Nonseparable Costly Attention and Revealed Preference

Christopher P. Chambers, Ce Liu, and John Rehbeck

cpchambers@ucsd.edu cel013@ucsd.edu jrehbeck@ucsd.edu

Department of Economics

University of California-San Diego

Last Updated: April 12, 2017

Abstract

We provide revealed preference characterizations for choices made under costly information acquisition. We provide non-separable generalizations and a special case of Caplin and Dean (2015). Our techniques parallel those involved in the duality characterization of utility maximization.

1 Introduction

The recent contribution of Caplin and Dean (2015) provides a revealed preference test for optimal costly information acquisition. Caplin and Dean (2015) consider a decision maker facing actions with state-contingent payoffs. In this environment, the decision maker chooses an information structure and then makes stochastic choices conditioned on the signal received from the information structure. If a researcher is able to observe a decision maker's stochastic choices conditioned on the true state of the world, then there is a natural revealed information structure from the decision maker's stochastic choices. Caplin and Dean (2015) establish that an acyclicity condition on the revealed information structure and an optimality condition on stochastic choices are necessary and sufficient for stochastic choice data to be consistent with the hypothesis of optimal costly information acquisition. The cost of information in Caplin and Dean (2015) is additively separable in the decision maker's utility function.

Our work generalizes that of Caplin and Dean (2015) by allowing for preferences with nonseparable costs of information. The only assumption we impose on preferences is that they are increasing in the gross expected utility derived from stochastic choices. As a special case, we characterize representations with multiplicative costs of information. Multiplicative cost is relevant, for example, when the information costs a fixed share of the decision maker's gross payoff.

Our primary characterization generalizes the acyclicity condition of Caplin and Dean (2015) in the same way that the generalized axiom of revealed preference (see Houthakker (1950); Richter (1966); Chambers and Echenique (2016)) generalizes the cyclic monotonicity condition of Rockafellar (1966) or the condition of Koopmans and Beckmann (1957).[1] Of note is

[1] See also Brown and Calsamiglia (2007) and Chambers, Echenique and Saito (2016)

that the generalized axiom of revealed preference is *ordinal*.

Our proofs exploit a duality between direct and indirect utility which has recently been fruitfully studied by Chateauneuf and Faro (2009), Cerreia-Vioglio, Maccheroni, Marinacci and Montrucchio (2011a), and Cerreia-Vioglio, Maccheroni, Marinacci and Montrucchio (2011b). Roughly, *any* quasiconcave and monotone function can be written as a "dual" of direct utility. The indirect utility function plays the role of the preferences with a nonseparable costly information representation in our framework.

Let us now explain our results in more detail. The contribution of Caplin and Dean (2015) is, roughly, to characterize a decision maker by their expected utility from experimentation less a cost of the experiment. An experiment, or signal, is a probability distribution over posteriors (as in Blackwell (1953)). Mathematically, probability measures and normalized price vectors are the same object. In this sense, if we treat the expected utility of experimentation as *wealth* and the probability distribution over posteriors as *prices*, we are dealing with an object that closely resembles an *indirect utility function*. This resemblance can in fact be made formal. The indirect utility that is considered in Caplin and Dean (2015) is one for which there are no "wealth effects."

Using the indirect utility, we can equivalently represent such a decision maker by their "direct utility" function. When there are no wealth effects, there are well-known revealed preference tests for rationalizability. These are based on the cyclic monotonicity condition of Rockafellar (1966). We go a step further, and observe that a *general* indirect utility function (that is, one in which wealth effects can be present) has an direct utility representation. In this case, the test for rationalizability by a utility function is well-known: the generalized axiom of revealed preference.

The only distinction between what we are doing here and the classical

---

for variants of this condition in an explicit revealed preference framework.

theory is that here, the direct utility is *increasing* and *convex*, whereas the indirect utility is quasiconcave. This owes to the fact that the direct utility here is derived from the indirect utility via maximization (whereas in the standard case, it results from minimization).

With this approach, results which are otherwise nontrivial to obtain fall directly out of the representation. For example, it is well-known that an indirect utility function has a quasiconcave representation; this is almost by definition. So, the fact that our nonlinear aggregator can be taken to be quasiconcave is without loss. Similarly, by taking this approach, it is relatively simple to show that the "cost function" in the Caplin and Dean (2015) setup can be taken to be convex without loss.

As we see it, the main objective of this note is to establish simple characterizations of costly information-acquisition models by leveraging existing revealed preference and duality techniques. As such, the note should also serve as a didactic exercise.

The paper proceeds as follows. Section 2 lays down the model and notation. Section 3 introduces and characterizes the nonseparable generalization of Caplin and Dean (2015). Section 4 presents a variant of the model whereby choice of information structure is costless, but is restricted to lie in some unknown set. Section 5 presents a multiplicative version of the model. Finally, section 6 concludes. Proofs are relegated to an appendix.

2 Preliminaries and Revealed Information Structures

2.1 Notation

We study a decision maker facing actions with state-contingent payoffs.[2] Notation is consistent with Caplin and Dean (2015) whenever possible for ease of comparison. We study a variety of models that are increasing in gross expected utility and satisfy Bayes' law. A decision maker chooses actions whose outcome depends on a finite number of states of the world. Let Ω denote a finite set of states. Let X denote a set of outcomes. Therefore, the set of all actions (state-contingent outcomes) is X^{Ω}.

The set of all finite decision problems is given by $\mathcal{A} = \{A \subset X^{\Omega} \mid |A| < \infty\}$. As in Caplin and Dean (2015) we investigate when a researcher has a state dependent stochastic choice dataset from decision problems in $\mathcal{A}$. For $A \in \mathcal{A}$, $\Delta(A)$ refers to the set of probability distributions over actions in A.

Definition 1 A *state dependent stochastic choice dataset* is a finite collection of decision problems $\mathcal{D} \subset \mathcal{A}$ and related set of state dependent stochastic choice functions $\mathcal{P} = \{P_A\}_{A \in \mathcal{D}}$ where $P_A : \Omega \to \Delta(A)$. Denote the probability of choosing an action a conditional on state ω in decision problem A as $P_A(a \mid \omega)$.

We assume that the prior beliefs of the decision maker $\mu \in \Gamma = \Delta(\Omega)$ are known. Moreover, we assume that the utility index $u : X \to \Re$ is a known function.

We take an abstract approach to modeling the choice of the information structure. Each subjective signal is identified with it's associated posterior beliefs $\gamma \in \Gamma$. Thus, an information structure is given by a finite support

[2]The ideas discussed here are broader if one considers general mappings over posteriors.

distribution over Γ that satisfies Bayes' law.

Definition 2 The set of *information structures*, Π, comprises all Borel probability distributions over Γ, $\pi \in \Delta(\Gamma)$, that have finite support and satisfy Bayes' law. A distribution over posteriors satisfies Bayes' law if the distribution over posteriors is a mean-preserving spread of the prior μ denoted as $E_\pi[\gamma] = \mu$.

