Discussion Papers
Department of Economics
University of Copenhagen

No. 20-01

How McFadden met Rockafellar and learnt to do more with less

by

Jesper R.-V. Sørensen and Mogens Fosgerau

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

How McFadden met Rockafellar and learnt to do more with less*

Jesper R.-V. Sørensen[†] Mogens Fosgerau[‡]

April 14, 2020

Abstract

We study the additive random utility model of discrete choice under minimal assumptions. We make no assumptions regarding the distribution of random utility components or the functional form of systematic utility components. Exploiting the power of convex analysis, we are nevertheless able to generalize a range of important results. We characterize demand with a generalized Williams-Daly-Zachary theorem. A similarly generalized version of Hotz-Miller inversion yields constructive partial identification of systematic utilities. Estimators based on our partial identification result remain well defined in the presence of zeros in demand. We also provide necessary and sufficient conditions for point identification.

Keywords: Additive random utility model; Discrete choice; Convex duality; Demand inversion; Partial identification.
JEL classification: C25, C6, D11.

1 Introduction

In this paper, we formulate a highly general additive random utility model (ARUM) of discrete choice and generalize a number of results fundamental to the discrete-choice literature. Proofs of our main results are by and large obtained as applications of existing results from convex analysis.

*We thank Emerson Melo and Matt Shum for useful comments and discussion. Mogens Fosgerau and Jesper R.-V. Sørensen have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 740369).

[†]University of Copenhagen; jrvs@econ.ku.dk
[‡]University of Copenhagen; mogens.fosgerau@econ.ku.dk

By definition, an ARUM models choice as guided by utility maximization among of a set of mutually exclusive alternatives, each of which is characterized by a utility that is the sum of a systematic and a random component. In this paper, the systematic component is an arbitrary nonparametric function of conditioning variables, and the random component follows an arbitrary conditional distribution.

We impose no restrictions on the distribution of the random utility component, which implies that utility ties may occur with positive probability. Our general ARUM allows any tie-breaking rule. Consequently, the model describes demand (formally, conditional mean choice) in the form of a conditional choice probability (CCP) correspondence, which need not be single-valued.

In the classical ARUM formulation random utility components are taken to be absolutely continuous with full support (Anderson, De Palma, and Thisse, 1992). These classical assumptions imply that demand is the gradient of the surplus function, the latter being defined as the expected maximum utility. This result is the famous Williams-Daly-Zachary (WDZ) theorem (McFadden, 1981), which is a discrete-choice analog of Roy's identity.[1] Presuming differentiability, the WDZ theorem does not apply to the general case, where the surplus function need not be differentiable. However, we show that surplus of a general ARUM is always *sub*differentiable and that its *sub*differential is exactly the CCP correspondence. Our finding does not require any smoothness or regularity and thus strictly generalizes the WDZ theorem.

Our generalized WDZ theorem has several uses. First, the theorem operationalizes the computation of the ARUM CCP correspondence based on a surplus function. Moreover, as the CCP correspondence coincides with the subdifferential of a convex function, it is necessarily *cyclically monotone*. Cyclical monotonicity yields conditions closely related to the Afriat (1967) construction and may be used to check whether observed behavior is consistent with the hypothesis of utility maximization. They also imply moment inequalities that can be exploited for identification and estimation as in Shi, Shum, and Song (2018), but without the smoothness and regularity imposed on the random utility components by these authors.

Empirical work in applied microeconomics often utilizes models of discrete choice. Using the classical ARUM formulation within a dynamic discrete choice setup, Hotz and Miller (1993) show that the CCP function is invertible. When the distribution of the random utility components is known, the Hotz-Miller inversion result implies that the vector of deterministic utility components is identified from the choice probability vector. This *invertibility* is often employed as a first step in two-(or multi-)step procedures for estimation of structural model

[1]The WDZ theorem is named after Williams (1977) and Daly and Zachary (1978). See McFadden (1978) and Rust (1994, Theorem 3.1) for detailed dicussions.

parameters. Specifically, in such procedures, one first obtains utility estimates by evaluating the inverse CCP function at estimated CCPs. Structural parameters of the model are then subsequently estimated, for example, by a minimum distance or generalized method of moments procedure. Essentially the same idea was exploited in a static discrete choice setup in the seminal contributions by Berry (1994) and Berry, Levinsohn, and Pakes (1995) [henceforth: BLP], adding the ability to exploit instruments in the second step in order to deal with endogenous prices. Chiong, Galichon, and Shum (2016) apply convex analysis to the surplus and its gradient in the WDZ theorem to obtain the inverse ARUM demand. These authors use conjugate duality to provide a computationally efficient approach to carrying out this inversion.

This paper establishes partial identification of systematic utilities in the case of a general ARUM, obtained through a dual version of our generalized WDZ theorem. Just as the CCP correspondence is the subdifferential of the surplus, we find that the *inverse* CCP correspondence is the subdifferential of the *conjugate* surplus. The conjugate surplus may be obtained from the surplus by means of convex optimization (see Rockafellar, 1970, Chapter 12). Our identification result is constructive and characterizes the identified set as the solutions to a convex minimization problem involving the surplus function. The closed-form nature of our partial-identification result also allows us to express necessary and sufficient conditions for *point* identification.

While Hotz and Miller (1993) focus on establishing invertibility of the CCP function, Norets and Takahashi (2013) point out the importance of *surjectivity* for estimation purposes, in particular. The latter authors observe that if the CCP function does not have full range and CCPs are obtained with error (e.g. estimated), then a procedure involving CCP inversion may involve evaluation of the inverse CCP function at points where it is undefined. Assuming continuity of the random utility components, Norets and Takahashi show that the image of the CCP function does in fact cover the interior of the probability simplex. Additionally, this image coincides with the simplex interior if the random utility components are also distributed with full support. Their main take-away is therefore that, under the Hotz-Miller assumptions, procedures based CCP inversion are unproblematic.

In this paper we show that the image of the CCP correspondence covers the interior of the probability simplex without making any assumption on the distribution of random utility components and allowing utility ties to occur with positive probability. Under the Norets-Takahashi/Hotz-Miller assumptions, the CCP correspondence reduces to a function. We furthermore show that the assumptions of these authors imply the stronger result that the CCP function is not only a bijection between the space of utilities and the interior of the probability simplex—it is also *continuous in both directions* (i.e., a homeomorphism). An implication of our latter finding is that, under the Hotz-Miller assumptions, CCP inversion is not only well defined, but also *well behaved*.

When estimating differentiated-products demand systems, one often encounters zeros in demand. Zeros are especially prevalent in "big data," which allow for a highly detailed view of consumers, firms and markets (see, e.g., Nurski and Verboven, 2016; Quan and Williams, 2018). Most existing demand-estimation techniques fail in the presence of zero demands. Specifically, estimation procedures building on the influential paper by BLP do not apply with zero demands as they fail to be well defined. Two approaches are commonly invoked to enforce nonzero demands: aggregation over products or markets or omission thereof. Neither approach can be viewed as fully satisfactory.[2] A notable exception to standard techniques can be found in Gandhi, Lu, and Shi (2019), who propose a solution to the zero-demands problem based on constructing bounds for the conditional expectation of inverse demand. In this paper the role of inverse demand (bounds) is played by the inverse CCP correspondence, which (as we establish) is always well defined. Our general partial identification result therefore readily allows for zero probabilities—there is no need for any aggregation or omission. Moreover, given the constructive nature of this identification, our result may therefore be used as a basis for estimation even if one encounters zero probabilities empirically.

In addition to generalizing a range of important results from the literature, this paper also contributes on a technical level. Specifically, we provide relatively simple proofs that rely to a large extent on mathematical results that have long been consolidated in standard convex analysis, thus avoiding the reinvention and reverification of mathematical arguments.

Our identification of the deterministic utility components parallels the analysis in Chiong et al. (2016) in that it treats the distribution of the random utility components as known.[3] Other authors rely on different model assumptions in order to obtain identification. Notably, in pioneering work, Rosa Matzkin establishes nonparametric identification of both the deterministic utility components and the distribution of random utility components under various independence and/or shape restrictions motivated by economic theory (see, e.g., Matzkin, 1992, 1993, 1994). Using a special regressor, Fosgerau and Kristensen (2019) show identification in a class of index models that comprise ARUM with very little structure imposed and without continuity of the deterministic utility component. This paper is also related to Allen and Rehbeck (2019) who study a class of latent utility models with additively separable unobservable heterogeneity assumed to be independent of conditioning variables (cf. their

[2]Specifically, aggregating data over markets or products may smooth over the heterogeneity of interest (Quan and Williams, 2018). In addition, omitting products without sales implicitly assumes that there is no demand for these products and may create a selection bias (Berry, Levinsohn, and Pakes, 2004; Gandhi, Lu, and Shi, 2019).

[3]We stress that our generalized WDZ and Hotz-Miller inversion theorems continue to hold even if the distribution of random utility components is unknown. It is only in interpreting these results as statements concerning identification that the knowledge of this distribution is relevant.

pp. 1027, 1046). Specializing their framework to the ARUM of discrete choice, their general results produce nonparametric point identification of systematic utilities. In contrast, we establish partial identification of systematic utilities using arguments which rely on neither independence assumptions nor the presence of a special regressor. The present paper should thus be viewed as complementary to the work of the above-mentioned authors.

We formulate the general ARUM in Section 2 and generalize the WDZ theorem in Section 3. Section 4 dualizes our generalized WDZ theorem to produce a generalized Hotz-Miller inversion result, which, in turn, produces partial identification of systematic utilities. Section 5 presents assumptions that are just sufficient to ensure that the CCP correspondence and its inverse are single-valued, thus yielding point identification. Section 6 concludes. Appendix A gives a quick tour of convex analysis, Appendix B contains all proofs, and Appendix C includes supplementary material.

Notation Vectors $\mathbf{x}$ and $\mathbf{y}$ are understood as columns. $\langle \mathbf{x}, \mathbf{y} \rangle$ denotes the usual scalar/inner product. $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the Euclidean and supremum norms, respectively. $\operatorname{int} S$ and $\operatorname{cl} S$ stand for the interior and closure, respectively, of a set S.

2 Discrete Choice Model Setup

In this section we set up a highly general version of the additive random utility model (ARUM) of discrete choice. In this model a single agent chooses from $J+1$ mutually exclusive alternatives labelled $\{0, 1, \ldots, J\} =: \mathcal{J}$. The utility the agent derives from choosing alternative j is

$$\pi_j(\mathbf{X}) + \varepsilon_j, \quad j \in \mathcal{J},$$

where $\{\pi_j(\mathbf{X})\}_{j\in\mathcal{J}}$ denotes the *systematic components* of utility, and $(\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_J)$ the *random components* (random elements of $\mathbb{R}^{1+J}$). Each function π_j translates conditioning variables $\mathbf{X}$, having support $\mathcal{X}$, into 'utils.' Conditioning variables $\mathbf{X}$ are observable to both the agent and the researcher. While in empirical applications, the functions $\pi_j : \mathcal{X} \to \mathbb{R}, j \in \mathcal{J}$, are typically parametrically specified, our treatment of these will be *non*parametric. The ε_j's are random variables observed only by the agent, defined on a given probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with ω denoting a generic element of Ω.

