Econometrica, Vol. 87, No. 3 (May, 2019), 1021–1054

IDENTIFICATION WITH ADDITIVELY SEPARABLE HETEROGENEITY

ROY ALLEN
Department of Economics, University of Western Ontario

JOHN REHBECK
Department of Economics, The Ohio State University

This paper provides nonparametric identification results for a class of latent utility models with additively separable unobservable heterogeneity. These results apply to existing models of discrete choice, bundles, decisions under uncertainty, and matching. Under an independence assumption, such models admit a representative agent. As a result, we can identify how regressors alter the desirability of goods using only average demands. Moreover, average indirect utility ("welfare") is identified without needing to specify or identify the distribution of unobservable heterogeneity.

KEYWORDS: Nonparametric identification, representative agent, symmetry, welfare.

1. INTRODUCTION

POPULARIZED IN DISCRETE CHOICE (McFadden (1973)), latent utility models with additively separable unobservable heterogeneity have been used in a variety of literatures. Recent examples include matching (Fox, Yang, and Hsu (2018)), bundles (Gentzkow (2007)), and school choice (Agarwal and Somaini (2018)).

This paper shows that these models share a tractable common structure when we focus on their average demands (formally, conditional means).[1] In contrast with much of the existing literature, we treat the distribution of unobservable heterogeneity as an unknown nuisance object with unrestricted dimension. Nonetheless, average demands are sufficient to identify how the desirability of goods depends on regressors, to identify differences in average indirect utility ("welfare"), and to provide nontrivial bounds on counterfactual demands.

Our identification results apply to models with a particular form of additively separable unobservable heterogeneity. This encompasses the aforementioned examples, which share the structure

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

Roy Allen: rallen46@uwo.ca
John Rehbeck: rehbeck.7@osu.edu

A previous version of this paper was circulated with the title, "Identification of Average Demand Models." We are indebted to three anonymous referees for valuable suggestions. We thank Nikhil Agarwal, Victor Aguiar, Andres Aradillas-Lopez, Brendan Beare, Marc Henry, Keri Hu, Sung Jae Jun, Nail Kashaev, Ivana Komunjer, Lixiong Li, Mark Machina, Francesca Molinari, David Rivers, Andres Santos, Jeffrey Shrader, Matt Shum, Gary Solon, Yixiao Sun, Kaspar Wüthrich, and seminar participants at UC San Diego for helpful comments. Any remaining errors are our own.

[1] In McFadden (1973), these are the conditional choice probabilities. In Fox, Yang, and Hsu (2018), these are the conditional match probabilities for each pair. In Gentzkow (2007), these are the average quantities conditional on observables. In Agarwal and Somaini (2018), these are the conditional school assignment probabilities.

 https://doi.org/10.3982/ECTA15867

where Y is the demand vector for K goods,[2] B is a fixed budget, u_k is a utility index that depends on the regressors X_k, and ε is unobservable heterogeneity, which enters through the disturbance function D. For our results, the form of D is not specified and the distribution of ε is not known and may not be identified. Observable regressors cannot enter D, which is a key separability condition.

This paper shows that when regressors and unobservable heterogeneity are independent, utility indices can be identified up to location and scale. The variation we exploit for identification comes from substitution and complementarity patterns between goods. Within this large class of models, identification up to location and scale has only been established for binary choice or problems whose identification analysis may be converted to a binary choice problem by the technique of identification at infinity (Matzkin (1993)).[3] Next, we show that average indirect utility is identified even when the distribution of unobservable heterogeneity is not. To our knowledge, this result is new outside of discrete choice (McFadden (1981), Small and Rosen (1981)) for general (non-price) regressors. Finally, we provide counterfactual bounds on average demand using the full structure of the model.

A key piece of our analysis is that when regressors and unobservable heterogeneity are independent, latent utility models with additively separable unobservable heterogeneity admit a representative agent. Formally,

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

for certain $\overline{D}$ and $\overline{B}$ arising from the individual problem. Utility maximization at the individual level implies that the conditional mean maximizes a function involving the same utility indices u_k. This aggregation result generalizes results in discrete choice (McFadden (1981), Hofbauer and Sandholm (2002)) to other settings. This aggregation result has several implications.

First, it is not necessary to identify the entire conditional distribution of demand given regressors in order to identify utility indices or average indirect utility. This clarifies data requirements for models in this class. The fact that marginal conditional distributions contain identifying power means that in practice it is not necessary to observe whether goods are consumed together or not. The fact that conditional means are sufficient, rather than needing full marginal distributions, means that perturbed utility models (PUMs) can be used with market-level data.[4] Finally, we note that our analysis applies when the demand vector is latent but its conditional mean is identified. One example with a latent choice is a moral hazard model with unobservable effort. We show that conditional outcome

[2] In McFadden (1973), Y is a vector of indicators denoting which alternative is chosen. In Fox, Yang, and Hsu (2018), it is a match matrix denoting who matches with whom. In Gentzkow (2007), it is the quantity vector. In Agarwal and Somaini (2018), it is a probability distribution over schools an individual might attend.

[3] For multinomial choice, identification at infinity means that there exist covariate values such that only two alternatives are chosen with positive probability. Identification may then be established by leveraging binary choice identification results.

[4] In discrete choice, the conditional distribution of $Y \mid X = x$ and conditional means $\mathbb{E}[Y \mid X = x]$ contain the same information, and so these distinctions are not important. See Berry and Haile (2014) for identification of discrete choice models from market-level data. In contrast, in the bundles model of Gentzkow (2007), conditional means contain potentially less information and so the distinctions we mention are nontrivial.

probabilities are sufficient for identification of a moral hazard model despite this unobservability.[5]

Second, using the aggregation result, we harness the optimizing structure for identification by applying the envelope theorem to the representative agent problem. Classical recoverability results such as Shephard's and Hotelling's lemmas also exploit the envelope theorem (see Varian (1992) for a textbook treatment). Our main technical contribution is not to integrate using the envelope theorem as in classical recoverability, but to differentiate and use asymmetries to identify utility indices.

In more detail, our identification technique uses asymmetries of cross-partial derivatives of the conditional mean to recover utility indices. This technique cancels out the distribution of unobservable heterogeneity. We use exclusion restrictions to establish that utility indices are identified up to location and scale. Symmetry is a well-known testable implication of additively separable heterogeneity in the discrete choice literature if utility indices are *known* (e.g., Anderson, De Palma, and Thisse (1992)). This paper shows that this result extends to a large class of models and shows how to use this testable implication for identification.

Next, we study the average indirect utility function, which is referred to as the social surplus function in discrete choice (McFadden (1978)). We show that this function is equal to the value function of the representative agent's problem. We provide two constructive identification results. The first shows that the average indirect utility function is an integral of identified functions of the conditional mean. This result is analogous to the usual consumer surplus formula but with a change of variables. The second result characterizes the derivative of average indirect utility in terms of the conditional mean. It is a local formula in the spirit of the "sufficient statistics" approach to welfare analysis (Chetty (2009)). It highlights that differences in average indirect utility are identified *without* full identification of all primitives of the model.

Finally, we provide sharp bounds on counterfactual average demands for a change in regressors. Unlike our main identification results, for this step we assume u_k has been identified or is known a priori, while maintaining that unobservable heterogeneity is unspecified. We provide bounds on counterfactual demands by exploiting an underlying shape restriction of the model that follows from the optimizing structure.[6] These bounds hold for any distribution of unobservable heterogeneity and parallel revealed preference bounds in the standard consumer problem (Varian (1982), Blundell, Browning, and Crawford (2003, 2008)).

Our work is complementary to the existing literature, which has largely focused on identification of the distribution of unobservable heterogeneity. Identification of the full distribution of unobservable heterogeneity allows one to identify the distribution of welfare differences given a change in the economic environment and allows one to conduct a variety of counterfactual exercises.[7] In order to identify this distribution, in the existing literature u_k is typically restricted by assuming either it is known in advance, is paramet-

[5]Data limitations arise as well in stochastic choice models when an individual chooses a probability distribution over outcomes, but the econometrician only observes outcomes (Fudenberg, Iijima, and Strzalecki (2015), Allen and Rehbeck (2019)).

[6]The shape restriction is cyclic monotonicity, which is a multivariate version of monotonicity. See Chiong, Hsieh, and Shum (2017) for recent work estimating counterfactuals using this shape restriction.

[7]As mentioned previously, we provide bounds on counterfactual average demands, but these results only hold for changes in regressors, not fundamental changes in the economic environment such as a change in the budget set B.

ric, or is quasilinear with respect to a "special regressor."[8] The identification results we present for utility indices show these assumptions can be relaxed. Once identification of utility indices has been established, identification of features of the distribution of unobservable heterogeneity may be established using results from the existing literature. Examples include Theorems 2 and 3 in Matzkin (1993) and Lewbel (2000) for discrete choice; Fox and Lazzati (2017) for bundles; Theorem 1 in Agarwal and Somaini (2018) for decisions under uncertainty; and Fox, Yang, and Hsu (2018) for matching.

This paper proceeds as follows. Section 2 introduces the model, establishes aggregation, and shows basic properties used in our subsequent analysis. Section 3 provides conditions for nonparametric identification of utility indices. Section 4 constructively identifies changes in the average indirect utility function. Section 5 characterizes counterfactual bounds on average demand using the model. Section 6 provides examples and further literature review. Section 7 concludes the paper.

2. THE MODEL

We consider the identification of latent utility models with a form of additively separable unobservable heterogeneity. The choice vector Y satisfies

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

We call this the *perturbed utility model* (PUM).[9] We interpret Y as the quantity demanded, $B \subseteq \mathbb{R}^K$ as a budget that does not vary with regressors, k as the index of a good, and $X_k = (X_{k,1}, \dots, X_{k,d_k})'$ as the regressors for good k. We collect these regressors in $X = (X_1', \dots, X_K')'$. The utility indices $\vec{u} = (u_1, \dots, u_K)$ encode how regressors vary the desirability of each good. Unobservable heterogeneity is denoted ε and enters through the unknown *disturbance function* D. The interpretation of D differs across examples, but it, together with the budget B, encode complementarity/substitutability information between the goods.

We illustrate PUM with two examples. We provide more examples and a literature review in Section 6.

EXAMPLE 1: The discrete choice *additive random utility model* (ARUM) (McFadden (1981)) specifies a latent utility for each alternative of the form

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

and assumes that individuals choose one alternative that maximizes latent utility. To see that this model is a PUM, let B be the probability simplex

$$B = \Delta^{K-1} := \left\{ y \in \mathbb{R}^K \mid y_k \geq 0, \sum_{k=1}^{K} y_k = 1 \right\}$$

[8]To our knowledge, the only exception is Theorem 4 in Matzkin (1993), which provides fully nonparametric identification of utility indices and the distribution of unobservable heterogeneity for discrete choice.

[9]Focusing on nonlinear budget variation in the consumer problem, McFadden and Fosgerau (2012) studied a similar representation but without regressors (X_k). They did not study identification.

and set $D(y, \varepsilon) = \sum_{k=1}^{K} y_k \varepsilon_k$. Note that typically the optimizing vector Y is a vector of indicators denoting which alternative is chosen. The only time Y may not be a vertex of the probability simplex is when there is a utility tie between several vertices.

EXAMPLE 2: This is a two-good version of the model studied in Gentzkow (2007) and Fox and Lazzati (2017). Assume an individual can buy either 0 or 1 unit of each good. Let $v_{j,k}$ denote utility obtained from quantity j of good 1 and quantity k of good 2. In the setting of Gentzkow (2007), the two goods are online news (good 1) and print news (good 2). We assume utilities are given by

$$\begin{aligned} v_{0,0} &= 0, \\ v_{1,0} &= u_1(X_1) + \varepsilon_{1,0}, \\ v_{0,1} &= u_2(X_2) + \varepsilon_{0,1}, \\ v_{1,1} &= v_{1,0} + v_{0,1} + \varepsilon_{1,1}. \end{aligned} \tag{2}$$

The vector X_1 contains regressors that shift the desirability of online news such as internet speed. Shifters for print news are denoted X_2 and may include the price of print news.

The unobservable random variable $\varepsilon_{1,1}$ determines whether the goods are complements ($\varepsilon_{1,1} > 0$) or substitutes ($\varepsilon_{1,1} < 0$). If we assume that the individual chooses the vector of quantities Y to maximize latent utility, this is a PUM. This may be seen by writing the utility of a quantity vector $\vec{y} = (y_1, y_2)$ as

$$\sum_{k=1}^{2} y_k u_k(X_k) + (y_1 \varepsilon_{1,0} + y_2 \varepsilon_{0,1} + y_1 y_2 \varepsilon_{1,1}).$$

The budget is $B = \{0, 1\}^2$.

Note that in these examples, the budget B is taken as a physical/quantity budget, not a price-income budget as in the standard consumer problem.

PUM embeds two layered assumptions. First, unobservable heterogeneity is additively separable from observable regressors.[10] Second, the quantity vector and utility indices $\vec{u}$ enter linearly. This structure is general enough to encompass many existing models, yet we show it is specific enough to identify utility indices and average indirect utility up to location and scale.

We make the following assumptions.

ASSUMPTION 1: *Assume the following*:

(i) *The random variables* Y, X, *and* ε *satisfy* (1).

(ii) $B \subseteq \mathbb{R}^K$ *is a nonempty, closed set*.

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

(iv) $D : \mathbb{R}^K \times E \to \mathbb{R} \cup \{-\infty\}$ *is an extended real-valued function*, *where* E *denotes the set in which* ε *lies*.

[10]Alternative forms of separability have been studied in the recent literature. Examples include random coefficients models (Ichimura and Thompson (1998), Gautier and Kitamura (2013)), quantile models (Matzkin (2003)), simultaneous equations models (Matzkin (2008, 2015), Berry and Haile (2018)), and triangular models (Imbens and Newey (2009), Torgovitsky (2015), Hoderlein, Holzmann, Kasy, and Meister (2017)).

