Rationality and Error in Individual Choice Data:
A Revealed Preferences Approach

Aluma Dembo*

November 20, 2019

- Preliminary Draft -

For current version [click here: www.alumadembo.com/JMP.html]

Abstract

The assumption of rationality is used to infer preferences from observed choices, but classic economic theory provides scant guidance when choice data has error. Methods to estimate preferences from noisy data inevitably invoke additional assumptions on those very preferences. This paper presents a procedure to detect and measure error in an individual's observed choice data when the individual has an underlying rational choice process that has been contaminated with random implementation errors. Using a single individual's choices over many menus, I construct an observed revealed preference relation, and prove it is a random graph whose acyclicity is equivalent to rationality. Exploiting the structure in the graph produced by the contaminating errors, I devise a classifier able to detect which observations are errors and an estimator to measure the rate at which errors occur. These two methods can be applied to any dataset in which an individual makes constrained choices from a sequence of overlapping non-identical menus, regardless of the choice environment. I apply the method to a benchmark dataset of choices observed in the lab (Choi et al. 2007) and show that most individuals have error rates between 5% and 14.5% (interquartile range). I show that three existing measures of goodness-of-fit for rationality, which are often used as proxies for error estimates, are not robust, not identified, or biased when choices are observed with error.

Introduction

The defining feature of economic choice models is that they make use of observed choices, and any available alternatives that were *not chosen*. Repeated observations of choices and unchosen alternatives can be reconciled as long as certain consistency conditions hold. The decision-maker is then called rational, and preferences can be inferred from their observed choices. Most modern economic theory is built on a foundation of rational choice, and rationality drives results in a number of other fields including industrial organization, finance, and welfare analysis.

Yet, in both lab and field studies, many individuals are observed making choices that fail the necessary conditions for rationality (Choi et al. 2007b; Halevy et al. 2018; Dean and Martin 2016; Echenique et al. 2011). This is unsurprising since the classic test for rational choice treats observations as deterministic, and does not account for the random error inherent in empirical data. Observed failures of individual rationality are interpreted as empirical error, and economists continue to apply economic choice models, often by appending an error term to some parametric choice function. There is certainly a practicality to this approach, despite the sometimes shaky theoretical groundings. Perhaps more problematic is that a parametric model is not suitable when the set of available alternatives cannot be summarized in an obvious way such as by prices or by characteristics. In these situations a nonparametric method like the revealed preference relation is better for incorporating the set of available but unchosen alternatives into the model of choice.

In this paper I use revealed preference relations to separate errors from rational choices for an individual choosing from generic menus of alternatives. The *revealed preference relation* assumes that by making a choice the individual declares that option to be weakly preferred to all other available options. When the researcher observes an individual's choices from a sequence of overlapping (non-identical) menus, the researcher can create a binary relation to capture this notion of preference. I develop a classifier based on the revealed preference relation to recover a subset of observed choices that are rational and that can be used to infer preferences. The classifier also flags possible errors; these can be tested for correlation with other observables from the time of the decision that may affect behavior. I also develop an estimator for how much error is needed to explain observed deviations from rationality in the choice data. This measure is interpretable and can be used to assess the quality of the choice data more generally. It can be compared across individuals and correlated with other traits, or it can be compared across choice contexts for the same individual to test behavioral hypotheses. The methods I develop are designed to be easily implemented in generic choice environments including choices over discrete menus.

The direct revealed preference relation was first introduced by Samuelson (1938) in the context of consumption bundles chosen under a price-restricted budget constraint. A large literature has developed conditions necessary for choice data to be rational under a variety of settings and assumptions.[1] I use the consistency condition from Nishimura, Ok, and Quah (2017) which allows my method to be applied to a general choice environment with minimal restrictions on the observed data.

