Coarse Revealed Preference*

Gaoji Hu[†] Jiangtao Li[‡] John K.-H. Quah[§] Rui Tang[¶]

December 3, 2025

Abstract

We propose a novel concept of rationalization, called *coarse rationalization*, tailored for the analysis of datasets where an agent's choices are imperfectly observed. We characterize those datasets which are rationalizable in this sense and present an efficient algorithm to verify the characterizing condition. We then demonstrate how our results can be applied through a duality approach to test the rationalizability of datasets with perfectly observed choices but imprecisely observed linear budget sets. For datasets that consist of both perfectly observed feasible sets and choices but are inconsistent with perfect rationality, our results could be used to measure the extent to which choices or prices have to be perturbed to recover rationality.

KEYWORDS: Coarse dataset, rationalization, revealed preference, Afriat's theorem, perturbation index, price misperception index

* We are grateful to the handling editor and three referees for their insightful guidance. We would also like to thank Mark Dean, Eddie Dekel, David Dillenberger, Faruk Gul, Bart Lipman, Pietro Ortoleva, Wolfgang Pesendorfer, Satoru Takahashi, Jingni Yang, and numerous participants at various seminars and conferences for helpful discussions. Hu gratefully acknowledges the financial support from the National Natural Science Foundation of China (No. 72473087, No. 72394391, No. 72033004) and the Key Laboratory of Mathematical Economics (SUFE), Ministry of Education. Li gratefully acknowledges the financial support from the Singapore Ministry of Education through its Academic Research Fund Tier 2 (T2EP40122-0018) and the Lee Kuan Yew Fellowship. Tang gratefully acknowledges the Lee Heng Fellowship.

† School of Economics and Dishui Lake Advanced Finance Institute, Shanghai University of Finance and Economics. Email: *hu.gaoji@gmail.com*.

‡ School of Economics, Singapore Management University. Email: *jtli@smu.edu.sg*.

§ Department of Economics, National University of Singapore. Email: *ecsqkhj@nus.edu.sg*.

¶ Department of Economics, Hong Kong University of Science and Technology. Email: *ruitang.economics@gmail.com*.

1 Introduction

The seminal results in revealed preference analysis begin with a dataset collected from a consumer and find conditions that are necessary and sufficient for it to be consistent with rationality. To be specific, suppose that the consumer chooses bundles from the consumption space $\mathbb{R}^n_+$. The dataset is a finite set of observations $\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$ indexed by t, where A^t and B^t are nonempty subsets of $\mathbb{R}^n_+$ with $A^t \subseteq B^t$. With the interpretation that A^t is the set of bundles which the consumer has been observed to choose from the *budget set* B^t, there are two natural concepts of rationalization.

The **first concept** requires a preference $\succsim$, i.e., a complete and transitive binary relation on the consumption space, with its strict part denoted by $\succ$, such that the optimal choices within B^t consist precisely of A^t, i.e.,

for all $x \in A^t$, we have $x \succsim y$ for all $y \in B^t$ and $x \succ y$ for all $y \in B^t \setminus A^t$.

Richter's Theorem (Richter, 1966) characterizes those datasets $\mathcal{O}$ which are rationalizable in this sense. The **second concept** requires a preference $\succsim$ such that every bundle in A^t is optimal (with respect to $\succsim$) but allows for the possibility that bundles in $B^t \setminus A^t$ are optimal; in other words, it simply requires that

for all $x \in A^t$, we have $x \succsim y$ for all $y \in B^t$.

Afriat's Theorem (and its generalizations to nonlinear domains in Forges and Minelli (2009) and Nishimura, Ok and Quah (2017)) characterize datasets that satisfy this concept of rationalization. Loosely speaking, the first notion is the one commonly used in the theoretical revealed preference literature, while empirical work using revealed preference have mostly relied on the second (weaker) notion, which is unsurprising since the second concept does not posit that the observer has observed all the optimal choices, but only one, or some, of them.

A **third concept** of rationalization has been studied by Fishburn (1976). This concept involves a *different* interpretation of the dataset, where the observations are thought to be *coarse*. By this we mean that the observer knows that, when presented with the budget B^t, the agent has chosen from among the bundles in

A^t, but does not know precisely which bundle in A^t was chosen. To be precise, Fishburn requires the existence of a preference $\succsim$ such that for every $t \in T$,

$$\textit{there exists } x^t \in A^t \textit{ with } x^t \succsim y \textit{ for all } y \in B^t \textit{ and } x^t \succ y \textit{ for all } y \in B^t \setminus A^t.$$

In other words, Fishburn's concept of rationalization relaxes the first concept (of rationalization) by allowing some elements of A^t to be non-optimal, but it retains the requirement that nothing outside of A^t is optimal. This suggests that a **fourth concept** of rationalization may be useful in empirical applications: one that allows for the possibility that some elements in A^t are non-optimal (following Fishburn) and also that some elements outside of A^t are optimal (following Afriat). Formally, this concept simply requires the existence of a preference $\succsim$ such that for all $t \in T$,

$$\textit{there exists } x^t \in A^t \textit{ with } x^t \succsim y \textit{ for all } y \in B^t.$$

This rationalization concept, which we shall refer to as *coarse rationalization*, is the focus of our paper.

The revealed preference literature since the 1970s has by and large neglected Fishburn's rationalization concept. We think that Fishburn's concept, as well as the relaxation of that concept which we just proposed, deserves notice because they are relevant to empirical applications of revealed preference. These concepts are applicable whenever the observer knows (or hypothesizes) that there is an optimal choice found in A^t, but is agnostic about precisely which alternatives within A^t are optimal. There are at least three scenarios in which it is useful to think of coarse rationalization.

(1) The most obvious cases are those where the bundles chosen are known to be imprecise. For example, a researcher may have information on how much is spent on broad categories of goods, without knowing the allocation within each category. An economist can estimate a worker's total income based on his hourly wage, but may only have a rough idea of his choices regarding leisure and consumption goods. Alternatively, a researcher may have records on a consumer's credit card purchases, which put a *lower* bound on how much is spent each month on different goods. Since there could be goods bought with cash, relying merely on such records, the researcher may not be able to recover the consumer's precise breakdown of monthly

expenditure on each good.

(2) There could be situations where some alternative x^t is recorded as the choice from B^t but, in testing for rationality or estimating the preference, the researcher may wish to accommodate the possibility that choices were observed with error; this could be accomplished by defining a neighborhood A^t around x^t (in some sense appropriate to the specific context) and then checking if $\mathcal{O}$ is coarsely rationalizable.

(3) In experimental settings, it is common to find subjects whose choice behavior is not exactly consistent with rationality. Since the choices x^t are typically observed perfectly, the rationality violations are not due to observational errors. Nonetheless, one could still use the size of the neighborhood A^t around x^t (suitably measured) as a way of comparing the rationality of different experimental subjects; those who require larger A^ts to rationalize their behavior can be deemed less rational.

In Section 2 of this paper we formulate a condition called the *never-covered property* (NCP) which is necessary and sufficient for a dataset $\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$ to admit a coarse rationalization (i.e., rationalization according to the fourth concept) by a continuous and strictly increasing utility function.[1] This result could be thought of as a generalization of Afriat's Theorem, which characterizes rationalizability by a continuous and strictly increasing utility function in the case where A^t is a singleton.[2] It is well-known that in Afriat's Theorem, rationalization is characterized by the generalized axiom of revealed preference (GARP); this concept coincides with the never-covered property when A^ts are singleton sets. In the case where B^ts are linear budget sets, i.e.,

$$B^t = \{x \in \mathbb{R}^n_+ : p^t \cdot x \leq 1\} \tag{1}$$

for some strictly positive price vector p^t, GARP can be easily checked; in this case, we show that there is a computationally efficient way of checking NCP as well, which facilitates the use of the concept in empirical applications.

Our concept of coarse rationalization pertains to situations in which a

[1] By this we mean that $\mathcal{O}$ has a coarse rationalization by a preference that can be represented by such a utility function.

[2] In this case, the second and fourth concepts of rationalization coincide.

consumer's choices are not perfectly observed. A related and natural question is how to test the rationalizability of a consumer's choices when the budget set is not precisely known, perhaps because prices are imperfectly observed. In this case a dataset has the form $\mathcal{O}^* = \{(x^t, \{B^{t,s}\}_{s\in G_t})\}_{t\in T}$ where for each observation $t \in T$, the bundle x^t is the observed choice made by the consumer, while the true budget set from which x^t is chosen is only known to be a set in the collection $\{B^{t,s}\}_{s\in G_t}$. We could then ask whether $\mathcal{O}^*$ has a *dual coarse rationalization* in the sense that there is a selection $s_t \in G_t$, for every observation $t \in T$, and a preference such that x^t is optimal in B^{t,s_t} for all $t \in T$.

In Section 3, we show that when the budget sets $B^{t,s}$ are linear budget sets, then the dual coarse rationalizability of $\mathcal{O}^*$ is equivalent to the coarse rationalizability of some dataset $\mathcal{O}^{**} = \{(\underline{A}^t, \underline{B}^t)\}_{t\in T}$ (which can be straightforwardly constructed from $\mathcal{O}^*$). Therefore, the results developed in Section 2 can be applied to ascertain if $\mathcal{O}^*$ admits a dual coarse rationalization by a continuous and strictly increasing utility function.