We formalize the budget set only for familiarity, but it may be absorbed into D. This may be seen by recognizing that B may be set to $\mathbb{R}^K$ and constraints may be interpreted as the points y where $D(y, \varepsilon)$ attains $-\infty$. Note that the set of points y such that $D(y, \varepsilon) = -\infty$ may depend on ε, and so our results allow random constraints. In general, we do not directly place assumptions on E except in our examples. We introduce this set only to describe the domain of definition of D.

2.1. *Aggregation*

An important feature of PUM is that it aggregates. This aggregation result allows us to tractably use the optimizing structure for identification.

THEOREM 1: *Let Assumption 1 hold. Let $x \in \operatorname{supp}(X)$.*[11] *Suppose Y is (X, ε)-measurable*[12] *and both $\mathbb{E}[D(Y(x, \varepsilon), \varepsilon)]$ and $\mathbb{E}[Y(x, \varepsilon)]$ are finite, where the expectations are over the marginal distribution of ε.*

(i) *It follows that*

$$\mathbb{E}\big[Y(x, \varepsilon)\big] \in \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k u_k(x_k) + \overline{D}(y)$$

for $\overline{B}$ the convex hull of B, and $\overline{D}(y) = \sup_{\tilde{Y} \in \mathcal{Y}: \mathbb{E}[\tilde{Y}] = y} \mathbb{E}[D(\tilde{Y}, \varepsilon)]$, where $\mathcal{Y}$ is the set of ε-measurable functions that map to B.

(ii) *Define the indirect utility function for the representative agent problem,*

$$V(\vec{v}) = \max_{y \in \overline{B}} \sum_{k=1}^{K} y_k v_k + \overline{D}(y).$$

It follows that

$$V\big(\vec{u}(x)\big) = \mathbb{E}\left[\max_{y \in B} \sum_{k=1}^{K} y_k u_k(x_k) + D(y, \varepsilon)\right].$$

This result states how the mean $\mathbb{E}[Y(x, \varepsilon)]$ varies when x changes, provided the measure over ε is fixed. Note that at this point, $\mathbb{E}[Y(x, \varepsilon)]$ is not necessarily the conditional mean of Y given X, since we have not described the joint distribution of X and ε; later we will assume independence, to more properly interpret this as an aggregation theorem. Part (i) characterizes the average structural function $\mathbb{E}[Y(x, \varepsilon)]$ as the maximizer of a function, and characterizes how the aggregate disturbance function $\overline{D}$ relates to the individual disturbance functions and the distribution of unobservable heterogeneity. An important feature for our subsequent analysis is that x does not enter the function $\overline{D}$. Part (ii) states that the representative agent's indirect utility function is the average of the individual indirect utility functions. We use this result to identify the average indirect utility from the conditional mean.

[11]The support of a random variable such as X, denoted $\operatorname{supp}(X)$, is the smallest closed set S such that $P(X \in S) = 1$.

[12]We provide sufficient conditions for existence of a measurable selector in Appendix S.2 of the Supplemental Material (Allen and Rehbeck (2019)).

A key step in the proof of Theorem 1 is to establish that the average indirect utility satisfies

$$\mathbb{E}\left[\sum_{k=1}^{K} Y_k(x,\varepsilon)u_k(x_k) + D\big(Y(x,\varepsilon),\varepsilon\big)\right]$$
$$= \sum_{k=1}^{K} \mathbb{E}\big[Y_k(x,\varepsilon)\big]u_k(x_k) + \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=\mathbb{E}[Y(x,\varepsilon)]} \mathbb{E}\big[D(\tilde{Y},\varepsilon)\big].$$

This equality illustrates how the utility indices pass outside of the expectation for a fixed x. The second feature of this equality is that the additive separability in the latent utility model (between x and ε) is preserved upon aggregation (between x and the distribution of ε). For intuition on why the supremum arises, note that because $Y(x,\varepsilon)$ is ε-measurable for fixed x,

$$\mathbb{E}\big[D\big(Y(x,\varepsilon),\varepsilon\big)\big] \leq \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=\mathbb{E}[Y(x,\varepsilon)]} \mathbb{E}\big[D(\tilde{Y},\varepsilon)\big].$$

We show this is an equality. If it were not an equality, then there would be some way to improve the "match" between quantities and unobservable heterogeneity, while keeping the mean quantity constant. Recall that the mean quantity is all that matters for describing the average "match" between quantities and regressors. Thus, if this inequality were strict, then we would reach a contradiction because there would be a way to obtain a better match between quantities and unobservables, holding the match between quantities and covariates fixed.

Theorem 1 is related to other representative agent arguments. For example, in the standard consumer problem, separability in the individual problem is inherited upon aggregation for the Gorman polar form of indirect utility (Gorman (1953)). The simplest proof of aggregation for the Gorman form proceeds by differentiating the indirect utility, applying Roy's identity, and then taking expectations. Our proof does not require differentiability of the value function and allows multiple maximizers.[13] We note that a special case of the Gorman form is quasilinear utility. Theorem 1 covers quasilinear utility by considering $u_k(x_k)$ as the negative of the price of good k, and $B = \mathbb{R}^K_+$ as an unrestricted budget. Theorem 1 also covers aggregation of price-taking, profit maximizing firms as a special case. For brevity, we present the details after the proof of Theorem 1 in Appendix A.

In order to apply Theorem 1 for identification, we now link the aggregation theorem to the joint distribution of Y and X. The identification results in this paper require identification of $\mathbb{E}[Y(x,\varepsilon)]$ at certain values of x. When this function is continuous in x and X and ε are independent, a continuous version of the conditional expectation of Y given X exists.[14] This version satisfies $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y(x,\varepsilon)]$ for each x in the support of X. Thus, we identify $\mathbb{E}[Y(x,\varepsilon)]$ from the conditional mean of Y given X. In the sequel, we place assumptions directly on $\mathbb{E}[Y \mid X = x]$. We formalize these assumptions below.

[13]This differs from the proof of the Williams–Daly–Zachary theorem (McFadden (1981)) for additive random utility models where differentiability holds over a suitable set of points so aggregation can still be proven by Roy's identity.

[14]We provide sufficient conditions for continuity of $\mathbb{E}[Y(x,\varepsilon)]$ in Lemma A.3.

ASSUMPTION 2: *Assume the following*:

(i)

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

holds for every $x \in \operatorname{supp}(X)$.

(ii) $\overline{B} \subseteq \mathbb{R}^K$ *is a nonempty, closed, and convex set.*

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

(iv) $\overline{D} : \mathbb{R}^K \to \mathbb{R} \cup \{-\infty\}$ *is concave, upper semi-continuous,*[15] *and finite at some* $y \in \overline{B}$.[16]

Note that (i) imposes that there is a unique maximizer. This is not implied by Theorem 1, but holds if $Y(x, \varepsilon)$ in the latent utility model is the unique maximizer for almost every ε and additional technical conditions hold.[17] In ARUM or the bundles model, a sufficient condition is that ε has a density with respect to Lebesgue measure.

REMARK 1—Maximum Score: The representative agent formulation allows us to use tools from convex analysis and does not require us to work explicitly with integrals of latent distributions. In Appendix S.4 of the Supplemental Material (Allen and Rehbeck (2019)), we use the representative agent formulation to derive a generalization of the maximum score inequalities from the discrete choice literature (Manski (1975), Matzkin (1993), Goeree, Holt, and Palfrey (2005), Fox (2007)). We derive these general inequalities using revealed preference techniques and a symmetry condition on $\overline{D}$.

2.2. *Model Structure*

Our identification results make use of several lemmas that have analogues in classical producer and consumer theory.

First, we present a version of Roy's identity. Assumption 2 ensures that the conditional mean is the unique maximizer of the representative agent's utility function, and so the conditional mean is a function of the indices $\vec{u}(x)$. We characterize this function by using the envelope theorem. The following result is a generalization of the Williams–Daly–Zachary theorem (McFadden (1981)) from discrete choice.

LEMMA 1—"Roy's Identity": *Let Assumption 2 hold. Then*

$$\mathbb{E}[Y \mid X = x] = \nabla V\big(\vec{u}(x)\big).$$

The function V is the average indirect utility function defined in Theorem 1, and ∇V denotes its gradient. Differentiability of V follows from the assumption of a unique maximizer. We also make use of the following lemma.

[15] A function $f : \mathbb{R}^K \to \mathbb{R} \cup \{-\infty, \infty\}$ is upper semi-continuous if for each α, the set $\{y \in \mathbb{R}^K \mid f(y) < \alpha\}$ is open.

[16] The aggregate disturbance function in Theorem 1 may not satisfy concavity, but this is not essential for our results. If $\overline{D}$ does not satisfy condition (iv), but part (i) still holds, then it is mild to assume there is some other disturbance function $\tilde{D}$ such that parts (i) and (iv) hold. See Theorem S.7.1 in the Supplemental Material.

[17] See Lemma A.1.

LEMMA 2—"Slutsky Conditions": *Let Assumption 2 hold and let V be twice continuously differentiable in a neighborhood of $\vec{v} \in \mathbb{R}^K$. Then*

(i) $\lambda' \nabla^2 V(\vec{v})\lambda \geq 0, \forall \lambda \in \mathbb{R}^K$.

(ii) $\partial_{k,\ell} V(\vec{v}) = \partial_{\ell,k} V(\vec{v})$ *for* $k, \ell = 1, \dots, K$.

We refer to (i) as positive semi-definiteness and (ii) as symmetry. These conditions are well-known in the discrete choice literature,[18] but to our knowledge have not been exploited for identification within this class of models.[19] Symmetry provides a cross-equation *equality* that is the foundation for our identification results for utility indices.

A dual version of Roy's identity holds here as well.

LEMMA 3: *Let Assumption 2 hold. Define*

$$\overline{D}_{\overline{B}}(y) = \begin{cases} \overline{D}(y), & \textit{if } y \in \overline{B}, \\ -\infty, & \textit{otherwise}. \end{cases}$$

If $\overline{D}_{\overline{B}}$ is differentiable at $\mathbb{E}[Y \mid X = x]$, then

$$\vec{u}(x) = -\nabla \overline{D}_{\overline{B}}\big(\mathbb{E}[Y \mid X = x]\big).$$

A closely related result has previously been exploited for identification in matching games (Galichon and Salanié (2015)) and dynamic discrete choice (Chiong, Galichon, and Shum (2016)) when $\overline{D}_{\overline{B}}$ is known. A related technique has been used for identification in a class of games (Kline (2017), Larsen and Zhang (2018)). Theorem 1 shows how to construct $\overline{D}_{\overline{B}}$ for any PUM whenever D, B, and the distribution of ε are known, and thus Lemma 3 provides a constructive identification formula when these are known. Knowledge of $\overline{D}_{\overline{B}}$ is not required for our subsequent identification results.

3. IDENTIFICATION OF UTILITY INDICES

This section provides conditions under which utility indices $\vec{u}$ are nonparametrically identified. We combine Lemma 1 ("Roy's identity") and Lemma 2(ii) ("Slutsky symmetry") to identify utility indices from asymmetries of partial derivatives of the conditional mean.

We treat the aggregate disturbance $\overline{D}$ as an unknown function. We assume the average demand $\mathbb{E}[Y \mid X = x]$ is identified over $x \in \operatorname{supp}(X)$. This approach abstracts from sampling error in order to understand whether utility indices can be uniquely determined with ideal, population-level data. For example, knowledge of $\mathbb{E}[Y \mid X = x]$ may be obtained from an "infinite" number of independent and identically distributed draws of (Y, X).

Our results require that each good have a continuous regressor that is excluded from the other equations.[20] In Example 2, internet speed satisfies this assumption for online

[18]These conditions have also been recognized in the matching literature (Decker, Lieb, McCann, and Stephens (2013), Galichon and Salanié (2015)).

[19]Recently, Abaluck and Adams (2018) used asymmetries to identify consideration set probabilities in a discrete choice model. Their model is not in the PUM class and their use of asymmetries differs from ours.

[20]The primary contribution of this paper is to establish point identification, and our technique requires continuous regressors. In Appendix S.7 of the Supplemental Material, we provide a characterization of the identified set for utility indices. This result applies with discrete regressors.

news, and the price of print news satisfies this assumption for print news. Formally, partition the regressors for each good into $X_k = (Z_k', W')'$. The d_z^k-dimensional vector Z_k consists of regressors that are *excluded* from each function u_ℓ for $\ell \neq k$. We require $d_z^k > 0$. The d_w-dimensional vector W contains regressors that are common across goods such as age, education, and household income.

Identification of utility indices is established as follows. First, we assume all regressors are continuous and that there are no common regressors. Using the analogue of Slutsky symmetry, we constructively identify utility indices up to a location/scale normalization. We then use this result to identify utility indices when there are common or discrete regressors using Lemma 3.

3.1. *Identification for Good-Specific Regressors*

We provide identification results when there are no common regressors. In particular, each regressor shows up in exactly one utility index u_k. These identification results hold as well if we condition on a fixed value w of common regressors. We later use this fact to identify utility indices when there are common regressors.

We informally outline how asymmetries identify utility indices. For simplicity, suppose that x_k is a scalar for $k = 1, \dots, K$. Using Lemma 1 and the chain rule, for arbitrary k, ℓ we have

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} = \partial_{\ell,k} V(\overline{u})\big|_{\overline{u}=\vec{u}(x)} \frac{\partial u_\ell(x_\ell)}{\partial x_\ell}. \tag{3}$$

This relies on the fact that x_ℓ is excluded from $u_j(\cdot)$ for $j \neq \ell$. In addition, x_ℓ must be continuous so we can take a derivative. An analogous equation holds with k and ℓ interchanged. Assuming all involved derivatives are nonzero, combining (3) with symmetry of second-order mixed partial derivatives of V (Lemma 2(ii)), we obtain

$$\frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell} \bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_k} = \frac{\partial u_\ell(x_\ell)}{\partial x_\ell} \bigg/ \frac{\partial u_k(x_k)}{\partial x_k}. \tag{4}$$

Thus, we identify the ratio of certain partial derivatives of utility indices.