In keeping with all classic papers on the topic, Nishimura et al. (2017) take the choice data as given, and do not address random error in empirical data. A number of papers incorporate error as an index measuring how far observed data is from rational choice (see Apesteguia and Ballester (2015); Halevy et al. (2018) for summaries). Inherent in this approach is the assumption that minimizing a distance measure to a rational dataset is meaningful, but this lacks a statistical foundation. I show that three of the most commonly used indices (Afriat 1972; Houtman and Maks 1985; Varian 1990) do not satisfy basic statistical robustness and consistency properties.[2] There is a small literature that incorporates a statistical foundation by explicitly modeling random error in the data (Epstein and Yatchew 1985; Varian 1985; Gross 1995; Fleissig and Whitney 2005; Echenique, Lee, and Shum 2011; Hjertstrand 2013). These papers all assume the data takes the form of consumption bundles chosen from budget constraints defined by prices. However, the error structure underpinning the statistical results is either additive in the consumption space or multiplicative in prices, and therefore cannot be extended to generic choice spaces.

I extend the framework of general choice data to explicitly have errors in choices. The researcher observes a choice drawn from a random distribution characterized by a measure of dispersion. The individual has a deterministic true choice, which is observed by the researcher if dispersion is zero. When choices are observed with error and the menus vary, an observation is a single draw from one of a sequence of non-identical distributions.[3] To relate these distributions I impose structure on the true choices by assuming only that they are rational.

---

[1] There are two strands in the literature; one focuses on finite consumption data (Houthakker 1950; Gale 1960; Afriat 1967; Hanoch and Rothschild 1972; Diewert 1973; Varian 1983), and the other looks at choices in abstract settings with restrictive assumptions on the dataset (Uzawa 1956; Arrow 1959; Richter 1966; Sen 1971). Chambers and Echenique (2016) provide an excellent review of the literature.

[2] When errors are baked in to the choices, I show that (i) Afriat's efficiency index is not robust to outliers, (ii) Varian's index is not well-identified, and (iii) the Houtman-Maks index is biased and underestimates the error.

[3] In contrast, stochastic choice models use a representative agent with random utility which captures the population heterogeneity in preferences. Empirical estimation requires the empirical distribution of choices to be observed (McFadden 2005; Kitamura and Stoye 2018). This is also unlike the related recent literature on consideration sets in which the menus considered by the individual are not necessarily those observed by the researcher(Manzini and Mariotti 2014; Abaluck and Adams 2018). In both contexts the choices are observed without error. An exception is Aguiar and Kashaev (2018) who develop a test of rationality with measurement error, which requires pooling across individuals and is only applicable to consumption data.

Additionally, I assume the choice distribution follows a contaminated data model (Huber 1964; Horowitz and Manski 1995); with probability π the observed choice is a randomly distributed error, and with probability $1 - \pi$ it is the true choice.[4] The contaminated data model is well suited for situations where error is thought to occur sporadically. While I remain agnostic as to the source of these mistakes, some types of errors persist for the duration of an experiment and are not well described by the contamination model. These include underreporting, Hawthorne effects, and rounding errors. The contaminated data model works well for implementation mistakes on the part of the individual, or mis-recording of observations on the part of the researcher or experiment. The model puts a probability point mass on the true rational choice, and distributes the remaining probability freely on the menu. It can be used in general choice settings since the error distribution can take any form, including discrete probabilities on finite unordered sets.

An observed revealed preference relation is a data construct, so it will inherit the random error present in the data. Coupling the contaminated data model with rationality of the true choices produces a random graph with a specific structure called the stochastic block model (Snijders and Nowicki 1997). A random graph is a collection of vertices and a set of edges connecting pairs of vertices. In the context of this paper, each observation is a vertex and it connects to another observation if they violate a pairwise consistency condition. From the contaminated data model there are two types of observations, either "error" (with probability π) or "error-free" (with probability $1 - \pi$). The probability of two observations forming a violation depends on their types. The researcher does not observe any observation's type, only a realization of the graph. Correctly classifying the observations into types recovers the true revealed preference relation, and an estimate for the error rate π. I show that the way a revealed preference relation is constructed causes it to have a unique dependency structure. Most existing estimation methods in the networks literature rely on independence and are therefore not applicable to this setting.[5]

I develop a classifier that leverages this dependency structure to uncover a subset of the error observations. I also develop an estimator for the error rate π which requires additional assumptions on the distribution of errors. The methods I develop operate on the observed revealed relation and so do not suffer from dimensionality or computational complexity issues. I show through monte carlo simulations that as the number of observations increase the methods perform better.