We now make some definitions required to define gross expected utility. Given a utility index, each decision problem $A \in \mathcal{A}$ induces a posterior value function, $f_A : \Gamma \to \Re$, which maps posterior beliefs γ to the maximal utility from A under posterior γ. Formally, for any decision problem A and posterior belief γ

$$f_A(\gamma) = \max_{a \in A} \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)).$$

Definition 3 We denote the *gross expected utility* induced by an information structure $\pi \in \Pi$ as

$$\pi \cdot f_A = \sum_{\gamma \in \text{Supp}(\pi)} \pi(\gamma) f_A(\gamma)$$

where $\pi(\gamma) = \Pr(\gamma \mid \pi) = \sum_{\omega \in \Omega} \mu(\omega)\pi(\gamma \mid \omega)$.

This representation of gross expected utility is intuitive since f_A is a continuous function on Γ and the set of continuous functions on Γ is topologically dual to the set of countably additive Borel measures on Γ (Aliprantis and Border (2006), Theorem 14.15).

2.2 Revealed Information Structures

While we present several models of rational inattention, the analysis relies on the recovery of a *revealed information structure* from the state dependent stochastic choice data. Using the procedure from Caplin and Dean (2015), we can associate each chosen action to a subjective information state. The revealed information structure may not be identical to the true information structure. However, the revealed information structure will be a garbling of the true information structure as defined in Blackwell (1953). The relationship between the true information structures and revealed information structures allows us to order the information structures and deduce conditions on revealed information. Without any further delay, we define *revealed posteriors* and *revealed information structures*.

Definition 4 Given $\mu \in \Gamma$, $A \in \mathcal{D}$, $P_A \in \mathcal{P}$, and $a \in \text{Supp}(P_A)$, the *revealed posterior* $\bar{\gamma}_A^a \in \Gamma$ is defined as

$$\begin{aligned}\bar{\gamma}_A^a(\omega) &= \Pr(\omega \mid a \text{ is chosen from } A)\\ &= \frac{\mu(\omega)P_A(a \mid \omega)}{\sum_{\nu\in\Omega}\mu(\nu)P_A(a \mid \nu)}.\end{aligned}$$

Definition 5 Given $\mu \in \Gamma$, $A \in \mathcal{D}$, and $P_A \in \mathcal{P}$, the revealed information structure $\bar{\pi}_A \in \Pi$ is defined by

$$\bar{\pi}_A(\gamma \mid \omega) = \sum_{\{a\in\text{Supp}(P_A) \mid \gamma=\bar{\gamma}_A^a\}} P_A(a \mid \omega)$$

and induces a revealed distribution on posteriors $\bar{\pi}_A$ such that

$$\bar{\pi}_A(\gamma) = \sum_{\omega\in\Omega}\mu(\omega)\bar{\pi}_A(\gamma \mid \omega).$$

The revealed information structure for decision problem A is a finite probability measure over the revealed posteriors. As mentioned before, we can use the notion of garbling to partially order information structures.

Definition 6 The information structure $\pi \in \Pi$ (with posteriors γ^j) is a garbling of $\rho \in \Pi$ (with posteriors η^i) if there exists a $|\text{Supp}(\rho)| \times |\text{Supp}(\pi)|$

matrix $\mathbf{B}$ with non-negative entries such that for all $i \in \{1, \dots, |\text{Supp}(\rho)|\}$ we have $\sum_{\gamma^j \in \text{Supp}(\pi)} b^{i,j} = 1$ and for all $\gamma^j \in \text{Supp}(\pi)$ and $\omega \in \Omega$ that

$$\pi(\gamma^j \mid \omega) = \sum_{\eta^i \in \text{Supp}(\rho)} b^{i,j} \rho(\eta^i \mid \omega).$$

In other words, π a garbling of ρ if there is a stochastic matrix $\mathbf{B}$ that can be applied to ρ that yields π. We present two important properties about garblings that will be used extensively in the analysis of different models of costly information acquisition.

Lemma 1 For $\pi \in \Pi$ and $P_A \in \mathcal{P}$, we say π is *consistent* with P_A if there exists a choice function $C_A : \text{Supp}(\pi) \to \Delta(A)$ such that for all $\gamma \in \text{Supp}(\pi)$,

$$C_A(a \mid \gamma) > 0 \quad \Rightarrow \quad \sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all } b \in A$$

and for all $\omega \in \Omega$ and $a \in A$

$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi)} \pi(\gamma \mid \omega) C_A(a \mid \gamma).$$

If π is consistent for P_A, then $\bar{\pi}_A$ is a garbling of π.

This lemma only depends on the definition of f_A and is proved in Caplin and Dean (2015). The lemma says that if an information structure is consistent with the state dependent stochastic choice dataset, then the revealed information structure is a garbling. The models we examine will all depend on the same definition of f_A and so this property holds throughout the rest of the analysis.

The second property of garblings we use is Blackwell's theorem (Blackwell, 1953) that establishes the notion that some information structures are "more valuable" than others. In particular, if π is a garbling of ρ, then ρ yields weakly higher gross expected utility in any decision problem.

Remark 1 Given a decision problem $A \in \mathcal{A}$ and $\pi, \rho \in \Pi$ with π a garbling of ρ, then

$$\rho \cdot f_A \geq \pi \cdot f_A.$$

From these two properties, we can make statements regarding the value of the gross expected utility at the revealed information structure. In particular, for all decision problems $A, B \in \mathcal{D}$ if π_A is an information structure consistent with choice data P_A, then the gross expected utility satisfies $f_B \cdot \pi_A \geq f_B \cdot \bar{\pi}_A$. Lastly, we have that $f_A \cdot \pi_A = f_A \cdot \bar{\pi}_A$ since the two information structures induce the same state dependent choices.

3 Nonseparable Costly Information

We place minimal restrictions on a decision maker's preferences on information structures. The only condition we impose is other than that it is monotone increasing in gross expected utility.

Definition 7 Given $\mu \in \Gamma$ and $u : X \to \Re$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *nonseparable costly information representation* if there exists a function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$, information structures $\{\pi_A\}_{A\in\mathcal{D}}$, and choice functions $\{C_A\}_{A\in\mathcal{D}}$ such that:

1. Monotonicity: For all $\pi \in \Pi$ and for all $t, s \in \Re$, if $t < s$ and $V(t, \pi) > -\infty$, then $V(t, \pi) < V(s, \pi)$.

2. Non-triviality: For all $t \in \Re$, there exists $\pi_t \in \Pi$ such that $V(t, \pi_t) > -\infty$.

3. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi\in\Pi} V(\pi \cdot f_A, \pi)$.

4. Choices are optimal: For all $A \in \mathcal{D}$, the choice function C_A : $\text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with

$C_A(a \mid \gamma) = \Pr(a \mid \gamma) > 0$, then

$$\sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega)u(b(\omega)) \quad \text{for all} \quad b \in A$$

.

5. The data is matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,

$$P_A(a \mid \omega) = \sum_{\gamma\in\text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega)C_A(a \mid \gamma)$$

.

We now define the properties that will completely characterize the model. The first condition is similar to the generalized axiom or revealed preference.

Condition 1 (Generalized Axiom of Costly Information (GACI)) We say the dataset $(\mathcal{D}, \mathcal{P})$ satisfies *GACI* if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D}$ for which $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for all i (with addition modulo k), then equality holds throughout.

Comparing this condition to GARP, we see that the $\bar{\pi}$ play a role similar to prices and the f terms play a role similar to consumption bundles albeit with the inequality reversed. The GACI condition rules out the possibility of there being cycles of gross expected utility across different decision problems. Using this condition, we can invoke a version of Afriat's theorem.