We now consider the general case where x_k is not a scalar and formalize the arguments leading to (4). Our identification results build on local variation, and only require certain smoothness assumptions to hold locally.[21] Thus, we make explicit the points where partial derivatives are evaluated. For the following definition, $x_{\ell,p}$ denotes the pth regressor for the ℓth good.

DEFINITION 1—Pairs: Say that $\frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}\big|_{x_\ell = x_\ell^*}$ and $\frac{\partial u_k(x_k)}{\partial x_{k,q}}\big|_{x_k = x_k^*}$ are paired if they exist and the following conditions hold:

(i) There exists a known value $x^* = (x_1^{*'}, \dots, x_K^{*'})' \in \operatorname{supp}(X)$ that has ℓth component x_ℓ^* and kth component x_k^*.

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

(iii) V is twice continuously differentiable in a neighborhood of $\vec{u}(x^*)$.

(iv) $\partial_{\ell,k} V(\overline{u})|_{\overline{u}=\vec{u}(x^*)} \neq 0$.

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

[21] We do not need (4) to hold as a functional relationship for all x. Instead, our results rely on a version of (4) holding at a rich set of points.

Part (i) is stated in terms of x^* being known so that 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 regressors. Part (iii) ensures symmetry of mixed partials of V. Part (iv) is a behavioral restriction requiring that if good ℓ becomes more attractive, then there need to be "spillovers" to good k.[22] These occur precisely when goods satisfy a local form of substitutability or complementarity. If two goods ℓ and k are everywhere substitutes, then $\partial_{\ell,k}V < 0$.[23] If they are everywhere complements, then $\partial_{\ell,k}V > 0$.

PROPOSITION 1: *Let Assumption 2 hold and assume $x_{\ell,p}$ and $x_{k,q}$ are regressors specific to ℓ and k, respectively. If the points $\frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}|_{x_\ell=x_\ell^*}$ and $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x_k^*}$ are paired and $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x_k^*} \neq 0$, then there is some known $x^* \in \operatorname{supp}(X)$ such that*

$$\frac{\partial \mathbb{E}[Y_k \mid X=x]}{\partial x_{\ell,p}}\bigg|_{x=x^*} \Bigg/ \frac{\partial \mathbb{E}[Y_\ell \mid X=x]}{\partial x_{k,q}}\bigg|_{x=x^*} = \frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}\bigg|_{x_\ell=x_\ell^*} \Bigg/ \frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=x_k^*}, \tag{5}$$

where x_ℓ^* and x_k^* are components of x^* specific to goods ℓ and k.[24] *In particular, the right-hand side of (5) is identified.*

Recall that we define $\mathbb{E}[Y \mid X=x]$ as a continuous version of the conditional expectation of Y given X. This function is uniquely defined for $x \in \operatorname{supp}(X)$. Thus, the derivatives of conditional expectations are identifiable.

This constructive identification result formalizes how asymmetries provide identifying information on the utility indices. When the goods are the same, $k=\ell$, this formula only identifies the slope of the level sets of u_k. However, when the goods are distinct, this formula provides information on the relative derivatives of different utility indices. Note that the left-hand side of (5) involves the regressors for all goods x^*, whereas the right-hand side only involves the regressors for alternatives ℓ and k. Thus, there is overidentifying information on utility indices when $K > 2$.

In order to identify utility indices, we require a normalization. To see this, let $\vec{c} \in \mathbb{R}^k$ and let $\lambda > 0$ be a scalar. Then we have the equality

$$\operatorname*{argmax}_{y\in\overline{B}} \sum_{k=1}^{K} y_k u_k(x_k) + \overline{D}(y) = \operatorname*{argmax}_{y\in\overline{B}} \sum_{k=1}^{K} y_k\big(\lambda u_k(x_k) + c_k\big) + \left(\lambda\overline{D}(y) - \sum_{k=1}^{K} y_k c_k\right). \tag{6}$$

Because we do not specify the aggregate disturbance function $\overline{D}$, identification requires two normalizations to handle both λ and $\vec{c}$. This is because if $\vec{u}$ is consistent with the model, then $\lambda\vec{u} + \vec{c}$ is as well.

The following theorem is our main result for identification of utility indices. Its assumptions ensure that the points $\frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}|_{x_\ell=x_\ell^*}$ and $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x_k^*}$ are paired whenever $k \neq \ell$ and these derivatives are nonzero. Identification of utility indices follows using Proposition 1 and the fundamental theorem of calculus.

[22] Recall (3) illustrates how mixed partial derivatives of V determine how changes in one utility index affect the demand for *another* good.

[23] Berry, Gandhi, and Haile (2013) considered a notion of substitutes that does not require differentiability. Their connected substitutes condition involves all goods while the local condition $\partial_{\ell,k}V < 0$ involves just goods ℓ and k.

[24] Note that $\frac{\partial \mathbb{E}[Y_\ell|X=x]}{\partial x_{k,q}}|_{x=x^*}$ is nonzero from (3) and the other assumptions in the proposition.

THEOREM 2: *Let Assumption 2 hold. Assume all regressors are good-specific and $K \geq 2$. Assume X has a rectangular support,*[25] *$\vec{u}$ is differentiable, $\vec{u}(\operatorname{supp}(X))$ contains an open ball,*[26] *and V is twice continuously differentiable with nonzero second-order mixed partial derivatives. It follows that $\vec{u} = (u_1, \dots, u_K)$ is identified over $\operatorname{supp}(X)$ under the following normalization:*

(i) *(Scale)* $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=x_k^*} \in \{-1, 1\}$ *for a tuple (k, q, x_k^*) such that $x_k^* \in \operatorname{supp}(X_k)$.*

(ii) *(Location) $u_\ell(0_{d_\ell}) = 0$ for each $\ell = 1, \dots, K$, where 0_{d_ℓ} denotes a d_ℓ-dimensional vector of zeros and $0_{d_\ell} \in \operatorname{supp}(X_\ell)$.*

This identification result requires that the model involve *multiple* goods in a nontrivial way. If $K = 1$ or there is no complementarity/substitutability across goods, then without additional assumptions each u_k is at best only identified up to a monotonic transformation.[27] In this sense, use of a cross-good restriction such as symmetry is necessary to identify utility indices up to location and scale, unless additional restrictions are imposed.

The assumptions of Theorem 2 are overly strong in some contexts. In Example 2, it is not innocuous to assume that all mixed partials of V are *everywhere* nonzero. Ruling out zero mixed partial derivatives does not allow goods to switch from being complements ($\partial_{j,k}V > 0$) to being substitutes ($\partial_{j,k}V < 0$) by continuity of derivatives. In Appendix S.5 of the Supplemental Material, we show identification of utility indices under a weaker set of assumptions that accommodates these cases. The basic idea is that Proposition 1 identifies derivative ratios of utility indices. We multiply these derivative ratios to identify new derivative ratios. If all derivative ratios of $\vec{u}$ are identified, then we identify $\vec{u}$ up to location and scale.

3.2. *Identification for Common Regressors*

In this section, we provide sufficient conditions for identification when there are discrete regressors or regressors that are not good-specific. We require that each good have at least one good-specific, continuous regressor. All other regressors can be discrete.

Recall we assume $X_k = (Z_k', W')'$, where W consists of regressors that are common across alternatives, and Z_k consists of regressors specific to alternative k. For simplicity, we assume Z_k is a scalar continuous regressor specific to good k.[28] All other regressors are absorbed into W. Let $Z = (Z_1, \dots, Z_K)'$.

ASSUMPTION 3: *There exists a known $\tilde{w} \in \operatorname{supp}(W)$ such that*

(i) $\vec{u}(\operatorname{supp}(Z \mid W = \tilde{w}), \tilde{w}) = \vec{u}(\operatorname{supp}(Z, W))$.

(ii) *$\vec{u}(z, \tilde{w})$ is identified for each $z \in \operatorname{supp}(Z \mid W = \tilde{w})$.*

Part (i) states that, conditional on a value $\tilde{w}$, Z can move sufficiently to trace out the full variation of utility indices. This is a relevance and support condition. Using the assumption that the conditional mean is the unique maximizer, (i) implies that for each $(z', w')' \in \operatorname{supp}(Z, W)$, there exists $\tilde{z} \in \operatorname{supp}(Z \mid W = \tilde{w})$ such that

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

[25] Formally, $\operatorname{supp}(X)$ is the Cartesian product of $\sum_{k=1}^K d_k$ intervals, each with nonempty interior.

[26] Given the other conditions, $\vec{u}(\operatorname{supp}(X))$ contains an open ball provided, for each good k, $u_k(\operatorname{supp}(X_k))$ is not a singleton.

[27] This is formalized in Remark S.7.1 in Appendix S.7 of the Supplemental Material.

[28] The results in this section can easily be adapted to handle multiple continuous or discrete regressors specific to each good, provided there is at least one continuous regressor for each good.

Condition (i) is not essential, and may be relaxed as discussed in Appendix S.3 of the Supplemental Material. Sufficient conditions for (ii) are given in Theorem 2. Recall that the previous results also hold *conditional* on $W = \tilde{w}$.

For the following theorem, recall

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

THEOREM 3: *Let Assumptions 2 and 3 hold. If the derivative of* $\overline{D}_{\overline{B}}$ *exists at* $\mathbb{E}[Y \mid Z = z, W = w]$ *and* $(z, w) \in \operatorname{supp}(Z, W)$, *then* $\vec{u}(z, w)$ *is identified.*

The assumptions on $\overline{D}_{\overline{B}}$ allow us to use Lemma 3 to establish that

$$\mathbb{E}[Y \mid Z = z, W = w] = \mathbb{E}[Y \mid Z = \tilde{z}, W = \tilde{w}] \quad \Longrightarrow \quad \vec{u}(z, w) = \vec{u}(\tilde{z}, \tilde{w}). \tag{7}$$

From this implication, we identify $\vec{u}(z, w)$ by "matching" it with a value $\vec{u}(\tilde{z}, \tilde{w})$ that is already identified. With the maintained assumptions that the aggregate disturbance $\overline{D}$ is concave and the budget $\overline{B}$ is convex, differentiability of $\overline{D}_{\overline{B}}$ is actually *necessary* for this implication without further restrictions.[29]

REMARK 2: Differentiability of $\overline{D}_{\overline{B}}$ at $\mathbb{E}[Y \mid Z = z, W = w]$ implies that this value cannot be on the boundary of $\overline{B}$. In particular, differentiability requires that $\overline{B}$ 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 $u_1(\cdot) = 0$, which is sensible if the first good is an outside option such as "buy nothing." With this normalization, the problem is reparameterized to eliminate the first good.[30] If there is no outside option, an alternative location normalization is that for some z_1, $u_1(z_1, w) = 0$ for each w. This assumption rules out the possibility that a change in w can change all utility indices by the same amount. Such a change would not affect choice probabilities, and so we would not be able to distinguish the change in the level of each index.

4. IDENTIFICATION OF AVERAGE INDIRECT UTILITY

We now study identification of V, which from Theorem 1 may be interpreted as average indirect utility,

$$V\big(\vec{u}(x)\big) = \mathbb{E}\left[\max_{y \in B} \sum_{k=1}^{K} y_k u_k(x_k) + D(y, \varepsilon)\right].$$

We show that differences in average indirect utility are identified from local information from the conditional mean. In particular, we do not need the entire function $\vec{u}$ to be identified, nor do we need the disturbance D or the distribution of unobservable heterogeneity to be identified.

[29] See Lemma B.1.

[30] See Appendix S.6 of the Supplemental Material for more details.

One may interpret V as a welfare object when we assume that the units of individual indirect utility are comparable across people. This follows assuming $\vec{u}$ has the same scale for each individual. Specialized to ARUM, V becomes

$$V(\vec{u}(x)) = \mathbb{E}\Big[\max_k u_k(x_k) + \varepsilon_k\Big],$$

which has been widely used in discrete choice to quantify welfare changes (McFadden (1981), Small and Rosen (1981)).[31]

Depending on the interpretation of the disturbance function D, alternative objects may be more natural measures of welfare. In ARUM, if the additive error is interpreted as mistakes, then differences in $\sum_{k=1}^{K} \mathbb{E}[Y_k \mid X = x]u_k(x_k)$ for different values of x may be welfare relevant. We refer to V as a welfare measure, with the caveat that its interpretation depends on the context.

This section proceeds as follows. First, we show how to identify differences in average indirect utility and the aggregate disturbance function once utility indices are identified. Next, we drop the assumption that utility indices are already identified and directly identify derivatives of V as functions of the conditional mean. This highlights that welfare analysis is possible without identifying all functions in the model.

4.1. *An Integral Representation for Average Indirect Utility*

We now establish point identification of differences in average indirect utility and the aggregate disturbance function. Recall Lemma 1 yields a version of Roy's identity,

$$\mathbb{E}[Y \mid X = x] = \nabla V(\vec{u}(x)).$$

We show that differences in average indirect utility are identified by integrating this identity.

Our previous results studied identification of $\vec{u}$. We now assume $\vec{u}$ is known, which allows us to relax Assumption 2. We present this weaker assumption because a researcher may be interested in welfare measurement with more structure on utility indices, such as $u_k(x_k) = x_k$.

ASSUMPTION 4: *Let Assumption 2 hold except* (i) *allows multiple maximizers*.

