Inference of Preference Heterogeneity
from Choice Data

Annie Liang*

October 4, 2016

Abstract

Suppose that an analyst observes inconsistent choices from a decision maker. Can the analyst determine whether this inconsistency arises from choice error (imperfect maximization of a single preference) or from preference heterogeneity (deliberate maximization of multiple preferences)? I model choice data as generated from context-dependent preferences, where contexts vary across observations, and the decision maker errs with small probability in each observation. I show that (a) simultaneously minimizing the number of inferred preferences and the number of unexplained observations can exactly recover the correct number of preferences with high probability; (b) simultaneously minimizing the *richness* of the set of preferences and the number of unexplained observations can exactly recover the choice implications of the decision maker's true preferences with high probability. These results illustrate that selection of simple models, appropriately defined, is a useful approach for recovery of stable features of preference.

1 Introduction

Let X be a finite set of choice alternatives, and consider an analyst who observes a decision maker's choices from various subsets of X. The standard approach for recovering the decision maker's preferences over X is to find a single ordering that "best fits" this choice data. In practice, however, such a method can obscure important heterogeneity in the decision maker's preferences. In particular:

- choice observations are often pooled from related but distinct contexts, across which preferences vary. For example, Einav, Finkelstein, Pascu & Cullen

(2012) study the stability of risk preferences across different kinds of financial decisions, and find that just over 30% of their subject pool makes decisions across all six domains that can be rationalized using a common risk preference.

- choice observations are often pooled from individuals in a population, across whom preferences vary. For example, Crawford & Pendakur (2012) study the consumption decisions of 500 Danish households over different kinds of milk, and find that no single utility function can simultaneously explain the choice behaviors of more than two-thirds of these households.

In such problems, the analyst may instead prefer to infer a set of preferences, corresponding to the different choice domains or different subpopulations present in the data. But without knowledge of which preferences are maximized in which observations, it is challenging not only to recover the preferences themselves, but even to recover the *number* of different preferences. At extremes, we can trivially assign each observation its own preference—thus inferring as many preferences as there are observations—or follow the classical approach, determining a single preference and interpreting the remaining observations as choice errors. This indeterminacy in the number of preferences is unsatisfying, because choice error and preference heterogeneity are clearly distinct sources of choice inconsistency, with different implications for welfare assessment and prediction.

An important alternative approach, proposed in Kalai, Rubinstein & Spiegler (2002), focuses on the smallest number of preferences that perfectly rationalizes the choice data (so that every observed choice is consistent with maximization of one of these preferences). This approach has since been applied in Crawford & Pendakur (2012) and Dean & Martin (2010) to determine the number preference types in household consumption decisions and individual choices over lotteries. The minimal number of preferences necessary to rationalize *every* choice observation, however, is extremely sensitive to the presence of choice error. To see this, consider a decision maker with a single preference, who errs in every choice observation with small probability. Then, a large number of preferences may be needed to explain all of the observed choices, even while a single preference explains almost all of these choices. The approach proposed by Kalai, Rubinstein & Spiegler (2002) thus exhibits a strong tendency to overfit to choice data, with important consequences for its application. In conducting welfare assessments, we may misinterpret choice errors as genuine preferences; and in predicting future choices of the decision maker, we may fail to distinguish choice inconsistencies that are stable (corresponding to intentional maximization) from choice inconsistencies that are not (corresponding to stochastic error).

The purpose of this paper is to develop a method to consistently recover the "correct" number of preferences from inconsistent choice data. The proposed ap-

proach minimizes a weighted sum of the number of recovered preferences and the number of unexplained observations (choices that cannot be rationalized by any of the recovered preferences). Intuitively, the approach imposes a cost on each recovered preference, so that a preference is recovered if and only if it explains sufficiently many observations that would otherwise be considered error. The classic Houtman & Maks (1985) and Kalai, Rubinstein & Spiegler (2002) solutions are returned for special choices of weights—the former is returned when the cost of preferences relative to unexplained observations is sufficiently high, and the latter is returned when the cost of preferences relative to unexplained observations is sufficiently low.

The main results interpret and justify this approach using a statistical model of choice based on Bernheim & Rangel (2009) and Rubinstein & Salant (2008). I assume that the decision-maker possesses a set of preferences $\{\succ_f\}_{f \in F}$, where F is a set of choice frames (corresponding, for example, to different choice domains or different subpopulations). In each choice problem, the decision-maker (DM) is presented with a choice set $A \subseteq X$ and choice frame $f \in F$, sampled from a fixed distribution over choice sets and choice frames. Given choice set A and frame f, the DM chooses the $\succ_f$-maximal observation in A with probability $1 - p$, but errs with probability p. I do not impose any additional assumptions on the distribution of the error, allowing in particular for this distribution to vary across observations (this paper thus takes a nonparametric, or "model-free" approach—see prior work in Varian (1982), Famulari (1995), Houtman & Maks (1985), and Kalai, Rubinstein & Spiegler (2002)).

Theorem 1 shows that under certain conditions on this choice model, the proposed approach exactly recovers the true number of preferences given sufficiently many observations.[1] Informally, these conditions require that: (1) both the probability of error p and the true number of preferences are sufficiently small; and (2) the DM's preferences have sufficiently differentiated choice implications. These conditions ensure that choice inconsistencies that arise from genuine preference heterogeneity will resemble other inconsistencies in the choice data, whereas choice inconsistencies that arise from error will appear idiosyncratic. The Kalai, Rubinstein & Spiegler (2002) approach is shown to recover the correct number of preferences when the probability of choice error is zero. To the best of my knowledge, this is the first statistical justification for the Kalai, Rubinstein & Spiegler (2002) approach.

Section 4 revisits an analysis conducted in Crawford & Pendakur (2012), in which the Kalai, Rubinstein & Spiegler (2002) approach is used to discover the number of preference types among 500 subjects. Crawford & Pendakur (2012) find that no more than five preferences are needed to perfectly rationalize their dataset, but the marginal benefit of the fifth ordering is low (eight out of 500 observations). I show how the proposed approach can be used to identify this fifth preference

---

[1]I assume that the DM may be presented with the same choice problem multiple times.

as fitting noise, and describe a sense in which (for purposes of welfare analysis or prediction) it should not be treated as a genuine preference.

Section 5 turns to the question of recovering the preferences themselves. Inference of multiple preferences from choice data alone is an ill-posed problem, and Section 5.1 presents several negative results that help to clarify the reasons for this. In Proposition 1, I show that most sets of orderings are indistinguishable based on their choice implications alone, so that even in the absence of choice error, most sets of multiple preferences cannot be recovered. This result is very much in the spirit of Ambrus & Rozen (2013), which studies a broad (but different) class of multi-self models and shows that these models have no testable implications without prior restrictions on the number of selves involved in a decision.[2]

In view of these results, I suggest that a more appropriate object of recovery is the set of *choice implications* of the decision maker's preferences—that is, the choice observations that are consistent with maximization of one of these preferences. I define equivalence classes for sets of preferences, where two sets belong to the same equivalence class if they have the same choice implications, and ask whether we can recover the equivalence class to which the true set of preferences belongs. Section 5 shows that this is indeed possible, but that penalizing the number of inferred preferences is not the appropriate criterion for this goal. Intuitively, penalizing only the number of preferences results in inference of sets of preferences whose choice implications are as diverse as possible. Section 5.2 proposes minimizing instead a weighted sum of the number of unexplained observations and the *richness* of the set of preferences, as measured through the number of unique choice implications. Proposition 2 shows that under certain conditions on the choice model described above, this approach will exactly recover the equivalence class of choice implications containing those of the true model. Loosely, these conditions require that the probability of error is sufficiently small, and that "informative" choice problems are sampled with sufficient probability.

Taken together, these results suggest that an appropriate penalty on the complexity of the inferred choice model—for example, as measured through the number of preferences used or the number of choice implications—can be leveraged towards recovering stable features of preference from inconsistent choice data.

Section 6 presents the related literature and Section 7 concludes.

---

[2]Ambrus & Rozen (2013) study several different choice-set independent aggregation rules over preferences, whereas I consider a specific aggregation rule (in which one preference is assigned "dictator") that varies across choice problems.

2 Approach

2.1 Preliminaries

Let $A \subseteq X$ denote a typical choice set, and let 2^X denote the set of possible choice sets. A choice observation is a pair (x, A), representing choice of alternative x from choice set A, and a dataset is a *multiset*[3] of choice observations

$$D = \{(x_1, A_1), \ldots, (x_n, A_n)\}.$$

A multiple preference rationalization is any set R of orderings over X.[4] The number of implied choice errors given data D (defined as the number of observations inconsistent with maximization of any preference in R)[5] is given by

$$\varepsilon(D, R) := \#\{(x, A) \in D : x \text{ is not } \succ\text{-maximal in } A \text{ for any } \succ \text{ in } R\}.$$

Say that R constitutes a *perfect rationalization* of D if $\varepsilon(D, R) = 0$. If we restrict to rationalizations that consist of k orderings, the minimal number of implied choice errors in D is

$$\varepsilon_k := \min_{|R|=k} \varepsilon(D, R).$$

Dependence of ε_k on the dataset D will be dropped throughout. Say that D is k-*rationalizable* (Kalai, Rubinstein & Spiegler 2002) if there is some set of k orderings that perfectly rationalizes D, so that $\varepsilon_k = 0$.

Remark 2.1. *For every dataset D, there exists a constant $L \leq \min\{|D|, |X|\}$ such that D is k-rationalizable for every $k \geq L$.*[6]

2.2 Choice Model

I consider choice data that is generated in the following way. Following Rubinstein & Salant (2008) and Bernheim & Rangel (2009), let F be an abstract set of *choice frames*. I will take $K = |F|$ to be the number of choice frames, and (without loss of generality) identify F with the set of integers $\{1, \ldots, K\}$.

A *choice problem* is a pair $(A, f) \in P_X \times F$ consisting of a choice set A and a choice frame f. Each choice frame f cues a (distinct) strict preference relation

---

[3]A multiset is a generalization of a set, in which the same element may appear multiple times.

[4]I will write as if the choice data were generated from a single individual, but the decision maker can equally well be interpreted as a population of subjects.

[5]Naturally, this is only one of many possible definitions for choice error. In particular, the number of implied choice errors will be different under different aggregations of preferences in R, for example as proposed in the multiple-ordering models of Rubinstein & Salant (2006), Fudenberg & Levine (2006), Green & Hojman (2007), Manzini & Mariotti (2007), or Manzini & Mariotti (2009).

[6]See Kalai, Rubinstein & Spiegler (2002) for a better bound than this one.

$\succ_f$ on X, which the DM imperfectly maximizes. Formally, the DM's choice rule is a map P from choice problems (A, f) to distributions over the elements of A. I will use $x^*_{(A,f)}$ to denote the $\succ_f$-maximal element in A (the alternative that the decision maker would choose under perfect maximization). The constant

$$p := \min \left\{ p' \; : \; P(A, f)[x^*_{(A,f)}] \geq 1 - p' \quad \forall \, (A, f) \right\}$$

is the smallest uniform upper bound on the DM's probability of error. That is, given any choice problem (A, f), the probability that the DM chooses the optimal choice $x^*_{(A,f)}$ is at least $1 - p$. No parametric assumptions are imposed on the distribution of error (which may also vary across different choice problems).