The agent chooses an alternative j that maximizes $\pi_j(\mathbf{X}) + \varepsilon_j$. Given that the solution to this optimization problem only depends on utility differences (calculated with respect to any fixed alternative), we may normalize $\pi_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{X}$ and $\varepsilon_0 = 0$ and, thus, treat alternative 0 as the "outside option." This normalization allows us to formulate the agent's problem in a notationally convenient manner. For this purpose, let $\{\mathbf{u}_j\}_1^J$ be the canonical basis for $\mathbb{R}^J$ and write $\mathbf{u}_0 := \mathbf{0}$ for the origin in $\mathbb{R}^J$. The latter vector "points" towards the outside option. The agent's problem may then be reformulated as

$$\max_{j\in\mathcal{J}}\left\{\pi_j\left(\mathbf{X}\right)+\varepsilon_j\right\}=\max_{j\in\mathcal{J}}\left\{\left\langle\boldsymbol{\pi}\left(\mathbf{X}\right)+\boldsymbol{\varepsilon},\mathbf{u}_j\right\rangle\right\},$$

where $\boldsymbol{\pi}\left(\mathbf{x}\right):=\left(\pi_1\left(\mathbf{x}\right),\ldots,\pi_J\left(\mathbf{x}\right)\right)$ collects the (normalized) systematic components and $\boldsymbol{\varepsilon}:=\left(\varepsilon_1,\ldots,\varepsilon_J\right)$. The solutions to the latter problem is

$$\mathcal{J}\left(\boldsymbol{\pi}\left(\mathbf{X}\right)+\boldsymbol{\varepsilon}\right):=\operatorname*{argmax}_{j\in\mathcal{J}}\left\{\left\langle\boldsymbol{\pi}\left(\mathbf{X}\right)+\boldsymbol{\varepsilon},\mathbf{u}_j\right\rangle\right\},$$

a random subset of $\mathcal{J}$.

We denote the cumulative distribution function (CDF) of $\boldsymbol{\varepsilon}$ conditional on $\mathbf{X}$ by $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$.[4] We initially impose no conditions on this CDF. Our analysis thus differs from most of the literature, which assumes that this conditional distribution is absolutely continuous (with respect to Lebesgue measure) with finite means. For the purpose of identification, we follow Chiong et al. (2016) and treat $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ as known to the researcher.

A fundamental entity for our analysis is the *surplus function* $W\left(\cdot|\,\mathbf{x}\right):\mathbb{R}^J\to\mathbb{R}$ (McFadden, 1978, 1981), defined for each fixed $\mathbf{x}\in\mathcal{X}$ by

$$W\left(\mathbf{v}|\,\mathbf{x}\right):=\mathrm{E}\Big[\max_{j\in\mathcal{J}}\left\langle\mathbf{v}+\boldsymbol{\varepsilon},\mathbf{u}_j\right\rangle-\max_{j\in\mathcal{J}}\left\langle\boldsymbol{\varepsilon},\mathbf{u}_j\right\rangle\Big|\mathbf{X}=\mathbf{x}\Big].\tag{1}$$

The function $W\left(\cdot|\,\mathbf{x}\right)$ may be interpreted as the expected surplus of the agent when faced with utilities $\mathbf{v}$ relative to zero utilities. In the definition of $W\left(\cdot|\,\mathbf{x}\right)$, we follow McFadden (1981) and subtract the maximum of the random utility components. This subtraction ensures that the expectation is well defined, even if $\boldsymbol{\varepsilon}$ is not integrable (i.e., even if some ε_j has no mean). Indeed, the vector-max function $\max_{j\in\mathcal{J}}\langle\cdot,\mathbf{u}_j\rangle$ is 1-Lipschitz with respect to the supremum norm, which implies

$$\mathrm{E}\Big[\Big|\max_{j\in\mathcal{J}}\left\langle\mathbf{v}+\boldsymbol{\varepsilon},\mathbf{u}_j\right\rangle-\max_{j\in\mathcal{J}}\left\langle\boldsymbol{\varepsilon},\mathbf{u}_j\right\rangle\Big|\Big]\leqslant\max_{1\leqslant j\leqslant J}|v_j|<+\infty.$$

It follows from the Radon-Nikodym theorem that the surplus function is always well defined; no conditions need to be placed on the distribution of $\boldsymbol{\varepsilon}$. Beyond ensuring the well-definedness of the surplus, subtracting the maximum of the random utility components has no substantial effect on its shape.[5]

The utility maximizing choice of the agent is represented by $\mathbf{Y}:=\left(Y_1,\ldots,Y_J\right)$, a random element of $\left\{\mathbf{u}_j\right\}_0^J$ whose conditional support given $\left(\mathbf{X},\boldsymbol{\varepsilon}\right)$ is contained in $\left\{\mathbf{u}_j\right\}_{j\in\mathcal{J}\left(\boldsymbol{\pi}\left(\mathbf{X}\right)+\boldsymbol{\varepsilon}\right)}$.

[4] To keep notation at a minimum, throughout this paper we identify distributions with their CDFs.

[5] One may alternatively define the surplus by $W\left(\mathbf{v}|\,\mathbf{x}\right):=\mathrm{E}\left[\max_{j\in\mathcal{J}}\left\langle\mathbf{v}+\boldsymbol{\varepsilon},\mathbf{u}_j\right\rangle\middle|\,\mathbf{X}=\mathbf{x}\right]$. However, this definition requires the ε_j's to be integrable, which rules out fat-tailed distributions (e.g. the Cauchy family).

Most of the existing literature assumes $F_{\boldsymbol{\varepsilon}|\mathbf{X}}$ is absolutely continuous, an assumption which allows for choices to be expressed in terms of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, namely $\mathbf{Y} = \mathbf{u}_{\mathcal{J}(\boldsymbol{\pi}(\mathbf{X})+\boldsymbol{\varepsilon})}$.[6] In contrast, we do not specify how the agent breaks utility ties but allow any rule consistent with utility maximization. Consequently, our approach allows for *incomplete* models.[7] Based on the utility-maximizing choice, the *conditional choice probability* (CCP) *function* $\mathbf{p} : \mathcal{X} \to \mathbb{R}^J$ is defined by

$$\mathbf{p}(\mathbf{x}) := \mathrm{E}[\mathbf{Y}|\, \mathbf{X} = \mathbf{x}]. \tag{2}$$

By definition of the choice vector $\mathbf{Y}$, the CCP function $\mathbf{p}$ takes values in the unit simplex,

$$\Delta := \operatorname{conv}(\{\mathbf{u}_j\}_{j\in\mathcal{J}}) = \Big\{\mathbf{q} \in \mathbb{R}^J; q_j \geqslant 0, \sum_{j=1}^{J} q_j \leqslant 1\Big\}, \tag{3}$$

where we write $\operatorname{conv}(S)$ for the convex hull of a set S, i.e. all convex combinations of elements of S.[8]

For each $\mathbf{x} \in \mathcal{X}$, we define the *CCP correspondence* $\boldsymbol{\mathcal{P}}(\cdot|\, \mathbf{x}) : \mathbb{R}^J \rightrightarrows \mathbb{R}^J$ as the set of all CCP vectors that are consistent with utility-maximizing choices. This is the set expectation

$$\begin{aligned}\boldsymbol{\mathcal{P}}(\mathbf{v}|\, \mathbf{x}) &:= \mathrm{E}\left[\operatorname{conv}\left(\{\mathbf{u}_j\}_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})}\right)\Big|\, \mathbf{X} = \mathbf{x}\right] \\ &= \Big\{\boldsymbol{\mu} \in \mathbb{R}^J; \boldsymbol{\mu} = \mathrm{E}[\boldsymbol{\xi}|\, \mathbf{X} = \mathbf{x}] \text{ for } \boldsymbol{\xi} : \Omega \to \mathbb{R}^J \text{ measurable} \\ &\qquad \text{satisfying } \mathbf{P}\left(\boldsymbol{\xi} \in \operatorname{conv}\left(\{\mathbf{u}_j\}_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})}\right)\Big|\, \mathbf{X} = \mathbf{x}\right) = 1\Big\}.\end{aligned}$$

In this definition, we may think of different $\boldsymbol{\xi}$'s as choice rules consistent with utility maximization but employing different tie-breaking rules. An element of $\boldsymbol{\mathcal{P}}(\mathbf{v}|\, \mathbf{x})$ is a choice probability vector, which is the expected value of the choice rule given $\mathbf{X} = \mathbf{x}$.[9] Then in particular, $\mathbf{p}(\mathbf{x}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{\pi}(\mathbf{x})|\mathbf{x})$ for each $\mathbf{x} \in \mathcal{X}$. In Section 5, we provide necessary and sufficient conditions for $\boldsymbol{\mathcal{P}}(\cdot|\, \mathbf{x})$ and its inverse to reduce to functions (see Theorems 4 and 5, respectively).

---

[6] Absolute continuity ensures that $\mathcal{J}(\mathbf{v} + \boldsymbol{\varepsilon})$ is singleton with probability one for all $\mathbf{v}$.

[7] Koning and Ridder (2003) refer to incomplete models as *incoherent*.

[8] We calculate the probability of the outside option in a residual manner, i.e. $\mathbf{P}(\mathbf{Y} = \mathbf{u}_0|\, \mathbf{X} = \mathbf{x}) = 1 - \sum_{j=1}^{J} p_j(\mathbf{x})$, which explains why the elements $\mathbf{q}$ of Δ may sum to less than one.

[9] Note that $\mathbf{P}(\boldsymbol{\xi} \in \operatorname{conv}(\{\mathbf{u}_j\}_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})})|\mathbf{X} = \mathbf{x}) = 1$ implies that $\boldsymbol{\xi}$ is confined to the simplex Δ with probability one. Given that Δ is bounded, such $\boldsymbol{\xi}$'s are therefore automatically integrable, in the sense that each coordinate of $\boldsymbol{\xi}$ is integrable. Therefore $\boldsymbol{\mathcal{P}}(\cdot|\, \mathbf{x})$ is a well-defined correspondence.

3 The Williams-Daly-Zachary Theorem

Our first result is a generalization of the famous Williams-Daly-Zachary Theorem to the present general case where the conditional distribution of the random utility components is unrestricted. Even in the absence of any continuity of the conditional distribution of random utilities, the surplus is everywhere *sub*differentiable and its *sub*differential is the CCP *correspondence*. Being the subdifferential of a finite convex function, $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ must be cyclically monotone (see Appendix A).