---

[4]A version of the contaminated data model has been used in empirical work testing expected utility in choices over lotteries (Hey and Orme 1994).

[5]Existing methods to recover the classifications rely on either a known labeling probability, known same-type and cross-type edge probabilities, or independence (Bickel and Chen 2009; Rohe et al. 2011; Channarond et al. 2012; Zhao et al. 2012; Fishkind et al. 2013).

My paper uses consumption choices over two goods as its running example, but the methods are designed to work in generic choice settings. I apply the classifier and estimator to a benchmark dataset of choice data on risky assets collected in the lab (Choi, Fisman, Gale, and Kariv 2007a). Among the 47 subjects I find that most have error rates between 5% and 14.5%.

This paper is divided into three parts (1 Theory, 2 Model, and 3 Estimation) which can be read separately, with minimal referencing. In Section 1, I outline the framework for the choice data and observed revealed preference relation. I prove that acyclicity of the observed revealed preference relation is equivalent to rationality (under certain sampling properties of the menus). I also prove that when the choice distribution has a positive dispersion, as the number of observations increases data will violate acyclicity almost surely. Related to this are negative results on existing indices of rationality (Afriat 1972; Houtman and Maks 1985; Varian 1990), which can be found in Appendix A.

Section 2 defines the contaminated data model in the context of individual choice data. I prove that the observed revealed preference relation is identified by the true choice and the error rate π, for a given error distribution. I show how the observed revealed preference relation can be used to construct a graph with identifiable partitions. I provide monte carlo simulations and examples to illustrate the properties of this random graph.

Section 3 presents the classifier and estimator. I use monte carlo results to demonstrate the performance of these methods. Section 4 contains the empirical results of applying the classifier and estimator to data from Choi et al. (2007a).[6] Images and tables are interspersed throughout, but proofs are in Appendix B. Related literature is discussed as it arises. For the unfamiliar reader, an overview of graph notation used in this paper can be found in Appendix D.

A note on transparency: all results and images are reproducible, please contact the author for code files. Simulations were run in R version 3.5.3 using the Mersenne-Twister pseuso-random number generator with arbitrary seeds based on astronomical distances. The running example in the paper uses a seed of 252088, the distance in miles from the earth to the moon at its furthest point.

Regarding font style and notation: in this paper x denotes a number or element in a set and $\mathbf{v}$ denotes a column vector of length n. Each vector consists of an ordered list of numbers $\mathbf{v} = (v_1, v_2, ..., v_n)^\top$. The dot product of two vectors is $\mathbf{v} \cdot \mathbf{x} = \sum_i v_i x_i$ with the transpose operator $^\top$ implied. The object $\mathbf{X}$ denotes a matrix of size $n \times n$ unless otherwise

---

[6]As of this draft only one empirical example is provided. Future drafts will include examples of discrete choices, high-dimensional basket data, and mixed menu-types Hey and Orme (1994); Echenique et al. (2011); Halevy et al. (2018).

specified. A matrix may be written as an ordered list of column vectors $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n]$. An arbitrary collection of objects is denoted $\mathfrak{X}$, which may or may not be indexed. A collection of collections is $\boldsymbol{\mathfrak{X}}$. A distribution function F is often specified by a subscript clarifying the random variable it generates. I have attempted to use standard notation, but occasionally context may dictate the meaning of a symbol.

1 Theory

In this section I incorporate stochastic elements into existing theory. I begin by defining the dataset of a single individual's choices from a sequence of budget sets. I formalize the notion that this individual's choices are observed with random error, and further that they are drawn from a distribution with some measure of dispersion. When the dispersion is zero, the researcher observes classic deterministic rational choices (Section 1.1). I define what is meant by rationality for generic choice settings in Section 1.2.