Section 4 applies our results to the computation of rationality indices. We assume that $A^t = \{x^t\}$ is a singleton and that B^t is a linear budget set, so the dataset could be written as $\mathcal{D} = \{(x^t, B^t)\}_{t\in T}$. Afriat's Theorem tells us that GARP is a necessary and sufficient condition for $\mathcal{D}$ to be rationalized by a continuous and strictly increasing utility function. However, it is common in empirical applications for subjects to fail GARP. Various indices have been proposed to measure the severity of a subject's departure from rationality and we focus on two such indices.

One index, which we call the *perturbation index*, measures the severity of rationality violations by measuring the extent to which the observed bundles x^t have to be perturbed for the dataset to be rationalizable (while keeping the budget set at observation t fixed at B^t). Such an approach is intuitive and similar to the one adopted by Varian (1985) to measure deviations from cost-minimizing factor demand in a production model (see Online Appendix B for a fuller discussion). Another index, the *price misperception index* was proposed by de Clippel and Rozen (2023) and measures the extent to which p^t (see (1)) have to be altered (which can be interpreted as price misperception by the consumer) in order to restore rationality. Our main results enable both indices to be calculated with ease. In Online Appendix A, we also provide an illustration of the performance of our algorithm by computing the perturbation indices of subjects in a portfolio choice experiment carried out by Choi et al. (2007).

2 Coarse Rationalizability

In this section, we formulate the notion of coarse rationalization and show that a dataset admits such a rationalization if and only if it satisfies the never-covered property. This property requires each subset of observations to contain an observation that is not (in a sense we make specific) revealed dominated by other elements in the subset.

2.1 Basic Concepts

Let the consumption space be $\mathbb{R}^n_+$. A consumer's *preference* $\succsim$ is a binary relation on $\mathbb{R}^n_+$ that is complete and transitive. We use $\succ$ to denote the asymmetric part of $\succsim$, and refer to it as a *strict preference*.[3] A preference $\succsim$ is *continuous* if, whenever $x \succ y$, there are open neighborhoods N_x and N_y of x and y respectively, such that for all $x' \in N_x$ and $y' \in N_y$, $x' \succ y'$. It is well-known that any continuous preference on $\mathbb{R}^n_+$ admits a continuous utility representation $U : \mathbb{R}^n_+ \to \mathbb{R}$, i.e., $U(x) \geq U(y)$ if and only if $x \succsim y$ (see Debreu (1954)). A preference $\succsim$ is *locally nonsatiated* if, for every $x \in \mathbb{R}^n_+$ and every open neighborhood N_x of x, there is $x' \in N_x$ such that $x' \succ x$. Local nonsatiation holds if $\succsim$ is *increasing*; by this we mean that $x \succsim y$ if $x \geq y$ and $x \succ y$ if $x \gg y$.[4] The preference $\succsim$ is *strictly increasing* if $x \succ y$ whenever $x > y$. Clearly, if $\succsim$ admits a utility representation $U : \mathbb{R}^n_+ \to \mathbb{R}$, then U will be an increasing (strictly increasing) function if $\succsim$ is increasing (strictly

[3] A *binary relation* R on X is a nonempty subset of $X \times X$. We write xRy to mean that $(x, y) \in R$. The binary relation R is *reflexive* if for all $x \in X$, xRx, *transitive* if for all $x, y, z \in X$, xRy and yRz imply xRz, and *complete* if for all $x, y \in X$, either xRy or yRx holds. The *asymmetric part of* R is the binary relation P on X such that xPy if xRy but not yRx.

[4] For any two n-dimensional vectors x and y, $x \geq y$ means that for each $i = 1, ..., n$, $x_i \geq y_i$, $x > y$ means that $x \geq y$ and $x \neq y$, and $x \gg y$ means that for each $i = 1, ..., n$, $x_i > y_i$.

increasing).[5] We refer to U as *regular* if it is continuous and strictly increasing.

For $x \in \mathbb{R}^n_+$ and $B \subseteq \mathbb{R}^n_+$, we write $x \succsim B$ if $x \succsim y$ for all $y \in B$ and $x \succ B$ if $x \succ y$ for all $y \in B$. We say that x is $\succsim$*-optimal* in B if $x \in B$ and $x \succsim B$; the set of $\succsim$-optimal elements in B is denoted by $\max(B; \succsim)$. Obviously, if $\succsim$ admits a utility function U, then for all $x \in \max(B; \succsim)$ and $y \in B$, we have $U(x) \geq U(y)$.

Suppose a researcher has collected a finite set of observations $\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$ (indexed by t), where A^t and B^t are nonempty subsets of $\mathbb{R}^n_+$ and $A^t \subseteq B^t$. We interpret B^t as the budget set from which the consumer chooses at observation t and A^t as the subset of B^t from which the choice was made. Following Forges and Minelli (2009), we assume that, for each t,

$$B^t = \{x \in \mathbb{R}^n_+ : g^t(x) \leq 0\}$$

where $g^t : \mathbb{R}^n_+ \to \mathbb{R}$ is a continuous and increasing function. Clearly, this formulation covers the case where B^t is a classical linear budget set; in this case, there is a price vector $p^t \gg 0$ prevailing at observation t, such that B^t consists of those bundles that cost less than the agent's wealth (normalized at 1), i.e., $B^t = L(p^t)$ where, for any $p \gg 0$,[6]

$$L(p) = \{x \in \mathbb{R}^n_+ : p \cdot x \leq 1\}. \tag{2}$$

In this case, $g^t(x) = p^t \cdot x - 1$, which is a continuous and strictly increasing function of x.

The following rationalization concept on $\mathcal{O}$ is the focus of our paper.

Definition 1. *$\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$ has a* coarse rationalization *by the preference $\succsim$ on $\mathbb{R}^n_+$ if for all $t \in T$,*

$$\max(B^t; \succsim) \cap A^t \neq \emptyset.$$

When we refer to $\mathcal{O}$ as being coarsely rationalized by a utility function U, we mean

[5] To be precise, we refer to a function $g : \mathbb{R}^n_+ \to \mathbb{R}$ as increasing if for all $x, y \in \mathbb{R}^n_+$, $x \geq y$ implies $g(x) \geq g(y)$ and $x \gg y$ implies $g(x) > g(y)$. We refer to g as being *strictly increasing* if, for all $x, y \in \mathbb{R}^n_+$, $x > y$ implies $g(x) > g(y)$.

[6] It is without loss of generality to normalize the agent's wealth to 1: for any price vector p^t and wealth $I^t > 0$ with which the associated budget set is $B^t = \{x \in \mathbb{R}^n_+ : p^t \cdot x \leq I^t\}$, we can consider an alternative price vector $\hat{p}^t = p^t / I^t$ such that $B^t = L(\hat{p}^t)$.

that it is being rationalized by the preference U induces.

When does a dataset $\mathcal{O} = \{(A^t, B^t)\}_{t\in T}$ admit a coarse rationalization? As stated, our question has a trivial answer, since any dataset can be coarsely rationalized by a preference where all bundles are deemed to be indifferent. However, the answer is no longer trivial once we require the preference to be locally nonsatiated. Our main result is a generalization of the well-known theorem of Afriat (1967), which, in the special case where A^t is singleton at each $t \in T$, characterizes those datasets that are rationalizable by a locally nonsatiated preference.

2.2 Afriat's Theorem and Revealed Dominated Observations

We refer to a dataset where A^t is a singleton (with $A^t = \{x^t\}$) as a *standard dataset* and write it as $\mathcal{D} = \{(x^t, B^t)\}_{t\in T}$. Afriat's Theorem states that $\mathcal{D}$ is rationalizable by a locally nonsatiated preference $\succsim$ (in the sense that $x^t \succsim B^t$ for all $t \in T$) if and only if $\mathcal{D}$ obeys the *generalized axiom of revealed preference* (GARP).[7] To understand this property, let $Y = \{x^t\}_{t\in T}$. For x^t and $x^{t'}$ in Y, we say that x^t is *revealed preferred to* $x^{t'}$, and denote it by $x^t \, R \, x^{t'}$, if $x^{t'} \in B^t$. We say that x^t is *revealed strictly preferred to* $x^{t'}$, and denote it by $x^t \, P \, x^{t'}$, if $x^{t'} \in \mathring{B}^t$ (the interior of B^t). The bundles $x^{t_1}, x^{t_2}, \ldots, x^{t_n}$ in Y form a *revealed preference cycle* if

$$x^{t_1} \, R \, x^{t_2} \, R \cdots R \, x^{t_n} \text{ and } x^{t_n} \, P \, x^{t_1}. \tag{3}$$

GARP requires that Y contain no revealed preference cycles. It is straightforward to check that if $\mathcal{D}$ is rationalizable by a locally nonsatiated preference $\succsim$, then $x^t \succsim x^{t'}$ if x^t is revealed preferred to $x^{t'}$ and $x^t \succ x^{t'}$ if x^t is strictly revealed preferred to $x^{t'}$; given this, it is clear that GARP is a necessary condition for $\mathcal{D}$ to be rationalizable by a locally nonsatiated preference.