Remark 2.2. *Special cases of this model include the following:*[7]

(a) $p = 0$, $K = 1$ *returns perfect maximization of a single ordering.*

(b) $p = 0$, $K > 1$ *returns perfect maximization of multiple orderings. If, more-over, every choice set A is identified with a unique frame f_A, and the data includes only choice problems of form (A, f_A), then we return the multiple preference model considered in Kalai, Rubinstein & Spiegler (2002).*

Choice problems are repeatedly sampled from a distribution π over the set of unique choice problems $2^X \times F$. I will use $\mathcal{P}$ to denote a typical multiset of sampled choice problems, and π^n to mean the product measure on multisets $\mathcal{P}$ of size n, where

$$\pi^n(\mathcal{P}) = \pi^n(\{(A_1, f_1), \ldots, (A_n, f_n)\}) = \prod_{i=1}^n \pi(A_i, f_i).$$

Given choice problems $\mathcal{P}$, the DM's stochastic choice rule induces a multiset of choice realizations

$$D^* = \{(x, (A, f)) \,|\, (A, f) \in \mathcal{P}\},$$

where $(x, (A, f))$ represents choice of alternative x given choice problem (A, f). Instead of observing $(x, (A, f))$ the analyst sees only (x, A), so that the observable choice data is

$$D = \{(x, A) \,|\, (x, (A, f)) \in D^*\}.$$

(Henceforth, I will refer to this more simply as a *set* of choice observations, although it should be understood that the same observation may appear multiple times.) Thus, the analyst observes the DM's choices and choice sets, but not his choice frames. Can the analyst recover the number of distinct preferences K?

---

[7]Additionally, $p > 0$ and $K = 1$ returns a classic random utility model, in which the decision maker perfectly maximizes a preference drawn from a distribution over orderings (where the distribution is fixed across choice problems). But this paper will focus on the case in which p is small, which translates (under the random utility interpretation) to a restriction that the distribution over orderings places large weight on a single ordering.

2.3 Solution

The proposed approach simultaneously minimizes the number of orderings k and the number of implied choice errors ε_k. These two goals are traded off linearly, resulting in a family of estimates parameterized by a tradeoff λ.

Definition 2.1. *For every $\lambda \in \mathbb{R}_+$, define*

$$K_\lambda^* = \operatorname*{argmin}_{k \in \mathbb{N}} k + \lambda \varepsilon_k. \tag{1}$$

When this solution is not unique, I will take K_λ^* to mean the smallest value of k in the minimizing set.

Intuitively, $1/\lambda$ is the "cost" of each ordering, so that an ordering is attributed to the DM if and only if it explains at least $1/\lambda$ observations that would otherwise be interpreted as choice error. As $\lambda \to 0$, the cost of errors becomes increasingly small relative to the cost of orderings, so that the analyst prefers to attribute a single ordering to the DM and interpret the unexplained observations as choice errors. As $\lambda \to \infty$, the cost of choice errors becomes increasingly large relative to the cost of orderings, so that the analyst prefers to use as many orderings as necessary to perfectly rationalize the data. In particular, if $\lambda < \frac{1}{\varepsilon_1}$ (where, recall, ε_1 is the necessary number of unexplained observations if the DM is modeled with a single preference), then the approach returns the Houtman & Maks (1985) solution, and if $\lambda > 1$, then the approach returns the Kalai, Rubinstein & Spiegler (2002) solution.

Observation 1. *For every dataset D,*

  (a) $K_\lambda^* = 1$ for every $\lambda < \frac{1}{\epsilon_1}$, and

  (b) $K_\lambda^* = L$ for every $\lambda > 1$, where L is the smallest integer such that the dataset is L-rationalizable.

The main results below provide conditions on the choice model described in Section 2.2, and values of λ, under which we find the proposed approach to recover the number of orderings K.

3 Main Results

3.1 No error benchmark ($p = 0$)

Suppose first that $p = 0$, so that given any choice problem (A, f), the DM chooses the $\succ_f$-maximal observation from A with probability 1. The following example illustrates that recovery of the number of orderings K can be infeasible even in this idealized benchmark.

Example 3.1. *Let* $X = \{x_1, x_2, x_3\}$ *and suppose that the DM's preferences are* $R = \{\succ_1, \succ_2, \succ_3\}$, *where*

$$x_1 \succ_1 x_2 \succ_1 x_3$$
$$x_1 \succ_2 x_3 \succ_2 x_2$$
$$x_3 \succ_3 x_2 \succ_3 x_1$$

The number of frames is 3, but there is no value of λ *given which* $K_\lambda^* = 3$. *To see this, define the alternative set* $R' = \{\succ_1', \succ_2'\}$, *where*

$$x_1 \succ_1' x_2 \succ_1' x_3$$
$$x_3 \succ_2' x_2 \succ_2' x_1$$

Every choice observation consistent with maximization of some ordering in R *is also consistent with maximization of an ordering in* R'.[8] *Therefore, any dataset* D *that can be perfectly rationalized using* R *can also be perfectly rationalized using* R'. *Since we have assumed no choice error* $(p = 0)$, *it must be that* $K_\lambda^* \leq 2$ *for every* $\lambda \in \mathbb{R}_+$.

A key feature of the example above is that the choice implications of the true set of preferences R are a subset of the choice implications of a strictly smaller set of preferences. Let us exclude this possibility by assuming:

Assumption 1. *The dataset*

$$D = \left\{ \left(x_{(A,f)}^*, A \right) \, : \, \pi(A, f) > 0 \right\}$$

cannot be perfectly rationalized using fewer than K *orderings.*

That is, if the analyst observes the DM's (perfectly maximizing) choice from every choice problem that is sampled under π with positive probability, then the resulting choice data cannot be perfectly rationalized using fewer than K orderings. Under this assumption, we can choose any $\lambda > 1$ (corresponding to the Kalai, Rubinstein & Spiegler (2002) solution) to return the correct number of orderings in the limit as the number of observations grows large.

Claim 1. *Suppose Assumption 1. Then, if* $p = 0$,

$$\Pr(K_\lambda^* = K) \to 1 \quad as \quad n \to \infty$$

for every $\lambda > 1$.

The proof is clear and omitted.

If the DM imperfectly maximizes, however, then $\lambda > 1$ will not generally be the best choice for recovery of K. Let us move on now to the case in which $p > 0$.

---

[8]The choice observations consistent with maximization of some ordering in R are: $(x_1, \{x_1, x_2, x_3\})$, $(x_3, \{x_1, x_2, x_3\})$, $(x_2, \{x_2, x_3\})$, $(x_3, \{x_2, x_3\})$, $(x_2, \{x_1, x_2\})$, $(x_1, \{x_1, x_2\})$, $(x_1, \{x_1, x_3\})$, $(x_3, \{x_2, x_3\})$, and $(x_3, \{x_1, x_3\})$. It is easy to verify that each of these observations is also consistent with maximization of some ordering in R'.

3.2 Error ($p > 0$)

The possibility of choice error complicates our problem in two ways. First, imperfect maximization may result in inconsistent choices that should not be mistaken to represent heterogeneity in preference. For example, consider a decision maker who has the single preference ordering $\succ$, and errs with some small but positive probability p. Then, a perfect rationalization will require multiple orderings, even while the single ordering $\succ$ explains almost all observations. This suggests that $1/\lambda$ must be chosen sufficiently large that an ordering which only explains errors will not be attributed to the DM. What "sufficiently large" means will depend on the probability of error p—the larger the probability of error p is, the larger the cost of each ordering $1/\lambda$ must be.

Second, choice error may "undo" choice inconsistencies that are intentional, in the sense that they are the consequence of maximization of different preferences. In this case, it may be that an imperfectly maximized dataset can be rationalized using fewer orderings than its perfectly maximized counterpart. Consider for example a decision maker with preferences $\{\succ_1, \succ_2\}$, where

$$x_1 \succ_1 x_2 \cdots \succ_1 x_{99} \succ_1 x_{100}$$

$$x_1 \succ_2 x_2 \cdots \succ_2 x_{100} \succ_2 x_{99}$$

Observe that every choice observation that results from maximizing $\succ_2$ is also consistent with maximizing $\succ_1$, except for the single choice observation

$$(x_{99}, \{x_{99}, x_{100}\}).$$

If the DM is either never observed choosing from the choice set $\{x_{99}, x_{100}\}$, or if he errs in these observations, then the solution to (1) will return a single ordering. This suggests that the DM's preferences must be "sufficiently" differentiated, so that evidence of preference heterogeneity is preserved even in the presence of choice error. As a preliminary step to defining differentiation, let us first construct a generalization of IIA:[9]

Definition 3.1. *Say that choice observations $\{(x_i, A_i)\}_{i=1}^K$ are in K-violation of IIA if*

(1) $x_i \neq x_j$ for every $i \neq j$,

(2) $x_i \in \bigcup_{j=1}^K A_j$ for every $i \in \{1, \ldots, K\}$.

The first condition requires that every chosen alternative is distinct, and the second condition requires that each of these K alternatives is available from each of the

---

[9]Observations (x, A) and (x', B) are in violation of the Independence of Irrelevant Alternatives (IIA) axiom if $x, x' \in B$ and $x, x' \in A$.

K choice sets. An immediate implication is that the set of choice observations $\{(x_i, A_i)\}_{i=1}^{K}$ cannot be rationalized using fewer than K orderings. (Notice also that every pair of choice observations from $\{(x_i, A_i)\}_{i=1}^{K}$ constitutes a (standard) violation of IIA.) These observations therefore suggest that the decision maker possesses at least K different preferences.

Now let us fix a set of choice problems $\mathcal{P}$ and consider the number of K-violations of IIA that would be observed if the DM perfectly maximized.

Definition 3.2. *The (multi-)set of choice problems*

$$\{(A_1, f_1), \ldots, (A_K, f_K)\},$$

induces a K-violation of IIA if the choice observations

$$\left\{ \left(x^*_{(A_1, f_1)}, A_1 \right), \ldots, \left(x^*_{(A_K, f_K)}, A_K \right) \right\}$$

are in K-violation of IIA.

Definition 3.3. *Let $\mathcal{P}$ be a set of choice problems, and let R be a set of orderings. Then, the differentiation parameter $\delta(R, \mathcal{P})$ is the largest integer δ' such that $\mathcal{P}$ contains δ' disjoint subsets of size K, each of which induces a K-violation of IIA.*

That is, the set of choice problems $\mathcal{P}$ includes at least $\delta(R, \mathcal{P})$ disjoint subsets of K choice problems, each of which induces (under perfect maximization) a set of choice observations in K-violation of IIA. A large value of $\delta(R, \mathcal{P})$ means that there are many subsets of choice problems in $\mathcal{P}$ that induce choice observations that cannot be rationalized using fewer than K orderings. A small value of $\delta(R, \mathcal{P})$ means that most subsets of choice problems in $\mathcal{P}$ induce choice observations that can be rationalized using fewer than K orderings. Thus, $\delta(R, \mathcal{P})$ is a measure of the extent to which multiplicity in R can be elicited using choice problems in $\mathcal{P}$.

To gain further insight into this notion of differentiation, let us consider a few examples. In the first example, the set of orderings R is "minimally differentiated," in the sense that the differentiation parameter $\delta(R, \mathcal{P})$ is zero for every set of choice problems $\mathcal{P}$. The second example defines a "maximally differentiated" set of orderings R, and shows that there exists a set of choice problems $\mathcal{P}$ such that the differentiation parameter $\delta(R, \mathcal{P})$ is one-half of the total number of choice problems in $\mathcal{P}$.