Theorem 1 (Generalized Williams-Daly-Zachary)**.** *For any fixed* $\mathbf{x} \in \mathcal{X}$*,*

1. $W(\cdot|\,\mathbf{x})$ *is finite, convex and 1-Lipschitz continuous with respect to* $\|\cdot\|_\infty$*;*
2. $W(\cdot|\,\mathbf{x})$ *is everywhere subdifferentiable, i.e.* $\operatorname{dom}\partial W(\cdot|\,\mathbf{x}) = \mathbb{R}^J$*;*
3. $\partial W(\cdot|\,\mathbf{x}) = \boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$*; and*
4. $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ *is cyclically monotone.*

The main insight is that the CCP correspondence is equal to the subdifferential of the surplus. To put that insight into perspective, we may note that, e.g., Shi et al. (2018, Lemma 2.1) show that, if the random utility components $\boldsymbol{\varepsilon}$ are (conditionally) absolutely continuously distributed, then the surplus $W(\cdot|\,\mathbf{x})$ is everywhere differentiable and its gradient is the CCP function $\mathbf{v} \mapsto \mathrm{E}[\mathbf{u}_{\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})}|\mathbf{X} = \mathbf{x}]$.[10] Theorem 1.3 generalizes this property to the case where the random utility components may follow a completely arbitrary distribution and where the utility maximizing agent may use any tie-breaking rule. It remains the case that the surplus encodes utility-maximizing behavior with the CCP correspondence now being equal to the subdifferential of the surplus.

Cyclical monotonicity yields a set of inequalities that capture the fact that the epigraph (the collection of points on or above the graph) of a convex function is also convex. As the subdifferential of a proper convex function is necessarily cyclically monotone, then so is the CCP correspondence. Cyclical monotonicity of the CCP correspondence $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ means that

$$\langle \mathbf{q}_0, \mathbf{v}_0 - \mathbf{v}_1 \rangle + \langle \mathbf{q}_1, \mathbf{v}_1 - \mathbf{v}_2 \rangle + \cdots + \langle \mathbf{q}_n, \mathbf{v}_n - \mathbf{v}_0 \rangle \geqslant 0$$

for any integer $n \geqslant 1$ and any sequence $\{(\mathbf{v}_i, \mathbf{q}_i)\}_{i=0}^n$ of pairs such that $\mathbf{q}_i \in \boldsymbol{\mathcal{P}}(\mathbf{v}_i|\,\mathbf{x})$. The CCP correspondence $\boldsymbol{\mathcal{P}}(\boldsymbol{\pi}(\mathbf{x})|\,\mathbf{x})$ contains *all* CCPs consistent with utility-maximizing

[10] As mentioned in the Introduction, the differential version of the WDZ theorem is a discrete-choice analog of Roy's identity. Allen and Rehbeck (2019, Lemma 1) shows that a "Roy's Identity" holds for an entire class of latent utility models with additively separable heterogeneity taken to be independent of conditioning variables.

behavior. The property of cyclical monotonicity is thus informative for identifying utilities regardless of how ties are broken.

Shi et al. (2018) show that the cyclical monotonicity may be used to derive identifying inequalities for the structural parameter parameterizing the ARUM utility functions $\boldsymbol{\pi}$ and construct an estimator based on these inequalities. In a *cross-sectional* setting, their treatment takes the distribution of the random utility component to be independent of the conditioning variables. In the present context, independence implies that the CCP correspondence does not depend on $\mathbf{x}$ and we may write $\boldsymbol{\mathcal{P}}(\cdot|\mathbf{x}) = \boldsymbol{\mathcal{P}}(\cdot)$. Theorem 1.4 then shows that no assumption need to be placed on the (marginal) distribution of random utility components in order to arrive at the same identifying inequalities as Shi et al. (2018). Specifically, cyclical monotonicity of $\boldsymbol{\mathcal{P}}$ implies that inequalities hold for any collection $\{(\boldsymbol{\pi}(\mathbf{x}_i), \mathbf{p}(\mathbf{x}_i))\}_{i=0}^{n}$ of pairs such that $\mathbf{p}(\mathbf{x}_i) \in \boldsymbol{\mathcal{P}}(\boldsymbol{\pi}(\mathbf{x}_i))$ and thus extends the applicability of their estimator based on cross-sectional data. In a *panel* setting with additive fixed effects, Shi et al. (2018) assume absolute continuity and strict stationarity conditional on the fixed effect of the distribution of random utility components in order to make use of cyclical monotonicity for each individual i. They may then integrate out the individual-specific fixed effect to arrive at identifying inequalities. Again, by appealing to Theorem 1.4, the applicability of their panel-data estimator may be extended to allow for arbitrary distributions, not necessarily absolutely continuous.

4 Hotz-Miller Inversion

The general WDZ theorem provides a (possibly set-valued) mapping from utilities to CCPs consistent with utility maximization. The issue of identification is an inverse problem: Given the CCPs $\mathbf{p}(\mathbf{x})$, can one recover the systematic utilities $\boldsymbol{\pi}(\mathbf{x})$? Appealing to Theorem 1.3, in our search for the correspondence inverse $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$ at a vector $\mathbf{q} \in \mathbb{R}^J$, we are asking for the set of vectors $\mathbf{v}$ satisfying $\mathbf{q} \in \partial W(\mathbf{v}|\,\mathbf{x})$. As it turns out, the correspondence inverse is conveniently expressed by means of *Fenchel conjugation*.

Given an $\mathbf{x} \in \mathcal{X}$, we define the *conjugate surplus function* $W^*(\cdot|\,\mathbf{x}) : \mathbb{R}^J \to (-\infty, +\infty]$ as the *convex conjugate* of $W(\cdot|\,\mathbf{x})$:

$$W^*(\mathbf{q}|\,\mathbf{x}) := \sup_{\mathbf{v} \in \mathbb{R}^J} \{\langle \mathbf{v}, \mathbf{q} \rangle - W(\mathbf{v}|\,\mathbf{x})\}, \quad \mathbf{q} \in \mathbb{R}^J. \tag{4}$$

Finiteness of the surplus (Theorem 1.1) ensures that its conjugate is never $-\infty$. When the ARUM is a multinomial logit model, the conjugate surplus function is the (negative) Shannon entropy (Anderson et al., 1988).[11] For any distribution of random utility, the

[11] With alternative 0 being the outside option, the Shannon entropy is the function $H : \mathbb{R}^J \to \mathbb{R} \cup \{+\infty\}$

(negative) conjugate surplus function may thus be viewed as a *generalized entropy* (Galichon and Salanie, 2015; Fosgerau, Melo, de Palma, and Shum, 2018).

The next theorem relates the conjugate surplus to the CCP correspondence. We denote the *image* (or *range*) of a correspondence $\boldsymbol{\rho} : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ as

$$\operatorname{Im} \boldsymbol{\rho} := \{\mathbf{y} \in \mathbb{R}^m ; \mathbf{y} \in \boldsymbol{\rho}(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^m\} = \bigcup_{\mathbf{x} \in \mathbb{R}^m} \boldsymbol{\rho}(\mathbf{x}).$$

Theorem 2 (Generalized Hotz-Miller)**.** *For any* $\mathbf{x} \in \mathcal{X}$*,*

1. $W^*(\cdot|\,\mathbf{x})$ *is lower semi-continuous*[12] *and proper convex with effective domain* $\operatorname{int} \Delta \subseteq \operatorname{dom} W^*(\cdot|\,\mathbf{x}) \subseteq \Delta$*; in particular,* $W^*(\cdot|\,\mathbf{x})$ *is continuous on* $\operatorname{int} \Delta$*;*
2. $\partial W^*(\cdot|\,\mathbf{x}) = \boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$*;*
3. $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$ *is cyclically monotone;*
4. $\operatorname{int} \Delta \subseteq \operatorname{dom} \partial W^*(\cdot|\,\mathbf{x}) = \operatorname{Im} \boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x}) \subseteq \Delta$ *and* $\operatorname{Im} \partial W^*(\cdot|\,\mathbf{x}) = \operatorname{dom} \boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x}) = \mathbb{R}^J$.

The properties stated in Theorem 2 are derived using 'zero-order' convex analysis alone and do not rely on any differential theory. They correspond via convex duality to the properties in Theorem 1.

Lower semi-continuity, properness and convexity are generic to convex conjugates of proper convex functions (Rockafellar, 1970, Theorem 12.2). However, continuity is not. Yet, Theorem 2.1 shows that continuity on the simplex interior is a feature common to all conjugate surplus functions.

The effective domain $\operatorname{dom} W^*(\cdot|\,\mathbf{x})$ does not include the simplex boundary in general. In the context of binary choice ($J = 1$), in Appendix C we show that whether or not the boundary is included boils down to whether or not the (scalar) random utility is integrable. More broadly, one may show that $\operatorname{dom} W^*(\cdot|\,\mathbf{x})$ *is* the entire simplex whenever all the ε_j's are integrable. Continuity (relative to Δ) then follows from Rockafellar (1970, Theorem 10.2).

When utility ties have positive probability, the classical Hotz-Miller function

$$\mathbf{v} \mapsto \boldsymbol{\phi}(\mathbf{v}|\,\mathbf{x}) := \mathrm{E}[\mathbf{u}_{\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})}|\mathbf{X} = \mathbf{x}] \tag{5}$$

given by $H(\mathbf{q}) = -\sum_{j=1}^{J} q_j \ln q_j - (1 - \sum_{j=1}^{J} q_j) \ln(1 - \sum_{j=1}^{J} q_j)$ if $\mathbf{q} \in \Delta$ and $H(\mathbf{q}) = +\infty$ otherwise (where $0 \ln 0$ is understood as zero).

[12] A function $f : \mathbb{R}^m \to (-\infty, +\infty]$ is called *lower semi-continuous* at a point $\mathbf{x} \in \mathbb{R}^m$ if for every sequence $\{\mathbf{x}_n\}_1^\infty$ in $\mathbb{R}^m$ converging to $\mathbf{x}$, one has $\liminf_{n \to \infty} f(\mathbf{x}_n) \geqslant f(\mathbf{x})$. The function f is called *lower semi-continuous* if this conclusion holds true at every point.

is not well defined as a map into the probability simplex Δ. For this reason, one cannot speak of "CCP inversion" in the Hotz-Miller sense, in general. In contrast, the inverse CCP correspondence $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$ is always well defined and nonempty-valued everywhere on (at least) $\operatorname{int}\Delta$ (cf. Theorem 2.4). Given that the inverse CCP correspondence maps choice probabilities into utilities, we refer to it as a *generalized Hotz-Miller inverse*.

From Theorem 2.2 and the variational representation of subdifferentials of proper conjugate pairs [see (A4) in Appendix A], we see that the generalized Hotz-Miller inverse $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$ may be evaluated as

$$\boldsymbol{\mathcal{P}}^{-1}(\mathbf{q}|\,\mathbf{x}) = \partial W^{*}(\mathbf{q}|\,\mathbf{x}) = \operatorname*{argmax}_{\mathbf{v}\in\mathbb{R}^{J}}\left\{\langle\mathbf{v},\mathbf{q}\rangle - W(\mathbf{v}|\,\mathbf{x})\right\}.$$

This observation leads to a *partial identification* result.

Theorem 3 (Partial Identification)**.** *Fix $\mathbf{x}\in\mathcal{X}$. Then the utility values $\boldsymbol{\pi}(\mathbf{x})$ are partially identified with identified set being the closed convex set*

$$\Pi(\mathbf{x}) := \partial W^{*}(\mathbf{p}(\mathbf{x})|\,\mathbf{x}) = \operatorname*{argmax}_{\mathbf{v}\in\mathbb{R}^{J}}\left\{\langle\mathbf{v},\mathbf{p}(\mathbf{x})\rangle - W(\mathbf{v}|\,\mathbf{x})\right\}. \tag{6}$$

The identified set is both nonempty and bounded (hence compact) if and only if $\mathbf{p}(\mathbf{x})\in\operatorname{int}\Delta$.

