic Review 105,* 7, 2183–2203.

Cerreia-Vioglio, S., Dillenberger, D., Ortoleva, P., and Riella, G. 2015. Deliberately stochastic. *Manuscript Columbia University*.

Chambers, C. P. and Echenique, F. 2009. Profit maximization and supermodular technology. *Economic Theory 40,* 2, 173–183.

Chen, L. L. 2013. Identification and estimation of linear transformation models with endogeneity: an instrumental variables method. *Working Paper*.

Chernozhukov, V., Galichon, A., Henry, M., and Pass, B. 2014. Single market nonparametric identification of multi-attribute hedonic equilibrium models. *Available at SSRN 2392549*.

Chiong, K. X., Galichon, A., and Shum, M. 2016. Duality in dynamic discrete-choice models. *Quantitative Economics 7,* 1, 83–115.

Chu, C. 1981. Structural issues and sources of bias in residential location and travel choice models. Ph.D. thesis, Northwestern University.

Chu, C. 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.

Dunker, F., Hoderlein, S., and Kaido, H. 2015. Nonparametric identification of endogenous and heterogeneous aggregate demand models: complements, bundles and the market level. *Working Paper*.

Escanciano, J. C., Salanié, B., and Yildiz, N. 2016. Testing for moral hazard when adverse selection is present. *Working Paper*.

Fosgerau, M. and de Palma, A. 2015. Demand systems for market shares. *Working Paper*.

Fox, J. T. 2007. Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *RAND Journal of Economics*, 1002–1019.

Fox, J. T. and Lazzati, N. 2015. Identification of discrete choice models for bundles and binary games. *Working Paper*.

Fudenberg, D., Iijima, R., and Strzalecki, T. 2015. Stochastic choice and revealed perturbed utility. *Econometrica 83,* 6, 2371–2409.

Galichon, A. and Salanié, B. 2015. Cupid's invisible hand: Social surplus and identification in matching models. *Available at SSRN 1804623*.

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

Han, A. K. 1987. Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics 35,* 2, 303–316.

Hofbauer, J. and Sandholm, W. H. 2002. On the global convergence of stochastic fictitious play. *Econometrica 70,* 6, 2265–2294.

Huber, J., Payne, J. W., and Puto, C. 1982. Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research 9,* 1, 90–98.

Koppelman, F. S. and Wen, C.-H. 2000. The paired combinatorial logit model: properties, estimation and application. *Transportation Research Part B: Methodological 34,* 2, 75–89.

Kos, N. and Messner, M. 2013. Extremal incentive compatible transfers. *Journal of Economic Theory 148,* 1, 134–164.

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

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

LOUVIERE, J. J., CARSON, R. T., BURGESS, L., STREET, D., AND MARLEY, A. 2013. Sequential preference questions factors influencing completion rates and response times using an online panel. *Journal of Choice Modelling 8*, 19–31.

MACHINA, M. J. 1985. Stochastic choice functions generated from deterministic preferences over lotteries. *The Economic Journal 95,* 379, 575–594.

MANSKI, C. F. 1975. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3,* 3, 205–228.

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

MATEJKA, F. AND MCKAY, A. 2014. Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review 105,* 1, 272–298.

MATTSSON, L.-G. AND WEIBULL, J. W. 2002. Probabilistic choice and procedurally bounded rationality. *Games and Economic Behavior 41,* 1, 61–78.

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

MATZKIN, R. L. 1993. Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58,* 1, 137–168.

MATZKIN, R. L. 1994. Restrictions of economic theory in nonparametric methods. *Handbook of Econometrics 4*, 2523–2558.

MATZKIN, R. L. 2007. Nonparametric identification. *Handbook of Econometrics 6*, 5307–5368.

Matzkin, R. L. 2008. Identification in nonparametric simultaneous equations models. *Econometrica 76,* 5, 945–978.

Matzkin, R. L. 2013. Nonparametric identification in structural economic models. *Annu. Rev. Econ. 5,* 1, 457–486.

Matzkin, R. L. 2015. Estimation of nonparametric models with simultaneity. *Econometrica 83,* 1, 1–66.

McFadden, D. 1973. Conditional logit analysis of qualitative choice behavior. New York: Academic Press.

McFadden, D. 1978. Modelling the choice of residential location. In *Spatial Interaction Theory and Residential Location*, A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, Eds. North Holland Pub., Co.