Example 3.2. *Let $X = \{x_1, x_2, x_3\}$ and define $R = \{\succ_1, \succ_2, \succ_3\}$, where*

$$x_1 \succ_1 x_2 \succ_1 x_3$$
$$x_1 \succ_2 x_3 \succ_2 x_1$$
$$x_3 \succ_3 x_2 \succ_3 x_1$$

as in Example *3.1*. Then, for every set of choice problems $\mathcal{P}$, the differentiation parameter satisfies

$$\delta(R, \mathcal{P}) = 0$$

To see this, suppose to the contrary that there exists some set of choice problems $(A_1, f_1), (A_2, f_2), (A_3, f_3)$ in 3-violation of IIA. Condition (1) in Definition *3.1* requires that

$$x^*_{(A_1, f_1)} \neq x^*_{(A_2, f_2)} \neq x^*_{(A_3, f_3)}. \tag{2}$$

Condition (2) requires moreover that all three distinct choice alternatives appear in every choice set A_1, A_2, and A_3. Since $|X| = 3$, it must be that $A_1 = A_2 = A_3 = \{x_1, x_2, x_3\}$. But $(x_2, \{x_1, x_2, x_3\})$ is not consistent with maximization of any choice ordering in R, so (*2*) is contradicted. Thus, there are no sets of choice observations (assuming no choice error) whose perfect rationalization requires more than two orderings.

Example 3.3. *Consider* $X = \{x_1, \ldots, x_n\}$ *and define* $R = \{\succ_1, \succ_2\}$ *where*

$$x_1 \succ_1 x_2 \succ_1 \cdots \succ_1 x_n$$
$$x_n \succ_2 x_{n-1} \succ_2 \cdots \succ_2 x_1$$

Let $\mathcal{P}$ *consist of every choice problem* $(A, f) \in 2^X \times F$, *where* A *consists of at least two choice alternatives. There are* $2(2^n - n - 1)$ *such choice problems. Then,*

$$\delta(R, \mathcal{P}) = 2^n - n - 1. \tag{3}$$

To see this, observe that for every non-singleton choice set $A \in 2^X$, *the choice observations*

$$\left\{ \left(x^*_{(A,1)}, A \right), \left(x^*_{(A,2)}, A \right) \right\}$$

are in 2-violation of IIA, since $x^*_{(A,1)} \neq x^*_{(A,2)}$, *and also* $x^*_{(A,1)}, x^*_{(A,2)} \in A$. *The set of choice problems* $\mathcal{P}$ *can be partitioned into* $2^n - n - 1$ *such pairs; therefore, the differentiation parameter is* $2^n - n - 1$ *(or, one half of the total number of choice problems in* $\mathcal{P}$).

In fact, it can be shown that

$$0 \leq \delta(R, \mathcal{P}) \leq n/K \tag{4}$$

for every R, $\mathcal{P}$, and n. The upper bound for $\delta(R, \mathcal{P})$ is attained if $\mathcal{P}$ can be partitioned into disjoint subsets, each of which contains K choice problems and induces a K-violation of IIA—for example, as described in Example *3.3*. The lower bound is attained if $\mathcal{P}$ does not include any set of choice problems that induces a K-violation of IIA—for example, as described in Example *3.2*.

Finally, for every sampling distribution π over choice problems, define[10]

$$\delta_\pi = \liminf_{n \to \infty} \sum_{|\mathcal{P}|=n} \left(\frac{\delta(R, \mathcal{P})}{n} \right) \pi^n(\mathcal{P}). \tag{5}$$

We can understand the parameter δ_π as follows. Suppose that the DM is presented with an increasing number of choice problems, repeatedly sampled from the distribution π. Then, (5) says that the expected value of $\delta(R, \mathcal{P})/n$ is bounded below by δ_π when the number of observations n is sufficiently large. Recalling that $\delta(R, \mathcal{P})$ is the number of subsets of $\mathcal{P}$ that are useful for differentiating between orderings in R, we can interpret δ_π as the limiting fraction of choice observations that help to reveal preference heterogeneity.

The main theorem provides values of λ and conditions on the DM's choice model, under which the proposed approach recovers the true number of orderings as the number of observations grows large.

Theorem 1. *Fix any $\overline{p} > p$, and set $\lambda = 1/(\overline{p}n)$. If*

$$\delta_\pi > \frac{\overline{p}}{(1-\overline{p})^K} \tag{6}$$

then,

$$\Pr(K_\lambda^* = K) \to 1 \quad \text{as } n \to \infty.$$

Theorem 1 says that if the true probability of error is less than $\overline{p}$, and the limiting differentiation parameter δ_π is sufficiently large, then the proposed approach (taking $\lambda = 1/(\overline{p}n)$) will recover the correct number of orderings given sufficiently many choice observations.[11] Observe that the required lower bound on δ_π in (6) is increasing in both $\overline{p}$ and K. Thus, the smaller the probability of error, and the smaller the number of preferences, the less restrictive the condition. In practice, $\overline{p}$ should be chosen to overestimate the true probability of error p by as little as possible.

If either the number of orderings K or the probability of error p is too large, it may be that the condition in (6) is infeasible to satisfy for any choice of $\overline{p} > p$. Specifically, it follows from the bounds on the differentiation parameter described in (4) that

$$0 \le \delta_\pi \le 1/K, \tag{7}$$

---

[10]Throughout, the dependence of δ_π on R is dropped.

[11]Since the number of orderings K and the probability of error p are not known, the assumption in (6) cannot be directly tested. Nevertheless, we can carry out an approximate test by substituting the inferred number of orderings K_λ^* for K and the inferred fraction of choice errors $\hat{p} := \varepsilon_{K_\lambda^*}/n$ for $\overline{p}$, verifying that

$$1/K_\lambda^* > \frac{\hat{p}}{(1-\hat{p})^{K_\lambda^*}}.$$

12

so the condition in (6) implies

$$\frac{1}{K} > \frac{p}{(1-p)^K}.$$

For example, if K exceeds 10, then the probability of error cannot be larger than $p = 0.05$, and if K exceeds 5, then the probability of error cannot be larger than $p = 0.1$.

The proof of Theorem 1 is deferred to the appendix, but I will sketch the main ideas here. The key idea is to identify every dataset with an undirected hypergraph[12] (henceforth *graph*) in the following way: every node corresponds to a choice observation, and there is an edge between a set of nodes if and only if the corresponding observations are not consistent with maximization of any single ordering. The proof notes that a dataset is k-rationalizable if and only if the corresponding graph is k-colorable.[13,14] Thus, the problem in (1) can be re-cast as finding the smallest number of colors k such that a large subset of nodes are k-colorable.

Let us fix any multiset of choice problems $\mathcal{P}$ and consider the dataset that would be generated in the absence of choice error ($p = 0$). Since by construction, this dataset is generated by perfect maximization of K orderings, the corresponding graph must admit a K-coloring. Moreover, since every set of observations in a K-violation of IIA constitutes a complete K-partite subgraph, and the data includes at least one such set, the corresponding graph cannot be colored by fewer than K colors. The challenge is to show that even when the graph is perturbed by choice error, with high probability it will remain the case that a large subset of the nodes can be colored by K colors, but no fewer.

To show that K colors are sufficient to color most nodes, I use Hoeffding's inequality to bound the number of imperfectly maximized choice observations below $\bar{p}n$ (with high probability) as n gets large. To show that K colors are needed, I use the assumption in (6) and repeated applications of Hoeffding's inequality to bound the number of disjoint complete K-partite subgraphs below by $\bar{p}n$ (with high probability) as n gets large. Since each complete K-partite subgraph cannot be colored by fewer than K colors, the number of such subgraphs provides an approximate[15] lower bound on the number of nodes that are uniquely colored by each of the first K colors. Thus, each of the first K orderings uniquely explains

---

[12]A *hypergraph* is a generalization of a graph in which edges may connect more than two vertices.

[13]A *k-coloring* of a graph is a partition of its vertex set V into k color classes such that no edge in E connects two nodes of the same color. A graph is *k-colorable* if it admits an k-coloring.

[14]This equivalence is shown by taking each color class to represent consistency with a distinct ordering.

[15]The proof considers the number of such subgraphs that are additionally *perfectly maximized*; this is an exact lower bound on the number of nodes that are uniquely colored by each of the first K colors.

at least $\bar{p}n$ observations, and the marginal $(K + 1)$-st ordering explains strictly fewer than $\bar{p}n$ additional observations, so the proposed approach correctly returns K orderings.

3.3 Extension: Continuous Utility

So far, we have considered a decision maker whose preferences are orderings over a discrete set X. I will now show that the results extend easily to the case in which his preferences are instead a set of continuous utility functions $\{u_f\}_{f \in F}$, $u_f : \underline{X} \to \mathbb{R}$, where $(\underline{X}, \tau)$ is a topological space.

Formally, suppose that choice sets are compact subsets $\underline{A} \subseteq \underline{X}$, and choice problems are pairs $(\underline{A}, f)$. The DM's choice rule is a map P from choice problems $(\underline{A}, f)$ into distributions over $\underline{A}$[16]; that is, P takes choice sets $\underline{A}$ and choice frames f into distributions over the available alternatives in $\underline{A}$.

Choice problems $(\underline{A}, f)$ are repeatedly sampled from a (Borel-measurable) distribution π.[17] The typical set of sampled choice problems is denoted by $\underline{\mathcal{P}}$, and the DM's choices (induced by his stochastic choice rule P) are

$$\underline{D}^* = \{(x, (\underline{A}, f)) \,|\, (\underline{A}, f) \in \underline{\mathcal{P}}\}.$$

The analyst observes only

$$\underline{D} = \{(x, \underline{A}) \,|\, (x, (\underline{A}, f)) \in \underline{D}^*\}, \tag{8}$$

which includes the DM's choice sets and selected alternatives, but not his choice frames. The goal is to determine the number of utility functions $K := |F|$.

The proposed approach minimizes a weighted average of the number of inferred utility functions and the number of unexplained observations in $\underline{D}$. The latter quantity is defined as follows: For every set of utility functions $\mathcal{U}$, let

$$\underline{\varepsilon}(\underline{D}, \mathcal{U}) = \# \left\{ (x, \underline{A}) \in \underline{D} \,:\, x \neq \max_{x' \in \underline{A}} u(x') \text{ for any } u \in \mathcal{U} \right\}$$

be the number of choice observations in $\underline{D}$ that are not consistent with maximization of any utility function in $\mathcal{U}$. Then,

$$\underline{\varepsilon}_k = \min_{|\mathcal{U}|=k} \underline{\varepsilon}(\underline{D}, \mathcal{U}).$$

is the minimal number of observations in $\underline{D}$ that are unexplained if we rationalize the DM's choices using k utility functions.

The solution below simultaneously minimizes the number of utility functions k and the implied choice error $\underline{\varepsilon}_k$:

---

[16]More precisely, every $P(\underline{A}, f)$ is a Borel-measurable distribution over $\underline{X}$ with support contained in $\underline{A}$.

[17]Take the topology over choice problems to be the product topology of τ and the discrete topology on F.

Definition 3.4. *For every $\lambda \in \mathbb{R}_+$, define*

$$\underline{K}^*_\lambda \in \operatorname*{argmin}_{k \in \mathbb{N}} k + \lambda \underline{\varepsilon}_k. \tag{9}$$

Corollary 1 below shows that this solution recovers the "correct" number of utility functions K under conditions that directly parallel the previous section. Formally, define

$$\underline{x}^*_{(\underline{A},f)} = \max_{x \in \underline{A}} u_f(x)$$

to be the choice that a perfectly maximizing DM would make given choice problem $(\underline{A}, f)$, and let

$$\overline{p} = \min \left\{ p' \,:\, P(\underline{A}, f) \left[\underline{x}^*_{(\underline{A},f)} \right] \geq 1 - p' \quad \forall \, (\underline{A}, f) \right\}$$

be the smallest uniform upper bound on the probability that the DM errs in any given observation.[18] Then, the statement below follows as a corollary to Theorem 1 (where the differentiation parameter δ_π is defined as in Section 3.2):

Corollary 1. *Fix any $\overline{p} > p$, and set $\lambda = 1/(\overline{p}n)$. If*

$$\delta_\pi > \frac{\overline{p}}{(1 - \overline{p})^K}, \tag{10}$$

then,

$$\Pr(\underline{K}^*_\lambda = K) \to 1 \quad \text{as } n \to \infty.$$

That is, so long as the true probability of error is less than $\overline{p}$, and the limiting differentiation parameter δ_π is sufficiently large, then the proposed approach (taking $\lambda = 1/(\overline{p}n)$) will recover the true number of utility functions as the number of observed choices gets large.