The main result of this section is Proposition 2 in Section 1.4. It states that when dispersion is positive there is a probability, which is less than 1, of observing a rational dataset. Furthermore, as the number of observations increases, this probability converges to 0.

I introduces a number of assumptions that add restrictions on the choice space, the budget set sampling procedure, or the choice distribution. These are largely technical in nature, but are necessary for the theory to be applicable. Additional independence assumptions are used in the proofs, but it is possible with some effort future researchers may be able to relax these them. Lastly, in section 1.5 I introduce the random graph notation for the observed reveald preferenec relation. This will carry over into the remainder of the paper.

1.1 The observed dataset

We begin with the *choice space* $\mathfrak{X}$, a universal set of alternatives. An individual's choice is an element x from some *budget set* $\mathcal{B}$, a subset of the choice space . An *experiment* consists of n (≥ 2) budget sets indexed $i = 1, ..., n$ forming the collection $\boldsymbol{\mathcal{B}}_n = \{\mathcal{B}_i\}_{i=1}^n$. The researcher records the individual's *observed choice* from each budget set, $x_i \in \mathcal{B}_i$. The *observed dataset* is therefore $\mathcal{D}_n = \{\mathcal{B}_i, x_i\}_{i=1}^n$, which we sometimes write as the union $\{x_i\}_{i=1}^n \cup \boldsymbol{\mathcal{B}}_n$ where matching indices are retained. Each budget set is sampled randomly from a probability space on the power set $2^{\mathfrak{X}}$ with a probability measure $F_{\mathcal{B}}$.[7]

---

[7]The *probability space* consists of a triplet: the sample space $2^{\mathfrak{X}}$, the event space which is a σ-algebra on the collection of sets in $2^{\mathfrak{X}}$, and a probability measure on the event space (a countably additive set function which integrates to 1)

In addition to sampling variation in the budget sets, we also introduce the notion of random *error in the observed choices*. To do so, we distinguish between the observed choice x_i, and what would have been observed had there been no error.

Assumption 1. The *true choice*, x_i^*, is a deterministic function of the menu of alternatives in the budget set $\mathcal{B}_i$.

When there is random error, we can think of each observed choice x_i as a draw from a probability distribution F over the available alternatives in $\mathcal{B}_i$. In a given experiment two budget sets will not necessarily be identical (due to sampling variation in $F_{\mathcal{B}}$). So each observed choice is distributed $x_i \sim F(\cdot|\mathcal{B}_i)$, according to the conditional *choice distribution*.

If choice is observed without error, then the researcher should observe the true choice. To explicitly parameterize the classical error-free case, we use the choice distribution's measure of dispersion.

Assumption 2. Let the *choice distribution* be parameterized by the true choice x_i^* and some measure of dispersion σ. Conditional on the budget set $\mathcal{B}_i$, the choice $x_i \sim F(\cdot; x_i^*, \sigma|\mathcal{B}_i)$. When the dispersion is equal to 0, $x_i = x_i^*$.[8]

The assumption above is intentionally vague so researchers have the flexibility to use context-appropriate choice distributions. Depending on the type of distribution, the choice space, and the nature of budget sets, the dispersion could be a parameter, or it could be an aggregate function of a multi-dimensional parameter. Here are two common examples to help illustrate how a researcher could choose an appropriate measure of dispersion for their setting.

Example 1. The researcher observes choices, by the same individual, from a sequence of budget sets. The researcher expects the true choice to occur with high frequency, and for errors to occur infrequently but at the same rate for all budget sets. In this case the probability of an error is a measure of dispersion satisfying Assumption 2. ◄

Example 2. Suppose choices are made over bundles of L goods from a sequence of budget planes defined by randomly generated prices. The researcher expects each observed choice to fall close to the individual's true choice, and for further-away errors to be less likely. Suppose the choices are distributed as a truncated normal distribution with mean on the true choice. Then the standard deviation (averaged across all L dimensions) is a measure of dispersion satisfying Assumption 2. ◄

---

[8]The distribution need not be identified by the measure of dispersion; it is possible for σ to a be a function of other parameters of the choice distribution.