For our purposes, it is useful to develop a different but equivalent formulation of GARP. For any $T' \subseteq T$, we say that the observation $s \in T'$ is *revealed dominated*

[7] Proofs of this result when B^t are linear budget sets can be found in Afriat (1967) and Varian (1982). For the case of more general budget sets, see Forges and Minelli (2009) and Nishimura, Ok and Quah (2017). The term *generalized axiom of revealed preference* follows Varian (1982).

in T' (or that x^s is revealed dominated in $Y(T') = \{x^t\}_{t\in T'}$) if there is x^q in $Y(T')$ such that $x^q \, P \, x^s$ or there is x^q, $x^{q_1}, \ldots, x^{q_m}$ in $Y(T')$ such that

$$x^q \, P \, x^{q_1} \, R \, x^{q_2} \, R \cdots R \, x^{q_m} \, R \, x^s. \tag{4}$$

It is straightforward to check that if there is no revealed preference cycle made up of elements of $Y(T')$, then T' contains an observation that is not revealed dominated in T'. Notice also that if the observations $T' = \{t^1, t^2, \ldots, t^n\}$ form a revealed preference cycle in the sense of (3), then each bundle x^{t_k} is revealed dominated by x^{t_n} (including x^{t_n} itself). Thus GARP is equivalent to the following property: *every* $T' \subseteq T$ *contains an observation that is not revealed dominated in* T'. It is this property that, appropriately generalized, characterizes coarse rationalizability.

2.3 The Never-Covered Property

Given a coarse dataset $\mathcal{O} = \{(A^t, B^t)\}_{t\in T}$, we introduce an iterative procedure that allows us to build up the set of *revealed dominated observations* in any nonempty $T' \subseteq T$.

Let $\Phi^0(T') := \emptyset$, and let $\Phi^1(T')$ consist of $t \in T'$ such that A^t is contained in $\mathring{B}(T') = \cup\{\mathring{B}^t : t \in T'\}$,[8] i.e.,

$$\Phi^1(T') := \left\{t \in T' : A^t \subseteq \mathring{B}(T')\right\}.$$

Given $\Phi^1(T')$, we can construct

$$\Phi^2(T') := \left\{t \in T' : A^t \subseteq \mathring{B}(T') \bigcup B(\Phi^1(T'))\right\}.$$

and more generally,

$$\Phi^{m+1}(T') := \left\{t \in T' : A^t \subseteq \mathring{B}(T') \bigcup B(\Phi^m(T'))\right\}.$$

for $m = 2, 3, \ldots$. Obviously, $\Phi^m(T')$ is an increasing sequence in m (in the set inclusion sense). Since T' is finite, there is some m^* at which $\Phi^{m^*}(T') = \Phi^{m^*+1}(T')$. We define $\Phi(T') := \Phi^{m^*}(T')$ and refer to $\Phi(T')$ as the set of *revealed dominated observations* (or simply *dominated observations*) in T'.

To understand this terminology, suppose $\mathcal{O}$ admits a coarse rationalization

[8] More generally, for any collection of sets $\{C^t\}_{t\in T'}$ where $T' \subseteq T$ and $C^t \subseteq \mathbb{R}^n_+$, we use $C(T')$ to denote the set $\cup\{C^t : t \in T'\}$.

by a locally nonsatiated preference $\succsim$. For each $t \in T'$, pick $x^t \in A^t$ such that $x^t \succsim B^t$ and let $\hat{x} \in \max(\{x^t\}_{t \in T'}; \succsim) \subseteq A(T')$. Since $\succsim$ is locally nonsatiated, we must have $\hat{x} \succ \mathring{B}(T')$. It follows that $\hat{x} \succ A^t$ for each $t \in \Phi^1(T')$ and since $A^t \cap \max(B^t; \succsim) \neq \emptyset$, we know that $\hat{x} \succ B^t$ for each $t \in \Phi^1(T')$. In other words, $\hat{x} \succ B(\Phi^1(T'))$. Furthermore, if $A^{\underline{t}} \subseteq \mathring{B}(T') \bigcup B(\Phi^1(T'))$ for some $\underline{t} \in T'$, then $\hat{x} \succ A^{\underline{t}}$ and $\hat{x} \succ B^{\underline{t}}$; therefore, $\hat{x} \succ B(\Phi^2(T'))$. More generally, $\hat{x} \succ B(\Phi^m(T'))$ for $m \geq 1$, which justifies our terminology.

Since $\hat{x} \in A(T')$, there is $\hat{t} \in T'$ at which $\hat{x} \in A^{\hat{t}}$. Since $\hat{x}$ cannot be strictly preferred to itself, $\hat{t} \notin \Phi(T')$ and thus $T' \setminus \Phi(T')$ is nonempty. This motivates the following definition.

Definition 2. $\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$ *satisfies* the never-covered property *(NCP) if* $\Phi(T')$ *is a strict subset of* T' *for every nonempty* $T' \subseteq T$*; in other words, every nonempty* T' *contains an observation that is not revealed dominated in* T'*.*

Note that when $A^t = \{x^t\}$ for all $t \in T$, then $s \in \Phi(T')$ for $T' \subseteq T$ if and only if x^s is revealed dominated in $Y(T')$ in the sense given by (4). Thus NCP reduces to GARP in this case.

We have shown that NCP is a necessary condition for $\mathcal{O}$ to be coarsely rationalizable by a locally nonsatiated preference. Theorem 1 below shows that it is also sufficient. Indeed, whenever $\mathcal{O}$ satisfies NCP then it can be coarsely rationalized by a continuous and increasing utility function.

Theorem 1. *The following statements on* $\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$ *are equivalent:*

(1) *$\mathcal{O}$ can be coarsely rationalized by a locally nonsatiated preference.*

(2) *$\mathcal{O}$ satisfies NCP.*

(3) *$\mathcal{O}$ can be coarsely rationalized by a continuous and increasing utility function.*

To show the sufficiency of NCP, we first show that if $\mathcal{O}$ satisfies NCP, then we can find $x^t \in A^t$ such that $\mathcal{D} = \{(x^t, B^t)\}_{t \in T}$ satisfies GARP. The details of how this is done are found in the Appendix. A continuous and increasing utility function that rationalizes $\mathcal{D}$ is then guaranteed by Afriat's Theorem.[9]

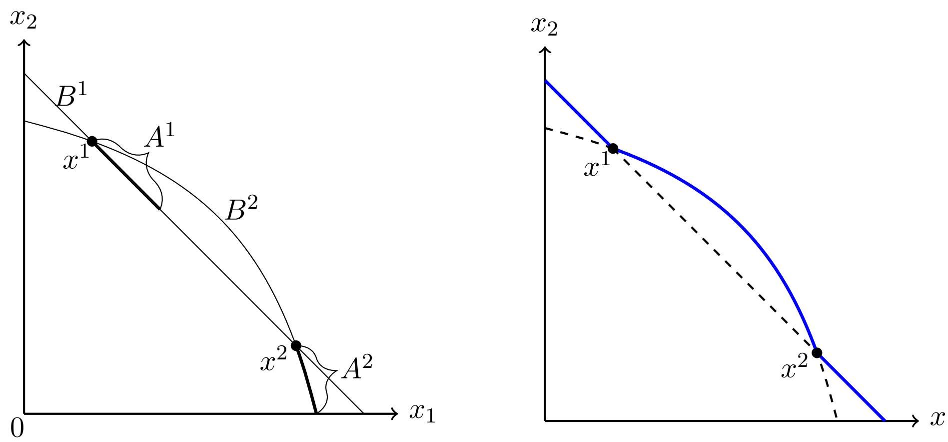
In the case where B^t are linear budget sets (see (2)) with $p^t \gg 0$, we know that if $\mathcal{D} = \{(x^t, B^t)\}_{t \in T}$ satisfies GARP, then it can be rationalized by a *strictly* increasing, continuous, *and concave* utility function (see Afriat (1967)). This leads immediately to the following, sharper, version of Theorem 1.

Theorem 2. *The following statements on $\mathcal{O} = \{(A^t, L(p^t))\}_{t \in T}$ are equivalent:*

(1) *$\mathcal{O}$ can be coarsely rationalized by a locally nonsatiated preference.*

(2) *$\mathcal{O}$ satisfies NCP.*

(3) *$\mathcal{O}$ can be coarsely rationalized by a strictly increasing, continuous, and concave utility function.*



(a) $\mathcal{O} = \{(A^1, B^1), (A^2, B^2)\}$ (b) Indifference curve through x^1 and x^2

Figure 1: Coarse Rationalization

Example 1. The coarse dataset depicted in Figure 1a consists of two observations, so we may write $T = \{1, 2\}$. For $T' = \{1\}$, we have $\Phi(T') = \emptyset$ since A^1 is on the boundary of B^1 and so $\Phi(T') \neq T'$ as required by NCP. Clearly, the same holds for $T' = \{2\}$. For $T' = \{1, 2\}$, notice that $x^1 \in A^1$ is not contained in $\mathring{B}^1 \cup \mathring{B}^2$ and neither is $x^2 \in A^2$. Thus $\Phi(T') = \Phi^1(T') = \emptyset \neq T'$. We conclude that this dataset satisfies NCP. It follows that there ought to be a selection from A^1 and A^2 so that

[9] More precisely, we appeal to the version of Afriat's Theorem in Forges and Minelli (2009) (see Appendix).

the resulting dataset obeys GARP; indeed $\mathcal{D} = \{(x^1, B^1), (x^2, B^2)\}$ obeys GARP. Figure 1b depicts the indifference curve passing through x^1 and x^2 of a preference that rationalizes the data; notice that this indifference curve is not convex. Indeed it is quite clear that *any* locally nonsatiated preference that coarsely rationalizes $\mathcal{O}$ must have both x^1 and x^2 as optimal bundles in B^1 (and in B^2) and that such a preference cannot be convex. This does not contradict Theorem 2 since B^2 is not a linear budget set. $\square$