Lemma 2 (Afriat's Theorem) Let $\mathcal{D}$ be finite. For all $(A, B) \in \mathcal{D}^2$, let $\alpha_{A,B} \in \Re$ so that for all $A \in \mathcal{D}$, $\alpha_{A,A} = 0$. Suppose that for a sequence $A_1, A_2, \ldots, A_k \in \mathcal{D}$ that whenever $\alpha_{A_i,A_{i+1}} \leq 0$ (with addition mod k) for all i in a sequence, then it follows that $\alpha_{A_i,A_{i+1}} = 0$ for all i. Then there exist numbers U_A and $\lambda_A > 0$ such that for all $(A, B) \in \mathcal{D}^2$, $U_A \leq U_B + \lambda_B \alpha_{B,A}$.

The other condition which characterizes the nonseparable costly information representation is the no improving action switches (NIAS) condition.

Condition 2 (NIAS) Given $\mu \in \Gamma$ and $u : X \to \mathbb{R}$, a dataset $(\mathcal{D}, \mathcal{P})$ satisfies NIAS if, for every $A \in \mathcal{D}$, $a \in \text{Supp}(P_A)$, and $b \in A$,

$$\sum_{\omega \in \Omega} \mu(\omega) P_A(a \mid \omega)(u(a(\omega)) - u(b(\omega))) \geq 0$$

The combination of GACI and NIAS completely characterizes the model.

Theorem 1 Given $\mu \in \Gamma$ and $u : X \to \Re$, the dataset $(\mathcal{D}, \mathcal{P})$ has a nonseparable costly information representation if and only if it satisfies GACI and NIAS.

This shows the basic result that GACI and NIAS are equivalent to the nonseparable costly information representation. However, one can impose additional properties on the nonseparable costly information representation. These conditions are monotonicity, quasiconcavity, and a normalization property on the function $V(\cdot, \cdot)$.

Condition 3 The function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ satisfies weak monotonicity in information if for any $t \in \Re$ and $\pi, \rho \in \Pi$ with π a garbling of ρ, then

$$V(t, \rho) \leq V(t, \pi).$$

The monotonicity condition says that if one can add noise to a signal ρ, then the noisier signal is cheaper. This is one definition of monotonicity and it agrees with the notion of informativeness introduced in Blackwell (1953).

Condition 4 The function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ is quasiconcave if for any $(t_1, \pi_1), (t_2, \pi_2) \in \Re \times \Pi$ and $\lambda \in [0, 1]$,

$$V(\lambda t_1 + (1 - \lambda)t_2, \lambda \pi_1 + (1 - \lambda)\pi_2) \geq \min\{V(t_1, \pi_1), V(t_2, \pi_2)\}.$$

The quasiconcavity condition is similar to the convexity of the cost function in Caplin and Dean (2015). This condition says if we have a mixture of a level of gross expected utilities and information structures, then the utility of the mixture is weakly higher than the minimum of the separate environments. In particular, this implies quasiconcavity in information structures if one sets $t_1 = \pi_1 \cdot f$ and $t_2 = \pi_2 \cdot f$.

Condition 5 Define π_0 as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. The function $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ satisfies the normalization $V(0, \pi_0) = 0$.

The normalization condition says we can normalize the utility to be zero when the gross expected utility is zero and the individual did not update the prior information. We now show that all of these conditions can be satisfied using the construction of $V(\cdot, \cdot)$ in Theorem 1.

Theorem 2 Given $\mu \in \Gamma$ and $u : X \to \Re$, the data set $(\mathcal{D}, \mathcal{P})$ satisfies GACI and NIAS if and only if it has a nonseparable costly information representation that satisfies Conditions 3, 4 and 5.

4 Constrained Costly Information

The previous section studied the nonseparable costly information representation, but there other representations that can be considered. We now consider when the individual is constrained to choose information structures within some fixed set of information structures. The interpretation is that the decision maker has a limited set of information structures to use when updating the prior, but use of these information structures is costless.

Definition 8 Given $\mu \in \Gamma$ and $u : X \to \Re$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *constrained costly information representation* if there exists a set $\Pi_c \subseteq \Pi$ of available information structures, information structures $\{\pi_A\}_{A\in\mathcal{D}}$, and choice functions $\{C_A\}_{A\in\mathcal{D}}$ such that:

1. Non-triviality: The set $\Pi_c \neq \emptyset$.

2. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi\in\Pi_c} \pi \cdot f_A$.

3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr(a \mid \gamma) > 0$, then

$$\sum_{\omega\in\Omega} \gamma(\omega)u(a(\omega)) \geq \sum_{\omega\in\Omega} \gamma(\omega)u(b(\omega)) \quad \text{for all} \quad b \in A$$

   .

4. The data is matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,

$$P_A(a \mid \omega) = \sum_{\gamma\in\text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega)C_A(a \mid \gamma)$$

   .

A constrained costly information structure will be characterized by a condition similar to the weak axiom of cost minimization for production from Varian (1984). In the condition below, there is a natural similarity between the information structures with inputs of production and f_A with prices of inputs.

Condition 6 (Weak Axiom of Costly Information (WACI)) The dataset $(\mathcal{D}, \mathcal{P})$ satisfies *WACI* if for all $A, B \in \mathcal{D}$ that

$$\bar{\pi}_A \cdot f_A \geq \bar{\pi}_B \cdot f_A.$$

Theorem 3 Given $\mu \in \Gamma$ and $u : X \to \Re$, the dataset $(\mathcal{D}, \mathcal{P})$ has a constrained costly information representation if and only if it satisfies WACI and NIAS.

Similar to the nonseparable case, we are able to place stronger restrictions on the constrained costly information representation using standard methods. In particular, the additional condition is that the constrained set Π_c is convex.

Theorem 4 Given $\mu \in \Gamma$ and $u : X \to \Re$, the dataset $(\mathcal{D}, \mathcal{P})$ has a constrained costly information representation with a convex Π_c if and only if it satisfies WACI and NIAS.

5 Multiplicative Cost of Information

We now study a multiplicative costly information representation. In this representation, one can think of the cost as losing a fraction of the gross expected utility.

Definition 9 Given $\mu \in \Gamma$ and $u : X \to \Re_+$, a state dependent stochastic choice dataset $(\mathcal{D}, \mathcal{P})$ has a *multiplicative costly information representation* if there exists a function $K : \Pi \to [0, 1]$, information structures $\{\pi_A\}_{A \in \mathcal{D}}$, and choice functions $\{C_A\}_{A \in \mathcal{D}}$ such that:

1. Non-triviality: There exists $\pi \in \Pi$ such that $K(\pi) < 1$.

2. Information is optimal: For all $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi \in \Pi} [(1 - K(\pi))(\pi \cdot f_A)]$.

3. Choices are optimal: For all $A \in \mathcal{D}$, the choice function $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ is such that given $a \in A$ and $\gamma \in \text{Supp}(\pi_A)$ with $C_A(a \mid \gamma) = \Pr(a \mid \gamma) > 0$, then

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \text{for all} \quad b \in A$$

.

4. The data is matched: For all $A \in \mathcal{D}$, given $\omega \in \Omega$ and $a \in A$,

$$P_A(a \mid \omega) = \sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma)$$

.