McFadden, D. 1981. Econometric models of probabilistic choice. In *Structural Analysis of Discrete Data with Econometric Applications*, C. F. Manski and D. McFadden, Eds. Mit Press Cambridge, MA.

McFadden, D. L. and Fosgerau, M. 2012. A theory of the perturbed consumer with general budgets. Tech. rep., National Bureau of Economic Research.

Myerson, R. B. 1981. Optimal auction design. *Mathematics of Operations Research 6,* 1, 58–73.

Riley, J. G. and Samuelson, W. F. 1981. Optimal auctions. *The American Economic Review 71,* 3, 381–392.

Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press.

Romano, J. P., Shaikh, A. M., and Wolf, M. 2014. A practical two-step method for testing moment inequalities. *Econometrica 82,* 5, 1979–2002.

Selten, R. 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory 4,* 1, 25–55.

Shi, X., Shum, M., and Song, W. 2016. Estimating semi-parametric panel multinomial choice models using cyclic monotonicity. *Available at SSRN 2748647*.

Small, K. A. and Rosen, H. S. 1981. Applied welfare economics with discrete choice models. *Econometrica*, 105–130.

Stinchcombe, M. B. and White, H. 1992. Some measurability results for extrema of random functions over random sets. *The Review of Economic Studies 59,* 3, 495–514.

Train, K. E. 2009. *Discrete Choice Methods with Simulation.* Cambridge University Press.

Varian, H. R. 1982. The nonparametric approach to demand analysis. *Econometrica*, 945–973.

Wen, C.-H. and Koppelman, F. S. 2001. The generalized nested logit model. *Transportation Research Part B: Methodological 35,* 7, 627–641.

Appendix A Preliminaries in Convex Analysis

Definition A.1 (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 A.2 (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 A.1 (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 A.2. *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 A.3 (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.*

Appendix B Proofs for Section 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 B.1 (Allen and Rehbeck (2016a))**.** *Let*

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

where ε is possibly infinite-dimensional. Suppose ε is independent of X, Y is mea-

surable, and $\mathbb{E}[\tilde{D}(Y,\varepsilon) \mid X = x]$ *is finite for each* $x \in \text{supp}(X)$.[33] *Then*

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

for $D(y) = \sup_{\tilde{Y} \in \mathcal{Y}: \mathbb{E}[\tilde{Y}(\varepsilon)] = y} \mathbb{E}\left[\tilde{D}(\tilde{Y}(\varepsilon), \varepsilon)\right]$, [34] *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.[35]

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 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 3.1. Let

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

By Lemma A.2, $\mathbb{E}[Y \mid X = x] \in \partial(-D_B)^*(\vec{v}(x))$, where $(-D_B)^*$ is the convex conju-

[33] 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.

[34] 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 $\mathbb{E}[Y \mid X = x]$.

[35] 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.

gate of $-D_B$. By Assumption 2 and Lemma A.3, $\mathbb{E}[Y \mid X = x] \in \partial(-D_B)^*(\vec{v}(x))$, 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. □

Appendix C Proofs for Section 4

C.1 Proofs for Section 4.1

Proof of Proposition 4.1. The proof is a direct extension of the arguments in the text. Under the assumptions of Definition 1, there is some $x \in \text{supp}(X)$ such that

$$\begin{aligned} \left.\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_{\ell,p}}\right|_{x=x^*} &= \left.\frac{\partial M_k(\overline{v})}{\partial \overline{v}_\ell}\right|_{\overline{v}=\vec{v}(x^*)} \left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=x_\ell^*} \\ \left.\frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_{k,q}}\right|_{x=x^*} &= \left.\frac{\partial M_\ell(\overline{v})}{\partial \overline{v}_k}\right|_{\overline{v}=\vec{v}(x^*)} \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=x_k^*}. \end{aligned} \tag{28}$$

Recall these equalities hold at the specified *points*. Under condition (iv) of Definition 1, we may take the ratio of the equations in (28) to yield the result. □

Proof of Theorem 4.1. This is implied by Corollary G.1 so we provide only a brief discussion.

Using Proposition 4.1, the conditions of the theorem provide constructive identification of

$$\left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell=\tilde{x}_\ell} \Bigg/ \left.\frac{\partial v_k(x_k)}{\partial x_{k,q}}\right|_{x_k=\tilde{x}_k} \tag{29}$$

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 (33). Thus, (29) is identified without restrictions.

Corollary G.1 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 4.1).