Example 2.[10] It is natural to ask if there is a characterization of coarse rationalizability that is closer in form to GARP. In particular, generalizing from the case where A^t is a singleton, we could define A^t as being revealed preferred (revealed strictly preferred) to A^s if $A^s \subseteq B^t$ ($A^s \subseteq \mathring{B}^t$); a no-cycling condition can then be imposed on these relations in a way analogous to GARP. Such a condition is *necessary* for $\mathcal{O}$ to be coarse rationalizable, but it is not sufficient. For example, consider the dataset depicted in Figure 2, which has three observations, with $A^1 = \{a^1, a^2\}$, and $A^2 = A^3 = \{x\}$. Notice that (i) $x \in \mathring{B}^1$, (ii) $A^1 \not\subseteq B^2$ and $A^1 \not\subseteq B^3$, and (iii) $A^1 \subseteq \mathring{B}^2 \cup \mathring{B}^3$. From (i) and (ii), we know that the revealed preference relations do not have a cycle, but these relations do not account for (iii). In fact, there is *no* locally nonsatiated preference $\succsim$ that coarsely rationalizes this dataset: if $\succsim$ exists, then (i) implies that either $a^1 \succ x$ or $a^2 \succ x$ while (iii) implies that $x \succ a^1$ and $x \succ a^2$, leading to a contradiction. Consistent with Theorem 1, since $\{a^1, a^2, x\} \subset \mathring{B}(T)$, we have $T = \Phi(T)$ and NCP is violated. $\square$

Variations on Coarse Rationalization. Nishimura, Ok and Quah (2017) extend Afriat's Theorem by characterizing $\mathcal{D} = \{(x^t, B^t)\}_{t \in T}$ that can be rationalized by a utility function that is continuous and increasing with respect to a given preorder (with the product order as a special case). For example, when an agent is choosing among bundles of contingent consumption, with states having commonly known probabilities, it would be natural to require a rationalizing utility function to be increasing with respect to first order stochastic dominance; there is a suitably

[10] We would like to thank the referee whose comments inspired this example.

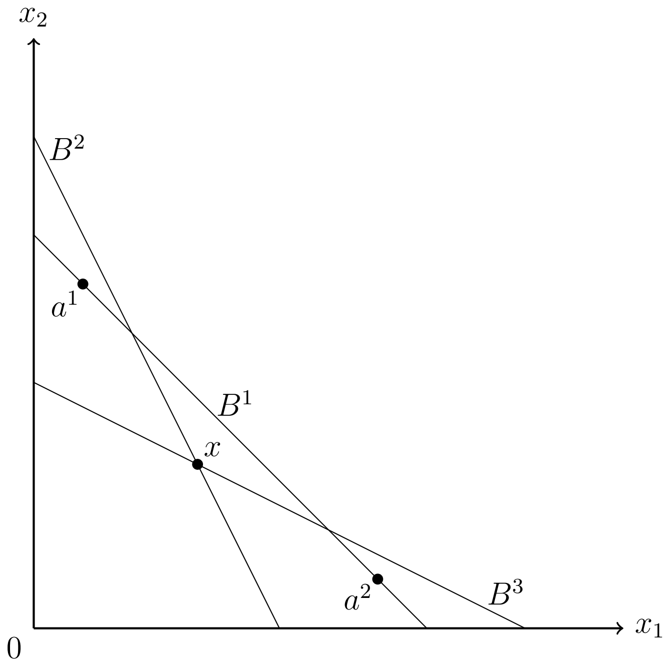


Figure 2: Violation of NCP

modified version of GARP that could test if $\mathcal{D}$ admits such a rationalization. Theorem 1 can be similarly extended to characterize, via suitably modified versions of NCP, coarse rationalizability with respect to the types of utility families studied in Nishimura, Ok and Quah (2017). For details, see Hu et al. (2022).

Partial Congruence Axiom. Fishburn (1976) studies when $\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$ can be rationalized by a preference $\succsim$ in the following sense: for all $t \in T$, there is $x^t \in A^t$ with $x^t \succsim B^t$ and $x^t \succ B^t \setminus A^t$. It shows that this holds if and only if $\mathcal{O}$ satisfies the *partial congruence axiom*, which requires

$$A(T') \not\subseteq \bigcup_{t \in T'} (B^t \setminus A^t) \text{ for all nonempty } T' \subseteq T. \tag{5}$$

Notice that Fishburn's rationalization concept is not comparable with coarse rationalization as characterized by Theorem 1. It is weaker in the sense that $\succsim$ is not required to be locally nonsatiated or strictly increasing, but it is stronger in the sense that it requires elements in $B^t \setminus A^t$ to be *non*-optimal. Correspondingly, the partial congruence axiom is neither stronger nor weaker than NCP. For example, a single observation (A^1, B^1) with A^1 in the interior of B^1 violates NCP but satisfies the partial congruence axiom. On the other hand, the dataset depicted in Figure 1a violates (5) (and hence the partial congruence axiom) for $T' = T$. This is consistent with the fact that x^1 and x^2 are in both B^1 and B^2 and thus there cannot be a preference $\succsim$ for which x^1 is optimal in B^1 but not x^2 and x^2 is optimal in B^2 but not x^1.[11]

Coarse Rationalization as an Alternative to Product Aggregation. In studies of consumer demand, a researcher would often not have information on the demand for every relevant good. One way of addressing this issue is to perform an aggregation procedure across goods, even though this approach is strictly valid only under stringent conditions on the utility function and/or the pattern of prices changes. To be specific, suppose that at observation t, the information available consists of the prices of all goods $p^t \in \mathbb{R}^n_{++}$, the demand for the first $m-1$ goods, and the total expenditure on the remaining goods (which we denote by $c^t_{m,n}$). In other words, the specific demands for goods m, $m+1, \dots, n$ are not observed. To get round this problem, the researcher could construct a price index for those goods, $\bar{p}^t_m$, which would be a function of their prices $(p^t_m, p^t_{m+1}, \dots, p^t_n)$, with the corresponding demand for the composite good being $\bar{x}^t_m = c^t_{m,n}/\bar{p}^t_m$. In this way, the researcher creates a dataset of the standard form, with observation t consisting of the price vector $(p^t_1, p^t_2, \dots, \bar{p}^t_m)$ and the demand $(x^t_1, x^t_2, \dots, \bar{x}^t_m)$ for the m goods.

Coarse rationalization offers an alternative approach to tackle this problem. At observation t, since x^t_i for $i = 1, \dots, m-1$ and $c^t_{m,n}$ are observed, the bundle chosen by the consumer must lie in the set

$$A^t = \left\{ x \in \mathbb{R}^n_+ : x_i = x^t_i \text{ for } i = 1, \dots, m-1 \text{ and } \textstyle\sum_{i=m}^n p^t_i x_i = c^t_{m,n} \right\}. \qquad (6)$$

The corresponding coarse dataset is $\mathcal{O} = \{(A^t, B^t)\}_{t \in T}$, where

$$B^t = \left\{ x \in \mathbb{R}^n_+ : p^t \cdot x \leq \sum_{i=1}^{m-1} p^t_i x^t_i + c^t_{m,n} \right\}.$$

Theorem 1 can be used to ascertain the rationalizability of this coarse dataset.

[11] Fishburn's result allows T to be an infinite set and A^t and B^t to be nonempty sets in an arbitrary space of alternatives. We follow Afriat's Theorem by requiring T to be finite and A^t and B^t to be in a Euclidean space because we want to impose topological conditions on the rationalization. In our setting, rationalization is by a *continuous* utility function; this guarantees important features such as the existence of optima on compact sets.

As an illustration, suppose that $\mathcal{O}$ consists of two observations where

$$p^1 = (2, 2.5, 3.5), \quad x_1^1 = 1.5, \quad c_{2,3}^1 = 9;$$
$$p^2 = (4, 3, 3), \quad x_1^2 = 3, \quad c_{2,3}^2 = 4.5.$$

This dataset is coarsely rationalizable. Indeed, $\tilde{x} = (1.5, 9/2.5, 0)$ is in A^1 but $p^2 \cdot \tilde{x} = 16.8 > 16.5$, so it is not in B^2. This guarantees that $\mathcal{O}$ satisfies NCP. On the other hand, suppose we aggregate goods 2 and 3 into a composite commodity, with the price of the composite being 3 (the average price of its constituent goods) at both observations 1 and 2. Then the demand for the composite good at these observations are $\bar{x}_2^1 = 9/3 = 3$ and $\bar{x}_2^2 = 4.5/3 = 1.5$. The corresponding two-good dataset has

$$p^1 = (2, 3), \quad x^1 = (1.5, 3), \quad I^1 = 12;$$
$$p^2 = (4, 3), \quad x^2 = (3, 1.5), \quad I^2 = 16.5.$$

It is straightforward to check that this dataset violates GARP.[12]

2.4 Checking NCP

Like Afriat's Theorem, Theorem 1 is potentially useful in empirical applications: to check if a dataset is coarsely rationalizable, we simply need to check whether the never-covered property (NCP) holds, i.e., whether $\Phi(T') \neq T'$ for all $T' \subseteq T$. However, on the face of it, these checks could be too numerous to be implementable since there are $2^{|T|} - 1$ nonempty subsets of T. Apart from this difficulty, there is also the issue of whether, for datasets that one is likely to encounter in applications, it is straightforward to check that $\Phi(T') \neq T'$ for a *given* T'. We address these two issues in turn.