We note that one difference in the statement of the multiplicative costly information representation is that the utility index u is required to be non-negative. While this is more restrictive than the other cases, this is a common property of representations such as in Chateauneuf and Faro (2009). The condition which characterizes the multiplicative costly information representation is a version of the homothetic axiom of revealed preference; see Varian (1983).[3]

Condition 7 (Homothetic Axiom of Costly Information (HACI)) Given data set $(\mathcal{D}, \mathcal{P})$, define $\mathcal{D}_0 = \{A \in \mathcal{D} | \sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) = 0$ for all $a \in A\}$. We say the dataset $(\mathcal{D}, \mathcal{P})$ satisfies *HACI* if for all sequences $(\bar{\pi}_{A_1}, f_{A_1}), \ldots, (\bar{\pi}_{A_k}, f_{A_k})$ with $A_i \in \mathcal{D} \backslash \mathcal{D}_0$, that $\prod_{i=1}^{k} \frac{\bar{\pi}_{A_i} \cdot f_{A_{i+1}}}{\bar{\pi}_{A_i} \cdot f_{A_i}} \leq 1$ (with addition modulo k).

The HACI is essentially the homothetic axiom of revealed preference restricted to decision problems that give positive gross expected utility. The decision problems that give zero gross expected utility are removed since they can be trivially rationalized and they would create an indeterminate fraction.

Theorem 5 Given $\mu \in \Gamma$ and $u : X \to \Re_+$, the dataset $(\mathcal{D}, \mathcal{P})$ has a multiplicative costly information representation if and only if it satisfies HACI and NIAS.

As in the case of the nonseparable costly information representation, we are able to put additional properties on the function K. We find that K

[3]It can also be derived as a relatively easy corollary from the general work of Rochet (1987).

respects monotonicity with respect to the Blackwell partial order, is convex, and satisfies a normalization property. We now define these properties and the give a statement of the theorem.

Condition 8 The function $K : \Pi \to [0, 1]$ satisfies weak monotonicity in information if $\rho, \pi \in \Pi$ with π a garbling of ρ,

$$K(\pi) \leq K(\rho).$$

Condition 9 The function $K : \Pi \to [0, 1]$ is convex in information structures if for for any $\pi_1, \pi_2 \in \Pi$ and $\lambda \in [0, 1]$,

$$K(\lambda\pi_1 + (1 - \lambda)\pi_2) \leq \lambda K(\pi_1) + (1 - \lambda)K(\pi_2).$$

Condition 10 Define π_0 as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. The function $K : \Pi \to [0, 1]$ satisfies normalization if $K(\pi_0) = 0$.

Theorem 6 Given μ and $u : X \to \Re_+$, data set $(\mathcal{D}, \mathcal{P})$ satisfies HACI and NIAS if and only if it has a non-separable costly information representation that satisfies Conditions 8, 9, and 10.

6 Conclusion

In this note, we have shown there are several natural revealed preference models of rational inattention. The characterization of these models follows directly from classical revealed preference theory.

Appendix A Proofs of Results

Proof of Theorem 1. ($\Rightarrow$) First, we show that a nonseparable costly information representation satisfies NIAS. Fix $A \in \mathcal{D}$, $\pi_A \in \arg\max_{\pi \in \Pi} V(\pi \cdot$

$f_A, \pi)$, and $C_A : \text{Supp}(\pi_A) \to \Delta(A)$ and $a \in \text{Supp}(P_A)$. By definition of a nonseparable costly information representation, we know that the $V(\pi_A \cdot f_A, \pi_A)$ is monotone in $\pi_A \cdot f_A$ and choices are optimal conditional on posteriors. Thus, if a was chosen when γ was realized, then the expected utility must be weakly higher for these γ. For γ such that $C_A(a \mid \gamma) > 0$,

$$\sum_{\omega \in \Omega} \gamma(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \gamma(\omega) u(b(\omega)) \quad \forall b \in A.$$

The proof now follows from arguments in Caplin and Dean (2015) that are reproduced here for completeness. Recall that

$$\gamma(\omega) = \frac{\mu(\omega) \pi_A(\gamma \mid \omega)}{\sum_{\nu \in \Omega} \mu(\nu) \pi(\gamma \mid \nu)},$$

which can be substituted on both sides and the denominator cancels so

$$\sum_{\omega \in \Omega} \mu(\omega) \pi_A(\gamma \mid \omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \mu(\omega) \pi_A(\gamma \mid \omega) u(b(\omega)) \quad \forall b \in A.$$

Therefore,

$$\begin{aligned} &\sum_{\gamma \in \text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega \in \Omega} \mu(\omega) \pi_A(\gamma \mid \omega) u(a(\omega)) \right] \\ \geq &\sum_{\gamma \in \text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega \in \Omega} \mu(\omega) \pi_A(\gamma \mid \omega) u(b(\omega)) \right] \quad \forall b \in A \end{aligned}$$

since $C_A(a \mid \gamma)$ are either zero or positive multiples of the earlier introduced inequalities. Next, recall from data matching that $P_A(a \mid \omega) =$

$\sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma)$. Therefore, we see that

$$\begin{aligned}
\sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) P_A(a \mid \omega) &= \sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) \left[\sum_{\gamma \in \text{Supp}(\pi_A)} \pi_A(\gamma \mid \omega) C_A(a \mid \gamma) \right] \\
&= \sum_{\gamma \in \text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega \in \Omega} \mu(\omega) u(a(\omega)) \pi_A(\gamma \mid \omega) \right] \\
&\geq \sum_{\gamma \in \text{Supp}(\pi_A)} C_A(a \mid \gamma) \left[\sum_{\omega \in \Omega} \mu(\omega) u(b(\omega)) \pi_A(\gamma \mid \omega) \right] \\
&= \sum_{\omega \in \Omega} \mu(\omega) u(b(\omega)) P_A(a \mid \omega)
\end{aligned}$$

where the first set of equalities follows from substitutions, the inequality follows from optimality conditional on γ, and the last equality follows from the same substitutions above. Rearranging this inequality shows that NIAS is satisfied.

Next, we show that a nonseparable costly information representation implies GACI. Suppose without loss of generality that $\bar{\pi}_{A_i} \cdot f_{A_i} \leq \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$ for $i = \{1, \ldots, k\}$ (with addition modulo k). Observe that $\arg\max_\pi V(\pi \cdot f_{A_i}, \pi) = V(\pi_{A_i} \cdot f_{A_i}, \pi_{A_i})$ by definition. It follows that

$$\begin{aligned}
V(\pi_{A_i} \cdot f_{A_i}, \pi_{A_i}) &= V(\bar{\pi}_{A_i} \cdot f_{A_i}, \pi_{A_i}) \\
&\leq V(\bar{\pi}_{A_i} \cdot f_{A_{i+1}}, \pi_{A_i}) \\
&\leq V(\pi_{A_{i+1}} \cdot f_{A_{i+1}}, \pi_{A_{i+1}}) = V(\bar{\pi}_{A_{i+1}} \cdot f_{A_{i+1}}, \pi_{A_{i+1}})
\end{aligned}$$

with strict inequality if $\bar{\pi}_{A_i} \cdot f_{A_i} < \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$. Consequently, we must have equality throughout for all i so $\bar{\pi}_{A_i} \cdot f_{A_i} = \bar{\pi}_{A_i} \cdot f_{A_{i+1}}$. If instead there is a strict inequality, then we obtain the contradiction $V(\pi_{A_1} \cdot f_{A_1}, \pi_{A_1}) < V(\pi_{A_1} \cdot f_{A_1}, \pi_{A_1})$ by strict monotonicity in the first component of V.