This weaker assumption allows the distribution of unobservable heterogeneity to be discrete in examples such as the additive random utility model. The following theorem identifies differences in V by an analogue of the consumer surplus formula with a change of variables. The change of variables is made with the function $x(t)$, which draws a path between x^0 and x^1 with utility indices beginning at $\vec{u}(x^0)$ and ending at $\vec{u}(x^1)$.

THEOREM 4: *Let Assumption 4 hold. Assume $\vec{u}$ 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 $\vec{u}(x(t)) = t\vec{u}(x^1) + (1-t)\vec{u}(x^0)$ and $x(t) \in \operatorname{supp}(X)$ for $t \in [0, 1]$. It follows that*

$$V(\vec{u}(x^1)) - V(\vec{u}(x^0))$$

[31] See also Bhattacharya (2015), who provided identification results in discrete choice with nonseparable heterogeneity when there is variation in prices and income. Our results apply without price or income variation and extend beyond discrete choice.

and

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

are identified. *In particular,*

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

This result shows that conditional means are sufficient for identification of average welfare. Specifically, differences in average indirect utility may be identified as the area under the average demand curve after a change of variables.

When $\vec{u}(\text{supp}(X))$ is convex, the function $x(t)$ may always be constructed and we deduce the following corollary.

COROLLARY 1: *Let Assumption 4 hold*. *Suppose $\vec{u}$ is known*, *V is everywhere finite*, *and the set $\vec{u}(\text{supp}(X))$ is convex. It follows that for every $x^0, x^1 \in \text{supp}(X)$,*

$$V\big(\vec{u}\big(x^1\big)\big) - V\big(\vec{u}\big(x^0\big)\big)$$

and

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

are identified.

Sufficient conditions for convexity of $\vec{u}(\text{supp}(X))$ are that $\text{supp}(X)$ is rectangular and $\vec{u}$ is continuous, since then $\vec{u}(\text{supp}(X)) = \prod_{k=1}^{K} u_k(\text{supp}(X_k))$ and each set $u_k(\text{supp}(X_k))$ is an interval.

4.2. *A "Sufficient Statistics" Formula for Welfare Changes*

Theorem 4 is a two-step identification result. It assumes utility indices are first identified by previous results, and then identifies differences in the average indirect utility function. In this section, we provide an explicit one-step mapping from conditional means to welfare changes.

For notational simplicity, we assume each good has one continuous regressor, which is excluded from the indices of the other goods. Recall that under smoothness conditions, Lemma 1 yields

$$\frac{\partial V\big(\vec{u}(x)\big)}{\partial x_\ell} = \mathbb{E}[Y_\ell \mid X = x]\frac{\partial u_\ell(x_\ell)}{\partial x_\ell}.$$ [32]

Proposition 1 provides conditions under which

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

[32] This equality between partial derivatives may be taken as a pointwise equality (i.e., holding fixed the points of evaluation), or a functional relationship.

Combining these equations, we have the following "sufficient statistics" (Chetty (2009)) formula for local changes in welfare:

$$\frac{\partial V(\vec{u}(x))}{\partial x_\ell} = \mathbb{E}[Y_\ell \mid X = x] \frac{\dfrac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell}}{\dfrac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_k}} \frac{\partial u_k(x_k)}{\partial x_k}. \tag{8}$$

By integrating over x_ℓ, holding other regressors fixed, this equality identifies differences in V up to the scale term $\frac{\partial u_k(x_k)}{\partial x_k}$. For example, if x^1 and x^0 only differ with respect to the regressor for good ℓ, then

$$V(\vec{u}(x^1)) - V(\vec{u}(x^0)) = \int_{x_\ell^0}^{x_\ell^1} \mathbb{E}[Y_\ell \mid X = x] \frac{\dfrac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_\ell}}{\dfrac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_k}} dx_\ell \frac{\partial u_k(x_k)}{\partial x_k}.$$

Note that the scale term is at a fixed value x_k. One may interpret this scale as a local conversion rate between utils and regressor x_k.

5. COUNTERFACTUAL BOUNDS

In this section, we show how to use the PUM framework to provide counterfactual bounds on the average structural function $\mathbb{E}[Y(x^0, \varepsilon)]$ for x^0 outside the support of X, where the expectation is over the marginal distribution of ε. We establish bounds assuming $\vec{u}$ is known a priori, but $\overline{D}$ is unknown. These theoretical results provide a foundation for a "semiparametric" approach to counterfactual analysis, wherein $\vec{u}$ is estimated in a way that facilitates extrapolation to out-of-sample values, but $\overline{D}$ is an unrestricted nuisance function.

The assumption we make on $\mathbb{E}[Y(x^0, \varepsilon)]$ at an out-of-sample value $x^0 \notin \operatorname{supp}(X)$ is that

$$\mathbb{E}[Y(x^0, \varepsilon)] \in \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k u_k(x_k^0) + \overline{D}(y).^{33} \tag{9}$$

By leveraging a shape restriction implied by this maximization problem, we show how knowledge of $\mathbb{E}[Y \mid X = x]$ over the support of X, together with knowledge of $\vec{u}$ at x^0 and over the support of X, delivers inequality restrictions on $\mathbb{E}[Y(x^0, \varepsilon)]$. This approach originates in the revealed preference literature (Varian (1982), Blundell, Browning, and Crawford (2003, 2008)).

We formalize sharp bounds maintaining Assumption 4, which relaxes Assumption 2 to allow multiple maximizers.

THEOREM 5—Sharp Counterfactual Bounds: *Let Assumption 4 hold and assume $\vec{u}$ is known. Let $x^0 \notin \operatorname{supp}(X)$ and assume*

$$\mathbb{E}[Y(x^0, \varepsilon)] \in \operatorname*{argmax}_{y \in \overline{B}} \sum_{k=1}^{K} y_k u_k(x_k^0) + \overline{D}(y).$$

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

Then $\mathbb{E}[Y(x^0, \varepsilon)] \in \overline{B}$ *and, for every integer* S *and sequence* $x^1, \dots, x^{S-1}$ *of points in* $\operatorname{supp}(X)$,

$$\mathbb{E}\big[Y\big(x^0, \varepsilon\big)\big]'\big(\vec{u}\big(x^0\big) - \vec{u}\big(x^{S-1}\big)\big) \geq \mathbb{E}\big[Y \mid X = x^1\big]'\vec{u}\big(x^0\big) - \mathbb{E}\big[Y \mid X = x^{S-1}\big]\vec{u}\big(x^{S-1}\big) - \sum_{s=1}^{S-2}\big(\mathbb{E}\big[Y \mid X = x^s\big] - \mathbb{E}\big[Y \mid X = x^{s+1}\big]\big)'\vec{u}\big(x^s\big). \quad (10)$$

These bounds are sharp if no further restrictions are placed on $\overline{D}$.

We note that in many examples, *more* structure is placed on $\overline{D}$. In such cases, these bounds may no longer be sharp. These bounds are easiest to interpret if we consider sequences consisting of a single point. In this case, (10) becomes

$$\big(\mathbb{E}\big[Y\big(x^0, \varepsilon\big)\big] - \mathbb{E}[Y \mid X = x]\big)'\big(\vec{u}\big(x^0\big) - \vec{u}(x)\big) \geq 0$$

for each $x \in \operatorname{supp}(X)$. This is a version of the multivariate law of demand (cf. Mas-Colell, Whinston, and Green (1995), p. 111). Utility indices are analogous to the negative of the price vector.

When $\operatorname{supp}(X)$ is finite, the counterfactual restrictions on $\mathbb{E}[Y(x^0, \varepsilon)]$ are given by a finite number of linear inequalities, intersected with the budget $\overline{B}$. This is because each cycle constitutes a linear restriction and there are finitely many such cycles. The number of cycles grows quickly with the cardinality of $\operatorname{supp}(X)$, however, and so directly operationalizing these inequalities can be computationally expensive.[34]

REMARK 3: This theorem has some important implications in practice. Suppose we have an estimate of $\vec{u}$ over its entire domain, for example, we specify $\vec{u}$ to be linear and estimate the slope coefficients. Suppose we also have estimates of $\mathbb{E}[Y \mid X = x]$ for values of x over a certain set, for example, over the observations of covariates X^i in a particular data set. To apply Theorem 5, the estimates of $\vec{u}$ and $\mathbb{E}[Y \mid X = x]$ must be consistent with the restrictions of the model. When these estimates are not consistent with the model, the restrictions of (10), with the estimates plugged in, rule out *any* possible conjectured value of $\mathbb{E}[Y(x^0, \varepsilon)]$. This highlights an advantage of using shape-restricted estimation methods: If estimates of $\vec{u}$ and $\mathbb{E}[Y \mid X = x]$ are constrained to satisfy the restrictions of the model, then Theorem 5 can directly be used to provide counterfactual bounds.

6. EXAMPLES

This section illustrates the applicability of our results in several examples. We first revisit the additive random utility model (Example 1) and the bundles model (Example 2). Finally, we turn to further examples: decisions under uncertainty, stochastic choice, moral hazard, and matching.

In several settings, existing work has shown that once utility indices are known, features of the distribution of unobservable heterogeneity are identified. Our results may be used to extend these identification results since we show utility indices are nonparametrically identified. We provide several examples in this section. However, our analysis shows

[34]Focusing on linear utility indices, Chiong, Hsieh, and Shum (2017) have recently demonstrated that it is tractable to calculate certain counterfactual bounds.

that if interest lies only in average indirect utility, then it is not necessary to identify the distribution of unobservable heterogeneity, which requires additional restrictions on the disturbance function D.

Our analysis also complements existing identification work on these models by showing the identifying power of means for PUMs. When one is only interested in utility indices and average indirect utility, conditional means contain sufficient identifying power. Thus, the results in this paper can be employed with market-level/aggregated data or other settings in which the full distribution of demand is not identified.[35]

6.1. *Additive Random Utility Model*

Discrete choice models have received considerable attention in the econometrics literature.[36] A special case is the additive random utility model (ARUM) of Example 1. Identification of ARUM has been extensively studied. Fully nonparametric identification of ARUM has been established by Theorem 4 in Matzkin (1993) and so we discuss how we differ in detail. We then relate our work with Shi, Shum, and Song (2018), who also used the envelope theorem for identification with linear utility indices.

Our assumptions and results differ from Theorem 4 in Matzkin (1993) in several ways. First, we do not require identification at infinity, which is a technique that reduces identification to a binary choice problem by assuming there are regressor values such that only two alternatives are chosen with positive probability.[37] Second, our results are constructive, providing an explicit mapping from conditional means to utility indices. Third, we show that differences in welfare are identified using local variation in the conditional mean, without identifying the distribution of unobservable heterogeneity.[38] Finally, our results do not rely on the full structure of the additive random utility model. These results also apply to more general discrete choice models allowing bounded rationality, which we discuss in Section 6.4. Note that we do not study identification of the distribution of unobservable heterogeneity, which was covered by Matzkin (1993).

Shi, Shum, and Song (2018) used the envelope theorem for identification in a panel setting. For a simplified presentation of their technique, begin with Lemma 1,

$$\mathbb{E}[Y \mid X = x] = \nabla V\big(\vec{u}(x)\big).$$

This is an application of the envelope theorem and relates the conditional mean to the gradient of a convex function. Gradients of convex functions possess a shape restriction called cyclic monotonicity (see Appendix S.7 of the Supplemental Material), which implies

$$\big(\mathbb{E}[Y \mid X = \tilde{x}] - \mathbb{E}[Y \mid X = x]\big)'\big(\vec{u}(\tilde{x}) - \vec{u}(x)\big) \geq 0. \tag{11}$$

Shi, Shum, and Song (2018) showed that when we assume regressors have the same dimension for each good and $u_k(x_k) = \beta' x_k$, then (11) may be used to point identify β up

[35]As mentioned in the Introduction, this connection is known for discrete choice since market shares and conditional choice probabilities are closely related.

[36]Examples include the seminal work of McFadden (1973), maximum score (Manski (1975)), the special regressor approach (Lewbel (2000)), and random coefficients models (Boyd and Mellman (1980), Cardell and Dunbar (1980)).

[37]We assume utility indices are differentiable with respect to at least one regressor, which is not assumed in Matzkin (1993). Therefore, our results are not strictly more general.

[38]When utility indices are known, this result may be deduced from the Williams–Daly–Zachary theorem (McFadden (1981)).

to scale under certain conditions. Our results allow utility indices to be nonlinear and also allow them to differ across goods. However, we do not study the problem of panel (or grouped) data as in Shi, Shum, and Song (2018), and so our nonparametric results do not nest theirs.

6.2. *Bundles*

The literature allowing multiple purchases is less well-developed than the literature allowing only a single purchase.[39] One technical challenge is that it is not obvious how to use techniques that exploit smoothness when the choices are discrete. The results in the present paper show that models in the spirit of Gentzkow (2007) are amenable to smoothness techniques by working with conditional expectations.

We now return to a simplified version of the model of Gentzkow (2007) (Example 2), in which demand satisfies

$$Y \in \operatorname*{argmax}_{y\in\{0,1\}^2} \sum_{k=1}^{2} y_k u_k(X_k) + (y_1\varepsilon_{1,0} + y_2\varepsilon_{0,1} + y_1 y_2 \varepsilon_{1,1}). \tag{12}$$

While this model can in principle be studied as a classical discrete choice problem if the distribution of Y given X is observed, we note that identification of this model does not immediately follow from existing identification results for discrete choice models. The reason is that there is not a shifter specific to each of the possible quantities $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$.

Fox and Lazzati (2017) formally studied identification in this model. They provided sufficient conditions for identification of the distribution of unobservable heterogeneity once utility indices are identified.[40] Combining our results with theirs, one can identify the distribution of unobservable heterogeneity when utility indices are not known in advance.