Checking $\Phi(T') \neq T'$ on selected T'. It turns out that we can verify that NCP

[12] Varian (1988) studies a related problem where, at each observation t, the price vector $p^t = (p_1^t, \ldots, p_n^t)$ of the goods is known, and the consumer's consumption $x_{-n}^t = (x_1^t, \ldots, x_{n-1}^t)$ of the first $n-1$ goods is known. Varian (1988) shows that any such dataset is rationalizable in the sense that we can always find $x_n^t \geq 0$ at each t such that the new dataset $\{(x_{-n}^t, x_n^t, p^t)\}_{t \in T}$ is rationalizable. This result relies crucially on there being no known upper bounds on the consumer's consumption of the nth good at each t; if such bounds exist, it is quite clear that the "anything goes" result does not hold. In our setting, while the observer may not know the precise breakdown of consumption in some subset of goods, the overall expenditure on those goods *is* known: this knowledge leads to observable restrictions.

holds simply by checking that $\Phi(T') \neq T'$ on a sequence of nested subsets of T. The algorithm below outlines the procedure.

Following the convention in the computer science literature, we use k' to denote the updated value of a variable k.

Algorithm I. Set $T^0 := T$. Set $k := 1$.

START. Derive $T^k := \Phi(T^{k-1})$. Consider the following mutually exclusive cases:

(a). $T^k = \emptyset$: Stop and output *NCP holds.*

(b). $\emptyset \neq T^k \subsetneq T^{k-1}$: Go to START with $k' = k + 1$.

(c). $\emptyset \neq T^k = T^{k-1}$: Stop and output *NCP fails.*

Algorithm I checks whether $\Phi(T^k) = T^k$ for an endogenous sequence of subsets of T. For a dataset with $|T|$ observations, this algorithm necessarily terminates within $|T|$ steps. The justification for Algorithm I is provided by Proposition 1.

Proposition 1. $\mathcal{O} = \{(A^t, B^t)\}_{t\in T}$ *satisfies NCP if and only if Algorithm I outputs* "NCP holds."

We turn now to the second potential obstacle to empirical applications of our characterization of coarse rationalizability.

Checking $\Phi(T') \neq T'$ for a given T'. Whether or not this is a simple task clearly hinges on the nature of the sets A^t and B^t, but we think that such checks should be computationally undemanding in many empirical applications,

In particular, let us consider the case where B^t are linear budget sets and A^t can be defined via linear constraints, i.e., $B^t = L(p^t) = \left\{x \in \mathbb{R}^n_+ : p^t \cdot x \leq 1\right\}$ and

$$A^t = \left\{x \in \mathbb{R}^n_+ : p^t \cdot x = 1 \text{ and } C^t \cdot x \leq c^t\right\},$$

where $C^t \in \mathbb{R}^{k\times n}$ and $c^t \in \mathbb{R}^k$, representing k linear restrictions imposed on the budget line. For instance, in the application at the end of Section 2.3,

$$C^t = (0, \ldots, 0, p^t_m, \ldots, p^t_n) \in \mathbb{R}^{1\times n}$$

and $c^t = c^t_{m,n} \in \mathbb{R}^1$ (see (6)). Recall that $\Phi^1(T') = \left\{ t \in T' : A^t \subseteq \bigcup_{t\in T'} \mathring{L}(p^t) \right\}$. We calculate $\Phi^1(T')$ by checking whether each A^s (for $s \in T'$) is contained in

$\bigcup_{t\in T'} \mathring{L}(p^t)$ and this in turn can be verified by checking if there is a solution $z \in \mathbb{R}^n_+$ to the following system of linear inequalities:

$$p^s \cdot z = 1$$

$$C^s \cdot z \leq c^s \tag{7}$$

$$p^t \cdot z \geq 1 \quad \text{for all } t \in T'. \tag{8}$$

Clearly, $A^s \subseteq \bigcup_{t\in T'} \mathring{L}(p^t)$ if and only if there is *no* solution to this system of linear inequalities. Having calculated $\Phi^1(T')$ we can then calculate $\Phi^2(T')$ (which contains $\Phi^1(T')$), where

$$\Phi^2(T') := \left\{ t \in T' : A^t \subseteq \left(\bigcup_{t\in T'} \mathring{L}(p^t)\right) \bigcup \left(\bigcup_{t\in\Phi^1(T')} L(p^t)\right) \right\}.$$

This can be done using a similar inequality system, with (8) being replaced by

$$p^t \cdot z \geq 1 \text{ for all } t \in T' \qquad \text{and} \qquad p^t \cdot z > 1 \text{ for all } t \in \Phi^1(T').$$

We continue in this manner until $\Phi^m(T') = \Phi^{m+1}(T')$ at which point we can check whether $\Phi(T') := \Phi^m(T')$ is a strict subset of T'.

It follows that checking whether $\mathcal{O}$ satisfies NCP can be accomplished in *polynomial time*. Indeed, for each $m = 1, 2, \ldots$, calculating $\Phi^m(T')$ requires us to solve at most $|T'| + 1 - m$ systems of linear inequalities. Thus checking whether $T' = \Phi(T')$ involves solving at most $|T'|(|T'|+1)/2$ linear problems and verifying whether $\mathcal{O}$ satisfies NCP involves solving no more than

$$\frac{|T|(|T|+1)}{2} + \frac{(|T|-1)|T|}{2} + \frac{(|T|-2)(|T|-1)}{2} + \ldots = \frac{|T|(|T|+1)(|T|+2)}{6}$$

linear problems.[13]

In Online Appendix E, we present an alternative method for checking the coarse rationalizability of a dataset (when A^t is a linearly constrained subset of linear B^t) using Afriat inequalities.[14]

[13] For algorithms to check the solvability of a system of linear inequalities, see, for example, Karmarkar (1984).

[14] We thank an anonymous referee for suggesting this approach.

3 Dual Coarse Rationalizability

The notion of coarse rationalizability is founded on the imperfect observability of a consumer's choice from a budget set, but it is also possible for price observations to be imperfect and, in this section, we explain how this leads to a dual notion of coarse rationalizability which we could also test.

We consider a dataset of the form

$$\mathcal{O}^* = \left\{(x^t, \{L(p^{t,s})\}_{s\in G_t})\right\}_{t\in T}$$

where $T \neq \emptyset$ is finite and, for every $t \in T$, $x^t \in \cap_{s\in G_t} L(p^{t,s})$. We interpret x^t as the choice observed to be made by a consumer in observation t and $\{L(p^{t,s})\}_{s\in G_t}$ as the collection of possible budget sets from which the consumer's choice was made. In other words, the researcher observes perfectly the bundle chosen by the consumer but may have imprecise information on the corresponding price vector. We are interested in identifying those datasets where there is a preference such that, at each t, the bundle x^t is optimal according to that preference for some budget set among those in $\{L(p^{t,s})\}_{s\in G_t}$. This is stated formally as follows.

Definition 3. $\mathcal{O}^* = \{(x^t, \{L(p^{t,s})\}_{s\in G_t})\}_{t\in T}$ *has a* dual coarse rationalization *by the preference* $\succsim$ *on* $\mathbb{R}^n_+$ *if there exists a selection* $s_t \in G_t$ *for every* $t \in T$ *such that for all* $t \in T$,

$$x^t \in \max\left(L(p^{t,s_t}); \succsim\right).$$

We refer to $\mathcal{O}^*$ as being dual-coarsely rationalized by a utility function U if it is dual-coarsely rationalized by the preference induced by U. We are interested in the non-trivial case in which the preference $\succsim$ that dual-coarsely rationalizes the dataset $\mathcal{O}^*$ is locally nonsatiated. Therefore, without loss of generality, we can consider a dataset such that for every $t \in T$ and $s \in G_t$, $x^t \cdot p^{t,s} = 1$: whenever we have $x^t \cdot p^{t,\bar{s}} < 1$, we can simply remove $L(p^{t,\bar{s}})$ from the set of candidate budget sets at observation t, since local nonsatiation implies that x^t cannot be the optimal bundle in $L(p^{t,\bar{s}})$ in such a case.

To motivate our next result, suppose that $\mathcal{O}^* = \{(x^t, \{L(p^{t,s})\}_{s\in G_t})\}_{t\in T}$

admits a dual coarse rationalization by a locally nonsatiated preference. This means that there is a selection $L(p^{t,s_t})$ from $\{L(p^{t,s})\}_{s\in G_t}$ such that the standard dataset $\mathcal{D} = \{(x^t, L(p^{t,s_t}))\}_{t\in T}$ can be rationalized by a locally nonsatiated preference. But Afriat's Theorem guarantees that $\mathcal{D}$ can also be rationalized by a regular function U; furthermore, with no loss of generality, we can choose U to be bounded above by $\bar{u} \in \mathbb{R}$.[15] For each $p \in \mathbb{R}^n_{++}$, we can associate it with an indirect utility level, denoted by $V(p)$, such that

$$V(p) = \max_{x\in L(p)} U(x).$$

We further extend V to the whole domain of $\mathbb{R}^n_+$ such that for all $\hat{p} \in \mathbb{R}^n_+\backslash\mathbb{R}^n_{++}$, $V(\hat{p}) = \bar{u} + 1$, and define $W : \mathbb{R}^n_+ \to \mathbb{R}$ such that $W = -V$.