□

Proof of Corollary 4.1. 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 4.1 hold. □

C.2 Proofs for Section 4.2

Proof of Theorem 4.2. By Assumption 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 2 and Proposition I.1 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 3(ii) completes the proof since $\vec{v}(\tilde{z}, \tilde{w})$ is identified. □

Appendix D Proofs for Section 5

Proof of Theorem 5.1. 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 H.1(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 $\tilde{x} \in \operatorname{supp}(X)$ (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 5.2. Step 1: Change of variables.

Define

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

where we recognize $\tilde{V}(\overline{v}) = V(x)$ when $\overline{v} = \vec{v}(x)$. Define

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

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

$$\mathbb{E}[Y \mid X = x] \in \partial\tilde{V}(\overline{v}). \tag{30}$$

This follows from Lemma A.1 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 (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

$$\tilde{V}(\vec{v}(x^1)) - \tilde{V}(\vec{v}(x^0)) = \int_0^1 \mathbb{E}[Y \mid X = x(t)] \cdot (\vec{v}(x^1) - \vec{v}(x^0))dt.$$ [36]

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.

□

Appendix E Proofs for Section 6

Proof of Theorem 6.1. Obviously, $\tilde{m}(x^0) \in B$ by use of the *a priori* knowledge of the budget.

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

[36] 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.

value. The conjectured value is consistent with the model if and only if the restrictions of Theorem H.1(iv) hold. By rearranging the cyclic monotonicity inequalities, we obtain (25). □

Appendix F Proofs for Section 7

Proof of Proposition 7.1. The first part of the lemma is known. It remains to show PCL is consistent with the perturbed utility model.

The PCL probabilities satisfy

$$P(\overline{v}, \lambda) = \nabla \ln \left(\sum_{j \neq k} (e^{\overline{v}_j/\lambda_{jk}} + e^{\overline{v}_k/\lambda_{jk}})^{\lambda_{jk}} \right).$$ [37]

By the arguments in Appendix J, 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

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

equals that of

$$-\left[P_k + \frac{\left(\frac{1}{\lambda_{jk}-1}\right)(P_j + P_k)(P_{j|jk})(P_{k|jk})}{P_j} \right]. \tag{31}$$ [38]

We suppress dependence of each P term on $\overline{v}$ and λ for simplicity. We show that (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 $\overline{v}_j = \overline{v}_k$. This

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

implies $P_{j|jk} = .5$, so the sign of (31) is determined by the sign of

$$-P_k P_j + \frac{1}{4}\left(\frac{1}{\lambda_{jk}-1}\right)(P_j + P_k). \tag{32}$$

By setting $\overline{v}_j = \overline{v}_k$ to sufficiently small values, (32) becomes positive.

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

As discussed previously, Theorem 4.1 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 4.1, allows us to integrate these derivatives and obtain constructive identification; recall Lemma 4.1. The second approach, taken in this section, is to use the mean value theorem to obtain non-constructive results. Recall 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 4.1.

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 4.1). If $x_\ell^1 = x_\ell^2$, we can multiply these derivative ratios to identify

$$\left.\frac{\partial v_j(x_j)}{\partial x_{j,r}}\right|_{x_j = x_j^2} \Bigg/ \left.\frac{\partial v_\ell(x_\ell)}{\partial x_{\ell,p}}\right|_{x_\ell = x_\ell^1}. \tag{33}$$

Note that this derivative ratio can be identified *even if* $j = k$ and $x_j^2 \neq x_\ell^1$. 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 G.1 (Paths). *There is a path from the point* $a := \frac{\partial v_k(x_k)}{\partial x_{k,q}}\Big|_{x_k = x_k^*}$ *to* $b := \frac{\partial v_j(x_j)}{\partial x_{j,r}}\Big|_{x_j=\tilde{x}_j}$ *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 are nonzero or that M is continuously differentiable everywhere

Theorem G.1. *Let Assumptions 1 and 2 hold and assume* $x_{k,q}$ *and* $x_{j,r}$ *are regressors specific to* k *and* j*, respectively. If there is a path from the point* $\frac{\partial v_k(x_k)}{\partial x_{k,q}}\Big|_{x_k=x_k^*}$ *to* $\frac{\partial v_j(x_j)}{\partial x_{j,r}}\Big|_{x_j=\tilde{x}_j}$*, then*

$$\frac{\partial v_j(x_j)}{\partial x_{j,r}}\Big|_{x_j=\tilde{x}_j} \Big/ \frac{\partial v_k(x_k)}{\partial x_{k,q}}\Big|_{x_k=x_k^*} \tag{34}$$

is identified.