The following independence assumptions are used in the proof of Proposition 2 as well as the negative results in Appendix A.

Assumption 3. The budget sets are sampled *independently* and *identically* such that $\mathcal{B}_i \sim_{iid} F_{\mathcal{B}}$ for all i.

Assumption 4. Conditional on any $\mathcal{B}_i, \mathcal{B}_j$, the choice distribution F is such that
1. for any $i \neq j$, $x_i \perp x_j$, and
2. if $\mathcal{B}_i = \mathcal{B}_j$, then $x_i \overset{d}{=} x_j$.

The first independence assumption requires that the budget sets in an experiment be sampled independently and identically (Assumption 3 is on $F_{\mathcal{B}}$). The second requires that observed choices be independent and identically distributed conditional on the budget sets (Assumption 4 is on F). In Section 2 we will replace the assumptions on the choice distribution with a semi-parametric model.

1.2 Rationality in generic choice environments

Rationality means there is some underlying weak order on the choice space that the individual uses in order to select a most preferred option from a set. In practice we often require that this order be consistent with an objective ordering on the choice space.[9]

To that end, we assume the choice space $\mathcal{X}$ is governed by an exogenous and objective dominance relation $\unrhd$, which is a partial order.[10] Choices are $\unrhd$-rationalizable, or *rational*, if there is some $\succsim$ preference relation that is consistent with (extends) $\unrhd$, and each choice x_i belongs to the subset of $\mathcal{B}_i$ that is maximal according to $\succsim$.[11] Formally $x_i \in \max(\mathcal{B}_i, \succsim) = \{x \in \mathcal{B}_i : x \succsim y \; \forall y \in \mathcal{B}_i\}$.

It is difficult to check $\unrhd$-rationalizability directly, since it would require searching over all possible preference relations (binary relations that are complete and transitive). So we instead use an equivalent and testable criteria that searches for sequences of choices which violate transitivity by forming cycles. For example, suppose $x_i \rhd x_j$, and that x_i is chosen from $\mathcal{B}_i$ when x_j is available, but that x_j is chosen from $\mathcal{B}_j$ when x_i is available. Any rationalizing preference relation must have x_i indifferent to x_j, but this would be inconsistent with $\unrhd$.

---

[9]There are a few different ways to formalize rationality in choices; this paper presents a framework which is designed for use in generic, possibly non-Euclidean, choice spaces (based on Nishimura, Ok, and Quah (2017)).

[10]A *partial order* is a binary relation that is reflexive ($x \unrhd x$), transitive ($x \unrhd y$ and $y \unrhd z$ implies $x \unrhd z$), and has the property that if $x \unrhd y$ and $y \unrhd x$ then $y = x$. Nishimura, Ok, and Quah (2017) use the slightly weaker preorder $\rhd$ requirement, however for most applications a partial order is easier to work with.

[11]A complete preorder $\succsim$ on X *extends* $\unrhd$ if (i) $x \succsim y$ whenever $x \unrhd y$ and (ii) $x \succ y$ whenever $x \rhd y$.

In a dataset with n finite observations, this idea is extended to sequences of length $k \geq 2$. Each budget set has a decreasing closure with respect to the partial order of $\mathcal{X}$, called the *interior* $\mathcal{B}_i^{\downarrow} = \{x \in \mathcal{X} : y \trianglerighteq x \text{ for some } y \in \mathcal{B}_i\}$. The budget set's *exterior* is the budget set excluding its strict interior $\partial \mathcal{B}_i^{\uparrow} = \mathcal{B}_i \setminus \mathcal{B}_i^{\downarrow\downarrow}$. The strict decreasing closure is the *strict interior*, defined as $\mathcal{B}_i^{\downarrow\downarrow} = \{x \in \mathcal{X} : y \triangleright x \text{ for some } y \in \mathcal{B}_i\}$.