It is straightforward to check that W is locally nonsatiated.[16] Furthermore, for every $t \in T$ and $p \in \mathbb{R}^n_{++}$ with $x^t \cdot p \leq 1$, we have

$$V(p^{t,s_t}) = U(x^t) \leq \max_{x\in L(p)} U(x) = V(p),$$

where the first equality follows from the fact that U rationalizes $\mathcal{D}$ and the inequality holds since $x^t \in L(p)$ whenever $x^t \cdot p \leq 1$. In other words, p^{t,s_t} is the vector that maximizes W in $\{p \in \mathbb{R}^n_+ : x^t \cdot p \leq 1\}$.[17]

Therefore, a *necessary* condition for $\mathcal{O}^*$ to admit a dual coarse rationalization by a locally nonsatiated preference is the existence of a locally nonsatiated utility function W that coarsely rationalizes the dataset

$$\mathcal{O}^{**} := \left\{\left(\{p^{t,s}\}_{s\in G_t}, \{p \in \mathbb{R}^n_+ : x^t \cdot p \leq 1\}\right)\right\}_{t\in T}$$

By Theorem 1, this is equivalent to $\mathcal{O}^{**}$ satisfying NCP. We claim that this is also *sufficient* for the dual coarse rationalization of $\mathcal{O}^*$. Indeed, by Afriat's Theorem, at each observation t, there must be selection p^{s_t} (with $s_t \in G_t$) such that

$$\mathcal{D}' = \left\{\left(p^{t,s_t}, \{p \in \mathbb{R}^n_+ : x^t \cdot p \leq 1\}\right)\right\}_{t\in T}$$

---

[15] The boundedness of U can be ensured by taking the monotone and continuous transformation $\arctan(\cdot)$ for U.

[16] It suffices to show that for all $p, \hat{p} \in \mathbb{R}^n_+$, if $p \gg \hat{p}$, then $W(p) > W(\hat{p})$. Indeed, $p \gg \hat{p}$ implies either (i) $\hat{p} \in \mathbb{R}^n_+\backslash\mathbb{R}^n_{++}$ and $p \in \mathbb{R}^n_{++}$, indicating $W(p) \geq -\bar{u} > -\bar{u} - 1 = W(\hat{p})$, or (ii) $\hat{p}, p \in \mathbb{R}^n_{++}$, which further implies $W(p) > W(\hat{p})$ as a result of the monotonicity of U.

[17] This notion of rationalization by indirect utility and its relation to Afriat's Theorem was studied in Brown and Shannon (2000), where it was employed to develop some results in general equilibrium comparative statics.

obeys GARP (with the revealed preference relations defined on the price vectors p^{t,s_t}). It is straightforward to check that, at observations t and q, the price $p^{t,s_t} \in G_t$ is directly revealed preferred (directly strictly revealed preferred) to $p^{q,s_q} \in G_q$ if and only if x^q is directly revealed preferred (directly strictly revealed preferred) to x^t.[18] From this it follows that $\mathcal{D}'$ obeys GARP if and only if $\mathcal{D} = \{(x^t, L(p^{t,s_t}))\}_{t\in T}$ (with its revealed preference relations defined on bundles) obeys GARP. Finally, Afriat's Theorem guarantees that $\mathcal{D}$ can be rationalized by a regular and concave utility function. The result below summarizes our finding.

Theorem 3. *The following statements on* $\mathcal{O}^* = \{(x^t, \{L(p^{t,s})\}_{s\in G_t})\}_{t\in T}$ *are equivalent:*

(1) $\mathcal{O}^*$ *can be dual-coarsely rationalized by a locally nonsatiated preference.*

(2) $\mathcal{O}^{**}$ *satisfies NCP.*

(3) $\mathcal{O}^*$ *can be dual-coarsely rationalized by a strictly increasing, continuous, and concave utility function.*

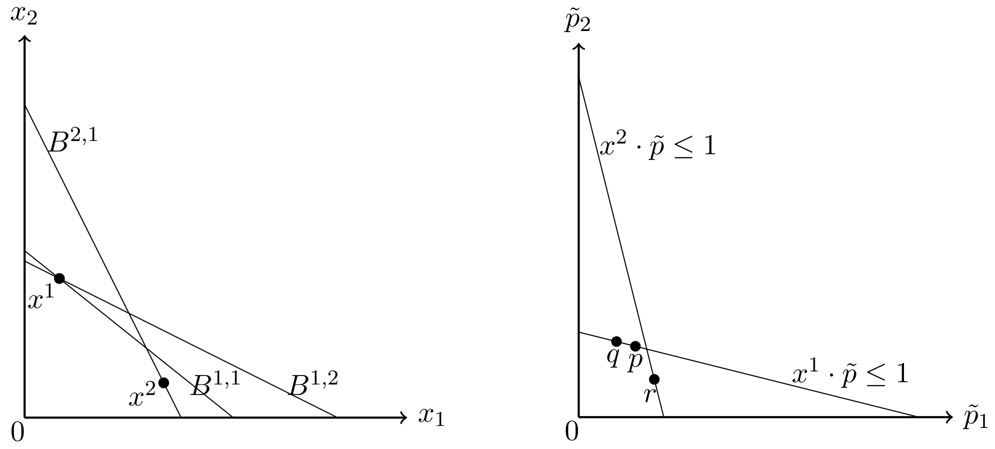
Of course, our characterization also gives us a test of the dual coarse rationalizability of $\mathcal{O}^*$ since we know (from Theorem 1) that the coarse rationalizability of $\mathcal{O}^{**}$ can be checked through the never-covered property.

Example 3. The dataset $\mathcal{O}^*$ depicted in Figure 3a contains two observations: $(x^1, \{B^{1,1}, B^{1,2}\})$ and $(x^2, \{B^{2,1}\})$, where

$$x^1 = (1,4), \quad B^{1,1} = L(p) = L(\tfrac{1}{6}, \tfrac{5}{24}), \quad B^{1,2} = L(q) = L(\tfrac{1}{9}, \tfrac{2}{9}); \text{ and}$$
$$x^2 = (4,1), \quad B^{2,1} = L(r) = L(\tfrac{2}{9}, \tfrac{1}{9}).$$

Note that no matter whether $B^{1,1}$ or $B^{1,2}$ is the true budget set, we can always reveal that x^1 is strictly better than x^2 and vice versa, indicating that $\mathcal{O}^*$ is not dual-coarsely rationalizable by a locally nonsatiated preference. The dual dataset $\mathcal{O}^{**}$ of $\mathcal{O}^*$ is depicted in Figure 3b. Since r is in the interior of $\{\tilde{p} \in \mathbb{R}^n_+ : x^1 \cdot \tilde{p} \le 1\}$ and both p and q are in that of $\{\tilde{p} \in \mathbb{R}^n_+ : x^2 \cdot \tilde{p} \le 1\}$, we conclude that NCP fails for $\mathcal{O}^{**}$. □

[18] This is because $1 = p^{t,s_t} \cdot x^t \ge (>)\, p^{q,s_q} \cdot x^t$ if and only if $1 = p^{q,s_q} \cdot x^q \ge (>)\, p^{q,s_q} \cdot x^t$.



(a) $\mathcal{O}^* = \{(x^1, \{B^{1,s}\}_{s=1}^2), (x^2, \{B^{2,1}\})\}$ (b) The dual dataset $\mathcal{O}^{**}$ of $\mathcal{O}^*$

Figure 3: Dual Coarse Rationalization

4 Rationality Indices

For a standard dataset of the form $\mathcal{D} = \{(x^t, L(p^t))\}_{t \in T}$, where $x^t \in \mathbb{R}^n_+$ is the observed choice from the budget set $L(p^t) = \{x \in \mathbb{R}^n_+ : p^t \cdot x \leq 1\}$, Afriat's Theorem tells us that GARP is a necessary and sufficient condition for $\mathcal{D}$ to be rationalized by a regular utility function. However, in most empirical applications, it is common for subjects to fail GARP. Various indices have been proposed to measure the severity of a subject's departure from rationality, with the most commonly used measure being the *critical cost efficiency index* (CCEI) due to Afriat (1973).[19] The CCEI is defined as

$$e^* := \sup\{e : \mathcal{D} \text{ is rationalized at efficiency level } e \text{ by a regular utility function}\},$$

where a utility function U *rationalizes* $\mathcal{D}$ *at cost efficiency level* $e \in (0, 1]$ if $U(x^t) \geq U(x)$ for $x \in \mathbb{R}^n_+$ that satisfies $p^t \cdot x \leq e$. Obviously, if $\mathcal{D}$ is rationalized by a regular utility function, then its CCEI is 1. One reason for the popularity of this index is its simplicity: it is easy to calculate because rationalizability at any

[19] Papers that use this concept or a related version due to Varian (1990) in empirical work include Harbaugh, Krause and Berry (2001), Andreoni and Miller (2002), Choi et al. (2007), Choi et al. (2014), Fisman, Kariv and Markovits (2007), Carvalho, Meier and Wang (2016), and Halevy, Persitz and Zrill (2018). See also Dziewulski (2020), which provides a behavioral foundation for the CCEI in terms of consumers who strictly prefer one bundle over another only when the utility difference between the two bundles is sufficiently large; a consumer with a higher CCEI is one who is better at noticing utility differences.

efficiency level e can be ascertained by a modified version of GARP.

In this section, we discuss two alternative rationality indices that, following from our results in the earlier sections, are easy to compute and which also constitute natural ways of measuring departures from rationality.

4.1 Perturbation Index

A natural way of measuring the severity of departures from rationality is to measure the extent to which the chosen bundles x^t have to be perturbed before the perturbed dataset becomes rationalizable. Such a measure was used by Varian (1985) in his analysis of production data, where the chosen bundles are bundles of factor demand. In that context, it is natural to assume that output levels at each bundle of factors are observable, which considerably simplifies the calculation of such an index, whereas the same exercise in the context of consumption is more complicated, essentially because it is *not* reasonable to assume that utility levels are observable (see Online Appendix B for more details). Nonetheless, our extension of Afriat's Theorem (with Algorithm I for checking NCP) makes it feasible to implement such a rationality measure.