In order to identify the distribution of unobservable heterogeneity, Fox and Lazzati (2017) imposed that either $\varepsilon_{1,1} \geq 0$ almost surely or $\varepsilon_{1,1} \leq 0$ almost surely. This means goods must either be complements for everyone or substitutes for everyone. When the researcher is interested in only the average indirect utility function, one may use (8) to identify differences in average indirect utility without imposing these conditions.[41] However, in order to identify the distribution of indirect utilities, it may be necessary to identify the distribution of unobservable heterogeneity, which is not covered by our results.

This model allows us to cleanly illustrate the importance of complementarity/substitutability, which is determined by the sign of the mixed partial derivatives of the average indirect utility function. To that end, assume that $\varepsilon_{1,1}$ is 0 almost surely, which implies that the goods are nowhere complements/substitutes. Given independence between ε and X, we have

$$Y_1 = 1\{u_1(X_1) + \varepsilon_{1,0} \geq 0\}$$

[39]See Berry et al. (2014) for a review of this literature.

[40]Fox and Lazzati (2017) assumed utility indices are quasilinear with respect to a special regressor, that is, $u_k(X_k) = X_{k,1} + \tilde{u}_k(X_{k,-1})$, where $X_k = (X_{k,1}, X'_{k,-1})$. They fixed $X_{k,-1}$ for each k and varied $X_{k,1}$ to identify the distribution of unobservable heterogeneity.

[41]To use (8), we require sufficiently rich aggregate complementarity/substitutability. This form of aggregate complementarity is reflected in the cross-partial derivatives of V. Goods can switch from being aggregate complements ($\partial_{1,2}V > 0$) to substitutes ($\partial_{1,2}V < 0$), so long as the denominator of (8) is nonzero except on a measure zero set.

almost surely, provided $\varepsilon_{1,0}$ has a density. We obtain

$$\mathbb{E}[Y_1 \mid X = x] = F_{1,0}\big(u_1(x_1)\big),$$

where $F_{1,0}$ denotes the cumulative distribution function of $-\varepsilon_{1,0}$. This is equivalent to a binary choice problem *without* variation in the regressors for both alternatives. It is well-known and easy to see that without additional restrictions, u_1 is at best identified up to a monotonic transformation (Matzkin (1992)). This example highlights that complementarity/substitutability is essential to identify utility indices up to location and scale unless additional assumptions are imposed.[42]

We note that this model immediately extends to more than two goods, regardless of whether the quantities are discrete or continuous. This is easy to see from the general structure of PUM.

6.3. *Decisions Under Uncertainty*

A number of econometric models have been proposed to measure risk preferences. See Barseghyan, Molinari, O'Donoghue, and Teitelbaum (2018) for a review of this literature with field data. An expected utility model considered in Agarwal and Somaini (2018) fits into the PUM framework. Let

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

denote the utility obtained in state k. Assume each individual chooses a lottery Y to maximize expected utility,

$$Y \in \operatorname*{argmax}_{y \in B} \sum_{k=1}^{K} y_k u_k(X_k) + \sum_{k=1}^{K} y_k \varepsilon_k. \tag{13}$$

In the utility function, y_k denotes the probability of state k occurring. In the setting of Agarwal and Somaini (2018), k is a school and B is a finite set of lotteries describing the probabilities of being assigned to each school. For concreteness, one can think of X_k as distance to school k, which implicitly depends on the individual–school pair. Individual heterogeneity in preferences for schools is denoted $\varepsilon = (\varepsilon_1, \dots, \varepsilon_K)$. Agarwal and Somaini (2018) showed that if utility indices and the budget B are known, then the distribution of unobservable heterogeneity is identified from the conditional distribution of Y given X. Our results show how to identify utility indices, thus augmenting the results of Agarwal and Somaini (2018). In particular, one may combine our results with theirs to identify the distribution of unobservable heterogeneity without assuming utility indices are known in advance.

It is not necessary to identify the conditional distribution of Y given X to identify utility indices. In Agarwal and Somaini (2018), the researcher does not observe the choice Y directly, only a report of rankings over schools. If the link from reports to lotteries is unspecified (so that Y is not directly observable), $\mathbb{E}[Y \mid X = x]$ is still identified since it is a vector of conditional assignment probabilities.[43] We reiterate that while the conditional

[42]Note that when $\varepsilon_{1,1}$ is 0 almost surely, it may not be possible to provide an ordinal welfare comparison between vectors x and $\tilde{x}$ if they change with respect to regressors for both goods.

[43]Agarwal and Somaini (2018) assumed equilibrium play and estimated the mapping from reports to lotteries. This allowed them to equate choice of report with choice of a lottery. Thus, with their assumptions, the conditional distribution of Y given X is identified. Note that identification of the conditional mean does not require observing reported rankings over schools, only outcomes.

mean is sufficient to identify utility indices, the conditional distribution of Y given X is required to identify the distribution of ε using the results of Agarwal and Somaini (2018).

6.4. *Stochastic Choice*

In discrete choice models, a growing literature studies choice that is stochastic at the *individual* level in ways that are not compatible with the random utility model (Block and Marschak (1960)). We studied the following stochastic choice PUM in Allen and Rehbeck (2019):

$$Y \in \underset{y \in \Delta^{K-1}}{\operatorname{argmax}} \sum_{k=1}^{K} y_k u_k(X_k) + D(y, \varepsilon), \tag{14}$$

where Y is the choice of lottery for a specific individual ε given regressors X, and Δ^{K-1} is the probability simplex in $\mathbb{R}^K$. This model strictly nests ARUM, but allows an individual to have a strict preference for randomization rather than to choose a vertex of the probability simplex.[44] Such strict preference is ruled out by ARUM, which takes the disturbance function to be $D(y, \varepsilon) = \sum_{k=1}^{K} y_k \varepsilon_k$. The present paper is the first to study identification of this general model.

In practice, we may not observe the choice of lottery Y, only the realizations of the lottery. These realizations are the discrete outcomes, while the underlying choice object is the lottery Y. Observing only realizations is sufficient for our identification results, since we only need the vector of conditional choice probabilities, $\mathbb{E}[Y \mid X = x]$.

The interpretation of stochastic choice as deliberate choice of lottery is due to Machina (1985) and has received renewed theoretical and empirical interest.[45] The representation (14) can capture forms of costly optimization, preference for variety, and ambiguity aversion arising from uncertainty over the true utility of a good (Mattsson and Weibull (2002), 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 (14) is a form of complementarity. For example, u_k can increase due to a change in x_k, and the conditional probability of choosing some *other* alternative can increase. This behavior may be natural in a model of mistakes or preference for variety. This form of complementarity is formally ruled out in ARUM.

6.5. *Moral Hazard*

We may also write a moral hazard problem as a PUM. Suppose that an individual can exert costly effort, e, to affect the distribution of possible outcomes, $\vec{p}(e)$. In state k, the individual attains utility

$$v_k = u_k(X_k) - c_k(e, \varepsilon),$$

where $c_k(e, \varepsilon)$ represents the cost of effort level e. If the individual maximizes expected utility, then the choice e^* satisfies

$$e^* \in \underset{e \in S}{\operatorname{argmax}} \sum_{k=1}^{K} p_k(e) u_k(X_k) - \sum_{k=1}^{K} p_k(e) c_k(e, \varepsilon), \tag{15}$$

[44] A vertex may be interpreted as a deterministic choice.

[45] Recent work includes Cerreia-Vioglio, Dillenberger, Ortoleva, and Riella (forthcoming) and Agranov and Ortoleva (2017).

where S is the budget for effort and $p_k(e)$ is the probability of state k occurring if effort is e. This problem is equivalent to choosing a lottery Y such that

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

where $B = \{y \in \Delta^{K-1} \mid y = \vec{p}(e) \text{ for some } e \in S\}$, and $\tilde{c}(y, \varepsilon) = \min_{e \in S | y = \vec{p}(e)} \sum_{k=1}^{K} p_k(e) \times c_k(e, \varepsilon)$.[46] Even if e^* and Y are not observed, $\mathbb{E}[Y \mid X = x]$ is identified from conditional outcome probabilities. Thus, we can identify the von Neumann–Morgenstern indices $\vec{u}$. In particular, one can check whether u_k depends on the state. For example, if X_k includes a payoff to the agent in state k, one can determine if individuals prefer to be paid more in certain states.

6.6. *Matching With Several Markets*

There has been considerable recent interest in econometric models of matching, including the seminal work of Choo and Siow (2006). For an overview of this literature, see Chiappori and Salanié (2016). Recently, Fox, Yang, and Hsu (2018) considered a one-to-one matching market for firms that may be written as a PUM. They exploited variation across multiple markets. We describe how our results may be used to generalize the semiparametric identification results of Fox, Yang, and Hsu (2018) to fully nonparametric identification.

Assume that if firms k and ℓ match, then they attain profit (or surplus)

$$v_{k,\ell} = u_{k,\ell}(X_{k,\ell}) + \varepsilon_{k,\ell},$$

where $X_{k,\ell}$ is a vector that comprises match-specific observables and $\varepsilon_{k,\ell}$ is a match-specific unobservable. Suppose there are K upstream and K downstream firms, where the ordered indices (k, ℓ) determine which is upstream (k) and which is downstream (ℓ). Let $\mathcal{M}$ be the set of symmetric $K \times K$ matrices with entries either 0 or 1 and such that the sum of each row or column cannot exceed 1. For $M \in \mathcal{M}$, $M_{k,\ell} = 1$ is interpreted as a match between firm k and ℓ. Under the assumptions outlined in Fox, Yang, and Hsu (2018), the observed match matrix Y maximizes total surplus. Specifically,

$$Y \in \operatorname*{argmax}_{M \in \mathcal{M}} \sum_{k,\ell} M_{k,\ell} u_{k,\ell}(X_{k,\ell}) + \sum_{k,\ell} M_{k,\ell} \varepsilon_{k,\ell}.$$

This model is thus a PUM, and so identification of utility indices may be established given $\mathbb{E}[Y \mid X = x]$, where X is a vector stacking each $X_{k,\ell}$. Fox, Yang, and Hsu (2018) provided identification results for features of the distribution of ε, including certain linear combinations termed unobservable complementarities. Their results apply when $u_{k,\ell}$ is either known in advance or is linear and $X_{k,\ell}$ is scalar, but the distribution of ε is unrestricted. Thus, we refer to their results as semiparametric identification results. Our results instead apply to nonlinear utility indices, and thus may be combined with the identification results of Fox, Yang, and Hsu (2018) to nonparametrically identify the distribution of unobservable complementarities.

[46] This minimum exists because e^* is a maximizer.

Note that an important assumption in our general analysis is that a good is the same when regressors change. In this setting, a good k, ℓ is interpreted as a match. We require that these subscripts be meaningful to apply our results. As emphasized in Fox, Yang, and Hsu (2018), it is important that the pair k, ℓ interact in multiple markets and that the distribution of $\varepsilon_{k,\ell}$ does not vary across these markets.

7. DISCUSSION

This paper provides nonparametric identification results for a class of latent utility models with additively separable unobservable heterogeneity. We show that these models contain identifying information on utility indices and average indirect utility by exploiting asymmetries in cross-partial derivatives of conditional means.

When covariates and unobservables are independent, a core feature of these models is a version of Roy's identity,

$$\mathbb{E}[Y \mid X = x] = \nabla V(\vec{u}(x)),$$

where V is a convex function. An implication of our analysis is that it is possible to estimate average indirect utility (V) and utility indices ($\vec{u}$) using mean demands, *without* observing or modeling whether goods are purchased together.

Providing numerically feasible estimation procedures motivated by Roy's identity that guarantee convexity of $\hat{V}$ is an interesting topic for future work.[47] We note that a number of estimators motivated via Roy's identity or related duality results have been proposed in related problems, including the firm problem (e.g., Diewert (1971)), consumer problem (e.g., Deaton and Muellbauer (1980)), and discrete choice (e.g., McFadden (1978)).

Roy's identity has been used recently in the semiparametric estimation approach of Shi, Shum, and Song (2018). While they used a version of Roy's identity for discrete choice (the Williams–Daly–Zachary theorem), their core arguments apply to all PUMs when we restrict u_k to be linear and to have the same coefficient across goods k.

APPENDIX A: PROOFS FOR SECTION 2

As a setup for the proof of Theorem 1, we introduce some additional notation. Let $\mathcal{Y}$ denote the set of ε-measurable functions $\tilde{Y}$ to $\mathbb{R}^K$ that satisfy $\tilde{Y}(\varepsilon) \in B$ for each ε. Recall Y is (X, ε)-measurable, and so we may define $Y_x(\cdot) = Y(x, \cdot) \in \mathcal{Y}$. Write

$$U(y; x, \varepsilon) = \sum_{k=1}^{K} y_k u_k(x_k) + D(y, \varepsilon),$$

which we have assumed to have a maximizer. We use $a'b$ to denote the inner product.

As an outline of the proof of Theorem 1, we first discuss several related optimization problems. These problems hold fixed the value of x. First, consider optimization over all ε-measurable functions that map to B:

$$\sup_{\tilde{Y} \in \mathcal{Y}} \mathbb{E}\big[U(\tilde{Y}; x, \varepsilon)\big].$$

[47] We note that our identification technique for utility indices $\vec{u}$ uses a symmetry condition that requires that V be twice continuously differentiable. The additional restriction of convexity of V allows us to interpret V and $\vec{u}$ in terms of an optimizing model.

Second, consider the expected utility from the pointwise optimization problem:

$$\mathbb{E}\Big[\sup_{y\in B} U(y; x, \varepsilon)\Big].$$

In general, one may not expect these two problems to attain the same value. One inequality is clear:

$$\mathbb{E}\Big[\sup_{y\in B} U(y; x, \varepsilon)\Big] \geq \sup_{\tilde{Y}\in\mathcal{Y}} \mathbb{E}\big[U(\tilde{Y}; x, \varepsilon)\big].$$

We formally show this is an equality by relying on the assumption that the pointwise solution Y is (X, ε)-measurable; hence $Y_x(\varepsilon) := Y(x, \varepsilon)$ is ε-measurable. Recall that Appendix S.2 of the Supplemental Material provides sufficient conditions for existence of a measurable selector.