Theorem 3 provides constructive partial identification of the utility values $\boldsymbol{\pi}(\mathbf{x})$. The theorem also shows that if the true CCPs lie on the boundary of the probability simplex, then systematic utilities *cannot* be point-identified as the identified set must be either empty or unbounded. When an analytic expression for $\partial W^{*}(\cdot|\,\mathbf{x})$ is available, (6) immediately suggests an estimator of the identified set $\Pi(\mathbf{x}) = \partial W^{*}(\mathbf{p}(\mathbf{x})|\,\mathbf{x})$ by substituting the true CCPs with an estimator $\widehat{\mathbf{p}}(\mathbf{x})$. When such an analytic expression is unavailable or difficult to come by, one may still use the "argmax" characterization in (6) to construct a set estimator based on an approximation of $W(\cdot|\,\mathbf{x})$. If no one chooses a particular alternative in the data, then the empirical CCPs lie on the boundary of the probability simplex. The resulting estimator of the identified set remains well defined. Our approach to identification is therefore robust to the "zeros-in-demand problem" mentioned in the introduction. Zero empirical probabilities imply that the resulting utility set estimate is either empty or unbounded. Either finding is informative in the sense of narrowing down the candidate utility space ($\mathbb{R}^{J}$).

Example 1 (Two-Point Support and Partial Identification)**.** *Consider the case of binary choice ($J=1$) with deterministic conditioning variables, such that $\mathbf{X}\equiv\mathbf{x}$ for some fixed vector $\mathbf{x}$. Then the systematic utility function, our object of interest, boils down to the scalar $\pi := \pi(\mathbf{x})$ at $\mathbf{x}$. Let ε be Rademacher distributed, such that $\mathrm{P}(\varepsilon=-1)=\mathrm{P}(\varepsilon=1)=\frac{1}{2}$, and let $\pi\in(-1,1)$. Then the probability of ties is zero and each alternative is chosen with equal* *probability in the population. The finite support of the random utility component implies that the surplus function is piecewise affine and therefore everywhere differentiable except possibly at line-segment endpoints. With the Rademacher distribution we get*

$$W(v) = \begin{cases} -\frac{1}{2}, & v < -1, \\ \frac{1}{2}v, & v \in [-1,1], \\ v - \frac{1}{2}, & v > 1. \end{cases}$$

Differentiating on the interior of each line segment and closing the gaps, we see that the surplus subdifferential (and thus CCP correspondence) takes the form:

$$\mathcal{P}(v) = \partial W(v) = \begin{cases} \{0\}, & v < -1, \\ \left[0, \frac{1}{2}\right], & v = -1, \\ \{\frac{1}{2}\}, & v \in (-1,1), \\ \left[\frac{1}{2}, 1\right], & v = 1, \\ \{1\}, & v > 1. \end{cases}$$

Inversion of this correspondence and evaluation of the inverse at the true choice probability $\frac{1}{2}$ produces the identified set. The desired inverse may here be obtained by mirroring the graph of the CCP correspondence in the 45 degree line. This process is illustrated in Figure 1, where the identified set is the compact and convex $\partial W^*\left(\frac{1}{2}\right) = [-1,1]$. If instead the true utility $\pi > 1$, then alternative $j = 1$ is chosen with probability one. The identified set is then $\partial W^*(1) = [1, +\infty)$, which is closed and convex but unbounded.

5 Point Identification

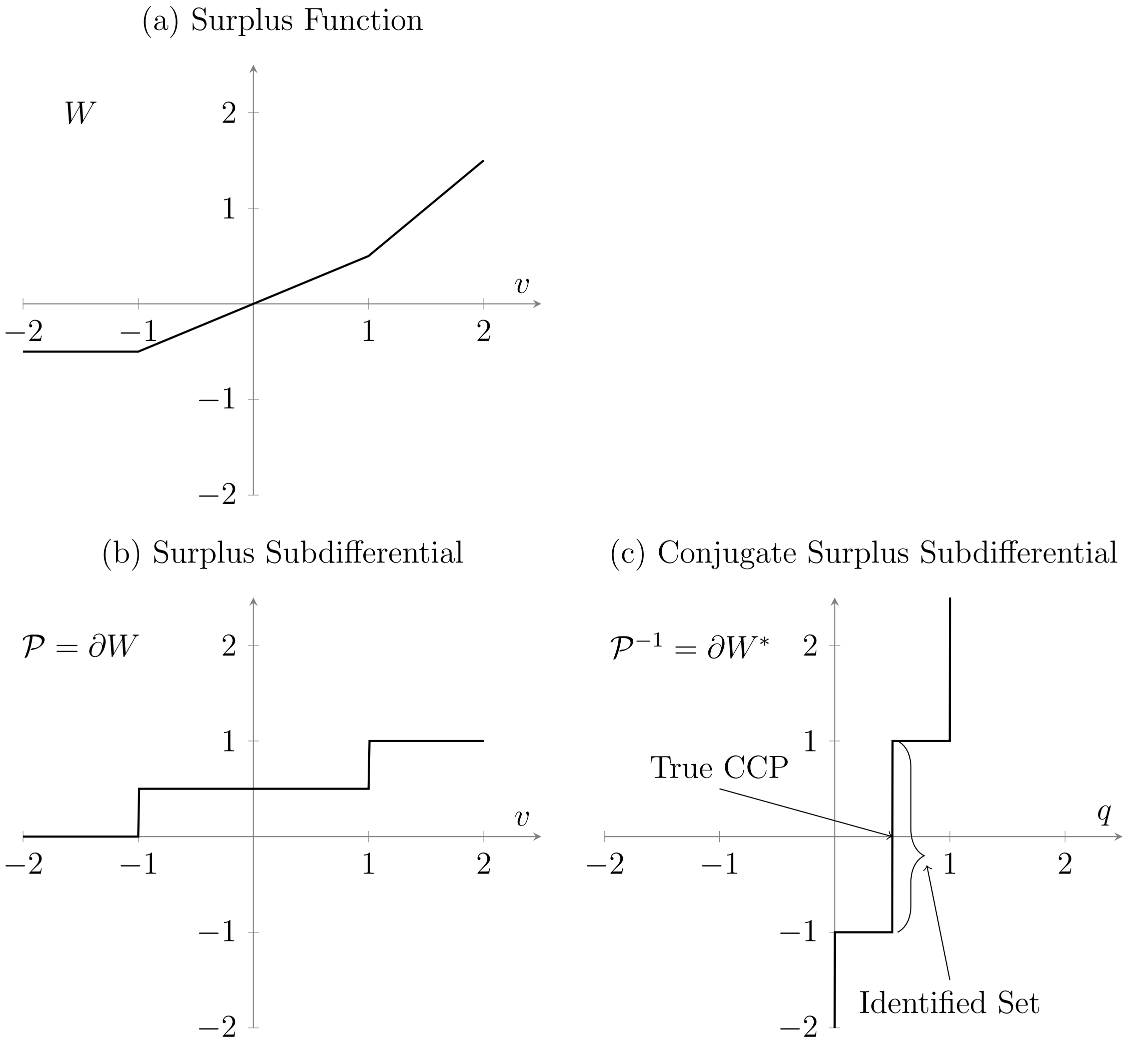
Up to this point, we have derived generic properties of the surplus function, its conjugate and the CCP correspondence, and translated these properties into a partial identification result. We next provide conditions under which stronger properties may be obtained and translate these properties into a *point* identification result.

Recall that a function f finite on a convex set $C \subseteq \mathbb{R}^J$ is called *strictly convex on C* if it satisfies

$$f((1-\lambda)\mathbf{x} + \lambda\mathbf{y}) < (1-\lambda) f(\mathbf{x}) + \lambda(\mathbf{y}) \text{ for all } 0 < \lambda < 1$$

and every pair of distinct points $\mathbf{x}$ and $\mathbf{y}$ in C. To state the desired properties, we require the related notion of *essential strict convexity.*

Figure 1: Identification in Graphics



Definition 1 (Essential Strict Convexity)**.** *A proper convex function f is* ***essentially strictly convex*** *if it is strictly convex on every convex subset of* $\operatorname{dom} \partial f$*.*

Essential strict convexity is dual to the notion of *essential smoothness* (Rockafellar, 1970, Theorem 26.3), defined as follows.

Definition 2 (Essential Smoothness)**.** *A proper convex function f is called* ***essentially smooth*** *if it satifies the following three conditions:*

1. $\operatorname{int}(\operatorname{dom} f)$ *is not empty;*

2. *f is differentiable on* $\operatorname{int}(\operatorname{dom} f)$*; and*

3. $\|\nabla f(\mathbf{x}_k)\| \to \infty$ *for every sequence* $\{\mathbf{x}_k\}_1^\infty$ *in* $\operatorname{int}(\operatorname{dom} f)$ *converging to a boundary point* $\mathbf{x}$ *of* $\operatorname{int}(\operatorname{dom} f)$.

For finite functions, essential smoothness reduces to differentiability ("ordinary smoothness"). To control the essential-smoothness/strict-convexity properties of the surplus and its conjugate, we introduce two conditions.

Condition (C($\mathbf{x}$)). *Fix* $\mathbf{x} \in \mathcal{X}$. *For all* $\mathbf{v}, \mathbf{h} \in \mathbb{R}^J$,

$$\mathbf{P}\left(\max_{j \in \mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})} \langle \mathbf{h}, \mathbf{u}_j \rangle > \min_{j \in \mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})} \langle \mathbf{h}, \mathbf{u}_j \rangle \,\middle|\, \mathbf{X} = \mathbf{x} \right) = 0. \tag{7}$$

Condition (S($\mathbf{x}$)). *Fix* $\mathbf{x} \in \mathcal{X}$. *For all* $\mathbf{v}, \mathbf{h} \in \mathbb{R}^J, \mathbf{h} \neq \mathbf{0}$,

$$\mathbf{P}\left(\max_{j \in \mathcal{J}} \langle \mathbf{v} + \mathbf{h} + \boldsymbol{\varepsilon}, \mathbf{u}_j \rangle > \max_{j \in \mathcal{J}} \langle \mathbf{v} + \boldsymbol{\varepsilon}, \mathbf{u}_j \rangle + \max_{j \in \mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})} \langle \mathbf{h}, \mathbf{u}_j \rangle \,\middle|\, \mathbf{X} = \mathbf{x} \right) > 0. \tag{8}$$

To interpret these conditions, recall that a finite function f is *one-sided directionally differentiable at* $\mathbf{x}$ *in the direction* $\mathbf{h}$ if the limit

$$\lim_{\lambda \to 0_+} \frac{f(\mathbf{x} + \lambda \mathbf{h}) - f(\mathbf{x})}{\lambda}$$

exists (in $\mathbb{R}$), where $\lambda \to 0_+$ is short for "as $\lambda > 0$ approaches zero." It is *directionally differentiable at* $\mathbf{x}$ *in the direction* $\mathbf{h}$ if limits exists for both $\mathbf{h}$ and $-\mathbf{h}$ and these limits coincide, and (fully) *differentiable at* $\mathbf{x}$ if it this conclusion holds true for all directions $\mathbf{h}$. The event inside the probability in (7) is equivalent to the max function $\max_{j \in \mathcal{J}} \langle \cdot + \boldsymbol{\varepsilon}, \mathbf{u}_j \rangle$ *failing* to be directionally differentiable at $\mathbf{v}$ in the direction $\mathbf{h}$. Equation (7) makes it explicit that *non*differentiability may occur only in the event of utility ties, as otherwise $\mathcal{J}(\mathbf{v} + \boldsymbol{\varepsilon})$ is a singleton and the involved "max" and "min" must coincide. Condition C($\mathbf{x}$) enforces that the conditional distribution of $\boldsymbol{\varepsilon}$ given $\mathbf{X} = \mathbf{x}$ views points leading to ties as negligible. It is therefore a *continuity* condition.