($\Leftarrow$) The converse is a direct application of Afriat's Theorem. Let $\alpha_{A,B} = -\bar{\pi}_A \cdot (f_B - f_A)$ for all $(A, B) \in \mathcal{D}^2$. Observe that by GACI, the condition in

Afriat's Theorem is satisfied. Conclude there is U_A and $\lambda_A > 0$ such that for all $(A, B) \in \mathcal{D}^2$, $U_A \leq U_B - \lambda_B \bar{\pi}_B \cdot (f_A - f_B)$. Taking negatives and letting $\tilde{U}_A = -U_A$, we have

$$\tilde{U}_B + \lambda_B \bar{\pi}_B (f_A - f_B) \leq \tilde{U}_A.$$

Most of the remaining construction follows Afriat's theorem directly. Let $C(\Gamma)$ be the space of continuous functions on Γ. Define $U : C(\Gamma) \to \Re$ by

$$U(f) = \max_{A \in \mathcal{D}} \tilde{U}_A + \lambda_A \bar{\pi}_A \cdot (f - f_A)$$

Clearly, U is convex, continuous, and monotone[4] (as the maximum of a finite number of continuous affine functionals). For every $A \in \mathcal{D}$, $U(f_A) = \tilde{U}_A$ by construction. Moreover, for every $A \in \mathcal{D}$, if $\bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A$, then $U(f) \geq U(f_A)$, which is also straightforward by construction.

Define $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ by $V(t, \pi) = \inf_{f \in C(\Gamma)} \{U(f) : \pi \cdot f \geq t\}$. Observe that the monotonicity condition is trivially satisfied for fixed π, since a greater t reduces the set of $f \in C(\Gamma)$ satisfying the inequality. The assumption that for each $t \in \Re$, there exists a $\pi_t \in \Pi$ such that $V(t, \pi_t) > -\infty$ is also satisfied. To see this, consider the constant function $f_t(\gamma) = t$. Therefore, the set $G_- = \{g \in C(\Gamma) \mid U(g) \leq U(f_t)\}$ is closed and convex. Moreover, the set $G_{++} = \{g \in C(\Gamma) \mid g > f_t\}$[5] is sup-norm open, convex, and disjoint from G_- by the monotonicity property on $U(\cdot)$. Note that $C(\Gamma)$ is dual to the set of Borel measures over Γ (Aliprantis and Border (2006) Theorem 14.15). Let $M(\Gamma)$ be the set of Borel (countably additive) measures on Γ. By Aliprantis and Border (2006) Theorem 7.19, the two sets can be separated by a non-zero hyperplane $\mu \in M(\Gamma)$. We show that μ can be chosen to be a probability measure. Choose μ so that for all $g \in G_{++}$ that

[4] Monotonicity of the functional U in f means that if $(f - g)(\gamma) > 0$ for all $\gamma \in \Gamma$, then $U(f) > U(g)$.

[5] Where $g > f$ denotes that for all $\gamma \in \Gamma$ $g(\gamma) > f(\gamma)$.

$g \cdot \mu \geq f_t \cdot \mu$ which can be done since f_t is on the boundary of G_{++} and G_-. This implies that $\mu \in M_+(\Gamma)$, the set of non-negative Borel measures.

To see $\mu \in M_+(\Gamma)$, suppose by way of contradiction that there exists a $\hat{\gamma} \in \Gamma$ such that $\mu(\hat{\gamma}) < 0$. Consider the function

$$\hat{g}(\gamma) = \begin{cases} t+1 & \text{if } \gamma = \hat{\gamma} \\ t & \text{otherwise} \end{cases}$$

so that $\hat{g} \in G_{++}$. Examining the inner product, we see that $(\hat{g} - f_t) \cdot \mu = \mu(\hat{\gamma}) < 0$ a contradiction to the choice of μ. In particular, this implies that for all $h \in \{g \in C(\Gamma) \mid g > 0\}$ that $h \cdot \mu \geq 0$. By continuity, this implies that all $h \in \{g \in C(\gamma) \mid g \geq 0\}$ that $h \cdot \mu \geq 0$. The last statement means that μ can be normalized to a probability measure since it maps non-negative functions to non-negative numbers and is a non-zero hyperplane. Lastly, μ supports G_- at f_t so in fact $V(f_t \cdot \mu, \mu) = U(f)$ where $f_t \cdot \mu = t$ from the probability normalization and f_t a constant function. Therefore, setting $\pi_t = \mu$ will give an information structure that gives $V(t, \pi_t) > -\infty$.

We now assert that for all $A \in \mathcal{D}$, $\bar{\pi}_A \in \arg\max_{\pi \in \Pi} V(\pi \cdot f_A, \pi)$. First, from the monotonicity property of the U function

$$\begin{aligned} V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) &= \inf_{f \in C(\Gamma)} \{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\} \\ &= U(f_A) \end{aligned}$$

Second, for any $\pi \in \Pi$, we have $V(\pi \cdot f_A, \pi) = \inf_{f \in C(\Gamma)} \{U(f) : \pi \cdot f \geq \pi \cdot f_A\} \leq U(f_A)$, since $\pi \cdot f_A \geq \pi \cdot f_A$. Therefore $V(\pi \cdot f_A, \pi) \leq V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$ for all $\pi \in \Pi$. Therefore, the revealed information structure is optimal for V.[6]

[6] We note that a version of Roy's identity holds (Roy (1947)). Observe that by definition of V, if $\pi \cdot f_A \geq w$ implies $U(f_A) \geq V(w, \pi)$. We conclude that $\pi \cdot f_A \geq \bar{\pi}_A \cdot f_A$ implies $U(f_A) \geq V(\bar{\pi}_A \cdot f_A, \pi)$. We have already shown that $U(f_A) = V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$. Thus, if $\pi \cdot f_A \geq \bar{\pi}_A \cdot f_A$, then $V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) \geq V(\pi_A \cdot f_A, \pi)$.

We now show data matching and choices are optimal by following Caplin and Dean (2015) and using NIAS. Next we show that there exists stochastic choice functions $\{C_A : \text{Supp}(\bar{\pi}_A) \to \Delta(A)\}_{A \in \mathcal{D}}$ that satisfy optimality and matches data.

For each $\gamma \in \text{Supp}(\bar{\pi}_A)$, define:

$$C_A(a|\gamma) = \begin{cases} \frac{P_A(a)}{\sum_{\{b \in A : \bar{\gamma}_A^b = \gamma\}} P_A(b)} & \text{if } \gamma = \bar{\gamma}_A^a \\ 0 & \text{otherwise} \end{cases}$$

where $P_A(a) = \sum_{\omega \in \Omega} P_A(a|\omega)\mu(\omega)$ is the unconditional probability of choosing action a from choice problem A. Note the $C_A(a|\gamma) > 0$ only if $\gamma = \bar{\gamma}_A^a$. The NIAS condition implies that

$$\sum_{\omega \in \Omega} \mu(\omega) P_A(a \mid \omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \mu(\omega) P_A(b \mid \omega) u(b(\omega))$$
$$\Rightarrow \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(a(\omega)) \geq \sum_{\omega \in \Omega} \bar{\gamma}_A^a(\omega) u(b(\omega))$$

The second line follows by dividing both sides by $P_A(a)$. Thus, NIAS ensures that the choices are optimal.