The final problem to consider is a two-step optimization problem where one first picks a vector y (thought of as a mean), and then optimizes over measurable functions in $\mathcal{Y}$ subject to this mean constraint:

$$\sup_{y\in\mathbb{R}} \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=y} \mathbb{E}\big[U(\tilde{Y}; x, \varepsilon)\big]. \tag{16}$$

Clearly, we have the bound

$$\sup_{\tilde{Y}\in\mathcal{Y}} \mathbb{E}\big[U(\tilde{Y}; x, \varepsilon)\big] \geq \sup_{y\in\mathbb{R}} \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=y} \mathbb{E}\big[U(\tilde{Y}; x, \varepsilon)\big].$$

Since the utility maximizing choices have finite mean $\mathbb{E}[Y(x, \varepsilon)]$, searching over general ε-measurable functions (the left-hand side) yields the same value to the optimization problem constrained to search over ε-measurable functions with finite mean (the right-hand side). Thus, the above inequality holds with equality.

Relating these optimization problems constitutes a technical hurdle, and our formal argument relates all three by means of a sequence of inequalities. The proof establishes that the value functions for these three optimization problems are equivalent, and moreover, that $\mathbb{E}[Y(x, \varepsilon)]$ is a mean vector that maximizes the first supremum in (16). Another key step is simplifying the optimization problem in (16). Additive separability in the utility function U is key to allowing us to break the expected utility at the maximizer, $\mathbb{E}[U(Y(x, \varepsilon); x, \varepsilon)]$, into two parts:

$$\begin{aligned}
\mathbb{E}\big[U\big(Y(x, \varepsilon); x, \varepsilon\big)\big] &= \mathbb{E}\big[Y(x, \varepsilon)'\vec{u}(x)\big] + \mathbb{E}\big[D\big(Y(x, \varepsilon), \varepsilon\big)\big] \\
&= \mathbb{E}\big[Y(x, \varepsilon)\big]'\vec{u}(x) + \mathbb{E}\big[D\big(Y(x, \varepsilon), \varepsilon\big)\big] \\
&= \mathbb{E}\big[Y(x, \varepsilon)\big]'\vec{u}(x) + \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=\mathbb{E}[\tilde{Y}(x,\varepsilon)]} \mathbb{E}\big[D\big(Y(x, \varepsilon), \varepsilon\big)\big].
\end{aligned}$$

The second equality exploits that quantities and utility indices interact in an affine manner, and so we can pull utility indices outside of an expectation. The third equality uses the fact that the expectations are computed with the same marginal distribution of ε, even as x changes. While in the main text we provide intuition why the second and third lines are equal via a proof by contradiction argument, we now provide our formal proof by a cycle of inequalities.

PROOF OF THEOREM 1: Recall that x is fixed so the following expectations are taken over the marginal distribution of ε. We have

$$\begin{aligned}
&\mathbb{E}\Big[\max_{y\in B} U(y;x,\varepsilon)\Big] && (17)\\
&\quad = \mathbb{E}\big[Y_x'\vec{u}(x)\big] + \mathbb{E}\big[D(Y_x,\varepsilon)\big] && (18)\\
&\quad = \mathbb{E}[Y_x]'\vec{u}(x) + \mathbb{E}\big[D(Y_x,\varepsilon)\big] && (19)\\
&\quad \le \mathbb{E}[Y_x]'\vec{u}(x) + \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=\mathbb{E}[Y_x]} \mathbb{E}\big[D(\tilde{Y},\varepsilon)\big] && (20)\\
&\quad \le \sup_{y\in\overline{B}}\Big\{y'\vec{u}(x) + \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=y} \mathbb{E}\big[D(\tilde{Y},\varepsilon)\big]\Big\} && (21)\\
&\quad = \sup_{y\in\overline{B}}\ \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=y} \mathbb{E}\big[y'\vec{u}(x) + D(\tilde{Y},\varepsilon)\big] && (22)\\
&\quad \le \sup_{\tilde{Y}\in\mathcal{Y}} \mathbb{E}\big[\tilde{Y}'\vec{u}(x) + D(\tilde{Y},\varepsilon)\big] && (23)\\
&\quad \le \mathbb{E}\Big[\max_{y\in B} U(y;x,\varepsilon)\Big]. && (24)
\end{aligned}$$

The equalities in lines (17)–(19) are straightforward. The inequality on line (20) follows since $Y_x \in \mathcal{Y}$ and we assume $\mathbb{E}[Y(x,\varepsilon)]$ exists. The inequality on line (21) uses the fact that $Y_x(\varepsilon) \in B$ for each ε and $\mathbb{E}[Y(x,\varepsilon)]$ is finite. Thus, $\mathbb{E}[Y_x] \in \overline{B}$, where $\overline{B}$ denotes the convex hull of B. The equality of line (22) follows since $y'\vec{u}(x)$ is a constant relative to $\tilde{Y}$ and ε. Line (23) uses the fact that mappings in $\mathcal{Y}$ with finite expectations are in particular a subset of $\mathcal{Y}$. Finally, line (24) follows since $\tilde{Y}'\vec{u}(x) + D(\tilde{Y},\varepsilon) \le \max_{y\in B} U(y;x,\varepsilon)$ for all $\tilde{Y} \in \mathcal{Y}$.

We conclude that all lines above are equal. In particular, $\mathbb{E}[Y(x,\varepsilon)] := \mathbb{E}[Y_x]$ maximizes the expression in line (21):

$$\begin{aligned}
&\mathbb{E}\big[Y(x,\varepsilon)\big]'\vec{u}(x) + \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=\mathbb{E}[Y(x,\varepsilon)]} \mathbb{E}\big[D(\tilde{Y},\varepsilon)\big]\\
&\quad = \max_{y\in\overline{B}}\Big\{y'\vec{u}(x) + \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=y} \mathbb{E}\big[D(\tilde{Y},\varepsilon)\big]\Big\}.
\end{aligned}$$

Thus, $\mathbb{E}[Y(x,\varepsilon)]$ is a maximizer, establishing the first part of the theorem. The second part follows from the additional fact that

$$\begin{aligned}
&\max_{y\in\overline{B}}\Big\{y'\vec{u}(x) + \sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=y} \mathbb{E}\big[D(\tilde{Y},\varepsilon)\big]\Big\}\\
&\quad = \mathbb{E}\Big[\max_{y\in B} U(y;x,\varepsilon)\Big].
\end{aligned}$$

Q.E.D.

Note that the theorem requires that Y be (X,ε)-measurable, and that $\mathbb{E}[Y(x,\varepsilon)]$ exists, where the expectation is taken over the marginal distribution of ε. This implicitly places a domain restriction on $Y(x,\cdot)$, since the expectation $\mathbb{E}[Y(x,\varepsilon)]$ must be defined.

We state this theorem treating x as a parameter and holding fixed the distribution of ε. Thus, when x changes, the same function $\overline{D}$ characterizes the optimizer. The form of the utility function that $\mathbb{E}[Y(x,\varepsilon)]$ maximizes thus does not depend on x except through the values of the utility indices $\vec{u}(x)$. We note that for our subsequent identification results, we treat X as a random variable instead of a parameter. Thus, in Appendix A.1, we add the assumption that X and ε are independent; independence allows us to identify $\mathbb{E}[Y(x,\varepsilon)]$ from $\mathbb{E}[Y \mid X = x]$ as described in Appendix A.1.

In the main text, we claim that this theorem covers aggregation across firms as a special case. Aggregation for price taking, profit maximizing firms arises by considering $u_k(x_k)$ as a price, $Y(x,\varepsilon)$ as a netput vector, $B = \mathbb{R}^K$ as unrestricted, and $D(y,\varepsilon)$ as a function that encodes feasible production plans. Specifically, $D(y,\varepsilon)$ is 0 if the plan y is feasible for firm type ε, and $-\infty$ otherwise. Textbook treatments (e.g., Acemoglu (2009), Kreps (2012)) prove firm aggregation by a proof by contradiction argument. To our knowledge, firm aggregation has only been proven for countably many firms (Acemoglu (2009)); our setup covers uncountably many firms because we allow general probability measures over ε.

A.1. *Linking Theorem 1 and Assumption 2*

We now link the conclusion of Theorem 1 with Assumption 2, which is our core maintained assumption for identification of utility indices in Section 3. We note that we do not need the full strength of Assumption 2 for all of the results. In particular, we relax the assumption of a unique maximizer (part (i)) for identification results on average indirect utility (Section 4) and counterfactual bounds (Section 5).

The primary goal is to link $\mathbb{E}[Y(x,\varepsilon)]$ to the distribution of Y given X. This appendix provides sufficient conditions for $\mathbb{E}[Y(x,\varepsilon)]$ to be continuous in x. Maintaining the assumption of independence between X and ε, continuity is a sufficient condition for a continuous version of the conditional mean of Y given X to exist. In turn, the unique continuous version of the conditional mean satisfies $\mathbb{E}[Y \mid X = x] = \mathbb{E}[Y(x,\varepsilon)]$ over the support of X.

We first introduce some new notation. For the latent utility model, define the indirect utility function

$$\tilde{V}(\vec{v},\varepsilon) = \sup_{y\in B} \sum_{k=1}^{K} y_k v_k + D(y,\varepsilon).$$

The following lemma provides sufficient conditions on $D(y,\varepsilon)$ for $\mathbb{E}[Y(x,\varepsilon)]$ to be the unique maximizer of the representative agent problem. We let $\nabla_{\vec{v}}\tilde{V}(\vec{v},\varepsilon)$ denote the gradient with respect to $\vec{v}$, provided it exists.

LEMMA A.1: *Let the assumptions of Theorem 1 hold and assume B is convex. Suppose that for almost every ε, $D(\cdot,\varepsilon)$ is upper semi-continuous and concave. In addition, suppose that for every $x \in \operatorname{supp}(X)$, $Y(x,\varepsilon)$ is the unique maximizer of the latent utility model for almost every ε. Suppose that for every $\vec{v}$ such that $\mathbb{E}[\nabla_{\vec{v}}\tilde{V}(\vec{v},\varepsilon)]$ exists, we have $\mathbb{E}[\nabla_{\vec{v}}\tilde{V}(\vec{v},\varepsilon)] = \nabla_{\vec{v}}\mathbb{E}[\tilde{V}(\vec{v},\varepsilon)]$. It follows that $\mathbb{E}[Y(x,\varepsilon)]$ is the unique maximizer of the representative agent problem, that is,*

$$\mathbb{E}\big[Y(x,\varepsilon)\big] = \operatorname*{argmax}_{y\in\overline{B}} \sum_{k=1}^{K} y_k u_k(x_k) + \overline{D}(y).$$

PROOF: By Lemmas S.1.1 and S.1.2, because $Y(x, \varepsilon)$ is the unique maximizer for almost every ε, $\tilde{V}$ is differentiable in $\vec{v}$ for almost every ε. Thus, we have

$$Y(x, \varepsilon) = \nabla_{\vec{v}} \tilde{V}(\vec{u}(x), \varepsilon)$$

for almost every ε. By the assumptions of Theorem 1, $\mathbb{E}[Y(x, \varepsilon)]$ exists and thus

$$\mathbb{E}[Y(x, \varepsilon)] = \mathbb{E}[\nabla_{\vec{v}} \tilde{V}(\vec{u}(x), \varepsilon)]$$

exists. By assumption, we can interchange differentiation and integration:

$$\mathbb{E}[\nabla_{\vec{v}} \tilde{V}(\vec{v}, \varepsilon)] = \nabla_{\vec{v}} \mathbb{E}[\tilde{V}(\vec{v}, \varepsilon)].$$

Recall that by Theorem 1, V is the average indirect utility function, $V(\vec{u}(x)) = \mathbb{E}[\tilde{V}(\vec{u}(x), \varepsilon)]$. Thus, V is differentiable at $\vec{u}(x)$. Again applying Lemmas S.1.1 and S.1.2, which show that differentiability of the value function is equivalent to having a unique maximizer, we obtain that $\mathbb{E}[Y(x, \varepsilon)]$ is the unique maximizer of the representative agent's problem. *Q.E.D.*

One (overly strong) sufficient condition for $Y(x, \varepsilon)$ to be a unique maximizer is that $D(y, \varepsilon)$ is finite over $y \in B$, strictly concave in y for all ε when B is nonempty, compact, and convex.

We now add the assumption of continuity of $\vec{u}$.

LEMMA A.2: *Let the assumptions of Lemma A.1 hold. If $\vec{u}$ is continuous, then $\mathbb{E}[Y(x, \varepsilon)]$ is continuous in x over* $\operatorname{supp}(X)$.

PROOF: From Lemma A.1, $\mathbb{E}[Y(x, \varepsilon)]$ is the unique maximizer. Thus, $\mathbb{E}[Y(x, \varepsilon)] = \nabla V(\vec{u}(x))$. (See Lemma 1 with $\mathbb{E}[Y(x, \varepsilon)]$ in place of $\mathbb{E}[Y \mid X = x]$.) The mapping ∇V is continuous over the set $\vec{u}(\operatorname{supp}(X))$ by Rockafellar (1970, Theorem 25.5). Since we have assumed $\vec{u}$ is continuous, $\mathbb{E}[Y(x, \varepsilon)]$ is the composition of two continuous functions and hence continuous in x. *Q.E.D.*

We now use independence between X and ε to link a continuous version of the conditional expectation $\mathbb{E}[Y \mid X = x]$ and $\mathbb{E}[Y(x, \varepsilon)]$.