Why does Theorem 1 extend, essentially without modification, to this more general setting? The key observation is that any choice data of the nature described in (8) can be mapped into discrete choice data, where we reduce $\underline{X}$ to the finite set

$$X = \{x \in \underline{X} \,:\, (x, \underline{A}) \in \underline{D} \text{ for some } \underline{A} \subseteq \underline{X}\},$$

consisting of the choice alternatives that are observed to be chosen from some choice set. For example, take $\underline{X} = \mathbb{R}$, and suppose we observe

$$\underline{D} := \{(3, [0, 4]), \ (2, [1, 4]), \ (8, [0, 10])\}.$$

Then, labelling '3' as x_1, '2' as x_2, and '8' as x_3, we can redefine the set of choice alternatives as $X = \{x_1, x_2, x_3\}$, and the choice data as

$$D := \{(x_1, \{x_1, x_2\}), \ (x_2, \{x_1, x_2\}), \ (x_3, \{x_1, x_2, x_3\})\}.$$

---

[18]Observe that as before, no parametric assumptions are made regarding the distribution of error; future work may include such assumptions to strengthen the recovery results.

This is a standard mapping used in the literature, and yields a dataset of the form introduced in Section 2.1.

The main lemma in the proof of Corollary 1 shows that the new problem posed in Definition 3.4 is equivalent to the original problem posed in Definition 2.1, in the sense that the solution to

$$\underset{k \in \mathbb{N}}{\operatorname{argmin}} \, k + \lambda \underline{\varepsilon}_k,$$

given data $\underline{D}$, is equivalent to the solution to

$$\underset{k \in \mathbb{N}}{\operatorname{argmin}} \, k + \lambda \underline{\varepsilon}_k$$

given data D. It immediately follows that the conditions for recovery stated in the previous section are also the conditions for recovery needed in the present setting.

4 Application

Let us now consider an example application of the proposed approach. Crawford & Pendakur (2012) study the consumption decisions of Danish households over six different kinds of milk. Each household's consumption decisions are aggregated over a month. The relevant choice information is the quantity of each kind of milk purchased during this time (written as a quantity vector $\mathbf{q} \in \mathbb{R}^6$), and the price index at which these purchases were made (written as a price vector $\mathbf{p} \in \mathbb{R}^6$). The main sample analyzed in Crawford & Pendakur (2012) consists of 500 households, so the choice data can be written

$$(\mathbf{p}_1, \mathbf{q}_1), \ldots, (\mathbf{p}_{500}, \mathbf{q}_{500}).[19]$$

Let us map these choice observations into the present framework, using the relabelling described above. Index the observations by $i = 1, \ldots, 500$, and for every observation i, define $\mathbf{x}_i = (\mathbf{p}_i, \mathbf{q}_i)$. Take $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_{500}\}$ to be the set of choice alternatives. Additionally, for every observation i, define the choice set

$$A_i = \{\mathbf{x}_j \, : \, \mathbf{p}_i \cdot \mathbf{q}_i \geq \mathbf{p}_j \cdot \mathbf{q}_j\}$$

to consist of every alternative $\mathbf{x}_j$ that is less costly than the selected alternative $\mathbf{x}_i$—that is, every alternative in X that could have been chosen when the alternative $\mathbf{x}_i$ was chosen.[20] The observed data from Crawford & Pendakur (2012) is now written

$$D = \{(\mathbf{x}_1, A_1), \ldots, (\mathbf{x}_{500}, A_{500})\}.$$

---

[19]Their data also includes a household indicator and covariates describing each household, but these are not relevant to the present discussion.

[20]That is, if $\mathbf{p}_i \cdot \mathbf{q}_i \geq \mathbf{p}_j \cdot \mathbf{q}_j$, then $\mathbf{x}_j \in A_i$.

This is equivalent to the original data in the following sense: if there exists a set of k utility functions such that m observations in the original data are consistent with maximization of a utility function in this set,[21] then we can find a set of k preference orderings on X such that m observations in the relabelled data are consistent with maximization of a preference in this set, and vice versa.[22]

There are many possible ways to rationalize the choice data D. For example, following the proposal of Houtman & Maks (1985), we can find the single ordering that explains the largest fraction of the data, and interpret the remaining observations as choice error. Crawford & Pendakur (2012) find that no single preference can explain more than two-thirds of the observations.

Alternatively—and this is the main approach taken in Crawford & Pendakur (2012)—we can seek the minimal set of preferences that explains every observation (thus following the proposal of Kalai, Rubinstein & Spiegler (2002)). Crawford & Pendakur (2012) find that no more than five orderings are needed to perfectly rationalize the data.

The present paper interprets the two solutions above as edge cases among a set of rationalizations of the data, each of which entails a different tradeoff between maximization of fit to the data and minimization of the number of orderings used. Formally, for every $\lambda \in \mathbb{R}_+$, we can define

$$K_\lambda^* = \underset{k \in \mathbb{N}}{\operatorname{argmin}} \, k + \lambda \varepsilon_k. \tag{11}$$

where ε_k is the minimal number of choice observations that must be left unexplained if we use k preferences to explain the observed data. Figure 1 provides approximations for ε_k that are taken from Crawford & Pendakur (2012) (for the purpose of illustration of the approach, I will treat these approximations as the exact values of ε_k).[23]

We see that using a single ordering, we must leave 179 observations unexplained; using two orderings, we must leave 79 observations unexplained; and if we allow ourselves five orderings, we can explain all of the observations. Each $k \in \{1, 2, 3, 4, 5\}$ is a solution to (11) for a particular tradeoff between the number of orderings and the number of choice errors—specifically, taking $\lambda < 1/100$ returns the Houtman & Maks (1985) solution, and taking $\lambda > 1/8$ returns the Kalai, Rubinstein & Spiegler

---

[21]That is,

$$u(\mathbf{q}_i) > u(\mathbf{q})$$

for all quantity vectors $\mathbf{q}$ satisfying $\mathbf{p}_i^T \mathbf{q}_i \geq \mathbf{p}_i^T \mathbf{q}_i$

[22]The argument is very similar to the argument for Lemma 3.

[23]These values correspond to an algorithm that computes an upper bound on the number of types needed to explain a given number of observations, so that the true values of ε_k are weakly smaller than those reported below. Crawford & Pendakur (2012) provide also lower bounds that show that the errors in this approximation are not large.

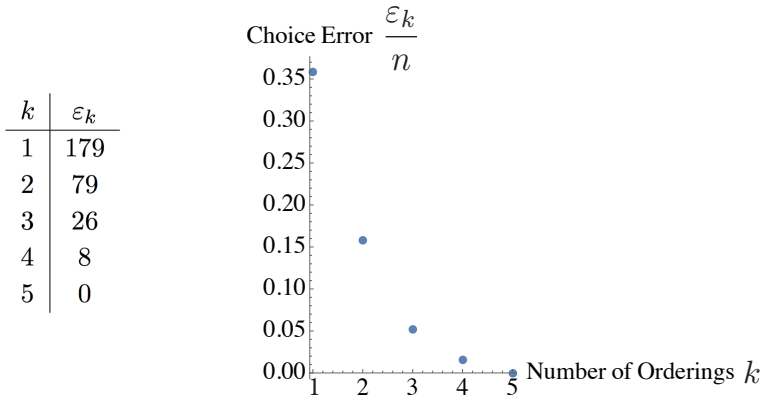| k	ε_k
1	179
2	79
3	26
4	8
5	0

Figure 1: *Left:* At least ε_k choice observations (out of 500) must be left unexplained, if we use k preferences to rationalize the choice data. *Right:* Choice errors (as a fraction of the total number of observations) are graphed against the number of preferences.

(2002) solution.[24]

But the intuition that this paper formalizes is that there is significant variation in the degree of evidence for each of the five orderings, which should be used to inform our judgment of whether these orderings reflect genuine preference or error. Observe, for example, that while the marginal explanatory value of the first ordering is 179 observations, and the marginal explanatory value of the second is 100 observations, the fifth ordering improves explanation by only eight observations. Let us therefore set the cost of each ordering as proposed in Theorem 1; that is, let $\lambda = \frac{1}{\overline{p}n}$, where $\overline{p}$ is our best (over-)approximation of the probability with which the DM errs, and $n = 500$ is the total number of observations. Then, supposing the probability of error is at least 0.016, we will have that $\lambda \geq \frac{1}{8}$, from which it follows that the recovered number of preferences is no more than four:

$$K_\lambda^* \leq 4.$$

This means that we should not interpret the fifth Kalai, Rubinstein & Spiegler (2002) ordering as reflecting genuine preference; instead, the eight observations that are left unexplained are better understood as the consequence of choice error.[25]

---

[24]Geometrically, we can understand the solution to (11) as follows: Let f be the linear interpolation of points $\{(k, \varepsilon_k), k \in \mathbb{N}\}$, as graphed in the right panel of Figure 1 and let $(F = \{(x, y) \mid y \geq f(x)\}$ be the epigraph of f. Fix any $\lambda \in \mathbb{R}_+$. Then, the problem in (1) returns a solution with k orderings if and only if the line with normal vector $(-1, -\lambda)$ properly supports F at (k, ε_k).

[25]Given more precise information about the probability of error, we can do better: the recovered

18

5   Can we Recover More?

Section 4 described conditions under which the problem in (1) recovers the correct number of frames with high probability. Let us now ask: is it possible to recover the orderings themselves?

5.1   Negative Results

Say that the set of orderings R is *identifiable* if there is at least one dataset D such that R is the unique set of $|R|$ orderings that perfectly explains every observation in D.

Definition 5.1. *The set of orderings R is* identifiable *if there exists some dataset D such that*

$$\varepsilon(D, R) = 0,$$

and moreover,

$$\varepsilon(D, R') > 0$$

for every $R' \neq R$ with $|R'| \leq |R|$.

The following proposition says that most sets of orderings are not identifiable.

Proposition 1. *No set of orderings R with $|R| \geq 3$ is identifiable. Suppose $R = \{\succ_1, \succ_2\}$; then, R is identifiable if and only if the $\succ_1$-maximal alternative in X is the $\succ_2$-minimal alternative in X, and vice versa. Every singleton set $R = \{\succ\}$ is identifiable.*

The proof constructs, for every set R of three or more orderings, and for every pair of orderings that do not satisfy the condition above, an alternative set of orderings R' such that every dataset that can be perfectly rationalized using R can also be perfectly rationalized using R'.