It remains to show that the data is matched. In other words, P_A is generated from the information structure $\bar{\pi}_A$ and choices C_A. First, note that for any $b, b' \in A$ such that $\bar{\gamma}_A^b = \bar{\gamma}_A^{b'}$, implies that for any $\omega \in \Omega$ such that $\bar{\gamma}_A^b(\omega) > 0$, then

$$\frac{P_A(b \mid \omega)}{P_A(b' \mid \omega)} = \frac{P_A(b)}{P_A(b')}.$$

Thus, for every $\omega \in \Omega$ and $a \in A$ such that $P_A(a) > 0$, then

$$\begin{aligned}\sum_{\gamma \in \text{Supp}(\bar{\pi}_A)} \bar{\pi}_A(\gamma|\omega) C_A(a|\gamma) &= \bar{\pi}_A(\bar{\gamma}_A^a|\omega) C_A(a|\bar{\gamma}_A^a) \\ &= \sum_{\{c \in A : \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a)}{\sum_{\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\}} P_A(b)} \\ &= \sum_{\{c \in A | \bar{\gamma}_A^c = \bar{\gamma}_A^a\}} P_A(c|\omega) \frac{P_A(a|\omega)}{\sum_{\{b \in A | \bar{\gamma}_A^b = \bar{\gamma}_A^a\}} P_A(b|\omega)} \\ &= P_A(a|\omega).\end{aligned}$$

Therefore, the data is matched. ■

Proof of Theorem 2. Let $V : \Re \times \Pi \to \Re \cup \{-\infty\}$ be the function constructed as in the proof of Theorem 1. The function

$$\tilde{V}(t, \pi) = V(t, \pi) - V(0, \pi_0)$$

satisfies all the additional properties. First, note that $\tilde{V}(0, \pi_0) = V(0, \pi_0) - V(0, \pi_0) = 0$ so the normalization condition is satisfied.

Since the difference of V and $\tilde{V}$ is a constant, we can check quasiconcavity and weak monotonicity of V. Next, we check weak monotonicity. If π is a garbling of ρ, then

$$\begin{aligned}V(t, \rho) &= \inf_{f \in C(\Gamma)} \{U(f) \mid \rho \cdot f \geq t\} \\ &\leq \inf_{f \in C(\gamma)} \{U(f) \mid \pi \cdot f \geq t\} \\ &= V(t, \pi)\end{aligned}$$

since $\pi \cdot f \geq t$ implies that $\rho \cdot f \geq t$ by Remark 1 so the infimum is taken over a weakly smaller set of functions. Thus, weak monotonicity holds.

Lastly, we examine quasiconcavity of V. Let $(t_1, \pi_1), (t_2, \pi_2) \in \Re \times \Pi$, then for $\lambda \in [0, 1]$

$$V(\lambda t_1 + (1-\lambda) t_2, \lambda \pi_1 + (1-\lambda)\pi_2) = \inf_{f \in C(\Gamma)} \{U(f) \mid \lambda \pi_1 \cdot f + (1-\lambda)\pi_2 \cdot f \geq \lambda t_1 + (1-\lambda) t_2\}.$$

Note that if $\lambda\pi_1 \cdot f + (1-\lambda)\pi_2 \cdot f \geq \lambda t_1 + (1-\lambda)t_2$, then either $\pi_1 \cdot f \geq t_1$ or $\pi_2 \cdot f \geq t_2$. Therefore, for $f \in C(\Gamma)$ we have

$$\{f \mid \lambda\pi_1 \cdot f + (1-\lambda)\pi_2 \cdot f \geq \lambda t_1 + (1-\lambda)t_2\} \subseteq \{f \mid \pi_1 \cdot f \geq t_1\} \cup \{f \mid \pi_2 \cdot f \geq t_2\}.$$

Therefore, the infmum of U over the first set, $V(\lambda t_1 + (1-\lambda)t_2, \lambda\pi_1 + (1-\lambda)\pi_2)$, is greater than or equal to the infimum over the second set, $\min\{V(t_1, \pi_1), V(t_2, \pi_2)\}$. Thus, quasiconcavity holds. ■

Proof of Theorem 3. We note that NIAS is equivalent to optimal choices and matched data. Therefore, we focus on non-triviality and optimal information.

($\Rightarrow$) Suppose the data is represented by a constrained costly information representation and for all $A \in \mathcal{D}$ that $\pi_A \in \arg\max_{\pi \in \Pi_c} \pi \cdot f_A$. Since the utility depends only on gross expected utility, then $\pi_A \cdot f_A = \bar{\pi}_A \cdot f_A \geq \pi_B \cdot f_A \geq \bar{\pi}_B \cdot f_A$. The first equality follows from equivalent choices, the next inequality follows from optimality, while the final inequality follows Remark 1.

($\Leftarrow$) Suppose WACI holds. Let $\bar{\Pi}_c = \bigcup_{A \in \mathcal{D}} \{\bar{\pi}_A\}$. For $\mathcal{D}$ nonempty, $\Pi_c \neq \emptyset$.[7] Moreover, for any $A, B \in \mathcal{D}$, we have

$$\bar{\pi}_A \cdot f_A \geq \bar{\pi}_B \cdot f_A.$$

In other words, for all $A \in \mathcal{D}$ we have $\bar{\pi}_A \in \arg\max_{\pi \in \bar{\Pi}_c} \pi \cdot f_A$. Therefore nontriviality and optimal information hold. ■

Proof of Theorem 4. From the proof of Theorem 3 we only need to prove that $WACI$ implies a convex constraint set. Let $\text{conv}(\bar{\Pi}_c) = \text{conv}\left(\bigcup_{A \in \mathcal{D}} \{\bar{\pi}_A\}\right)$. Here $\text{conv}(\cdot)$ represents the convex hull of information structures. For all $B \in \mathcal{D}$ let $\lambda_B \in [0,1]$ such that $\sum_{B \in \mathcal{D}} \lambda_B = 1$. Now for fixed $A \in \mathcal{D}$

$$\sum_{B \in \mathcal{D}} \lambda_B \bar{\pi}_B \cdot f_A \leq \sum_{B \in \mathcal{D}} \lambda_B \bar{\pi}_A \cdot f_A = \bar{\pi}_A \cdot f_A$$

[7] If $\mathcal{D} = \emptyset$, then let $\Pi_c = \Pi$.

where the inequality follows from WACI. The result holds for any fixed A and convex combination so that $\bar{\pi}_A \in \arg\max_{\pi \in \text{conv}(\bar{\Pi}_c)} \pi \cdot f_A$. Thus the constraint set can be convex. ■

Proof of Theorem 5. We note that NIAS is equivalent to optimal choices and matched data. Therefore, we focus on non-triviality and optimal information.