LEMMA A.3: *Let the assumptions of Lemma A.2 hold. It follows that a continuous version of the conditional expectation, denoted $\mathbb{E}[Y \mid X = x]$, exists, and for each $x \in \operatorname{supp}(X)$,*

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

PROOF: Let $x^0 \in \operatorname{supp}(X)$. For every $\delta > 0$, we note $P(X \in S(\delta, x^0)) > 0$ because $x^0 \in \operatorname{supp}(X)$. Using independence between X and ε and the fact that $\mathbb{E}[Y(x, \varepsilon)]$ exists over $x \in \operatorname{supp}(X)$, we obtain by Fubini's theorem that, for sufficiently small δ,

$$\mathbb{E}[Y \mid X \in S(\delta, x^0)] = \frac{1}{P(X \in S(\delta, x^0))} \int_{x \in S(\delta, x^0)} \mathbb{E}[Y(x, \varepsilon)] \mu(dx),$$

where μ is the probability measure associated with the marginal distribution of X. By Lemma A.2, $\mathbb{E}[Y(x,\varepsilon)]$ is continuous in x, and so the right-hand side exists for small δ. In addition,

$$\begin{aligned}
\limsup_{\delta\to 0}\mathbb{E}\big[Y \mid X\in S(\delta,x^0)\big] &= \limsup_{\delta\to 0}\frac{1}{P(X\in S(\delta,x^0))}\int_{x\in S(\delta,x^0)}\mathbb{E}\big[Y(x,\varepsilon)\big]\mu(dx)\\
&\le \limsup_{\delta\to 0}\frac{1}{P(X\in S(\delta,x^0))}\int_{x\in S(\delta,x^0)}\sup_{x\in S(\delta,x^0)}\mathbb{E}\big[Y(x,\varepsilon)\big]\mu(dx)\\
&= \limsup_{\delta\to 0}\sup_{x\in S(\delta,x^0)}\mathbb{E}\big[Y(x,\varepsilon)\big]\\
&= \mathbb{E}\big[Y(x^0,\varepsilon)\big].
\end{aligned}$$

The final equality follows because $\mathbb{E}[Y(x,\varepsilon)]$ is continuous in x. A similar argument yields $\liminf_{\delta\to 0}\mathbb{E}[Y \mid X\in S(\delta,x^0)]\ge \mathbb{E}[Y(x^0,\varepsilon)]$, and so we obtain $\mathbb{E}[Y\mid X=x^0]=\mathbb{E}[Y(x^0,\varepsilon)]$. As a byproduct, this establishes that $\mathbb{E}[Y\mid X=x]$ is well-defined for every $x\in \operatorname{supp}(X)$. *Q.E.D.*

Lemma A.3 builds on our previous lemmas in order to provide sufficient conditions on primitives to establish the result. We note that the conclusion of Lemma A.3 holds with these alternative assumptions: the assumptions of Theorem 1 hold, $\mathbb{E}[Y(x,\varepsilon)]$ is the unique maximizer for every $x\in\operatorname{supp}(X)$, and $\vec{u}$ is continuous. Thus, the conditions of Lemma A.1 were only used to ensure $\mathbb{E}[Y(x,\varepsilon)]$ is the unique maximizer.

We now provide a sufficient condition to ensure that the aggregate disturbance function $\bar{D}(y)=\sup_{\tilde{Y}\in\mathcal{Y}:\mathbb{E}[\tilde{Y}]=y}\mathbb{E}[D(\tilde{Y},\varepsilon)]$ is concave.

LEMMA A.4: *Let the assumptions of Theorem 1 hold, suppose B is convex, and that $D(y,\varepsilon)$ is concave in y for almost every ε. It follows that $\bar{D}(y)$ is concave.*

PROOF: Let $\lambda\in[0,1]$ and $y,\tilde{y}\in\bar{B}$. It follows that

$$\begin{aligned}
\bar{D}(\lambda y+(1-\lambda)\tilde{y}) &= \sup_{\substack{\hat{Y}\in\mathcal{Y}\\ \mathbb{E}[\hat{Y}]=\lambda y+(1-\lambda)\tilde{y}}}\mathbb{E}\big[D(\hat{Y},\varepsilon)\big]\\
&\ge \sup_{\substack{\hat{Y},Y,\tilde{Y}\in\mathcal{Y}\\ \hat{Y}=\lambda Y+(1-\lambda)\tilde{Y}\\ \mathbb{E}[Y]=y,\,\mathbb{E}[\tilde{Y}]=\tilde{y}}}\mathbb{E}\big[D(\hat{Y},\varepsilon)\big]\\
&= \sup_{\substack{Y,\tilde{Y}\in\mathcal{Y}\\ \mathbb{E}[Y]=y,\,\mathbb{E}[\tilde{Y}]=\tilde{y}}}\mathbb{E}\big[D(\lambda Y+(1-\lambda)\tilde{Y},\varepsilon)\big]\\
&\ge \sup_{\substack{Y,\tilde{Y}\in\mathcal{Y}\\ \mathbb{E}[Y]=y,\,\mathbb{E}[\tilde{Y}]=\tilde{y}}}\mathbb{E}\big[\lambda D(Y,\varepsilon)+(1-\lambda)D(\tilde{Y},\varepsilon)\big]\\
&= \lambda\bar{D}(y)+(1-\lambda)\bar{D}(\tilde{y}).
\end{aligned}$$

The first equality is by definition. The first inequality holds because the supremum is over a smaller set. The second equality holds by plugging in $\hat{Y}$ since $\mathcal{Y}$ is convex (because

B is convex). The second inequality holds from concavity for almost every ε. The final equality is immediate as it is the supremum of two functions over separable sets. *Q.E.D.*

A.2. *Proofs for Section 2.2*

PROOF OF LEMMA 1: This follows from Lemmas S.1.1 and S.1.2 once we identify f^* with V, $\mathbb{E}[Y \mid X = x]$ with $\tilde{z}$, and $\tilde{w}$ with $\vec{u}(x)$. *Q.E.D.*

PROOF OF LEMMA 2: The function V is convex because it is a convex conjugate. The result then follows from Rockafellar (1970, Theorem 4.5). *Q.E.D.*

PROOF OF LEMMA 3: This follows from Lemmas S.1.1 and S.1.2 once we identify f with $-\overline{D}_{\overline{B}}$, $\mathbb{E}[Y \mid X = x]$ with $\tilde{z}$, and $\tilde{w}$ with $\vec{u}(x)$. *Q.E.D.*

APPENDIX B: PROOFS FOR SECTION 3

B.1. *Proofs for Section 3.1*

PROOF OF PROPOSITION 1: The proof is a direct extension of the arguments in the text. Under the assumptions of Definition 1, there is some $x^* \in \operatorname{supp}(X)$ such that

$$\begin{aligned} \frac{\partial \mathbb{E}[Y_k \mid X = x]}{\partial x_{\ell,p}}\bigg|_{x=x^*} &= \partial_{\ell,k} V(\overline{u})|_{\overline{u}=\vec{u}(x^*)} \frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}\bigg|_{x_\ell = x_\ell^*}, \\ \frac{\partial \mathbb{E}[Y_\ell \mid X = x]}{\partial x_{k,q}}\bigg|_{x=x^*} &= \partial_{k,\ell} V(\overline{u})|_{\overline{u}=\vec{u}(x^*)} \frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k = x_k^*}. \end{aligned} \tag{25}$$

Recall these equalities hold at the specified *points*. Under condition (iv) of Definition 1, we may take the ratio of the equations in (25) to yield the result. *Q.E.D.*

PROOF OF THEOREM 2: This is implied by Corollary S.5.1 so we provide only a brief discussion.

Using Proposition 1, the conditions of the theorem provide identification of

$$\frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}\bigg|_{x_\ell=\tilde{x}_\ell} \bigg/ \frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=\tilde{x}_k} \tag{26}$$

whenever $\ell \neq k$, the denominator is nonzero, $\tilde{x}_\ell \in \operatorname{int}(\operatorname{supp}(X_\ell))$, and $\tilde{x}_k \in \operatorname{int}(\operatorname{supp}(X_k))$. We let $\operatorname{int}(A)$ denote the interior of a set A. Since $\vec{u}(\operatorname{supp}(X))$ contains an open ball, by the mean value theorem we can choose a value x_k^* and some q^* such that the denominator is nonzero and identify (26) whenever $\ell \neq k$; the tuple (k, q^*, x_k^*) thus describes where the scale is set in the statement of the theorem. Since the numerator is identified for each good $\ell \neq k$, characteristic p, and covariate value $\tilde{x}_\ell$ over $\operatorname{int}(\operatorname{supp}(X_\ell))$, we obtain that differences in u_ℓ over $\operatorname{int}(\operatorname{supp}(X_\ell))$ (and hence $\operatorname{supp}(X_\ell)$ by continuity) are identified up to the scale of $\frac{\partial u_k(x_k)}{\partial x_{k,q^*}}|_{x_k=x_k^*}$. Thus, all that remains is to identify ratios such as (26) for the same good when the covariate values differ.

Recall that we have assumed that $\vec{u}(\operatorname{supp}(X))$ contains an open ball and $\operatorname{supp}(X)$ is rectangular. Because of these assumptions, we can always find a "path" between the partial derivatives, as in the discussion leading to (S.1).

In more detail, note that since $\vec{u}(\mathrm{supp}(X))$ contains an open ball, we can find $\tilde{x}_\ell \in \mathrm{supp}(X_\ell)$ and $\tilde{x}_k \in \mathrm{supp}(X_k)$ such that (26) is identified for some nonzero partial derivatives, where $k \neq \ell$. Consider some other $\hat{x} \in \mathrm{supp}(X)$ such that $\hat{x}_\ell = \tilde{x}_\ell$ but $\hat{x}_k \neq \tilde{x}_k$. We thus identify

$$\frac{\partial u_k(x_k)}{\partial x_{k,r}}\bigg|_{x_k=\hat{x}_k} \bigg/ \frac{\partial u_\ell(x_\ell)}{\partial x_{\ell,p}}\bigg|_{x_\ell=\tilde{x}_\ell},$$

where we recall that the denominator is nonzero by construction. We can multiply this with (26) to identify

$$\frac{\partial u_k(x_k)}{\partial x_{k,r}}\bigg|_{x_k=\hat{x}_k} \bigg/ \frac{\partial u_k(x_k)}{\partial x_{k,q}}\bigg|_{x_k=\tilde{x}_k}.$$

We can integrate this as $\hat{x}_k$ varies to identify differences in u_k over $\mathrm{supp}(X_k)$ up to the scale $\frac{\partial u_k(x_k)}{\partial x_{k,q}}|_{x_k=\tilde{x}_k}$. This follows from the fundamental theorem of calculus; recall we have assumed $\mathrm{supp}(X)$ is rectangular. This holds in particular for $\tilde{x}_k = x_k^*$ and $q = q^*$.

We have thus identified differences in all utility indices $\vec{u}$ over $\mathrm{supp}(X)$ up the scale $\frac{\partial u_k(x_k)}{\partial x_{k,q^*}}|_{x_k=x_k^*}$. The proof of Corollary S.5.1 shows the sign of $\frac{\partial u_k(x_k)}{\partial x_{k,q^*}}|_{x_k=x_k^*}$ is identified. Using the scale and location normalization, $\vec{u}$ is identified over $\mathrm{supp}(X)$. *Q.E.D.*

B.2. *Proofs for Section 3.2*

PROOF OF THEOREM 3: By Assumption 3(i), there exists $(\tilde{z}, \tilde{w}) \in \mathrm{supp}(Z, W)$ such that

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

Since $\overline{D}_{\overline{B}}$ is differentiable at $\mathbb{E}[Y \mid Z = z, W = w]$, Lemma 3 establishes that

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

Assumption 3(ii) completes the proof since $\vec{u}(\tilde{z}, \tilde{w})$ is identified. *Q.E.D.*

In the main text, we claim that differentiability of $\overline{D}_{\overline{B}}$ is essential for the key injectivity implication in the proof of Theorem 3. We introduce some additional notation to formalize this claim. For $y^* \in \mathbb{R}^K$, define

$$\rho^{-1}(y^*) = \left\{ \vec{v} \in \mathbb{R}^K \;\middle|\; y^* \in \operatorname*{argmax}_y \sum_{k=1}^K y_k v_k + \overline{D}_{\overline{B}}(y) \right\}.$$

First, we note that if $\overline{B}$ is convex and $\overline{D}$ is concave, then $\overline{D}_{\overline{B}}$ is concave.

LEMMA B.1: *Assume that $\overline{D}_{\overline{B}}$ is a concave function that is finite at some point and $y^* \in \mathbb{R}^K$. The following are equivalent*:

1. *$\rho^{-1}(y^*)$ is a singleton.*
2. *$\overline{D}_{\overline{B}}$ is differentiable at y^*.*

PROOF: We can prove the result using Lemmas S.1.1 and S.1.2 once we identify $-\overline{D}_{\overline{B}}$ with f, y^* with $\tilde{z}$, and $\vec{v}$ with $\tilde{w}$.

By Lemma S.1.1,

$$\rho^{-1}(y^*) = \partial\big(-\overline{D}_{\overline{B}}(y^*)\big).$$

From Lemma S.1.2, we conclude that $\rho^{-1}(y^*)$ is a singleton if and only if $(-\overline{D}_{\overline{B}})$ is differentiable at y^*. *Q.E.D.*

To interpret this result in our setting, consider $\vec{u}(x)$ in place of $\vec{v}$ and $\mathbb{E}[Y \mid X = x]$ in place of y^*. We emphasize that to nontrivially apply this result, $\overline{B}$ needs to have a nonempty interior.

APPENDIX C: PROOFS FOR SECTION 4

PROOF OF THEOREM 4: By Lemma S.1.1,

$$\mathbb{E}[Y \mid X = x] \in \partial V\big(\vec{u}(x)\big), \tag{27}$$

where ∂V denotes the subgradient of V. Note that this holds even when $\mathbb{E}[Y \mid X = x]$ is not the unique maximizer.