One may also show that the event in (8) is equivalent to the max function $\max_{j \in \mathcal{J}} \langle \cdot + \boldsymbol{\varepsilon}, \mathbf{u}_j \rangle$ being nonconstant at $\mathbf{v}$ when moving in the direction $\mathbf{h} \neq \mathbf{0}$. Condition S($\mathbf{x}$) is therefore a condition of sufficiently rich *support*.

The commonly-invoked assumptions of absolute continuity and full support of the conditional distribution of $\boldsymbol{\varepsilon}$ given $\mathbf{X} = \mathbf{x}$ suffice for Conditions C($\mathbf{x}$) and S($\mathbf{x}$). However, Condition C($\mathbf{x}$) does *not* require the existence of a Lebesgue density, i.e. *absoluteness* is not necessary. Moreover, Condition S($\mathbf{x}$) may be satisfied with less than full support. For example, the support may have holes of finite diameter.[13]

The next two theorems translate these two conditions into properties of and relations between the surplus, its conjugate and the CCP correspondence. Starting with Condition

[13] Note that Condition S($\mathbf{x}$) does not require the support to be a connected set. Connectedness is part of the premise of the bijectivity result of Norets and Takahashi (2013, Corollary 1).

$\mathrm{C}(\mathbf{x})$, we obtain the following theorem.

Theorem 4 (Differential Williams-Daly-Zachary)**.** *Fix $\mathbf{x} \in \mathcal{X}$. Then Condition C($\boldsymbol{x}$) is equivalent to each of the following:*

1. *$W(\cdot|\,\mathbf{x})$ is everywhere differentiable.*
2. *$W^*(\cdot|\,\mathbf{x})$ is essentially strictly convex.*
3. *$\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ is single-valued.*

In this case, $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ reduces to the gradient mapping $\nabla W(\cdot|\,\mathbf{x})$, i.e. for all $\mathbf{v} \in \mathbb{R}^J$, $\boldsymbol{\mathcal{P}}(\mathbf{v}|\,\mathbf{x})$ consists of the vector $\nabla W(\mathbf{v}|\,\mathbf{x})$ alone.

Theorem 4 shows that the surplus function is differentiable if and only Condition $\mathrm{C}(\mathbf{x})$ holds, which, in turn, is equivalent to $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ reducing to a function. Condition $\mathrm{C}(\mathbf{x})$ constitutes a necessary and sufficient condition for the Hotz-Miller CCP function $\boldsymbol{\phi}(\cdot|\,\mathbf{x})$ in (5) to be well defined. The result of the classical WDZ theorem—that demand may be obtained by differentiation of the surplus function—is a discrete-choice analog of Roy's Identity. Evaluating $\mathbf{v} = \boldsymbol{\pi}(\mathbf{x})$, Theorem 4 extends the classical, differential WDZ theorem slightly by not requiring the existence of a Lebesgue density. That *absolute* continuity is not necessary for the conclusion of the differential WDZ theorem has been remarked by Norets and Takahashi (2013). The necessity of Condition $\mathrm{C}(\mathbf{x})$ appears to be new.

We next analyze the consequences of Condition $\mathrm{S}(\mathbf{x})$.

Theorem 5 (Differential Hotz-Miller)**.** *Fix $\mathbf{x} \in \mathcal{X}$. Then Condition S($\boldsymbol{x}$) is equivalent to each of the following:*

1. *$W(\cdot|\,\mathbf{x})$ is everywhere strictly convex.*
2. *$W^*(\cdot|\,\mathbf{x})$ is essentially smooth.*
3. *$\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$ is single-valued.*

In this case, $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$ reduces to the gradient mapping $\nabla W^*(\cdot|\,\mathbf{x})$, i.e. $\boldsymbol{\mathcal{P}}^{-1}(\mathbf{q}|\,\mathbf{x})$ consists of the vector $\nabla W^*(\mathbf{q}|\,\mathbf{x})$ alone for $\mathbf{q} \in \operatorname{int}\Delta$, while $\boldsymbol{\mathcal{P}}(\mathbf{q}|\,\mathbf{x}) = \emptyset$ for $\mathbf{q} \notin \operatorname{int}\Delta$. In particular, $\operatorname{dom}\partial W^*(\cdot|\,\mathbf{x}) = \operatorname{Im}\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x}) = \operatorname{int}\Delta$.

While the Hotz-Miller function $\boldsymbol{\phi}(\cdot|\,\mathbf{x})$ may not be well defined as a map from utilities to choice probabilities under Condition $\mathrm{S}(\mathbf{x})$, Theorem 5.3 shows that our support condition is both necessary and sufficient for the *inverse* of our generalized Hotz-Miller mapping $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x})$ to reduce to a function.

Theorems 4 and 5 are dual to each other. We have stated them as separate theorems to highlight the importance of Conditions $\mathrm{C}(\mathbf{x})$ and $\mathrm{S}(\mathbf{x})$, in turn. Combining the two theorems yields the following strong result.

Theorem 6 (Homeomorphic Demand)**.** *Fix* $\mathbf{x} \in \mathcal{X}$*. Then Conditions C(*$\boldsymbol{x}$*) and S(*$\boldsymbol{x}$*) are equivalent to each of the following:*

1. $W(\cdot|\,\mathbf{x})$ *is everywhere differentiable and strictly convex.*
2. $W^*(\cdot|\,\mathbf{x})$ *is essentially smooth and essentially strictly convex.*
3. $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ *is one-to-one.*

In this case, $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ *reduces to a one-to-one function from* $\mathbb{R}^J$ *onto* $\operatorname{int}\Delta$*, continuous in both directions, with* $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x}) = \{\nabla W(\cdot|\,\mathbf{x})\}$ *on* $\mathbb{R}^J$ *and* $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x}) = \{\nabla W^*(\cdot|\,\mathbf{x})\}$ *on* $\operatorname{int}\Delta$*.*

Theorem 6 has several implications. A straightforward consequence of Theorem 6 is the invertibility (or 'one-to-one'-ness) of our generalized Hotz-Miller mapping $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$, which is always well defined and reduces to the (usual) Hotz-Miller function $\boldsymbol{\phi}(\cdot|\,\mathbf{x})$ under Condition C($\mathbf{x}$). Theorem 6 therefore extends Hotz and Miller (1993, Proposition 1) to allow for possibly non-absolutely continuous stochastic utility distributions with less-than-full support.

As discussed in the Introduction, Norets and Takahashi (2013) point out the importance of surjectivity of the Hotz-Miller function $\boldsymbol{\phi}(\cdot|\,\mathbf{x})$ (when well defined) for estimation purposes. These authors show that the image of $\boldsymbol{\phi}(\cdot|\,\mathbf{x})$ coincides with the simplex interior $\operatorname{int}\Delta$ under Hotz and Miller's assumptions. In fact, under their assumptions, $\boldsymbol{\phi}(\cdot|\,\mathbf{x})$ is a *bijection* between $\mathbb{R}^J$ and $\operatorname{int}\Delta$ (Norets and Takahashi, 2013, Corollary 1). Both the surjectivity (or 'onto'-ness) $\operatorname{Im}\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x}) = \operatorname{int}\Delta$ and bijectivity ('one-to-one and onto'-ness) of our generalized Hotz-Miller mapping $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ may be deduced from Theorem 6. Theorem 6 therefore similarly generalizes the results of Norets and Takahashi (2013) to allow for possibly non-absolutely continuous stochastic utility distributions with less than full support.

In addition to generalizing these existing results, Theorem 6 reveals an added benefit to our convex analysis approach. Under Conditions C($\mathbf{x}$) and S($\mathbf{x}$), $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ reduces to a function between $\mathbb{R}^J$ and $\operatorname{int}\Delta$, which is continuous in *both* directions.[14] Hence, $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ reduces to not only a bijection, but in fact a *homeomorphism*. Consequently, under Conditions C($\mathbf{x}$) and S($\mathbf{x}$), the Hotz-Miller inverse function is not only *well defined*—it is also *well behaved*.

Both continuity conclusions follow from $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ and its inverse reducing to derivatives of convex functions, and the fact that a convex function differentiable on an open convex set is actually *continuously* differentiable (Rockafellar, 1970, Corollary 25.5.1).

The evaluation $\mathbf{v} = \boldsymbol{\pi}(\mathbf{x})$ now produces our second main result, which concerns *point* identification.

[14] A classical result shows that the inverse $\mathbf{f}^{-1}$ of a one-to-one, continuous mapping $\mathbf{f}$ from $X \subseteq \mathbb{R}^m$ onto $Y \subseteq \mathbb{R}^m$ is continuous if X is compact (Nikaido, 1968, Theorem 1.4). However, continuity of the inverse function need not hold in the absence of a compact domain [see Nikaido (1968, p. 9) for an example].

Theorem 7 (Point Identification)**.** *Fix* $\mathbf{x} \in \mathcal{X}$ *and let Conditions C(*$\boldsymbol{x}$*) and S(*$\boldsymbol{x}$*) hold. Then the utility values* $\boldsymbol{\pi}(\mathbf{x})$ *are point identified by*

$$\boldsymbol{\pi}(\mathbf{x}) = \nabla W^* \left(\mathbf{p}(\mathbf{x}) \middle| \mathbf{x} \right) = \underset{\mathbf{v} \in \mathbb{R}^J}{\operatorname{argmax}} \left\{ \langle \mathbf{v}, \mathbf{p}(\mathbf{x}) \rangle - W(\mathbf{v} | \mathbf{x}) \right\}. \tag{9}$$

Given that the differential WDZ theorem is the discrete-choice analog of Roy's identity, Theorem 7 may be interpreted as a *dual* version of Roy's identity. Being a special case of our constructive set-identification result (Theorem 3), our point-identification result is also constructive. Closely related duality results have been used for the purpose of identification in matching (Galichon and Salanie, 2015) and dynamic discrete choice (Chiong et al., 2016), who also treat the distribution(s) of unobserved heterogeneity as known.[15]

6 Conclusion

In this paper, we have obtained generalizations of a range of classical results for the additive random utility model of discrete choice, utilizing the power of convex analysis. Our generalization of the Williams-Daly-Zachary Theorem employs no assumptions on the structure of utility other than additivity, which means no further generalization is possible. Our generalized Hotz-Miller inverse always exists in the form of an inverse conditional choice probability (CCP) correspondence, mapping positive probability vectors to compact and convex sets of utilities. When the distribution of random utility components is known, this inverse CCP correspondence provides constructive partial identification of utilities. The identified set is the solution to a convex optimization problem, defined in terms of the surplus function. Estimators based on our constructive identification remain well defined even in the presence of zeros in empirical probabilities. Such estimators are therefore robust to the commonly encountered "zeros-in-demand" problem.