($\Rightarrow$) We show that a multiplicatively costly information representation satisfies HACI. For all $A \in \mathcal{D}$, let $\pi_A \in \arg\max_{\pi \in \Pi}[(1 - K(\pi))(\pi \cdot f_A)]$.

First, we show for $A \in \mathcal{D} \setminus \mathcal{D}_0$, that $(\pi \cdot f_A) > 0$ for all information structure π and $K(\pi_A) < 1$. Let π_0 denote the non-informative information structure with $\pi_0(\mu) = 1$. By assumption, $\pi_0 \cdot f_A > 0$ for any $A \in \mathcal{D} \backslash \mathcal{D}_0$. Since f_A is convex, $\pi \cdot f_A > 0$ for all information structure π. In particular, let $\pi' \in \Pi$ be an information structure such that $K(\pi') < 1$, then $\pi' \cdot f_A > 0$ as well. Note that such a $\pi' \in \Pi$ with $K(\pi') < 1$ exists by nontriviality. For any $A \in \mathcal{D} \setminus \mathcal{D}_0$ and for any $\pi \in \Pi$, we have $(1 - K(\pi_A))(\pi_A \cdot f_A) \geq (1 - K(\pi'))(\pi' \cdot f_A) > 0$ from π_A the optimal information structure. Therefore, for all $A \in \mathcal{D} \backslash \mathcal{D}_0$ we have $K(\pi_A) < 1$.

Next, for any pair $A_i, A_{i+1} \in \mathcal{D} \backslash \mathcal{D}_0$, we have

$$
\begin{aligned}
(1 - K(\pi_{A_i}))(\bar{\pi}_{A_i} \cdot f_{A_i}) &= (1 - K(\pi_{A_i}))(\pi_{A_i} \cdot f_{A_i}) \\
&\geq (1 - K(\pi_{A_{i+1}}))(\pi_{A_{i+1}} \cdot f_{A_i}) \\
&\geq (1 - K(\pi_{A_{i+1}}))(\bar{\pi}_{A_{i+1}} \cdot f_{A_i}) > 0
\end{aligned}
$$

where the equality follows from equivalent choices, the first inequality follows from optimality, the second inequality follows from Remark 1, and the last term is greater than zero by the earlier arguments. Rearranging the end terms of the inequalities,

$$
\frac{(1 - K(\pi_{A_{i+1}}))(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{(1 - K(\pi_{A_i}))(\bar{\pi}_{A_i} \cdot f_{A_i})} \leq 1.
$$

We can now take any cycle $A_1, \ldots, A_k \in \mathcal{D} \backslash \mathcal{D}_0$ and take products to see that costs will be removed so

$$\prod_{i=1}^{k} \frac{(1 - K(\pi_{A_{i+1}}))(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{(1 - K(\pi_{A_i}))(\bar{\pi}_{A_i} \cdot f_{A_i})} = \prod_{i=1}^{k} \frac{(\bar{\pi}_{A_{i+1}} \cdot f_{A_i})}{(\bar{\pi}_{A_i} \cdot f_{A_i})} \leq 1$$

where the indices are calculated with addition modulo k. Let $\sigma : \{1, \ldots, k,\} \to \{1, \ldots, k\}$ where $\sigma(1) = 1, \sigma(2) = k, \sigma(3) = k - 1, \ldots, \sigma(k) = 2$.[8] Therefore,

$$\prod_{i=1}^{k} \frac{\bar{\pi}_{A_{\sigma(i)}} \cdot f_{A_{\sigma(i+1)}}}{\bar{\pi}_{A_{\sigma(i)}} \cdot f_{A_{\sigma(i)}}} \leq 1$$

and HACI is satisfied.

($\Leftarrow$) Now we show from HACI that we can generate a non-trivial utility function. Following Varian (1983), for all $A \in \mathcal{D}$ let U_A be the maximum of

$$\prod_{i=1}^{k-1} \frac{\bar{\pi}_{A_i} \cdot f_{A_{i+1}}}{\bar{\pi}_{A_i} \cdot f_{A_i}}$$

where the maximization is taken over all finite sequences $\{A_i\}_{i=1}^{k-1} \subseteq \mathcal{D} \setminus \mathcal{D}_0$ with $A_k = A \in \mathcal{D}$. Note that if $A \in \mathcal{D}_0$ then $U_A = 0$. Since the number of cycles is finite, a maximum exists for each A. Moreover, each U_A is bounded by HACI. Note that $U_A > 0$ for all $A \in \mathcal{D} \setminus \mathcal{D}_0$, and for all $A, B \in \mathcal{D}$

$$U_B \geq U_A \frac{\bar{\pi}_A \cdot f_B}{\bar{\pi}_A \cdot f_A}$$

by definition. Rearranging the inequalities, we see that for all $A, B \in \mathcal{D}$,

$$U_B \geq U_A \frac{\bar{\pi}_A \cdot f_B}{\bar{\pi}_A \cdot f_A} \tag{1}$$

Define

$$U(f) = \begin{cases} \max_{A \in \mathcal{D}} \left[U_A \frac{\bar{\pi}_A \cdot f}{\bar{\pi}_A \cdot f_A} \right] & \text{if } f \in C_+(\Gamma) \\ +\infty & \text{otherwise} \end{cases}$$

[8] Note addition is still modulo k in the index so $\sigma(k+1) = 1$.

where $C_+(\Gamma)$ are nonnegative continuous functions on Γ. From the definition of U, it is obvious that $U(\cdot)$ is homogenous of degree 1 (as the supremum of a finite number of linear functionals), and $U(f) \geq 0$ for all $f \in C(\Gamma)$. In addition, inequality (1) implies that for all $A \in \mathcal{D}$ that $U(f_A) = U_A$. It is also straightforward that U is convex, continuous, and strictly monotonically increasing over $C_+(\Gamma)$. Finally, we have

$$U(f) \geq U(f_A) \text{ if } \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A \tag{2}$$

which is also straightforward by construction.

Let $m \in M_+(\Gamma)$ and define $V : \Re_+ \times M_+(\Gamma) \to \Re_+$ by $V(t,m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq t\}$. Now, we see that $V(\cdot,\cdot)$ is indeed of the multiplicative form. First, let $R : M_+(\Gamma) \to \Re$ be given by $R(m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq 1\}$. By the definition of V, for any $t > 0$ we have

$$\begin{aligned}
V(t,m) &= \inf_{f \in C(\Gamma)}\{U(f) : m \cdot \frac{f}{t} \geq 1\} \\
&= \inf_{tf' \in C(\Gamma)}\{U(tf') : m \cdot f' \geq 1\} \\
&= \inf_{f' \in C(\Gamma)}\{U(tf') : m \cdot f' \geq 1\} \\
&= t \inf_{f' \in C(\Gamma)}\{U(f') : m \cdot f' \geq 1\} \\
&= tR(m)
\end{aligned}$$

where the first equality comes from rearrangement, the second equality comes from $f' = f/t$, the third equality comes since any tf' can be expressed as a function, the fourth equality holds since U is homogeneous degree 1, and the final equality holds by definition of R.

Next, if $t = 0$ then $V(t,m) = 0$ which is consistent with the multiplicative form. To see this, consider the constant function $f_0(\gamma) = 0$ for all $\gamma \in \Gamma$ and see that $V(0,\pi) \leq U(f_0) = 0$. Lastly, recall that $U(f) \geq 0$ for all

$f \in C(\Gamma)$. Let $\tilde{R} : \Pi \to \Re_+$ be the restriction of R to Π. We later use the R function to show that the cost can be made convex without loss of generality.