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

$$S_{m-1,m} = \frac{\partial v_{\ell_m}(x_{\ell_m})}{\partial x_{\ell_m,p_m}}\Big|_{x_{\ell_m}=x_{\ell_m}^m} \Big/ \frac{\partial v_{\ell_{m-1}}(x_{\ell_{m-1}})}{\partial x_{\ell_{m-1},p_{m-1}}}\Big|_{x_{\ell_{m-1}}=x_{\ell_{m-1}}^{m-1}}. \tag{35}$$

This ratio is identified due to Proposition 4.1. 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,

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

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

The following corollary relaxes assumptions in Theorem 4.1.

Corollary G.1. *Let Assumptions 1 and 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 G.1. 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 and 2, it can be shown that

$$\mathbb{E}[Y \mid X = x] \neq \mathbb{E}[Y \mid X = \tilde{x}] \iff (\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'(\vec{v}(x) - \vec{v}(\tilde{x})) > 0. \tag{36}$$

(This is a straightforward extension of Lemma H.1.) 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

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

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 (36) this implies that there must be a change in $\mathbb{E}[Y_k \mid X = x]$ as well. Again using (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 G.1. 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 $\mathbb{E}[Y \mid X = x]$ may lie on the boundary of B and may not even be continuous. This illustrates how identification in Corollary G.1 is established using restrictions on pairs of partial derivatives. This is in contrast with Theorem 4.1, which places global restrictions that rule out boundary behavior and discontinuities as in the following example.

Remark G.1 (Boundaries and Discontinuities). 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.[39] Then

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

If $\vec{v}$ is differentiable, sufficient conditions for Corollary G.1 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.

Appendix H 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 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 3.1).

[39] These implicitly depend on x.

Characterizing the identifying power of the model is beneficial even when $\vec{v}$ is identified. Section 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 H.1. *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,

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

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 H.1 is the *monotonicity* condition

$$(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])'(\vec{v}(x) - \vec{v}(\tilde{x}) \geqslant 0. \tag{37}$$

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 (37), suppose that $\vec{v}(x)$ and $\vec{v}(\tilde{x})$ only differ with respect to their first component. Then (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 H.1 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 $\operatorname{supp}(X)$. By repeated application of Lemma H.1 we obtain,

$$\begin{aligned} &\sum_{m=0}^{M-1} (\mathbb{E}[Y \mid X = x^m] - \mathbb{E}[Y \mid X = x^{m+1}])'\vec{v}(x^m) \\ &\qquad \geqslant \sum_{m=0}^{M-1} D(\mathbb{E}[Y \mid X = x^{m+1}]) - D(\mathbb{E}[Y \mid X = x^m]) \\ &\qquad = 0. \end{aligned} \tag{38}$$

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 $\{x^m\}_{m=0}^{M-1} \subseteq \operatorname{supp}(X)$. Then for every permutation π we have

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

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 (38) capture the complete restrictions of the model for $\vec{v}$.

Theorem H.1 (Sharp Characterization of $\mathcal{V}_{ID}$)**.** *Let $\vec{v} \in \mathcal{V}$. The following are equiv-*

alent:

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

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

(ii) There is a concave function $D \in \mathcal{D}$ *such that*

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

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

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

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

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

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

By relating D_x with $D(\mathbb{E}[Y \mid X = x])$, the previous discussion shows (i) $\implies$ (iii) $\implies$ (iv). Note that while D may take on value $-\infty$ over some points, $D(\mathbb{E}[Y \mid X = x]) > -\infty$ for each $x \in \operatorname{supp}(X)$. 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) $\implies$ (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 $S = \{(\mathbb{E}[Y \mid X = x], \vec{v}(x))\}_{x \in \operatorname{supp}(X)}$, 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 H.1). 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 A.1 and the fact that

$\mathbb{E}[Y \mid X = x] \in B$ for $x \in \text{supp}(X)$, we have

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

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

Obviously, (ii) $\implies$ (i). □

This result is closely related to results in Brown and Calsamiglia (2007) and Chambers and Echenique (2009).[40] Related results that vary budgets are established in McFadden and Fosgerau (2012).

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 H.1 (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 et al. (2016). Shi et al. (2016) have previously demonstrated that additive random utility models (Section 3.1.2) imply the restrictions of (iv).