Proposition 1 suggests that recovery of the exact set of orderings from choice observations alone is not generally possible. But recovery of the exact set of orderings may not be the right goal, if we understand the real content of a set of orderings to be its choice implications. Define

$$c(R) = \{(x, A) \,:\, A \in P_X \text{ and } x \text{ is } \succ \text{-maximal in } A \text{ for some } \succ \in R\},$$

to be the set of all choice observations that are consistent with maximization of some ordering in R, and say that the sets of orderings R and R' are *choice-equivalent*

---

solution is $K_\lambda^* = 4$ if $\overline{p} \in [0.016, 0.036)$, $K_\lambda^* = 3$ for any $\overline{p} \in [0.036, 0.106)$, $K_\lambda^* = 2$ for any $\overline{p} \in [0.106, 0.2)$, and $K_\lambda^* = 1$ for any $\overline{p} > 0.2$.

if $c(R) = c(R')$. For example, the set of orderings $R = \{\succ_1, \succ_2, \succ_3\}$ with

$$x_1 \succ_1 x_2 \succ_1 x_3$$
$$x_2 \succ_2 x_1 \succ_2 x_3$$
$$x_3 \succ_3 x_1 \succ_3 x_2$$

and the set of orderings $R' = \{\succ'_1, \succ'_2, \succ'_3\}$ with

$$x_1 \succ'_1 x_2 \succ'_1 x_3$$
$$x_2 \succ'_2 x_3 \succ'_2 x_1$$
$$x_3 \succ'_3 x_2 \succ'_3 x_1$$

are choice-equivalent, since every observation that can be rationalized by an ordering in R can also be rationalized by an ordering in R', and vice versa.[26]

The following, weaker, notion of identifiability says that R is choice-identifiable if there exists a dataset D such that every set of ($|R|$ or fewer) orderings R' that perfectly rationalizes D is also choice-equivalent to R.

Definition 5.2. *Say that the set of orderings R is* choice-identifiable *if there exists some dataset D such that*

$$\varepsilon(D, R) = 0,$$

and moreover, for every set of orderings R' with $|R'| \leq |R|$,

$$\varepsilon(D, R') = 0 \quad \Rightarrow \quad c(R') = c(R).$$

It turns out that even this weaker property of choice-identifiability is not satisfied by many sets of orderings. The following example illustrates why.

Example 5.1. *Let $X = \{x_1, x_2, x_3, x_4, x_5\}$. Define $R = \{\succ_1, \succ_2\}$ to satisfy*

$$x_1 \succ_1 x_2 \succ_1 x_3 \succ_1 x_4 \succ_1 x_5$$
$$x_2 \succ_2 x_1 \succ_2 x_3 \succ_2 x_4 \succ_2 x_5$$

and the set of orderings $R' = \{\succ'_1, \succ'_2\}$ to satisfy

$$x_1 \succ'_1 x_2 \succ'_1 x_3 \succ'_1 x_4 \succ'_1 x_5$$
$$x_5 \succ'_2 x_4 \succ'_2 x_3 \succ'_2 x_2 \succ'_2 x_1$$

Then, every choice observation that is consistent with maximization of some ordering in R is also consistent with maximization of some ordering in R', so that if $\varepsilon(D, R) = 0$, then also $\varepsilon(D, R') = 0$.[27] But R and R' are not choice-equivalent; therefore, R is not choice-identifiable.

---

[26]Since $c(R) = c(R') = \{(x, A)\}_{x \in X, A \in 2^X}$.

[27]For example, (x_5, X) is consistent with maximization of $\succ'_2$, but not with maximization of either $\succ_1$ or $\succ_2$.

The example above highlights the general phenomenon that there is variation in the richness of choice implications across sets with a fixed number of orderings. Any approach that penalizes only the *size* of a set of orderings, as both the proposed approach in (1) and the approach suggested in Kalai, Rubinstein & Spiegler (2002) do, will be biased towards elicitation of sets with richer choice implications. Claim 2 characterizes choice-identifiability in two special cases: $|R| = 2$, and $|R| = |X|$. In the former, R is identifiable if and only if the two orderings "disagree maximally" (one is the reverse of the other); in the latter, R is identifiable if and only if the most preferred alternative according to each ordering is different.

Claim 2. *(a) The set of orderings $\{\succ_1, \succ_2\}$ is choice-identifiable if and only if*

$$x \succ_1 x' \quad \Leftrightarrow \quad x' \succ_2 x. \tag{12}$$

for every $x, x' \in X$. (b) Suppose $|R| = |X|$. Then, R is choice identifiable if and only if the $\succ$-maximal alternative in X is distinct for every $\succ \in R$.

5.2   A Revised Approach

These negative results suggest an alternative measure for the complexity of a set of orderings. That is, instead of penalizing the number of orderings, we can penalize the *number of choice implications*. Formally, take $\mathcal{R}$ to be the set of all finite sets of orderings over X (excluding the empty set), and define the function $l : \mathcal{R} \to \mathbb{Z}_+$ to count the number of unique choice observations (x, A) that are consistent with maximization of some ordering in R; that is,

$$l(R) = \#\{(x, A) \, : \, x \text{ is } \succ\text{-maximal in } A \text{ for some } \succ \in R\}.$$

The following observation points out trivial but useful features of this function.

Observation 2.   *(a) Fix any singleton set R. Then, $l(R) = 2^{|X|} - 1$.*

(b) There exist sets R, R' with $|R'| > |R|$ but $l(R') < l(R)$.

(c) $\max_{|R|=k} l(R)$ is weakly increasing in k.

That is: (a) $l(R)$ is constant across all singleton sets R; (b) there exist sets R', R, where R' consists of a larger number of orderings than R, but has fewer choice implications (so that the ranking on sets implied by l need not coincide with the ranking implied by cardinality); (c) increasing the permitted number of orderings weakly increases the number of possible choice implications.

The following definition modifies the proposed approach in Definition 2.1 by replacing the measure $|R|$ with the measure $l(R)$. The solution thus minimizes a weighted average of the number of choice implications and the number of unexplained choice observations.

Definition 5.3. *For every $\lambda \in \mathbb{R}_+$, define*

$$R^*_\lambda = \operatorname*{argmin}_{R \in \mathcal{R}} l(R) + \lambda \varepsilon(D, R). \tag{13}$$

In this case, $1/\lambda$ can be understood as the cost of each new choice implication, so that a choice implication is attributed to the DM only if it appears at least $1/\lambda$ times in the data. As $\lambda \to 0$, the cost of errors becomes increasingly small relative to the cost of new choice implications, so that the analyst prefers to attribute to the DM as few choice implications as possible. As $\lambda \to \infty$, the cost of choice errors becomes increasingly large relative to the cost of new choice implications, so that the analyst prefers to ascribe to the DM as many choice implications as is necessary to perfectly rationalize the data.

The following claim states that any choice of $\lambda < \frac{1}{\varepsilon_1}$ returns the Houtman & Maks (1985) solution. Below, let $\mathcal{R}_{HM} \subseteq \mathcal{R}$ be the set of all orderings that explain the maximal number of observations rationalizable by a single ordering; that is,

$$\varepsilon(D, \{\succ\}) = \varepsilon_1$$

for every $\succ \in \mathcal{R}_{HM}$. When the Houtman & Maks (1985) solution is unique, then R^*_λ is exactly the singleton set consisting of this ordering. When the solution is not unique, then R^*_λ is the set of all singleton sets consisting of different Houtman & Maks (1985) solutions. At the other extreme, $\lambda > 2^{|X|} - 1$ guarantees that $c(R^*_\lambda)$ includes (at least) every choice observation (x, A) that is observed in the data.

Claim 3. *For every dataset D,*

(a) $R^*_\lambda = \{\{\succ\} : \succ \in \mathcal{R}_{HM}\}$ *for every* $\lambda < \frac{1}{\epsilon_1}$.

(b) $\{(x, A) : (x, A) \in D\} \subseteq c(R^*_\lambda)$ *for every* $\lambda > 2^{|X|} - 1$.

Part (a) reveals an intuitive relationship between the solution to the modified problem in Definition 3.4 and the solution to the original problem in Definition 2.1, when $\lambda < \frac{1}{\varepsilon_1}$. Specifically, the solution to Definition 2.1 determines the size of the solution to Definition 3.4 (and both are consistent with the approach of Houtman & Maks (1985)). This relationship between the two problem need not hold for other values of λ. In particular, when λ is large, the approach in Definition 3.4 may return sets of orderings whose size is not described by Definition 2.1.

5.3   Recovery of Choice Implications

Let us first consider recovery of choice implications in the case of no error ($p = 0$). Take $R = \{\succ_f\}_{f \in F}$ to be the decision maker's true preferences, and suppose that:

Assumption 2. *For every $(x, A) \in c(R)$,*

$$\pi\left(\left\{(A, f) : x^*_{(A,f)} = x\right\}\right) > 0.$$

22

This restricts the sampling distribution π such that for every choice observation $(x, A) \in c(R)$, π places positive probability on a choice problem (A, f) with the property that x is the $\succ_f$-maximal alternative in A. Then, as in Claim 1, we can choose any $\lambda > 1$ to recover the set of choice implications $c(R)$.

Claim 4. *Suppose Assumption 2 is satisfied. Then, if $p = 0$,*

$$\Pr(c(R^*_\lambda) = c(R)) \to 1 \quad as \quad n \to \infty$$

for every $\lambda > 1$.

Intuitively, when $\lambda > 1$, the cost of an additional choice error is larger than the cost of an additional choice implication. This means that we want to infer as many choice implications as is necessary to explain all of the observations in the data (and no more). Under Assumption 2, and the assumption of perfect maximization, we further know that there exists a set of orderings whose choice implications are exactly those needed to rationalize the choice data—and this set is R.[28]

The proposition below introduces choice error, and provides conditions under which the proposed approach recovers the choice implications of R with probability arbitrarily close to 1 as the quantity of data increases.

Proposition 2. *Take any $\bar{p} > p$ and set $\lambda = 1/(\bar{p}n)$. Then, if*

$$\pi\left(\left\{(A, f) : x^*_{(A,f)} = x\right\}\right) > \frac{\bar{p}}{1 - \bar{p}}, \tag{14}$$

for every $(x, A) \in c(R)$, we have that

$$\Pr(c(R^*_\lambda) = c(R)) \to 1 \quad as \ n \to \infty.$$

The intuition for this result is as follows: the condition in (14) ensures that, for every choice implication $(x, A) \in c(R)$, sufficient probability is placed on choice problems that induce (x, A). This assumption is similar to condition (6) in Theorem 1, and its purpose is to guarantee that every choice observation in $c(R)$ is recovered. To see that no choice implication $(x, A) \notin c(R)$ will be incorrectly recovered, observe that the number of observations of (x, A) is upper bounded by the number of choice errors. We can show that with high probability, the number of choice errors concentrates below $1/(\bar{p}n)$, from which it follows that $c(R^*_\lambda)$ will include as few choice implications outside of $c(R)$ as possible.

---

[28]Observe that the reason why $\lambda > 1$ is not sufficient for part (b) of Claim 3 is because, generally, we do not have the flexibility to include "just one more" choice implication. In fact, we may need to include an additional ordering in order to obtain the new choice implication, which would increase the number of choice implications by up to $2^{|X|} - 1$.

6   Related Literature

6.1   Nonparametric Preference Recovery

This paper builds on a literature that seeks to nonparametrically identify multiple preferences from choice data. Most directly, it extends ideas in Kalai, Rubinstein & Spiegler (2002), which defines a set of orderings $R = \{\succ_k\}_{k=1}^L$ to be a *rationalization by multiple rationales* if for every observation (x, A), the choice alternative x is $\succ$-maximal in A for some ordering $\succ$ in the set R. Using the terminology of this paper, any set of orderings R with choice error $\varepsilon(D, R) = 0$ is a rationalization by multiple rationales of the dataset D. This set of orderings may not, however, correspond to a *best* multiple-ordering rationalization of the data as defined in (1). In particular, I suggest that the analyst may prefer an imperfect rationalization of the data using some $K < L$ orderings to a perfect rationalization of the data using L orderings. The key conceptual difference is that Kalai, Rubinstein & Spiegler (2002) is agnostic towards the degree of evidence for orderings, whereas the approach in this paper insists on sufficient evidence for each ordering in order to separate error from preference variation.