Without any restriction on the distribution of random utility components, utility maximization does not uniquely determine the choice probability. Without restriction we can similarly only have partial identification of utilities. In the paper, we have provided necessary and sufficient conditions for the choice probability to be unique as well as for point identification of utilities. Under these conditions, we show that the CCP correspondence reduces to a continuous function that is not only invertible and surjective—it also has a continuous inverse. The latter continuity property is desirable in that it may be used to establish consistency of estimators based directly on CCP inversion. In ongoing research, we utilize our CCP inversion results for the purpose of nonparametric estimation and inference.

[15]See also Allen and Rehbeck (2019, Lemma 3), who establish a dual version of "Roy's Identity" for an entire class of latent utility models with additively separable heterogeneity under an independence assumption.

Our analysis does not actually leverage the discreteness of choice. Consequently, all of our results carry over to a model in which the agent may choose among all elements of the probability simplex and not only its vertices. In such an ARUM of *stochastic choice* the agent is allowed to have weak preference for randomization. However, allowing the agent to have *strict* preference for randomization would lead to a choice model escaping the realm of convex analysis.

A Convex analysis

For easy reference, we here gather some key definitions and results from convex analysis.[16] A function $f : \mathbb{R}^m \to (-\infty, +\infty]$ is *convex* if

$$f\left(\left(1-\lambda\right)\mathbf{x}+\lambda\mathbf{y}\right) \leqslant \left(1-\lambda\right)f\left(\mathbf{x}\right)+\lambda f\left(\mathbf{y}\right), \quad 0<\lambda<1, \tag{A1}$$

for every $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^m$, and *strictly convex* if the inequality in (A1) is strict whenever $\mathbf{x}$ and $\mathbf{y}$ differ. Its *effective domain*, denoted $\mathrm{dom} f$, is the subset of $\mathbb{R}^m$ where f is finite,

$$\mathrm{dom} f := \left\{\mathbf{x}\in\mathbb{R}^m; f\left(\mathbf{x}\right) < +\infty\right\},$$

a convex set. Call a function $f : \mathbb{R}^m \to [-\infty, +\infty]$ *proper* provided that it is nowhere $-\infty$ and not identically $+\infty$. Call it *finite* if it is nowhere $\pm\infty$, i.e. if f actually takes values in $\mathbb{R}$.[17]

A vector $\mathbf{x}^*$ in $\mathbb{R}^n$ is called a *subgradient* of a convex function f at a point $\mathbf{x}$ if it satisfies the *subgradient inequality*

$$f\left(\mathbf{z}\right) \geqslant f\left(\mathbf{x}\right) + \langle \mathbf{x}^*, \mathbf{z}-\mathbf{x}\rangle \text{ for all } \mathbf{z}\in\mathbb{R}^m.$$

The collection of subgradients of f at $\mathbf{x}$ is called the *subdifferential of f at $\mathbf{x}$* and is denoted $\partial f\left(\mathbf{x}\right)$. The induced correspondence $\partial f : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is called the *subdifferential of f*. The set $\partial f\left(\mathbf{x}\right)$ may be empty or singleton, in general. If $\partial f\left(\mathbf{x}\right)$ is nonempty, f is said to be *subdifferentiable at $\mathbf{x}$*. The set of points where f is subdifferentiable,

$$\mathrm{dom}\,\partial f := \left\{\mathbf{x}\in\mathbb{R}^m; \partial f\left(\mathbf{x}\right) \neq \emptyset\right\},$$

is called the *domain of ∂f*. Unlike the effective domain of a convex function, the domain of

[16] We by and large adopt the terminology and notation of Rockafellar (1970), which we find to be an excellent reference on the topic.

[17] "Finite" is here synonymous with "real-valued." We avoid the latter term to avoid confusing real-valued and *extended* real-valued functions.

its subdifferential need not be convex.

A *finite* convex function is everywhere subdifferentiable, but not necessarily differentiable. If f is in fact differentiable at a point $\mathbf{x}$, then $\partial f(\mathbf{x})$ consists of the single vector given by the gradient $\nabla f(\mathbf{x})$ of f at $\mathbf{x}$. Conversely, if $\partial f(\mathbf{x})$ is the singleton $\{\mathbf{x}^*\}$, then f is differentiable at $\mathbf{x}$ with $\nabla f(\mathbf{x}) = \mathbf{x}^*$. The notion of subdifferentiability therefore extends the notion of differentiability for convex functions.

Subdifferentiability is tightly connected to the notion of cyclical monotonicity, which we now define.[18]

Definition A1 (Cyclical Monotonicity)**.** *A correspondence $\boldsymbol{\rho} : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is called* ***cyclically monotone*** *if*

$$\langle \mathbf{x}_0^*, \mathbf{x}_0 - \mathbf{x}_1 \rangle + \langle \mathbf{x}_1^*, \mathbf{x}_1 - \mathbf{x}_2 \rangle + \cdots + \langle \mathbf{x}_n^*, \mathbf{x}_n - \mathbf{x}_0 \rangle \geqslant 0 \tag{A2}$$

for any integer $n \geqslant 1$ and any set $\{(\mathbf{x}_i, \mathbf{x}_i^*)\}_0^n$ of pairs such that $\mathbf{x}_i^* \in \boldsymbol{\rho}(\mathbf{x}_i)$.

The concept of cyclical monotonicity generalizes the notion of monotonicity for univariate functions to possibly multivariate and/or multivalued mappings. To establish this connection, recall that a correspondence $\boldsymbol{\rho} : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is called *single-valued* if $\partial \boldsymbol{\rho}(\mathbf{x})$ contains at most one element for all $\mathbf{x}$. A single-valued correspondence $\boldsymbol{\rho}$ may be identified with a function $\mathbf{f} : D \to \mathbb{R}^m$ with domain $D := \operatorname{dom} \boldsymbol{\rho} = \{\mathbf{x} \in \mathbb{R}^m ; \boldsymbol{\rho}(\mathbf{x}) \neq \emptyset\}$. Cyclical monotonicity of $\boldsymbol{\rho}$ then boils down to the requirement that

$$\langle \mathbf{f}(\mathbf{x}_0), \mathbf{x}_0 - \mathbf{x}_1 \rangle + \langle \mathbf{f}(\mathbf{x}_1), \mathbf{x}_1 - \mathbf{x}_2 \rangle + \cdots + \langle \mathbf{f}(\mathbf{x}_n), \mathbf{x}_n - \mathbf{x}_0 \rangle \geqslant 0 \tag{A3}$$

for any integer $n \geqslant 1$ and any set $\{\mathbf{x}_i\}_0^n$ of points in D. In accordance with the definition employed by Shi et al. (2018), we therefore take the previous display to be the *definition* of cyclical monotonicity for functions.[19]

We will make use of the following result linking cyclic monotonicity and convexity.

Proposition A1 (Cyclical Monotonicity of Subdifferentials)**.** *Let f be proper convex. Then ∂f is cyclically monotone.*

The proposition is a consequence of the subgradient inequality Rockafellar (see, e.g., 1970, p. 238). Any univariate differentiable convex function must admit a monotone increasing

[18]Given the direction of the inequality in (A2), a map satisfying Definition A1 ought to be referred to as cyclically monotone *increasing*. As in Rockafellar (1970), we use "cyclically monotone" for the sake of brevity.

[19]To see that *cyclical* monotonicity implies *ordinary* monotonicity for univariate finite functions, let $m = 1, n = 1$ and let x_0, x_1 be any two points in $D \subseteq \mathbb{R}$. Then (A3) rearranges to $[f(x_1) - f(x_0)](x_1 - x_0) \geqslant 0$, so $x_1 > x_0$ implies $f(x_1) \geqslant f(x_0)$.

derivative. As shown in Proposition A1, cyclical monotonicity extends the notion of a monotone increasing derivative to multivariate convex functions not necessarily differentiable.

Given a convex function $f : \mathbb{R}^m \to (-\infty, +\infty]$, we define its *convex* (or *Fenchel*) *conjugate* f^* by

$$f^*(\mathbf{x}^*) := \sup_{\mathbf{x}\in\mathbb{R}^m} \{\langle \mathbf{x}, \mathbf{x}^* \rangle - f(\mathbf{x})\}, \quad \mathbf{x}^* \in \mathbb{R}^m.$$

The convex conjugate f^* is proper if and only if f itself is proper. When the latter holds, f^* may thus be viewed as a (convex) function from $\mathbb{R}^m$ to $(-\infty, +\infty]$.[20] The subdifferential $\partial f^* : \mathbb{R}^m \to \mathbb{R}^m$ of the convex conjugate f^* of a proper convex function f admits the variational characterization

$$\partial f^*(\mathbf{x}^*) = \operatorname*{argmax}_{\mathbf{x}\in\mathbb{R}^m} \{\langle \mathbf{x}, \mathbf{x}^* \rangle - f(\mathbf{x})\}. \tag{A4}$$

The inverse of a correspondence $\boldsymbol{\rho} : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is the correspondence $\boldsymbol{\rho}^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ defined by

$$\boldsymbol{\rho}^{-1}(\mathbf{y}) := \{\mathbf{x} \in \mathbb{R}^m; \mathbf{y} \in \boldsymbol{\rho}(\mathbf{x})\}.$$

Subdifferentials of proper conjugate pairs are inverse to each other in the correspondence sense, i.e. $\mathbf{x} \in \partial f^*(\mathbf{x}^*)$ if and only if $\mathbf{x}^* \in \partial f(\mathbf{x})$.

B Proofs

Proof of Theorem 1. (1) As argued in the main text, the fact that the vector-max function is 1-Lipschitz with respect to $\|\cdot\|_\infty$ establishes both existence and finiteness of the surplus function. The surplus inherits the Lipschitz property of the vector-max function by the conditional version of Jensen's inequality. For fixed $\omega \in \Omega$, the mapping

$$f(\cdot; \boldsymbol{\varepsilon}(\omega)) := \max_{j\in\mathcal{J}} \langle \cdot + \boldsymbol{\varepsilon}(\omega), \mathbf{u}_j \rangle - \max_{j\in\mathcal{J}} \langle \boldsymbol{\varepsilon}(\omega), \mathbf{u}_j \rangle \tag{B1}$$

is an affine transformation of a pointwise maximum of affine (hence convex) functions and is therefore convex (Rockafellar, 1970, Theorem 5.5). Integration preserves inequalities, so convexity of the surplus follows.

(2) A finite convex function is everywhere subdifferentiable (Rockafellar, 1970, Theorem 23.4).