We now convert to a single variable problem and invoke Rockafellar (1970, Corollary 24.2.1). For $t \in [0, 1]$, let

$$h(t) = V\big(t\vec{u}(x^1) + (1-t)\vec{u}(x^0)\big).$$

The function V is convex and $s(t) = t\vec{u}(x^1) + (1-t)\vec{u}(x^0)$ is affine so h is convex. Because h is convex, it is directionally differentiable. Let its left derivative be denoted $h'_-(t)$ and let $h'_+(t)$ denote the right derivative of h. The directional derivative of V at y in direction z is denoted $V'(y; z)$; see Rockafellar (1970) for the formal definition. From (27), we have

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

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

$$\begin{aligned}
h'_-(t) &= V'\big(t\vec{u}(x^1) + (1-t)\vec{u}(x^0); -\big(\vec{u}(x^1) - \vec{u}(x^0)\big)\big) \\
&\le \mathbb{E}\big[Y \mid X = x(t)\big] \cdot \big(\vec{u}(x^1) - \vec{u}(x^0)\big), \\
h'_+(t) &= V'\big(t\vec{u}(x^1) + (1-t)\vec{u}(x^0); \vec{u}(x^1) - \vec{u}(x^0)\big) \\
&\ge \mathbb{E}\big[Y \mid X = x(t)\big] \cdot \big(\vec{u}(x^1) - \vec{u}(x^0)\big).
\end{aligned}$$

From Rockafellar (1970, Corollary 24.2.1), we obtain

$$V\big(\vec{u}(x^1)\big) - V\big(\vec{u}(x^0)\big) = \int_0^1 \mathbb{E}\big[Y \mid X = x(t)\big] \cdot \big(\vec{u}(x^1) - \vec{u}(x^0)\big)\, dt.$$ [48] *Q.E.D.*

[48] 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 $\mathbb{E}[Y \mid X = x(t)] \cdot (\vec{u}(x^1) - \vec{u}(x^0))$ from 0 to 1 then follows from a sandwiching argument.

APPENDIX D: PROOFS FOR SECTION 5

PROOF OF THEOREM 5: Obviously, $\mathbb{E}[Y(x^0, \varepsilon)] \in \overline{B}$ by use of the a priori knowledge of the budget.

To prove the rest, let $\{\mathbb{E}[Y \mid X = x]\}_{x\in\text{supp}(X)} \cup \{\mathbb{E}[Y(x^0, \varepsilon)]\}$ be the original values of the conditional mean augmented with the counterfactual average structural function. The set just described is consistent with the model if and only if the inequality restrictions of Theorem S.7.1(iv) hold. By rearranging the cyclic monotonicity inequalities, we obtain (10).

For more details, note that the bounds are sharp in the following sense. If $\{\mathbb{E}[Y \mid X = x]\}_{x\in\text{supp}(X)} \cup \{\mathbb{E}[Y(x^0, \varepsilon)]\}$ satisfy the inequalities in the statement of the theorem, then by Theorem S.7.1(iv) there exists a concave, upper semi-continuous function $\tilde{D} : \mathbb{R}^K \to \mathbb{R} \cup \{+\infty\}$ that is finite at some $y \in \overline{B}$, which satisfies

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

for every $x \in \text{supp}(X)$ and similarly for $\mathbb{E}[Y(x^0, \varepsilon)]$. *Q.E.D.*

REFERENCES

ABALUCK, J., AND A. ADAMS (2018): "What Do Consumers Consider Before They Choose? Identification From Asymmetric Demand Responses," Technical report. Available at https://ssrn.com/abstract=2999545. [1029]

AGARWAL, N., AND P. SOMAINI (2018): "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism," *Econometrica*, 86 (2), 391–444. [1021,1022,1024,1040,1041]

AGRANOV, M., AND P. ORTOLEVA (2017): "Stochastic Choice and Preferences for Randomization," *Journal of Political Economy*, 125 (1), 40–68. [1041]

ALLEN, R., AND J. REHBECK (2019): "Supplement to 'Identification With Additively Separable Heterogeneity'," *Econometrica Supplemental Material*, 87, https://doi.org/10.3982/ECTA15867. [1026,1028]

——— (2019): "Revealed Stochastic Choice With Attributes," Technical report. Available at http://dx.doi.org/10.2139/ssrn.2818041. [1023,1041]

ACEMOGLU, D. (2009): *Introduction to Modern Economic Growth*. Princeton, NJ: Princeton University Press. [1046]

ANDERSON, S. P., A. DE PALMA, AND J. F. THISSE (1992): *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press. [1023]

BARSEGHYAN, L., F. MOLINARI, T. O'DONOGHUE, AND J. C. TEITELBAUM (2018): "Estimating Risk Preferences in the Field," *Journal of Economic Literature*, 56 (2), 501–564. [1040]

BERRY, S., A. GANDHI, AND P. HAILE (2013): "Connected Substitutes and Invertibility of Demand," *Econometrica*, 81 (5), 2087–2111. [1031]

BERRY, S., A. KHWAJA, V. KUMAR, A. MUSALEM, K. C. WILBUR, G. ALLENBY, B. ANAND, P. CHINTAGUNTA, W. M. HANEMANN, AND P. JEZIORSKI (2014): "Structural Models of Complementary Choices," *Marketing Letters*, 25 (3), 245–256. [1039]

BERRY, S. T., AND P. A. HAILE (2014): "Identification in Differentiated Products Markets Using Market Level Data," *Econometrica*, 82 (5), 1749–1797. [1022]

——— (2018): "Identification of Nonparametric Simultaneous Equations Models With a Residual Index Structure," *Econometrica*, 86 (1), 289–315. [1025]

BHATTACHARYA, D. (2015): "Nonparametric Welfare Analysis for Discrete Choice," *Econometrica*, 83 (2), 617–649. [1034]

BLOCK, H. D., AND J. MARSCHAK (1960): "Random Orderings and Stochastic Theories of Responses," in *Contributions to probability and statistics*, ed. by I. Olken. Stanford, CA: Stanford Univ. Press, 97–132. [1041]

BLUNDELL, R., M. BROWNING, AND I. CRAWFORD (2008): "Best Nonparametric Bounds on Demand Responses," *Econometrica*, 76 (6), 1227–1262. [1023,1036]

BLUNDELL, R. W., M. BROWNING, AND I. A. CRAWFORD (2003): "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71 (1), 205–240. [1023,1036]

BOYD, J. H., AND R. E. MELLMAN (1980): "The Effect of Fuel Economy Standards on the Us Automotive Market: An Hedonic Demand Analysis," *Transportation Research Part A: General*, 14 (5–6), 367–378. [1038]

CAPLIN, A., AND M. DEAN (2015): "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 105 (7), 2183–2203. [1041]

CARDELL, N. S., AND F. C. DUNBAR (1980): "Measuring the Societal Impacts of Automobile Downsizing," *Transportation Research Part A: General*, 14 (5–6), 423–434. [1038]

CERREIA-VIOGLIO, S., D. DILLENBERGER, P. ORTOLEVA, AND G. RIELLA (forthcoming): "Deliberately Stochastic," American Economic Review. [1041]

CHETTY, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1 (1), 451–488. [1023,1036]

CHIAPPORI, P.-A., AND B. SALANIÉ (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54 (3), 832–861. [1042]

CHIONG, K., Y.-W. HSIEH, AND M. SHUM (2017): "Counterfactual Estimation in Semiparametric Discrete-Choice Models," Technical report. Available at http://dx.doi.org/10.2139/ssrn.2979446. [1023,1037]

CHIONG, K. X., A. GALICHON, AND M. SHUM (2016): "Duality in Dynamic Discrete-Choice Models," *Quantitative Economics*, 7 (1), 83–115. [1029]

CHOO, E., AND A. SIOW (2006): "Who Marries Whom and Why," *Journal of Political Economy*, 114 (1), 175–201. [1042]

DEATON, A., AND J. MUELLBAUER (1980): "An Almost Ideal Demand System," *The American Economic Review*, 70 (3), 312–326. [1043]

DECKER, C., E. H. LIEB, R. J. MCCANN, AND B. K. STEPHENS (2013): "Unique Equilibria and Substitution Effects in a Stochastic Model of the Marriage Market," *Journal of Economic Theory*, 148 (2), 778–792. [1029]

DIEWERT, W. E. (1971): "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *Journal of Political Economy*, 79 (3), 481–507. [1043]

FOX, J. T. (2007): "Semiparametric Estimation of Multinomial Discrete-Choice Models Using a Subset of Choices," *The RAND Journal of Economics*, 38 (4), 1002–1019. [1028]

FOX, J. T., AND N. LAZZATI (2017): "A Note on Identification of Discrete Choice Models for Bundles and Binary Games," *Quantitative Economics*, 8 (3), 1021–1036. [1024,1025,1039]

FOX, J. T., C. YANG, AND D. H. HSU (2018): "Unobserved Heterogeneity in Matching Games," *Journal of Political Economy*, 126 (4), 1339–1373. [1021,1022,1024,1042,1043]

FUDENBERG, D., R. IIJIMA, AND T. STRZALECKI (2015): "Stochastic Choice and Revealed Perturbed Utility," *Econometrica*, 83 (6), 2371–2409. [1023,1041]

GALICHON, A., AND B. SALANIÉ (2015): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," Technical report. [1029]

GAUTIER, E., AND Y. KITAMURA (2013): "Nonparametric Estimation in Random Coefficients Binary Choice Models," *Econometrica*, 81 (2), 581–607. [1025]

GENTZKOW, M. (2007): "Valuing New Goods in a Model With Complementarity: Online Newspapers," *American Economic Review*, 97 (3), 713–744. [1021,1022,1025,1039]

GOEREE, J. K., C. A. HOLT, AND T. R. PALFREY (2005): "Regular Quantal Response Equilibrium," *Experimental Economics*, 8 (4), 347–367. [1028]

GORMAN, W. M. (1953): "Community Preference Fields," *Econometrica*, 21 (1), 63–80. [1027]

HODERLEIN, S., H. HOLZMANN, M. KASY, AND A. MEISTER (2017): "Corrigendum: Instrumental Variables With Unrestricted Heterogeneity and Continuous Treatment," *The Review of Economic Studies*, 84 (2), 964–968. [1025]

HOFBAUER, J., AND W. H. SANDHOLM (2002): "On the Global Convergence of Stochastic Fictitious Play," *Econometrica*, 70 (6), 2265–2294. [1022]

ICHIMURA, H., AND T. S. THOMPSON (1998): "Maximum Likelihood Estimation of a Binary Choice Model With Random Coefficients of Unknown Distribution," *Journal of Econometrics*, 86 (2), 269–295. [1025]

IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77 (5), 1481–1512. [1025]

KLINE, B. (2017): "Robust Identification in Mechanisms," Technical report. [1029]

KREPS, D. M. (2012): *Microeconomic Foundations I: Choice and Competitive Markets*. Princeton, NJ: Princeton university press. [1046]

LARSEN, B., AND A. L. ZHANG (2018): "A Mechanism Design Approach to Identification and Estimation," Technical report, National Bureau of Economic Research. [1029]

LEWBEL, A. (2000): "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97 (1), 145–177. [1024,1038]

MACHINA, M. J. (1985): "Stochastic Choice Functions Generated From Deterministic Preferences Over Lotteries," *The Economic Journal*, 95 (379), 575–594. [1041]

MANSKI, C. F. (1975): "Maximum Score Estimation of the Stochastic Utility Model of Choice," *Journal of Econometrics*, 3 (3), 205–228. [1028,1038]

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic Theory*. New York, NY: Oxford University Press. [1037]

MATEJKA, F., AND A. MCKAY (2014): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *The American Economic Review*, 105 (1), 272–298. [1041]

MATTSSON, L.-G., AND J. W. WEIBULL (2002): "Probabilistic Choice and Procedurally Bounded Rationality," *Games and Economic Behavior*, 41 (1), 61–78. [1041]

MATZKIN, R. L. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, 60 (2), 239–270. [1040]

——— (1993): "Nonparametric Identification and Estimation of Polychotomous Choice Models," *Journal of Econometrics*, 58 (1), 137–168. [1022,1024,1028,1038]

——— (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71 (5), 1339–1375. [1025]

——— (2008): "Identification in Nonparametric Simultaneous Equations Models," *Econometrica*, 76 (5), 945–978. [1025]

——— (2015): "Estimation of Nonparametric Models With Simultaneity," *Econometrica*, 83 (1), 1–66. [1025]

MCFADDEN, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," in *Frontiers in Econometrics*, ed. by P. Zarembka. New York, NY: Academic Press, 105–142. [1021,1022,1038]

——— (1978): "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Residential Location*, ed. by A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull. Amsterdam: North Holland Pub., Co., 75–96. [1023,1043]

——— (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data With Econometric Applications*, ed. by C. F. Manski and D. McFadden. Cambridge, MA: MIT Press, 198–272. [1022,1024,1027,1028,1034,1038]

MCFADDEN, D. L., AND M. FOSGERAU (2012): "A Theory of the Perturbed Consumer With General Budgets," Technical report, National Bureau of Economic Research. [1024]

ROCKAFELLAR, R. T. (1970): *Convex Analysis*. Princeton, NJ: Princeton University Press. [1047,1049,1051]

SHI, X., M. SHUM, AND W. SONG (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 86 (2), 737–761. [1038,1039,1043]

SMALL, K. A., AND H. S. ROSEN (1981): "Applied Welfare Economics With Discrete Choice Models," *Econometrica*, 49, 105–130. [1022,1034]

TORGOVITSKY, A. (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83 (3), 1185–1197. [1025]

VARIAN, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica*, 50 (4), 945–973. [1023,1036]

——— (1992): *Microeconomic Analysis*. New York: WW Norton. [1023]

Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 16 November, 2017; final version accepted 11 January, 2019; available online 13 February, 2019.