Ambrus & Rozen (2013) study multiple preference models in which choice is determined through maximization of a choice-set independent aggregation rule over preferences. They find that without prior restriction on the number of selves involved in a decision, many multiple preference ("multi-self") models have no testable implications. Although the class of models considered in their paper is different from the class studied in the present paper (in which the aggregation rule varies across choice problems), their lesson that restricting the number of preferences is important for constraining the available degrees of freedom holds in the setting considered in this paper as well (relating especially to the results in Section 5.1), and motivates in part the suggested criterion in (1).

Other nonparametric approaches for preference identification include Houtman & Maks (1985) and Varian (1982). These approaches differ from the present paper, and from Kalai, Rubinstein & Spiegler (2002), in finding a *single* best-fit ordering. A separate and sizable literature studies related questions under parametric assumptions—see, for example, Quandt (1956), McFadden & Richter (1970), and Train (1986)).

Finally, Crawford & Pendakur (2012) and Dean & Martin (2010) apply the approaches of Kalai, Rubinstein & Spiegler (2002) and Houtman & Maks (1985) towards recovery of preferences from real choice data. Deb (2009) and Dean & Martin (2010) provide computationally efficient approaches for approximating the Kalai, Rubinstein & Spiegler (2002) solution.

6.2 Testing Rationality

When choice data is *inconsistent*—meaning that it is incompatible with perfect rationalization by a single ordering—how should we measure the inconsistency of the observed choices? This question has been the subject of a large literature, with solutions proposed by Afriat (1967), Varian (1982), Echenique, Lee & Shum (2011), Houtman & Maks (1985), Gross (1989), Famulari (1995), Apesteguia & Ballester (2012), and Dean & Martin (2016) among others. See Apesteguia & Ballester (2012) for a summary and comparison of these approaches.

In view of this literature, one goal of the present paper is to distinguish between choice data that is inconsistent because of choice error, and choice data that is inconsistent because of multiplicity in preference. These two sources for error are confounded in many of the measures above—for example, Famulari (1995) counts the number of violations of a particular consistency property, and Houtman & Maks (1985) computes the size of the complement of the maximal set of consistent choice observations, but in neither case does the index provide guidance on whether to interpret the inconsistencies as arising from preference heterogeneity or from choice error.

The proposed approach offers a new perspective on this question. If the choice data is generated via approximately perfect rationalization of multiple preferences, the recovered number of preferences K^*_λ will be large (and the number of unexplained choice observations $\varepsilon_{K^*_\lambda}$ will be small), while if the data is generated by imperfect rationalization of a single preference, the approach will recover a single ordering (and the number of unexplained choice observations ε_1 will be large). These ideas are most similar to Halevy, Persitz & Zrill (2015), which suggests a novel approach for decomposing an incompatibility index into a measure of choice inconsistency and a measure of preference misspecification.

6.3 Context-Dependent Choice

The model of context-dependent choice that I consider throughout is based on Rubinstein & Salant (2008) and Bernheim & Rangel (2009). These papers extend the standard model to include a set F of contexts (called *frames* in Rubinstein & Salant (2008) and *ancillary conditions* in Bernheim & Rangel (2009)). A choice problem (called *extended choice problem* in Rubinstein & Salant (2008) and *generalized choice situation* in Bernheim & Rangel (2009)) is defined as a pair (A, f) where $A \subseteq X$ is a choice set and $f \in F$ is a frame. An extended choice function c assigns to every extended choice problem (A, f) an element of A. I consider an extension of this model with choice error, so that $c(A, f)$ is stochastic. I ask whether it is possible to recover the number of frames in F given observation of the realized choices $c(A, f)$ and the choice sets A, but not the choice frames f.

Appendix A: Preliminaries

Following, I collect definitions and results that will be used in the proofs of Theorem 1 and Proposition 1.

First, observe that K is the unique solution to the problem in (1) given data D if and only if

$$K + \lambda \varepsilon_K > k + \lambda \varepsilon_k \quad \forall \, k \in \mathbb{N}, k \neq K. \tag{15}$$

We can rewrite this as

$$\varepsilon_K - \varepsilon_k > (k - K)/\lambda.$$

Fix a number of observations n. The sampling distribution π over choice problems and the DM's stochastic choice rule P induce a distribution ν_n over datasets of size n, where

$$\nu_n(D) = \nu(\{(x_i, A_i)\}_{i=1}^n) = \prod_{k=1}^n \left[\sum_{f \in F} \pi(A_i, f) \, P(x_i \,|\, (A_i, f)) \right].$$

The probability that $K_\lambda^* = K$ is then the measure assigned by ν_n to the set of datasets D satisfying (15). To characterize this probability, it is useful to recast the problem in the following way.

Let $g : D \mapsto G_D$ be a map that identifies every dataset D with a (hyper-) graph $G_D = (V_D, E_D)$, where $V_D = \{1, \ldots, n\}$ indexes choice observations, and E_D consists of every set $T \subseteq V_D$ such that: (1) the observations $\{(x_i, A_i)\}_{i \in T}$ are 1-rationalizable, and (2) no proper subset of $\{(x_i, A_i)\}_{i \in T}$ are 1-rationalizable. These concepts are related to our problem as follows.

Claim 5. D is k-rationalizable $\iff$ G_D is k-colorable.

Proof. Take each color class to represent consistency with a distinct ordering, and the equivalence follows directly. $\square$

This claim further implies that ε_k is equal to the minimum number of vertices that must be removed from G_D in order for the graph to be k-colorable. From here on, I will refer to the vertices of G_D and the observations they represent interchangeably. Additionally, I will define the pushforward measure

$$g_*(\nu_n)(G) = \nu_n(D \,:\, D \in g^{-1}(G)),$$

to be the distribution over graphs induced by the measure ν_n and map g.

Finally, for a given set of choice problems $\mathcal{P} = \{(A_i, f_i)\}$ and observations $D = \{(x_i, A_i)\}$, say that observation i is perfectly maximized if x_i is $\succ_{f_i}$-maximal in A_i. For every set of choice problems $\mathcal{P}$, let $D^*(\mathcal{P})$ denote the data that would have been generated from the choice problems in $\mathcal{P}$ if the DM perfectly maximized in every choice problem, and define $G_{D^*(\mathcal{P})} = g(D^*(\mathcal{P}))$ to be the corresponding graph.

Observation 3. *(a) For any set of choice problems $\mathcal{P}$, the perfect maximization graph $G_{D^*(\mathcal{P})}$ is K-colorable. (b) Suppose $\delta(R, \mathcal{P}) > d$; then, the graph $G_{D^*(\mathcal{P})}$ contains at least d disjoint K-partite subgraphs.*

Part (a) follows from the definition of $D^*(\mathcal{P})$, and part (b) follows directly from the definition of the differentiation parameter $\delta(R, \mathcal{P})$.

Appendix B: Proofs from Main Text

.1 Proof of Theorem 1

The desired result follows from two lemmas.

Lemma 1. *There exists a constant $c_1 > 0$ (uniform across n) such that*

$$g_*(\nu_n)(\{D : k + \lambda\varepsilon_k > K + \lambda\varepsilon_k \ \forall \ k > K\}) \geq 1 - e^{-c_1 n} \quad \forall n$$

Proof. Fix any set of choice problems $\mathcal{P} = \{(A_i, f_i)\}_{i=1}^n$, and let $D = \{(x_i, A_i)\}_{i=1}^n$ describe the observed data. Suppose towards contradiction that there exists an integer $k > K$ such that

$$\varepsilon_K - \varepsilon_k > (k - K)/\lambda. \tag{16}$$

Observe that this implies

$$\varepsilon_K > 1/\lambda. \tag{17}$$

since $\varepsilon_k \geq 0$ and $k - K \geq 1$. Let $Y \sim \text{Bin}(n, p)$ be the number of observations in error. Removing each of the corrresponding Y nodes from G_D results in a subgraph of $G_{D^*(\mathcal{P})}$, which we know from Obs. 3 is K-colorable. Thus, $\varepsilon_K \leq Y$, and it follows from (17) that $\Pr(Y \leq 1/\lambda)$ is an upper bound on the probability that an integer $k > K$ exists such that (16) obtains.

Since by assumption $\frac{1}{\lambda} = \bar{p}n < pn$, it follows from Hoeffding's Inequality that

$$\Pr(Y \leq 1/\lambda) = \Pr(Y - pn \leq 1/\lambda - pn) \leq \exp\left(-\frac{2(1/\lambda - pn)^2}{n}\right) = e^{-2(\bar{p}-p)^2 n}$$

Moreover, the number of imperfectly maximized observations Y is independent of the set of choice problems $\mathcal{P}$ (conditional on fixed size n). Therefore,

$$g_*(\nu_n)(\{D : k + \lambda\varepsilon_k > K + \lambda\varepsilon_k \ \forall \ k > K\}) \geq 1 - \Pr(X \leq 1/\lambda)$$
$$\geq 1 - e^{-2(\bar{p}-p)^2 n}$$

Let $c_1 := 2(\bar{p} - p)^2 > 0$, and the desired inequality follows. $\qquad\square$

Lemma 2. *For every $n \geq 1$,*

$$g_*(\nu_n)(\{D : k + \lambda\varepsilon_k > K + \lambda\varepsilon_k \ \forall \ k < K\}) \to 1 \ as \ n \to \infty.$$

Proof. Fix any set of choice problems $\mathcal{P} = \{(A_i, f_i)\}_{i=1}^n$. Suppose towards contradiction that there exists some integer $k < K$ such that

$$\varepsilon_K - \varepsilon_k < (K - k)/\lambda.$$

Let us first condition on the event that a set of choice problems $\mathcal{P}$ is realized for which $\delta(R, \mathcal{P}) > \frac{\bar{p}n}{(1-\bar{p})^K}$. Then, from Obs. 3, the perfect maximization graph

$G_{D^*(\mathcal{P})}$ includes at least $\frac{\bar{p}n}{(1-\bar{p})^K}$ complete K-partite subgraphs. Label these subgraphs $j = 1, \ldots, \frac{\bar{p}n}{(1-\bar{p})^K}$, and define indicator variables X_j to take value 1 if and only if every node in the j-th complete K-partite subgraph is perfectly maximized.[29] Now define X to be the sum of the random variables X_j, $1 \leq j \leq \frac{\bar{p}n}{(1-\bar{p})^K}$. Since each $X_j \sim \text{Ber}((1-p)^K)$, the expected sum is $\mathbb{E}(X) = \frac{\bar{p}n}{(1-\bar{p})^K}(1-p)^K$. By construction, $\frac{1}{\lambda} = \bar{p}n < \frac{\bar{p}n}{(1-\bar{p})^K}(1-p)^K$, so it follows from Hoeffding's inequality that

$$\Pr(X < 1/\lambda) = \Pr\left(X - \mathbb{E}(X) < 1/\lambda - \mathbb{E}(X)\right)$$
$$\leq \exp\left(-2\bar{p}n\left((1-\bar{p})^K - (1-p)^K\right)^2\right)$$

Let $c_2 := 2\bar{p}\left((1-\bar{p})^K - (1-p)^K\right)^2$. Then,

$$\Pr(X \geq 1/\lambda) \geq 1 - e^{-c_2 n} \to 1. \tag{18}$$

Additionally, from assumption 10, we have that

$$\pi^n\left(\left\{\mathcal{P} : \delta(R, \mathcal{P}) > \frac{\bar{p}n}{(1-\bar{p})^K}\right\}\right) \to 1 \text{ as } n \to \infty. \tag{19}$$

Let E be the event in which $\{\mathcal{P} : \delta(R, \mathcal{P})\}$. Then, we have that

$$g_*(\nu_n)(\{D : k + \lambda\varepsilon_k > K + \lambda\varepsilon_k \ \forall \ k < K\}) \geq$$
$$\underbrace{\pi^n(E)}_{\to 1 \text{ by } (19)} \underbrace{\Pr\left(X \geq \frac{1}{\lambda} \mid E\right)}_{\to 1 \text{ by } (18)} \to 1 \text{ as } n \to \infty,$$

and the desired statement follows. $\qquad\square$

.2 Proof of Corollary 1

Fix any dataset $\underline{D} = \{(x_i, \underline{A}_i)\}_{i=1}^n$. Define $X = \{x_i\}_{i=1}^n$ to consist of the choice alternatives that were chosen in some observation, and let $A_i = X \bigcap \underline{A}_i$ for every $i = 1, \ldots, n$. Set $D = \{(x_i, A_i)\}_{i=1}^n$.