In addition, since $U(f) \geq 0$ for all $f \in C(\Gamma)$, we have $\tilde{R}(\pi) = \inf_{f \in C(\Gamma)}\{U(f) : \pi \cdot f \geq 1\} \geq 0$. Moreover, we show that $\tilde{R}(\pi) < \infty$. Consider the constant function $f_1(\gamma) = 1$ for all $\gamma \in \Gamma$ so that $\pi \cdot f_1 = 1$. Therefore, we deduce

$$\tilde{R}(\pi) \leq \max_{A \in \mathcal{D}} \frac{U_A}{\bar{\pi}_A \cdot f_1} < \infty.$$

We also prove that there are $\pi \in \Pi$ such that $\tilde{R} > 0$. For an arbitrary $A \in \mathcal{D} \backslash \mathcal{D}_0$, we have

$$\begin{aligned}
\tilde{R}(\bar{\pi}_A) &= \frac{V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)}{\bar{\pi}_A \cdot f_A} \\
&= \frac{\inf_{f \in C(\Gamma)}\{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\}}{\bar{\pi}_A \cdot f_A} \\
&= \frac{U(f_A)}{\bar{\pi}_A \cdot f_A} = \frac{U_A}{\bar{\pi}_A \cdot f_A} > 0
\end{aligned}$$

where from definitions and $\bar{\pi}_A \cdot f_A > 0$. Therefore, $\tilde{R}(\pi_A) > 0$.

We note that if π is a garbling of ρ then $\tilde{R}(\rho) \leq \tilde{R}(\pi)$ since $\tilde{R}(\pi) = \inf_{f \in C(\Gamma)}\{U(f) \mid \pi \cdot f \geq 1\}$ and $\pi \cdot f \geq 1$ implies $\rho \cdot f \geq 1$ so the infimum is over a weakly larger set. Let π_0 as the information structure with $\pi_0(\mu|\omega) = 1$ for all $\omega \in \Omega$. Since Π is the set of information sets consistent with Bayes' Law, π_0 is a garbling of any $\pi \in \Pi$. Thus, for all $\pi \in \Pi$, $\tilde{R}(\pi_0) \geq \tilde{R}(\pi) > 0$. Lastly, define $K : \Pi \to [0, 1]$ as

$$K(\pi) = 1 - \frac{\tilde{R}(\pi)}{\tilde{R}(\pi_0)}.$$

We now assert that for all $A \in \mathcal{D}$, $\bar{\pi}_A \in \arg\max_{m \in M_+(\Gamma)} V(\pi \cdot f_A, \pi)$. First, from inequality (2) we have

$$\begin{aligned}
V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A) &= \inf_{f \in C(\Gamma)}\{U(f) : \bar{\pi}_A \cdot f \geq \bar{\pi}_A \cdot f_A\} \\
&= U(f_A)
\end{aligned}$$

Second, for any $m \in M_+(\Gamma)$, we have $V(m \cdot f_A, m) = \inf_{f \in C(\Gamma)}\{U(f) : m \cdot f \geq m \cdot f_A\} \leq U(f_A)$, since $m \cdot f_A \geq m \cdot f_A$. Therefore $V(m \cdot f_A, m) \leq V(\bar{\pi}_A \cdot f_A, \bar{\pi}_A)$ for all $m \in M_+(\Gamma)$. From this, we have that

$$\bar{\pi}_A \in \arg\max_{\pi \in \Pi} \frac{V(\pi \cdot f_A, \pi)}{\tilde{R}(\pi_0)} = \arg\max_{\pi \in \Pi} \frac{\tilde{R}(\pi)}{\tilde{R}(\pi_0)}(\pi \cdot f_A)$$
$$= \arg\max_{\pi \in \Pi}(1 - K(\pi))(\pi \cdot f_A)$$

where π_A is am optimizer since π_A is optimal for V over a larger set and this is a positive scaling of V. Therefore π_A is optimal for the rescaled V and has the multiplicative costly representation. ■

Proof of Theorem 6. We note that the K generated in Theorem 5 satisfies the monotonicity property and the normalization property. The $R(m)$ defined in Theorem 5 is homogenous of degree one, increasing in m, and quasiconcave by the same arguments used in Theorem 2. By Theorem 1 in Prada (2011), we have that R is concave. Therefore, $\tilde{R}$ restricted to Π is the restriction of R to a convex set and is thus concave. Finally, K is convex as $K(\pi) = 1 - \frac{\tilde{R}(\pi)}{\tilde{R}(\pi_0)}$. ■

References

Aliprantis, Charalambos D and Kim Border, *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Springer Science & Business Media, 2006.

Blackwell, David, "Equivalent comparisons of experiments," *The annals of mathematical statistics*, 1953, *24* (2), 265–272.

Brown, Donald J and Caterina Calsamiglia, "The nonparametric approach to applied welfare analysis," *Economic Theory*, 2007, *31* (1), 183–188.

Caplin, Andrew and Mark Dean, "Revealed preference, rational inattention, and costly information acquisition," *The American Economic Review*, 2015, *105* (7), 2183–2203.

Cerreia-Vioglio, Simone, Fabio Maccheroni, Massimo Marinacci, and Luigi Montrucchio, "Complete monotone quasiconcave duality," *Mathematics of Operations Research*, 2011, *36* (2), 321–339.

_ , _ , _ , and _ , "Uncertainty averse preferences," *Journal of Economic Theory*, 2011, *146* (4), 1275–1330.

Chambers, Christopher P and Federico Echenique, *Revealed Preference Theory*, Vol. 56, Cambridge University Press, 2016.

_ , _ , and Kota Saito, "Testing theories of financial decision making," *Proceedings of the National Academy of Sciences*, 2016, *113* (15), 4003–4008.

Chateauneuf, Alain and José Heleno Faro, "Ambiguity through confidence functions," *Journal of Mathematical Economics*, 2009, *45* (9), 535–558.

Houthakker, Hendrik S, "Revealed preference and the utility function," *Economica*, 1950, *17* (66), 159–174.

Koopmans, Tjalling C and Martin Beckmann, "Assignment problems and the location of economic activities," *Econometrica*, 1957, pp. 53–76.

Prada, Juan David, "A note on concavity, homogeneity and non-Increasing returns to scale"," *Economics Bulletin*, 2011, *31* (1), 100–105.

Richter, Marcel K, "Revealed preference theory," *Econometrica*, 1966, pp. 635–645.

Rochet, Jean-Charles, "A necessary and sufficient condition for rationalizability in a quasi-linear context," *Journal of mathematical Economics*, 1987, *16* (2), 191–200.

Rockafellar, Ralph, "Characterization of the subdifferentials of convex functions," *Pacific Journal of Mathematics*, 1966, *17* (3), 497–510.

Roy, René, "La distribution du revenu entre les divers biens," *Econometrica*, 1947, pp. 205–225.

Varian, Hal R, "Non-parametric tests of consumer behaviour," *The review of economic studies*, 1983, *50* (1), 99–110.

_ , "The nonparametric approach to production analysis," *Econometrica*, 1984, pp. 579–597.