(3) The subdifferential of the vector-max function $[= f(\cdot; \mathbf{0})]$ is the convex hull of the directions $\mathbf{u}_j$ corresponding to coordinates achieving the maximum. Since $f(\mathbf{v}; \cdot)$ is integrable

[20] It is possible to give a geometric definition of convexity which allows for $-\infty$-valued functions [see Rockafellar (1970)]. Given that values of $-\infty$ can be rather easily ruled out in our framework, we adhere to the less general but more familiar definition of a convex function given in (A1).

with respect to $F_{\boldsymbol{\varepsilon}|\mathbf{X}}(\cdot|\,\mathbf{x})$ for any $\mathbf{v} \in \mathbb{R}^J$, the order of subdifferentiation and integration (expectation) may be interchanged (Bertsekas, 1973, Proposition 2.2), so

$$\partial W(\mathbf{v}|\,\mathbf{x}) = \mathrm{E}\Big[\partial f(\mathbf{v};\boldsymbol{\varepsilon})\,\Big|\mathbf{X}=\mathbf{x}\Big] = \mathrm{E}\Big[\mathrm{conv}\left(\{\mathbf{u}_j\}_{\,j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})}\right)\Big|\mathbf{X}=\mathbf{x}\Big] = \boldsymbol{\mathcal{P}}(\mathbf{v}|\,\mathbf{x}).$$

(4) By (1), $W(\cdot|\,\mathbf{x})$ is finite convex, hence proper convex. Cyclical monotonicity of $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ now follows from (3) and Proposition A1. ■

Proof of Theorem 2. (1) Lower semi-continuity, properness and convexity of $W^*(\cdot|\,\mathbf{x})$ follow from finiteness (hence properness) and convexity of $W(\cdot|\,\mathbf{x})$ (Rockafellar, 1970, Theorem 12.2). A proper convex function is continuous on any open subset of its effective domain. It therefore only remains to show the claimed effective domain containments. Recall that the *support function* σ_C of a convex set C is defined by $\sigma_C(\mathbf{x}) := \sup_{\mathbf{x}^*\in C}\langle \mathbf{x}, \mathbf{x}^*\rangle$. To show the claimed containments, it suffices to show that $\mathrm{cl}\,(\mathrm{dom}\,W^*(\cdot|\,\mathbf{x})) = \Delta$ (Rockafellar, 1970, Corollary 6.3.1), which, in turn, is equivalent to $\mathrm{dom}\,W^*(\cdot|\,\mathbf{x})$ and Δ having the same support function (Rockafellar, 1970, Theorem 13.1). Rockafellar (1970, Theorem 13.3) and Theorem 1 combine to show that the support function of $W^*(\cdot|\,\mathbf{x})$ is the *recession* (or *horizon*) function of $W(\cdot|\,\mathbf{x})$ [see Rockafellar (1970, Chapter 8) and Rockafellar and Wets (2009, Chapter 3)]. For any $\mathbf{h} \in \mathbb{R}^J$, it therefore follows from Rockafellar (1970, Corollary 8.5.2) and Theorem 1, that we may express the support function $\sigma_{\mathrm{dom}\,W^*(\cdot|\mathbf{x})}$ of $\mathrm{dom}\,W^*(\cdot|\,\mathbf{x})$ as the limit

$$\sigma_{\mathrm{dom}\,W^*(\cdot|\mathbf{x})}(\mathbf{h}) = \lim_{\tau\to\infty}\frac{W(\tau\mathbf{h}|\,\mathbf{x})}{\tau}, \quad \mathbf{h}\in\mathbb{R}^J.$$

Write the surplus function in terms of the support function of the simplex,

$$\begin{aligned} W(\mathbf{v}|\,\mathbf{x}) &= \mathrm{E}\Big[\max_{j\in\mathcal{J}}\langle \mathbf{v}+\boldsymbol{\varepsilon}, \mathbf{u}_j\rangle - \max_{j\in\mathcal{J}}\langle \boldsymbol{\varepsilon}, \mathbf{u}_j\rangle\Big|\mathbf{X}=\mathbf{x}\Big] \\ &= \mathrm{E}\left[\sigma_\Delta(\mathbf{v}+\boldsymbol{\varepsilon}) - \sigma_\Delta(\boldsymbol{\varepsilon})\,|\mathbf{X}=\mathbf{x}\right], \end{aligned}$$

fix $\mathbf{h}\in\mathbb{R}^J$, and let $\{\tau_m\}_1^\infty \subset \mathbb{R}_{++}$ be any monotone increasing sequence satisfying $\tau_n \to \infty$. By convexity and continuity of the support function σ_Δ, the sequence $\{f_m\}_1^\infty$ of functions $f_m : \mathbb{R}^J \to \mathbb{R}$ defined by

$$f_m(\mathbf{t}) := \frac{1}{\tau_m}\left[\sigma_\Delta(\tau_m\mathbf{h}+\mathbf{t}) - \sigma_\Delta(\mathbf{t})\right]$$

is monotone increasing ($f_1 \leqslant f_2 \leqslant \cdots$) with pointwise limit given by

$$\lim_{m\to\infty} f_m(\mathbf{t}) = \lim_{m\to\infty}\frac{1}{\tau_m}\sigma_\Delta(\tau_m\mathbf{h}+\mathbf{t}) = \sigma_\Delta(\mathbf{h}),$$

which does not depend on $\mathbf{t}$. It now follows from a monotone convergence theorem argument that

$$\sigma_{\operatorname{dom} W^*(\cdot|\mathbf{x})}(\mathbf{h}) = \lim_{m\to\infty} \mathrm{E}\left[f_m(\boldsymbol{\varepsilon})|\mathbf{X}=\mathbf{x}\right] = \mathrm{E}\left[\sigma_\Delta(\mathbf{h})|\mathbf{X}=\mathbf{x}\right] = \sigma_\Delta(\mathbf{h}).$$

(2) $\partial W^*(\cdot|\mathbf{x}) = (\partial W)^{-1}(\cdot|\mathbf{x})$ follows from finiteness (hence properness) of W (Theorem 1.1) and subdifferentials of conjugate pairs being inverse to each other in the correspondence sense (Rockafellar, 1970, Corollary 23.5.1). The claim that $\partial W^*(\cdot|\mathbf{x}) = \boldsymbol{\mathcal{P}}^{-1}(\cdot|\mathbf{x})$ therefore follows from Theorem 1.3.

(3) Cyclical monotonicity of $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\mathbf{x})$ follows from (2) by Proposition A1 in Appendix A.

(4) Both $\operatorname{dom}\partial W^*(\cdot|\mathbf{x}) = \operatorname{Im}\boldsymbol{\mathcal{P}}(\cdot|\mathbf{x})$ and $\operatorname{Im}\partial W^*(\cdot|\mathbf{x}) = \operatorname{dom}\boldsymbol{\mathcal{P}}(\cdot|\mathbf{x})$ follow from (2). The claimed containments $\operatorname{int}\Delta \subseteq \operatorname{dom}\partial W^*(\cdot|\mathbf{x}) = \operatorname{Im}\boldsymbol{\mathcal{P}}(\cdot|\mathbf{x}) \subseteq \Delta$ then follow from (1) and the fact that a proper convex function is subdifferentiable nowhere outside of its effective domain and everywhere on its (relative) interior (Rockafellar, 1970, Theorem 23.4). That $\operatorname{Im}\partial W^*(\cdot|\mathbf{x}) = \operatorname{dom}\boldsymbol{\mathcal{P}}(\cdot|\mathbf{x}) = \mathbb{R}^J$ is then a consequence of Theorem 1.2-1.3. ■

Proof of Theorem 3. Equation (6) is immediate from Theorem 1.3, Theorem 2.2 and the variational characterization of subdifferentials of conjugate pairs in (A4). That $\Pi(\mathbf{x})$ is a nonempty bounded set if and only if $\mathbf{p}(\mathbf{x}) \in \operatorname{int}\Delta$ follows from (Rockafellar, 1970, Theorem 23.4) and $\operatorname{int}\Delta = \operatorname{int}(\operatorname{dom} W^*(\cdot|\mathbf{x}))$ (Theorem 2.1). ■

Proof of Theorem 4. We suppress the "conditioning on $\mathbf{X}=\mathbf{x}$" throughout in order to ease notation. To show Condition $\mathrm{C}(\mathbf{x}) \Leftrightarrow$ (1), fix $\mathbf{v},\mathbf{h} \in \mathbb{R}^J$. We then want to show that W is directionally differentiable at $\mathbf{v}$ in the direction $\mathbf{h}$ if and only if (7) holds. Bertsekas (1973, Proposition 2.1) shows that $W'(\mathbf{v};\mathbf{h}) = \mathrm{E}\left[f'(\mathbf{v};\boldsymbol{\varepsilon};\mathbf{h})\right] \in \mathbb{R}$ where $f'(\mathbf{v};\boldsymbol{\varepsilon}(\omega);\mathbf{h})$ denotes the one-sided directional derivative of $f(\cdot;\boldsymbol{\varepsilon}(\omega))$ [see (B1)] at $\mathbf{v}$ in the direction $\mathbf{h}$. This one-sided directional derivative is two-sided if and only if the nonnegative $W'(\mathbf{v};\mathbf{h}) + W'(\mathbf{v};-\mathbf{h})$ is in fact zero. That is, if and only if,

$$W'(\mathbf{v};\mathbf{h}) + W'(\mathbf{v};-\mathbf{h}) = \mathrm{E}\left[f'(\mathbf{v};\boldsymbol{\varepsilon};\mathbf{h}) + f'(\mathbf{v};\boldsymbol{\varepsilon};-\mathbf{h})\right] = 0.$$

The integrand is itself nonnegative, so the previous display holds if and only if

$$\mathbf{P}\left(f'(\mathbf{v};\boldsymbol{\varepsilon};\mathbf{h}) + f'(\mathbf{v};\boldsymbol{\varepsilon};-\mathbf{h}) > 0\right) = 0.$$

A finite convex function is everywhere subdifferentiable with a compact-valued subdifferential (Rockafellar, 1970, Theorem 23.4). As established in the proof of Theorem 1, the finite convex function $f(\cdot;\boldsymbol{\varepsilon}(\omega))$ has subdifferential at $\mathbf{v}$ given by the nonempty compact set

$$\partial f(\mathbf{v};\boldsymbol{\varepsilon}(\omega)) = \operatorname{conv}\left(\{\mathbf{u}_j\}_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon}(\omega))}\right).$$

Its one-sided directional derivative at $\mathbf{v}$ in the direction $\mathbf{h}$ is therefore given by

$$f'(\mathbf{v};\boldsymbol{\varepsilon}(\omega);\mathbf{h}) = \sup_{\mathbf{q}\in\partial f(\mathbf{v};\boldsymbol{\varepsilon}(\omega))} \langle \mathbf{h},\mathbf{q}\rangle = \max_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon}(\omega))} \langle \mathbf{h},\mathbf{u}_j\rangle \tag{B2}$$

(Rockafellar, 1970, Theorem 23.4). The claim therefore follows from

$$f'(\mathbf{v};\boldsymbol{\varepsilon}(\omega);\mathbf{h}) + f'(\mathbf{v};\boldsymbol{\varepsilon}(\omega);-\mathbf{h}) = \max_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon}(\omega))} \langle \mathbf{h},\mathbf{u}_j\rangle - \min_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon}(\omega))} \langle \mathbf{h},\mathbf{u}_j\rangle\,.$$

(1) $\Leftrightarrow$ (2) follows from essential smoothness and essential strict convexity being dual properties (Rockafellar, 1970, Theorem 26.3) and $\operatorname{dom} W = \mathbb{R}^J$ (Theorem 1.1), such that essential smoothness reduces to differentiability ("ordinary smoothness").