Lemma 3. *For every $\lambda \in \mathbb{R}_+$,*

$$\min_{k \in \mathbb{N}} k + \lambda\varepsilon_k = \min_{k \in \mathbb{N}} k + \lambda\underline{\varepsilon}_k$$

Proof. I will show that $\varepsilon(D, R) = \delta$ for some set $R = \{\succ_j\}_{j=1}^k$ if and only if $\underline{\varepsilon}(D, \mathcal{U}) = \delta$ for some set $\mathcal{U} = \{u_j\}_{j=1}^k$.

---

[29]That is, the observation corresponding to this node is perfectly maximized.

Fix any set $R = \{\succ_j\}_{j=1}^k$ of k orderings defined on X, and take $\delta := \varepsilon(D, R)$. Every ordering $\succ_j$ admits representation via a utility function $u_j : X \to \mathbb{R}$.[30] Moreover, we can extend each u_j to a continuous function $\underline{u}_j$ on $\underline{X}$ satisfying $\mathrm{argmax}_{x \in \underline{X}} \underline{u}_j(x) = \mathrm{argmax}_{x \in X} u_j(x)$. Then, $x_j = \mathrm{argmax}_{x \in \underline{A}} \underline{u}_j(x)$ if and only if x_j is $\succ$-maximal in A. Set $\mathcal{U} = \{\underline{u}_j\}$; then,

$$\underline{\varepsilon}(\underline{D}, \mathcal{U}) = \#\left\{ (x, \underline{A}) \in \underline{D} \,:\, x \neq \mathrm{argmax}_{x' \in \underline{A}} \underline{u}_j(x') \text{ for all } j = 1, \ldots, k \right\} = \delta.$$

Thus, choice error δ is attainable using a set of k utility functions. It follows that

$$\min_{k \in \mathbb{N}} k + \lambda \varepsilon_k \geq \min_{k \in \mathbb{N}} k + \lambda \underline{\varepsilon}_k. \tag{20}$$

In the other direction, fix a set $\mathcal{U} = \{u_j\}_{j=1}^k$ of k continuous functions defined on $\underline{X}$, and take $\delta = \underline{\varepsilon}(\underline{D}, \mathcal{U})$. For every utility function u_j, let $\succ_j$ be an ordering on X that satisfies

$$x \succ_j x' \iff u_j(x) > u_j(x').$$

Then, $x_j = \mathrm{argmax}_{x \in \underline{A}} u_j(x)$ if and only if x_j is $\succ$-maximal in A. Setting $R = \{\succ_j\}_{j=1}^k$, we have that

$$\varepsilon(D, R) = \#\{(x, A) \in D \,:\, x \text{ is not } \succ_j \text{-maximal in } A \text{ for any } j = 1, \ldots, k\} = \delta.$$

Thus, also

$$\min_{k \in \mathbb{N}} k + \lambda \varepsilon_k \leq \min_{k \in \mathbb{N}} k + \lambda \underline{\varepsilon}_k.$$

$\square$

It follows from this lemma that $\underline{K}_\lambda^* = K_\lambda^*$ for every choice of λ, so the problem posed in Section 6.2 can be mapped directly into a corresponding problem involving choice over a discrete set. Directly apply Theorem 1, and we are done.

.3 Proof of Proposition 1

Let Π be the set of all permutations of $(1, 2, \ldots, |X|)$. It will be useful to identify every ordering $\succ$ on X with an ordinal vector $\mathbf{r}^\succ \in \Pi$ such that

$$\mathbf{r}^\succ(x) > \mathbf{r}^\succ(x') \iff x \succ x'.$$

First, I will show the following:

---

[30]That is, there exists a utility function u_j such that for every $x, x' \in X$, $x \succ_j x'$ if and only if $u_j(x) > u_j(x')$.

Claim 6. *If there exist orderings $\succ_1, \succ_2 \in R$ such that*

$$\operatorname{argmax}_{x \in X} \mathbf{r}^{\succ_1}(x) \neq \operatorname{argmin}_{x \in X} \mathbf{r}^{\succ_2}(x) \tag{21}$$

then R is not identifiable.

Proof. Consider any set of orderings R such that there exist orderings $\succ_1, \succ_2 \in R$ satisfying

$$x_1 := \operatorname*{argmax}_{x \in X} \mathbf{r}^{\succ_1}(x) \neq \operatorname*{argmin}_{x \in X} \mathbf{r}^{\succ_2}(x) := x_2, \tag{22}$$

Fix any dataset D such that $\varepsilon(D, R) = 0$. I will show by construction that there exists another set $R' \neq R$ with $|R'| \leq |R|$ such that also $\varepsilon(D, R') = 0$.

Define the ordering $\succ_2'$ to agree with $\succ_2$ everywhere, except that it ranks x_1 last.[31] Let $R' = R - \{\succ_2\} + \{\succ_2'\}$, where the operators denote set addition and subtraction. I will now show that $\varepsilon(D, R') = 0$.

Suppose towards contradiction that there is some choice observation $(x, A) \in D$ such that x not $\succ$-maximal in A for any $\succ \in R'$. By assumption, there is some ordering $\succ^* \in R$ such that x is $\succ^*$-maximal in A. If $\succ^* \neq \succ_2$, then also $\succ^* \in R'$, which immediately yields a contradiction. So it must be that x is $\succ_2$-maximal in A. Now, there are two possibilities: if $x \neq x_1$, then x must also be $\succ_2'$-maximal in A, and we are done. If instead $x = x_1$, then x is $\succ_1$-maximal in A (by definition in (22)). So we are again done. □

Since every set of three or more orderings satisfies the condition in (21), it immediately follows from Claim 6 that every set R with three or more orderings is not identifiable.

Now let us consider $R = \{\succ_1, \succ_2\}$. Index the alternatives such that x_1 is ranked first according to $\succ_1$ and x_2 is ranked first according to $\succ_2$. If (21) is satisfied, it again follows from Claim 6 that R is not identifiable. Suppose otherwise, so that x_1 is ranked last according to $\succ_2$ and x_2 is ranked last according to $\succ_1$. Define

$$D = \{(x, A) : x \text{ is } \succ_1\text{-maximal in } A, \ A \in 2^X\} \cup$$
$$\{(x, A) : x \text{ is } \succ_2\text{-maximal in } A, \ A \in 2^X\}.$$

Clearly there is no singleton set R' such that $\varepsilon(D, R') = 0$. Suppose towards contradiction that there exists some set $R' = \{\succ_1', \succ_2'\} \neq R$ such that $\varepsilon(D, R') = 0$. Without loss of generality, suppose that $\succ_1' \neq \succ_1$.[32] Then there exist alternatives x_i, x_j such that

$$x_i \succ_1 x_j \quad \text{and} \quad x_j \succ_1' x_i; \tag{23}$$

that is, x_i is preferred to x_j under $\succ_1$ but not under $\succ_1'$. Take

$$A := \{x : x_i \succsim_1 x\}$$

---

[31]That is, for every $x, x' \neq x_2$, $x \succ_2' x'$ if and only if $x \succ_2 x'$, and for every $x \neq x_2$, $x \succ_2' x_2$.

[32]Otherwise, $\succ_2' \neq \succ_2$ and the remainder of the proof is correspondingly mirrored.

to be the set of all alternatives that $\succ_1$ ranks weakly lower than x_i. From (23), necessarily $x_j \in A$. Also from (23), $x_j \succ_1' x_i$, so that x_i is not $\succ_1'$-maximal in A. Since $(x_1, X), (x_2, X) \in D$, either x_1 or x_2 must be $\succ_2'$-maximal in X. Suppose x_1. Then, x_1 is also $\succ_1'$-minimal in X, so $x_1 \in A$. But then x_i cannot be $\succ_2'$-maximal in A, since $x_2 \succ_2' x_i$. Suppose instead that x_2 is $\succ_2'$-maximal in X. Then, x_2 is $\succ_1'$-minimal in X, so that $x_2 \in A$. Again, x_i cannot be $\succ_2'$-maximal in A, since now $x_1 \succ_2' x_i$. Thus (x_i, A) is not consistent with maximization of either ordering in R', yielding the desired contradiction.

Finally, every singleton set $R = \{\succ\}$ is trivially identifiable, taking

$$D = \left\{ (x, A) \; : \; \text{is } \succ\text{-maximal in } A, \; A \in 2^X \right\}.$$

.4 Proof of Claim 2

(a) Suppose $R = \{\succ_1, \succ_2\}$ satisfies (12). Define

$$D = \left\{ \left(x_{(A,1)}^*, A \right), \left(x_{(A,2)}^*, A \right) \right\}_{A \in 2^X}$$

to both include observation of choice of the $\succ_1$-maximal alternative from every choice set A, and also the $\succ_2$-maximal alternative from every choice set A. Certainly $\varepsilon(D, R) = 0$. Consider any other R', with $|R'| \leq |R|$, such that $\varepsilon(D, R') = 0$.

Suppose $c(R') \subset c(R)$. Then, there is some (x, A) such that x is $\succ$-maximal in A for some $\succ$ in R, but not $\succ'$-maximal in A for any $\succ'$ in R'. But $(x, A) \in D$, so $\varepsilon(D, R') \neq 0$. This yields the desired contradiction. Now suppose that $c(R') \supset c(R)$. Then, there is some (x, A) such that x is $\succ$-maximal in A for some $\succ$ in R', but not $\succ'$-maximal in A for any $\succ'$ in R. Let x_1 denote the $\succ_1$-maximal alternative in A and x_2 denote the $\succ_2$-maximal alternative in A; by construction of R, these two alternatives are different. It must therefore be that $x \neq x_1 \neq x_2$. But $(x_1, A), (x_2, A) \in D$ and $\varepsilon(D, R') = 0$, so there must exist distinct orderings $\succ_1', \succ_2' \in R'$ such that x_1 is $\succ_1'$-maximal in A, and x_2 is $\succ_2'$-maximal in A. Since moreover $(x, A) \in c(R')$, there must exist also a third ordering $\succ_3'$ such that x is $\succ_3'$-maximal in A. This implies that $|R'| \geq 3 > |R|$, yielding the desired contradiction.

(b) Suppose $|R| = |X|$, and the $\succ$-maximal alternative in X is distinct for every ordering in R. Then, every alternative $x \in X$ is $\succ$-maximal in A for some ordering in R. Let us denote $\succ_x$ for the ordering in which x is ranked first. Consider any dataset D. For every $(x, A) \in D$, it must be that x is $\succ_x$-maximal in A. So $\varepsilon(D, R) = 0$ as desired, and R is choice-identifiable.