To be precise, suppose the researcher records the consumer choosing x^t from $L(p^t)$; to accommodate the possibility that x^t was observed with error, the researcher could allow the true consumption bundle to be in the set

$$A^{t,\kappa} := \left\{x \in L(p^t) : \; p^t \cdot x = 1 \text{ and } |p^t_i x_i - p^t_i x^t_i| \leq \kappa \text{ for all } i\right\}, \tag{9}$$

where $\kappa \in [0, 1]$. In other words, the true expenditure on good i is allowed to deviate from $p^t_i x^t_i$ but not by more than the fraction κ of income. In experimental settings, where there is no question that x^t is indeed the observed choice, we could interpret κ as a measure of the extent to which we allow the subject to make mistakes. Whatever the interpretation, we can ask if there is a regular utility function $U : \mathbb{R}^n_+ \to \mathbb{R}$ that coarsely rationalizes

$$\mathcal{O}^\kappa = \left\{\left(A^{t,\kappa}, L(p^t)\right)\right\}_{t\in T}.$$

If so, then there are bundles $\tilde{x}^t$, such that $|p^t_i \tilde{x}^t_i - p^t_i x^t_i| \leq \kappa$ for all i and t, and $\widetilde{\mathcal{D}} = \{(\tilde{x}^t, L(p^t)\}_{t\in T}$ is rationalized by U. This is illustrated in Figure 4, where the actual dataset $\{(x^1, L^1), (x^2, L^2)\}$ is not rationalizable, but $\{(A^1, L^1), (A^2, L^2)\}$ (as depicted) is coarsely rationalizable. Indeed, so long as we choose $y \in A^1 \setminus L^2$, then $\mathcal{D} = \{(y, L^1), (x^2, L^2)\}$ obeys GARP and is rationalizable by a regular utility function.

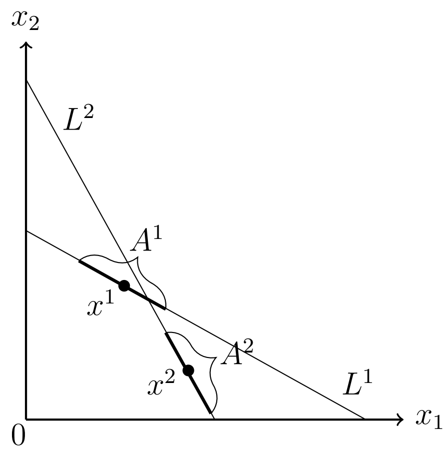


Figure 4: $\{(x^1, L^1), (x^2, L^2)\}$ is not rationalizable, but $\{(A^1, L^1), (A^2, L^2)\}$ (as depicted) is coarsely rationalizable.

We define the *critical perturbation index* (or *perturbation index*, for short) by

$$\kappa^* := \inf\{\kappa : \mathcal{O}^\kappa \text{ is coarsely rationalizable by a regular utility function}\}.$$

The index gives the smallest perturbation needed to guarantee that the coarsened dataset $\mathcal{O}^\kappa$ admits a coarse rationalization by a regular utility function. Obviously, if $\mathcal{D}$ is rationalizable to begin with, then its perturbation index equals zero.[20]

Our earlier results make the calculation of the perturbation index a straightforward task. By Theorems 1 and 2, $\mathcal{O}^\kappa$ can be coarsely rationalized by a regular utility function if and only if NCP holds. The latter could be ascertained using Algorithm I in Section 2.4, where to check whether A^s (for $s \in T'$) is contained in $\bigcup_{t \in T'} \mathring{L}(p^t)$, constraint (7) in the inequality system should be replaced by

$$|p_i^s z_i - p_i^s x_i^s| \leq \kappa \quad \text{for each good } i,$$

as specified by (9). The index κ^* can be obtained by binary search. In Online Appendix A, we illustrate the use of our algorithm by calculating the perturbation indices of subjects in the experimental data collected by Choi et al. (2007).

[20] If $\kappa = 1$, then $A^t = L(p^t) \setminus \mathring{L}(p^t)$ for all $t \in T$ and $\mathcal{O}^1$ is coarsely rationalizable. Thus the perturbation index is always well-defined.

As we indicated in Section 2.3, there are modified versions of NCP that could test if a coarse dataset such as $\mathcal{O}^\kappa$ admits a coarse rationalization by a utility function belonging to certain families. Thus perturbation indices with respect to such utility families could also be calculated.

In our definition of the perturbation index, the permissible perturbation to restore rationality is bounded by κ at every observation. A natural variation will remove this uniform bound, with the index modified to measure the average perturbation across observations. This modified version of the perturbation index (which is exactly analogous to the modification of the CCEI proposed in Varian (1990)) is discussed in greater detail in Online Appendix C.

4.2 Price Misperception Index

In Section 4.1, we show how our methods could be used to calculate the perturbation index, which measures how much bundles have to be perturbed in order to restore rationality. In this section, we demonstrate how our results can also be applied to compute the *price misperception index* (PMI) introduced by de Clippel and Rozen (2023), which evaluates a dataset's closeness to rationality by measuring the extent to which *prices* have to be perturbed to recover rationality. As shown by de Clippel and Rozen (2023), this index is also formally equivalent to an index that measures the extent to which the gradient of the utility function at chosen bundles differ from the prevailing price.

For a given dataset $\mathcal{D} = \{(x^t, L(p^t))\}_{t\in T}$, we say that $\mathcal{D}$ is rationalizable with price misperception ϵ (or simply ϵ-rationalizable) if the following holds:

$$\begin{aligned}
&\exists \text{ regular and quasiconcave } U \text{ and } \{\hat{p}^t\}_{t\in T} \subseteq \mathbb{R}^n_{++} \text{ such that} \\
&\forall t \in T,\ x^t \cdot \hat{p}^t = 1, \\
&\forall t \in T, \forall i, j \in \{1, ..., n\} \text{ with } i \neq j,\ 1-\epsilon \leq \frac{p_i^t/p_j^t}{\hat{p}_i^t/\hat{p}_j^t} \leq \frac{1}{1-\epsilon}, \\
&\forall y \in L(\hat{p}^t),\ U(x^t) \geq U(y).
\end{aligned} \tag{10}$$

In other words, the chosen bundle at each observation is optimal with respect to prices that are (in a certain sense) ϵ-perturbations of the true prices. The price misperception index is the minimum misperception we must allow the consumer in order for the data to be rationalizable, i.e.,

$$\text{PMI}(\mathcal{D}) := \inf\{\epsilon \in [0,1] : \mathcal{D} \text{ is } \epsilon\text{-rationalizable in the sense of (10)}\}.$$

Calculating PMI($\mathcal{D}$) is straightforward given our results. Indeed, for a given ϵ, checking whether condition (10) holds is equivalent to checking the dual coarse rationalizability of the dataset

$$\mathcal{O}^*_\epsilon = \left\{\left(x^t, \{L(\hat{p}^t)\}_{\hat{p}^t \in Z_{\delta,\epsilon}(p^t)}\right)\right\}_{t\in T}, \tag{11}$$

where for each t,

$$Z_{\delta,\epsilon}(p^t) := \left\{\hat{p}^t \in \mathbb{R}^n_{++} : \delta(p^t, \hat{p}^t) \le \frac{1}{1-\epsilon} \text{ and } x^t \cdot \hat{p}^t = 1\right\}.$$[21]

By Theorem 3, this is equivalent to the coarse rationalizability of

$$\mathcal{O}^{**}_\epsilon = \left\{\left(Z_{\delta,\epsilon}(p^t), L(x^t)\right)\right\}_{t\in T},$$

where $L(x^t) := \{p \in \mathbb{R}^n_+ : x^t \cdot p \le 1\}$, which is in turn equivalent to checking that $\mathcal{O}^{**}_\epsilon$ obeys NCP. We provide more details on checking NCP in this case in Online Appendix D.

Appendix

Proof of Theorem 1. Since every increasing utility function generates a increasing, thus locally nonsatiated, preference, (3) implies (1). We have already shown in the main text that (1) implies (2). It remains to show that (2) implies (3). We do this by first setting out the procedure with which x^t can be chosen from A^t so that $\mathcal{D} = \{(x^t, B^t)\}_{t\in T}$ obeys GARP.

Denote by $\mathcal{E}(T')$ the set of bundles that are revealed to be dominated through the procedure of iterated exclusion of dominated observations, i.e.,

$$\mathcal{E}(T') := \mathring{B}(T') \bigcup B\left(\Phi(T')\right).$$

Since $\mathcal{O}$ satisfies NCP, for any nonempty $T' \subseteq T$, $\Phi(T')$ is a strict subset of T', which implies that $A(T') \setminus \mathcal{E}(T') \neq \emptyset$.