(1) $\Leftrightarrow$ (3) follows from Theorem 1.3.

The final statement follows from Rockafellar (1970, Theorem 26.1) and Theorem 1.2-1.3.

■

Proof of Theorem 5. We again suppress the "conditioning on $\mathbf{X} = \mathbf{x}$" throughout in order to ease notation. A finite convex function f is everywhere subdifferentiable with a compact-valued subdifferential (Rockafellar, 1970, Theorem 23.4). It is *strictly* convex if and only

$$f(\mathbf{x}+\mathbf{h}) > f(\mathbf{x}) + \langle \mathbf{x}^*,\mathbf{h}\rangle$$

for all $\mathbf{x},\mathbf{h}\in\mathbb{R}^m, \mathbf{h}\neq\mathbf{0}$, and all $\mathbf{x}^*\in\partial f(\mathbf{x})$, i.e. if and only the graph of f lies strictly above every tangent line (except for the point of tangency). Since each $\partial f(\mathbf{x})$ is compact, strict convexity may equivalently be expressed as

$$f(\mathbf{x}+\mathbf{h}) > f(\mathbf{x}) + \max_{\mathbf{x}^*\in\partial f(\mathbf{x})} \langle \mathbf{x}^*,\mathbf{h}\rangle = f(\mathbf{x}) + f'(\mathbf{x};\mathbf{h})$$

for all $\mathbf{x},\mathbf{h}\in\mathbb{R}^J, \mathbf{h}\neq\mathbf{0}$, where we have again used Rockafellar (1970, Theorem 23.4). To show Condition S($\mathbf{x}$) $\Leftrightarrow$ (1), fix $\mathbf{v},\mathbf{h}\in\mathbb{R}^J, \mathbf{h}\neq\mathbf{0}$. The surplus W is finite convex (Theorem 1.1). We therefore want to that $W(\mathbf{v}+\mathbf{h}) > W(\mathbf{v}) + W'(\mathbf{v};\mathbf{h})$ is equivalent to (8). Bertsekas (1973, Proposition 2.1) shows that the order of one-sided directional differentiation and integration (expectation) may be interchanged. As established in the proof of Theorem 4, the finite convex function $f(\cdot;\boldsymbol{\varepsilon}(\omega))$ [see (B1)] has one-sided directional derivative at $\mathbf{v}$ in the direction $\mathbf{h}$ given by

$$f'(\mathbf{v},\boldsymbol{\varepsilon}(\omega);\mathbf{h}) = \max_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon}(\omega))} \langle \mathbf{h},\mathbf{u}_j\rangle$$

[cf. (B2)]. We may therefore express $W'(\mathbf{v};\mathbf{h})$ as

$$W'(\mathbf{v};\mathbf{h}) = \mathrm{E}\Big[\max_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})} \langle \mathbf{h}, \mathbf{u}_j\rangle\Big].$$

Combine the previous observations to see that

$$\begin{aligned} &W(\mathbf{v}+\mathbf{h}) - W(\mathbf{v}) - W'(\mathbf{v};\mathbf{h}) \\ &= \mathrm{E}\Big[\max_{j\in\mathcal{J}} \langle \mathbf{v}+\mathbf{h}+\boldsymbol{\varepsilon}, \mathbf{u}_j\rangle - \max_{j\in\mathcal{J}} \langle \mathbf{v}+\boldsymbol{\varepsilon}, \mathbf{u}_j\rangle - \max_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})} \langle \mathbf{h}, \mathbf{u}_j\rangle\Big]. \end{aligned}$$

The integrand is itself nonnegative, so $W(\mathbf{v}+\mathbf{h}) > W(\mathbf{v}) + W'(\mathbf{v};\mathbf{h})$ if and only if

$$\mathbf{P}\Big(\max_{j\in\mathcal{J}} \langle \mathbf{v}+\mathbf{h}+\boldsymbol{\varepsilon}, \mathbf{u}_j\rangle > \max_{j\in\mathcal{J}} \langle \mathbf{v}+\boldsymbol{\varepsilon}, \mathbf{u}_j\rangle + \max_{j\in\mathcal{J}(\mathbf{v}+\boldsymbol{\varepsilon})} \langle \mathbf{h}, \mathbf{u}_j\rangle\Big) > 0.$$

(1) $\Leftrightarrow$ (2) follows from essential smoothness and essential strict convexity being dual to each other (Rockafellar, 1970, Theorem 26.3) and $\operatorname{dom} W = \mathbb{R}^J$ (Theorem 1.1), such that essential strict convexity reduces to strict convexity on $\mathbb{R}^J$ ("ordinary strict convexity").

(2) $\Leftrightarrow$ (3) follows from Rockafellar (1970, Theorem 26.1) and $\partial W^* = \boldsymbol{\mathcal{P}}^{-1}$ (Theorem 2.2).

The final statement also follows from Rockafellar (1970, Theorem 26.1). ■

Proof of Theorem 6. The claimed equivalences follow from combining Theorems 4–5. The final statement then follows from Rockafellar (1970, Theorem 26.5). ■

Proof of Theorem 7. By construction, $\mathbf{p}(\mathbf{x}) \in \boldsymbol{\mathcal{P}}(\boldsymbol{\pi}(\mathbf{x})|\,\mathbf{x})$. Under the stated assumptions, $\boldsymbol{\mathcal{P}}(\cdot|\,\mathbf{x})$ reduces to a one-to-one function from $\mathbb{R}^J$ onto $\operatorname{int}\Delta$ given by $\boldsymbol{\mathcal{P}}^{-1}(\cdot|\,\mathbf{x}) = \{\nabla W^*(\cdot|\,\mathbf{x})\}$ on $\operatorname{int}\Delta$ (Theorem 6), so $\mathbf{p}(\mathbf{x}) \in \operatorname{int}\Delta$ and $\boldsymbol{\pi}(\mathbf{x}) = \nabla W^*(\mathbf{p}(\mathbf{x})|\,\mathbf{x})$. The last equality now follows from the variational characterization of subdifferentials of conjugate pairs in (A4). ■

C On the Domain of the Conjugate Surplus

In this appendix we illustrate by example that the inclusions of Theorem 2.1 cannot be improved in general. To see that the inclusion $\operatorname{dom} W^*(\cdot|\,\mathbf{x}) \subseteq \Delta$ may be strict, let $J = 1$ (binary choice), and let ε_1 be distributed independently of $\mathbf{X}$ according to a (for simplicity) continuous CDF F. Then W does not depend on $\mathbf{x}$ and may be calculated to be

$$W(v) = \begin{cases} v[1-F(-v)] + \int_{-v}^{0} t\mathrm{d}F(t), & v \geqslant 0, \\ v[1-F(-v)] - \int_{0}^{-v} t\mathrm{d}F(t) & v < 0. \end{cases}$$

It follows that

$$W^*(1) = \sup_{v\in\mathbb{R}} \{v \cdot 1 - W(v)\} \geqslant \sup_{v\geqslant 0} \left\{ vF(-v) - \int_{-v}^{0} t\mathrm{d}F(t) \right\}$$
$$\geqslant \sup_{v\geqslant 0} \left\{ \int_0^v t\mathrm{d}F(t) \right\} = \int_0^\infty t\mathrm{d}F(t).$$

Similarly, $W^*(0) \geqslant \int_0^\infty t\mathrm{d}F(t)$. Hence, if ε_1 isn't integrable, we have $W^*(0) = W^*(1) = +\infty$, and therefore $\Delta = [0,1] \subsetneq \operatorname{dom} W^*$. For example, if $\varepsilon_1 \sim \text{Cauchy}(0,1)$, a straightforward calculation shows that $\operatorname{dom} W^* = (0,1) = \operatorname{int} \Delta$.

References

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

Allen, R. and J. Rehbeck (2019): "Identification with additively separable heterogeneity," *Econometrica*, 87, 1021–1054.

Anderson, S. P., A. De Palma, and J.-F. Thisse (1988): "A Representative Consumer Theory of the Logit Model," *International Economic Review*, 29, 461–466.

Anderson, S. P., A. De Palma, and J.-F. Thisse (1992): *Discrete choice theory of product differentiation*, MIT press.

Berry, S. (1994): "Estimating discrete-choice models of market equilibrium," *The RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (2004): "Differentiated products demand systems from a combination of micro and macro data: The new car market," *Journal of political Economy*, 112, 68–105.

Berry, S. T., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Bertsekas, D. P. (1973): "Stochastic optimization problems with nondifferentiable cost functionals," *Journal of Optimization Theory and Applications*, 12, 218–231.

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

Daly, A. and S. Zachary (1978): "Improved multiple choice models," *Determinants of travel choice*, 335, 357.

Fosgerau, M. and D. Kristensen (2019): "Identification of a class of index models: A topological approach," *CEMMAP working paper.*

Fosgerau, M., E. Melo, A. de Palma, and M. Shum (2018): "Discrete Choice and Rational Inattention: A General Equivalence Result," *SSRN Electronic Journal.*

Galichon, A. and B. Salanie (2015): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *SSRN Electronic Journal.*

Gandhi, A., Z. Lu, and X. Shi (2019): "Estimating demand for differentiated products with zeroes in market share data," *Working Paper, Department of Economics, University of Wisconsin-Madison.*

Hotz, V. J. and R. A. Miller (1993): "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 60, 497–529.

Koning, R. H. and G. Ridder (2003): "Discrete choice and stochastic utility maximization," *The Econometrics Journal*, 6, 1–27.

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

——— (1993): "Nonparametric identification and estimation of polychotomous choice models," *Journal of Econometrics*, 58, 137–168.

——— (1994): "Restrictions of economic theory in nonparametric methods," *Handbook of econometrics*, 4, 2523–2558.

McFadden, D. (1978): "Modeling the choice of residential location," *Transportation Research Record.*

——— (1981): "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden, Cambridge, MA, USA: MIT Press, 198–272.

Nikaido, H. (1968): *Convex structures and economic theory*, Academic Press.

Norets, A. and S. Takahashi (2013): "On the surjectivity of the mapping between utilities and choice probabilities," *Quantitative Economics*, 4, 149–155.

Nurski, L. and F. Verboven (2016): "Exclusive dealing as a barrier to entry? Evidence from automobiles," *The Review of Economic Studies*, 83, 1156–1188, publisher: Wiley-Blackwell.

Quan, T. W. and K. R. Williams (2018): "Product variety, across-market demand heterogeneity, and the value of online retail," *The RAND Journal of Economics*, 49, 877–913, publisher: Wiley Online Library.

Rockafellar, R. T. (1970): "Convex Analysis (Princeton Mathematical Series)," *Princeton University Press*, 46, 49.

Rockafellar, R. T. and R. J.-B. Wets (2009): *Variational analysis*, vol. 317, Springer Science & Business Media.

Rust, J. (1994): "Structural estimation of Markov decision processes," *Handbook of econometrics*, 4, 3081–3143.

Shi, X., M. Shum, and W. Song (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 86, 737–761.

Williams, H. (1977): "On the formation of travel demand models and economic evaluation measures of user benefit," *Environment and planning A*, 9, 285–344.