[40] 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.

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

$$(\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = \tilde{x}])(\vec{v}(x) - \vec{v}(\tilde{x})) \geqslant 0.$$ [41]

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

$$\mathbb{E}[Y \mid X = x] > \mathbb{E}[Y \mid X = \tilde{x}] \implies \vec{v}(x) \geqslant \vec{v}(\tilde{x})$$
$$\mathbb{E}[Y \mid X = x] < \mathbb{E}[Y \mid X = \tilde{x}] \implies \vec{v}(x) \leqslant \vec{v}(\tilde{x}).$$ [42]

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.

Appendix I Injectivity

This section provides some injectivity results that are used in establishing identification of $\vec{v}$ in Section 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 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

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

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

[42] 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).

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 $u \in \vec{v}(\text{supp}(X))$.

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

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

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 function ρ^{-1} exists for every value in the support of y_i^*, so that we have

$$\rho_k^{-1}(y_i^*) = v_k(X_{k,i}) + \xi_{k,i} \tag{39}$$

almost surely. Using instruments that satisfy a completeness or conditional mean restriction, several recent papers contain identification results for equations similar to (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.[43]

We now turn to our formal results.

Assumption I.1. *$B \subseteq \mathbb{R}^K$ is convex and has nonempty interior.*

Assumption I.1 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 I.1.

[43]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.

Assumption I.2. *$D : B \to \mathbb{R} \cup \{-\infty\}$ is a concave function.*

Define

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

The following result is a consequence of Lemma A.3.

Proposition I.1. *Suppose $y^* \in T$. Let Assumptions I.1 and I.2 hold. Then the following are equivalent:*

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

(ii) D is differentiable at y^*.

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 lemmas in Appendix A once we identify $-D_B$ with f, y^* with z^*, and $\overline{v}$ with w.

By Lemma A.1,

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

From Lemma A.3, 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 4.2. We can state a global version of the proposition.

Corollary I.1 (Global Injectivity)**.** *Let the conditions of Proposition I.1 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 $y^* \in \tilde{T}$.

I.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 I.3. *$D : B \to \mathbb{R} \cup \{-\infty\}$ 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:[44]

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

[44] The relative interior of B is the set

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

Proposition I.2. *Suppose $y^* \in T_{\mathcal{U}}$. Assume $K \geqslant 2$. Let B be the probability simplex and let Assumption I.3 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 I.1 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 I.1.

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 $\mathbb{R}^K$ by

$$\rho_{\mathcal{U}}(u) = \operatorname*{argmax}_{y \in B} \sum_{k=2}^{K} y_k u_k + D\left(\left(1 - \sum_{k=2}^{K} y_k\right), y_2, \ldots, y_K\right). \tag{40}$$

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

$$\rho(u) = \operatorname*{argmax}_{y_2, \ldots, y_K \in \mathbb{R}^{K-1}} \sum_{k=2}^{K} y_k u_k + \tilde{D}(y_2, \ldots, y_K). \tag{41}$$

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 I.1, $\rho_{\mathcal{V}}^{-1}(y^*)$ is a singleton if and only if D is Fréchet differentiable at y^*. □

Appendix J 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 $\overline{v}$, the choice probability vector satisfies

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

where

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

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

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

from which we arrive at the multinomial logit probabilities,

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

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

Proposition J.1. *Let* $P(\overline{v}) = \nabla \ln \left(\sum_m \left(\sum_{k \in N_m} \alpha_{km} e^{\overline{v}_k/\lambda_m} \right)^{\lambda_m} \right)$. *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*

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

By setting $\overline{v} = \vec{v}(x)$ in Proposition J.1, 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 J.1. *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 $P(\overline{v}) = \nabla f(\overline{v})$ be a probability vector. Then*

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

where f^* denotes the convex conjugate of f.

Proof. This follows immediately from Lemma A.1. □

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 J.2. *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 J.3. *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 J.4. *The sum of log-convex functions is log-convex.*

Proof of Proposition J.1. By Lemma J.1, we only need to show that $\ln G_{GNL}(\cdot)$ is log-convex.

By Lemma J.3, the component

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

is log-convex in $\overline{v}$ due to the parameter restrictions $\alpha_{km} \geqslant 0$ and $\lambda_m > 0$. Since $G_{GNL}(\cdot)$ is the sum of such components, Lemma J.4 completes the proof. □