Let $T_1 := T$ and $S_1 := A(T_1) \setminus \mathcal{E}(T_1)$. We proceed by induction. Suppose that

[21] $\delta(p,g) = \max_{i,j\in\{1,\ldots,n\}: i\neq j} \left\{\frac{p_i/p_j}{g_i/g_j}, \frac{g_i/g_j}{p_i/p_j}.\right\}$.

we have constructed T_k and S_k for some $k \geq 1$. If $T_k \neq \emptyset$, construct T_{k+1} and S_{k+1}:

$$T_{k+1} := \Phi(T_k) = \{t \in T_k : A^t \subseteq \mathcal{E}(T_k)\} \text{ and } S_{k+1} := A(T_{k+1}) \setminus \mathcal{E}(T_{k+1}).$$

Since $\mathcal{O}$ satisfies NCP, if $T_k \neq \emptyset$, then $T_{k+1} = \Phi(T_k) \subsetneq T_k$ and $S_k = A(T_k) \setminus \mathcal{E}(T_k) \neq \emptyset$. The construction stops when $T_{k^*} \neq \emptyset$ and $T_{k^*+1} = \emptyset$ for some k^*.

We are now ready to select x^t in A^t for each $t \in T$ such that $\mathcal{D} = \{(x^t, B^t)\}_{t \in T}$ obeys GARP. For each $1 \leq k \leq k^*$, let $V_k := T_k \setminus T_{k+1}$ denote the collection of observations that are eliminated when constructing T_{k+1} from T_k. Clearly, $\{V_k\}_{k=1}^{k^*}$ is a partition of T. By definition, for each k and each $t \in V_k = T_k \setminus T_{k+1}$, we have $A^t \setminus \mathcal{E}(T_k) \neq \emptyset$ and hence $A^t \cap S_k = A^t \cap (A(T_k) \setminus \mathcal{E}(T_k)) \neq \emptyset$.

For each k and each $t \in V_k = T_k \setminus T_{k+1}$, we pick an arbitrary $x^t \in A^t \cap S_k$. We claim that $\mathcal{D} = \{(x^t, B^t)\}_{t \in T}$ obeys GARP. Let $k(t)$ be the corresponding index k such that $t \in V_k$. It suffices to show that (1) $x^t R x^{t'}$ implies that $k(t) \leq k(t')$; and (2) $x^t P x^{t'}$ implies that $k(t) < k(t')$. Suppose that $x^t R x^{t'}$ but $k(t) > k(t')$. Then $t \in \Phi(T_{k(t')})$ due to the construction of $\{V_k\}_{k=1}^{k^*}$. It follows that $A^t \subseteq \mathcal{E}(T_{k(t')})$ and $B^t \subseteq \mathcal{E}(T_{k(t')})$. Since $x^t R x^{t'}$, we have $x^{t'} \in B^t \subseteq \mathcal{E}(T_{k(t')})$, which contradicts with $x^{t'} \in S_{k(t')} = A(T_{k(t')}) \setminus \mathcal{E}(T_{k(t')})$. Hence, $x^t R x^{t'}$ implies $k(t) \leq k(t')$. Suppose that $x^t P x^{t'}$ but $k(t) \geq k(t')$. If $k(t) > k(t')$, then we have the same contradiction as argued above. If $k(t) = k(t') = k$ for some k, then both x^t and $x^{t'}$ belong to S_k. Since $S_k = A(T_k) \setminus \mathcal{E}(T_k)$ and $\mathring{B}(T_k) \subseteq \mathcal{E}(T_k)$, we have

$$x^t, x^{t'} \in S_k \subseteq A(T_k) \setminus \mathring{B}(T_k).$$

But this is impossible since $x^t P x^{t'}$ implies $x^{t'} \in \mathring{B}^t$. Hence, $x^t P x^{t'}$ implies $k(t) < k(t')$.

By Proposition 3 of Forges and Minelli (2009), there are numbers v^t and $\lambda^t > 0$ such that, for all t, $s \in T$,

$$v^s \leq v^t + \lambda^t g^t(x^s),$$

and the utility function

$$U(x) = \min_{t \in T} \{v^t + \lambda^t g^t(x)\} \tag{12}$$

rationalizes $\mathcal{D}$ and hence coarsely rationalizes $\mathcal{O}$. Since g^t are continuous, so is U, and since g^t are increasing, U is also increasing. □

Proof of Theorem 2. Given Theorem 1, the only part of this result that still needs proving is the claim that the utility function that coarsely rationalizes $\mathcal{O}$ can be chosen to be *strictly* increasing, continuous *and concave*. But this is clear from the form of the utility function (12). Indeed, $L(p^t) = \{x \in \mathbb{R}^n_+ : g^t(x) \leq 0\}$, where $g^t(x) = p^t \cdot x - 1$. Since g is continuous, strictly increasing, and linear, U is strictly increasing, continuous and concave. □

Proof of Proposition 1. "Only if": If $\mathcal{O}$ satisfies NCP, then $\Phi(T') \neq T'$ for any nonempty $T' \subseteq T$. Thus, Case (c) never occurs when we run Algorithm I on this dataset. Furthermore, T^k is strictly decreasing in k in the set inclusion sense, and $T^{k^*} = \emptyset$ for some k^*. Therefore, Algorithm I outputs *NCP holds*.

"If": We first claim that the operator $\Phi(\cdot)$ is monotonically increasing in the set inclusion sense, i.e., if $T' \subseteq T''$ then $\Phi(T') \subseteq \Phi(T'')$. Indeed, the iteration procedure in Section 2.3 that defines $\Phi(\cdot)$ satisfies, inductively, $\Phi^m(T') \subseteq \Phi^m(T'')$ for each $m = 1, 2, \ldots,$ which results in $\Phi(T') \subseteq \Phi(T'')$.

Now suppose that Algorithm I outputs *NCP holds*. We then have a sequence of subsets of T, $\{T^0, T^1, \ldots, T^{k^*}\}$, where $T^k = \Phi(T^{k-1}) \subsetneq T^{k-1}$ for all $k = 1, 2, \ldots, k^*$ and $T^{k^*} = \emptyset$. For any nonempty $T' \subseteq T$, there exists some k such that $T' \subseteq T^k$ and $T' \nsubseteq T^{k+1}$. Since $\Phi(\cdot)$ is monotonically increasing, $\Phi(T') \subseteq \Phi(T^k) = T^{k+1}$. Since $T' \nsubseteq T^{k+1}$, we have $\Phi(T') \neq T'$. Thus, the dataset satisfies NCP. □

References

Afriat, Sydney N. 1967. "The construction of utility functions from expenditure data." *International Economic Review*, 8(1): 67–77.

Afriat, Sydney N. 1973. "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method." *International Economic Review*, 460–472.

Andreoni, James, and John Miller. 2002. "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism." *Econometrica*, 70(2): 737–753.

Brown, Donald J, and Chris Shannon. 2000. "Uniqueness, stability, and comparative statics in rationalizable Walrasian markets." *Econometrica*, 68(6): 1529–1539.

Carvalho, Leandro S, Stephan Meier, and Stephanie W Wang. 2016. "Poverty and Economic Decision-Making: Evidence from Changes in Financial Resources at Payday." *American Economic Review*, 106(2): 260–284.

Choi, Syngjoo, Raymond Fisman, Douglas Gale, and Shachar Kariv. 2007. "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review*, 97(5): 1921–1938.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. "Who is (More) Rational?" *American Economic Review*, 104(6): 1518–1550.

Debreu, Gerard. 1954. "Representation of a preference ordering by a numerical function." *Decision Processes*, 3: 159–165.

de Clippel, Geoffroy, and Kareen Rozen. 2023. "Relaxed Optimization: How Close Is a Consumer to Satisfying First-Order Conditions?" *Review of Economics and Statistics*, 105(4): 883–898.

Dziewulski, Paweł. 2020. "Just-noticeable difference as a behavioural foundation of the critical cost-efficiency index." *Journal of Economic Theory*, 188: 105071.

Fishburn, Peter. 1976. "Representable Choice Functions." *Econometrica*, 44(5): 1033–1043.

Fisman, Raymond, Shachar Kariv, and Daniel Markovits. 2007. "Individual Preferences for Giving." *American Economic Review*, 97(5): 1858–1876.

Forges, Francoise, and Enrico Minelli. 2009. "Afriat's Theorem for General Budget Sets." *Journal of Economic Theory*, 144(1): 135–145.

Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2018. "Parametric Recoverability of Preferences." *Journal of Political Economy*, 126(4): 1558–1593.

Harbaugh, William T, Kate Krause, and Timothy R Berry. 2001. "GARP for Kids: On the Development of Rational Choice Behavior." *American Economic Review*, 91(5): 1539–1545.

Hu, Gaoji, Jiangtao Li, John K.-H. Quah, and Rui Tang. 2022. "A theory of revealed indirect preference." *Available at SSRN 3776049.*

Karmarkar, Narendra. 1984. "A new polynomial-time algorithm for linear programming." In *Proceedings of the sixteenth annual ACM symposium on Theory of computing.* 302–311.

Nishimura, Hiroki, Efe A. Ok, and John K.-H. Quah. 2017. "A Comprehensive Approach to Revealed Preference Theory." *American Economic Review*, 107(4): 1239–1263.

Richter, Marcel K. 1966. "Revealed Preference Theory." *Econometrica*, 34(3): 635–645.

Varian, Hal R. 1982. "The Nonparametric Approach to Demand Analysis." *Econometrica: Journal of the Econometric Society*, 945–973.

Varian, Hal R. 1985. "Non-parametric Analysis of Optimizing Behavior with Measurement Error." *Journal of Econometrics*, 30(1-2): 445–458.

Varian, Hal R. 1988. "Revealed preference with a subset of goods." *Journal of Economic Theory*, 46(1): 179–185.

Varian, Hal R. 1990. "Goodness-of-fit in optimizing models." *Journal of Econometrics*, 46(1-2): 125–140.