Now suppose $|R| = |X|$, but not every maximal alternative is distinct. Consider any dataset D such that $\varepsilon(D, R) = 0$. Now construct an alternative set of orderings R' with $|R'| = |X|$, such that the $\succ$-maximal alternative in X is distinct for every ordering in R'. By the previous argument, $\varepsilon(D, R') = 0$. But $|R'| = |R|$ and $c(R') \neq c(R)$, so R is not choice-identifiable.

Proof of Claim 3

Fix any $\lambda < \frac{1}{\varepsilon_1}$. Consider any ordering $\succ \in \mathcal{R}_{HM}$. Then,

$$l(\{\succ\}) + \lambda\varepsilon(D, \{\succ\}) = (2^{|X|} - 1) + \lambda\varepsilon_1. \tag{24}$$

Now consider any other set R'. Suppose $R' = \{\succ\}$ for some $\succ \notin \mathcal{R}_{HM}$. Then, $\varepsilon(D, R') > \varepsilon_1$, and we have

$$l(R') + \lambda\varepsilon(D, R') = (2^{|X|} - 1) + \lambda\varepsilon(D, \{\succ\}) > (2^{|X|} - 1) + \lambda\varepsilon_1 > l(\{\succ\}) + \lambda\varepsilon(D, \{\succ\})$$

from (24), so R' will not be returned as a solution. Now suppose that R' consists of more than one ordering. Then, $l(R') \geq 2^{|X|}$, so that

$$\begin{aligned}
l(R') + \lambda\varepsilon(D, R') &\geq 2^{|X|} + \lambda\varepsilon(D, R') \\
&\geq 2^{|X|} \\
&\geq (2^{|X|} - 1) + \left(\frac{1}{\varepsilon_1}\right)(\varepsilon_1) \\
&> l(\{\succ\}) + \lambda\varepsilon(D, \{\succ\}),
\end{aligned}$$

and again R' will not be returned as a solution. So $R \in \operatorname{argmin}_{R \in \mathcal{R}} l(R) + \lambda\varepsilon(D, R)$ if and only if $R = \{\succ\}$ for some $\succ \in \mathcal{R}_{HM}$.

(b) Fix any $\lambda > 2^{|X|} - 1$, and suppose towards contradiction that

$$\{(x, A) \mid (x, A) \in D\} \nsubseteq c(R_\lambda^*)$$

Then, there is some (x, A) such that $(x, A) \in D$ but $(x, A) \notin c(R_\lambda^*)$. Define $R = R_\lambda^* \bigcup \{\succ\}$ for any $\succ$ such that x is $\succ$-maximal in A. Then,

$$l(R) \leq l(R_\lambda^*) + 2^{|X|} - 1$$

and

$$\varepsilon(D, R) \leq \varepsilon(D, R_\lambda^*) - 1$$

so that

$$\begin{aligned}
l(R) + \lambda\varepsilon(D, R) &\leq (l(R_\lambda^*) + (2^{|X|} - 1)) + \lambda(\varepsilon(D, R_\lambda^*) - 1) \\
&\leq (l(R_\lambda^*) + (2^{|X|} - 1)) + \lambda\varepsilon(D, R_\lambda^*) - (2^{|X|} - 1) \\
&\leq l(R_\lambda^*) + \lambda\varepsilon(D, R_\lambda^*)
\end{aligned}$$

yielding the desired contradiction.

33

Proof of Claim 4

Condition on the event that a dataset D is realized for which $c(R) \subseteq D$, and suppose towards contradiction that $c(R_\lambda^*) \neq c(R)$. Then,

$$l(R') + \varepsilon(D, R') \underbrace{<}_{(1)} l(R') + \lambda\varepsilon(D, R') \underbrace{\leq}_{(2)} l(R) + \lambda\varepsilon(D, R) \underbrace{=}_{(3)} l(R)$$

using in (1) that $\lambda > 1$, in (2) the definition of R_λ^*, and in (3) that $p = 0$. This implies that

$$\varepsilon(D, R') < l(R) - l(R').$$

But then there must exist some (x, A) with $(x, A) \in c(R)$ but $(x, A) \notin D$, contradicting our assumption that $c(R) \subseteq D$. By Assumption 2, the probability that $c(R) \subseteq D$ converges to 1 in the quantity of data, and we are done.

Proof of Proposition 2

Consider any choice observation $(x, A) \in c(R)$. By assumption in (14), there exists a frame f such that x is $\succ_f$-maximal in A, and moreover $\pi(A, f) > \frac{\bar{p}}{1-\bar{p}}$. Let $Z \sim \text{Bin}\left(n, \frac{\bar{p}(1-p)}{1-\bar{p}}\right)$ be the number of times that (A, f) is both sampled and perfectly maximized. Since by assumption, $\frac{1}{\lambda} = \bar{p}n < \left(\frac{1-p}{1-\bar{p}}\right)\bar{p}n$, it follows from Hoeffding's inequality that

$$\Pr\left(Z < \frac{1}{\lambda}\right) = \Pr\left(Z - \frac{\bar{p}(1-p)n}{1-\bar{p}} < \frac{1}{\lambda} - \frac{\bar{p}(1-p)n}{1-\bar{p}}\right)$$

$$= \Pr\left(Z - \frac{\bar{p}(1-p)n}{1-\bar{p}} < \bar{p}n - \frac{\bar{p}(1-p)n}{1-\bar{p}}\right)$$

$$\leq \exp\left(-2\left(\bar{p} - \frac{\bar{p}(1-p)}{1-\bar{p}}\right)^2 n\right)$$

Setting $c_1 = 2\left(\bar{p} - \frac{\bar{p}(1-p)}{1-\bar{p}}\right)^2 > 0$, it follows that

$$\Pr\left(Z < \frac{1}{\lambda}\right) \leq e^{-c_1 n}.$$

and subsequently, the probability (x, A) appears fewer than $1/\lambda$ times in the realized dataset is no more than $e^{-c_1 n}$. Taking a union bound, the probability that any $(x, A) \in c(R)$ appears fewer than $1/\lambda$ times in the data is no more than $|c(R)|e^{-c_1 n}$. Thus, the probability that every $(x, A) \in c(R)$ appears at least $1/\lambda$ times in the realized dataset is at least

$$1 - |c(R)|e^{-c_1 n},$$

which converges to 1 as the quantity of data n increases. This immediately implies that

$$\Pr\left(c\left(R\right) \subseteq c(R_\lambda^*)\right) \to 1 \text{ as } n \to \infty. \tag{25}$$

In the other direction, let random variable $Y \sim \text{Bin}(n, p)$ be the total number of imperfectly maximized choices in the realized dataset. Then, $\Pr(Y \leq 1/\lambda)$ is an upper bound on the probability that any $(x, A) \notin c(R)$ appears $1/\lambda$ times in the realized data. Since moreover $\frac{1}{\lambda} = \bar{p}n < pn$, it follows from Hoeffding's Inequality that

$$\Pr(Y \leq 1/\lambda) = \Pr(Y - pn \leq 1/\lambda - pn) \leq e^{-2(\bar{p}-p)^2 n}. \tag{26}$$

Condition on the event in which $Y \leq 1/\lambda$, and suppose towards contradiction that $c(R_\lambda^*) \subseteq c(R)$. From the previous argument, with probability $1 - |c(R)|e^{-c_1 n}$,

$$c(R) \subseteq c(R_\lambda^*).$$

Moreover, from (26), with probability $1 - e^{-2(\bar{p}-p)^2 n}$, each of the choice implications in (the set difference) $c(R_\lambda^*) - c(R)$ explains no more than $1/\lambda$ unique new observations. Therefore,

$$l(R) + \lambda\varepsilon(D, R) \leq l(R) + \lambda(\varepsilon(D, R_\lambda^*) + (l(R_\lambda^*) - l(R))/\lambda)$$
$$= l(R_\lambda^*) + \lambda\varepsilon(D, R_\lambda^*)$$

This yields the desired contradiction. So $c(R_\lambda^*) \subseteq c(R)$ with probability

$$\left(1 - |c(R)|e^{-c_1 n}\right)\left(1 - e^{-2(\bar{p}-p)^2 n}\right),$$

implying that

$$\Pr(c(R_\lambda^*) \subseteq c(R)) \to 1.$$

Combine this with (25) and we are done.

References

Afriat, Sidney N. 1967. "The Construction of a Utility Function from Expenditure Data." *International Economic Review* 8(1):67–77.

Ambrus, Attila & Kareen Rozen. 2013. "Rationalizing Choice with Multi-Self Models." *Economic Journal* .

Apesteguia, Jose & Miguel A. Ballester. 2012. "A Measure of Rationality and Welfare." Working Paper.

Bernheim, B. Douglas & Antonio Rangel. 2009. "Beyond Revealed Preference: Choice Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* .

Crawford, Ian & Krishna Pendakur. 2012. "How Many Types Are There?" *Economic Journal* .

Dean, Mark & Dan Martin. 2010. "How Consistent are your Choice Data?" Working Paper.

Dean, Mark & Daniel Martin. 2016. "Measuring Rationality with the Minimum Cost of Revealed Preference Violations." *Review of Economics and Statistics* .

Deb, Rahul. 2009. "A testable model of consumption with externalities." *Journal of Economic Theory* 144.

Echenique, Federico, Sangmok Lee & Matthew Shum. 2011. "The Money Pump as a Measure of Revealed Preference Violations." *Journal of Political Economy* 119(6).

Einav, Liran, Amy Finkelstein, Iuliana Pascu & Mark Cullen. 2012. "How General are Risk Preferences? Choices under Uncertainty in Different Domains." *American Economic Review* .

Famulari, Melissa. 1995. "A Household-Based, Nonparametric Test of Demand Theory." *Review of Economics and Statistics* 77:372–383.

Fudenberg, Drew & David Levine. 2006. "A Dual-Self Model of Impulse Control." *American Economic Review* 96(5):1449–1476.

Green, Jerry & Daniel Hojman. 2007. "Choice, Rationality and Welfare Measurement.".

Gross, Andrew. 1989. Determining the number of violators of the weak axiom. Technical report University of Wisconsin–Milwaukee.

Halevy, Yoram, Dotan Persitz & Lanny Zrill. 2015. "Parametric Recoverability of Preferences." Working Paper.

Houtman, Martijn & J.A.H. Maks. 1985. "Determining all Maximal Data Subsets Consistent with Revealed Preference." *Kwantitatieve Methoden* 19:89–104.

Kalai, Gil, Ariel Rubinstein & Ran Spiegler. 2002. "Rationalizing Choice Functions by Multiple Rationales." *Econometrica* 70(6):2481–2488.

Manzini, Paola & Marco Mariotti. 2007. "Sequentially Rationalizble Choice." *American Economic Review* .

Manzini, Paola & Marco Mariotti. 2009. "Categorize Then Choose: Boundedly Rational Choice and Welfare.".

McFadden, Daniel & M.K. Richter. 1970. "Revealed Stochastic Preference.".

Quandt, Richard E. 1956. "A Probabilistic Theory of Consumer Behavior." *Quaterly Journal of Economics* 70:507–536.

Rubinstein, Ariel & Yuval Salant. 2006. "A model of choice from lists." *Theoretical Economics* 3(17).

Rubinstein, Ariel & Yuval Salant. 2008. "(A,f): Choice with Frames." *The Review of Economics Studies* .

Train, Kenneth. 1986. *Qualitative Choice Analysis*. Cambridge University Press.

Varian, Hal R. 1982. "The Nonparametric Approach to Demand Analysis." *Econometrica* 50(4):945–73.