Definition. Formally, a finite dataset satisfies cyclical $\trianglerighteq$-consistency, or *consistency*, if for any sequence of $k \geq 2$ budget sets, $\mathcal{B}_{(1)}, ... \mathcal{B}_{(k)} \subseteq \boldsymbol{\mathcal{B}}_n$, whenever

$$x_{(1)} \in \mathcal{B}_{(2)}^{\downarrow}, \ldots, x_{(k-1)} \in \mathcal{B}_{(k)}^{\downarrow}, x_{(k)} \in \mathcal{B}_{(1)}^{\downarrow}, \tag{1}$$

then

$$x_{(1)} \in \partial \mathcal{B}_{(2)}^{\uparrow}, \ldots, x_{(k-1)} \in \partial \mathcal{B}_{(k)}^{\uparrow}, x_{(k)} \in \partial \mathcal{B}_{(1)}^{\uparrow}. \tag{2}$$

When choices are consistent, they do not create strict cycles with themselves and they do not contradict the objective ordering on the choice space. From Theorem 1 in Nishimura et al. (2017), the cyclical $\trianglerighteq$-consistency is equivalent to $\trianglerighteq$-rationalizability.[12] We will refer to a dataset satisfying cyclical $\trianglerighteq$-consistency as rational or rationalizable interchangeably.

It is possible for the experiment to be generated in a way that satisfies cyclical $\trianglerighteq$-consistency trivially. If budget sets never intersect then (1) holds tautologically. Similarly, if budget sets are always contained entirely in each others' interiors, then sequences of overlapping budget sets can exist, but cycles can never form. We eliminate these two scenarios through assumptions on the experiment sampling distribution.

Assumption 5. The experiment is generated such that for any $i \neq j$:
1. $\mathbb{P}_{\mathcal{B}}(\mathcal{B}_i \cap \mathcal{B}_j^{\downarrow} \neq \emptyset) > 0$, and
2. $\mathbb{P}_{\mathcal{B}}(\mathcal{B}_i \cap \mathcal{B}_j^{\downarrow} \neq \mathcal{B}_i) > 0$.

The first part of Assumption 5 requires budget sets and interiors to intersect with positive probability. The second part requires there to be enough variation in the way budget sets are generated so part of the budget set might lie outside an overlapping budget set's interior.[13]

---

[12]If the choice space is continuous then it is possible to apply Theorem 2 in Nishimura et al. (2017) and further state that this preference relation is continuous.

[13]Part 1. is a technicality, while Part 2. omits, for example, certain yearly panel datasets of household consumption where wealth increases far outstrip price variation. For a discussion of how to address the problem of demand prediction in such cases see the literature related to Blundell, Browning, and Crawford (2003).

1.3 The observed revealed preference relation

To check consistency, we construct the *observed revealed preference relation*. The choice x_i is revealed preferred to x_j whenever x_i is chosen even though $x_j \in \mathcal{B}_i^{\downarrow}$. We write this as $x_i Q_n^0 x_j$, where Q_n^0 is a binary relation over $\{x_i\}_{i=1}^n \times \{x_i\}_{i=1}^n$.[14]

Definition. A *cycle of length k* (also called a *k-cycle*) is a sequence of observations $(1), \ldots, (k)$ such that:

$$x_{(2)} Q_n^0 x_{(1)}, ..., x_{(k)} Q_n^0 x_{(k-1)}, x_{(1)} Q_n^0 x_{(k)}$$

When the relation Q_n^0 has no cycles of any length, we say it is *acyclic*.

Cycles in the observed revealed preference relation are equivalent to cycles as in (1). We would like to equate acyclicity of Q_n^0 with rationality, but Q_n^0 does not check whether (2) holds. Formally, we add an assumption to the experiment sampling procedure which eliminates the existence of cycles as in (2).[15]

Assumption 6. Furthermore, for any $i \neq j$:
   1. $\mathbb{P}_{\mathcal{B}}(\mathcal{B}_i = \mathcal{B}_j) = 0$, and
   2. $\mathbb{P}_{\mathcal{B}_j}(\partial \mathcal{B}_j^{\uparrow} \cap x_i = \emptyset | x_i, \mathcal{B}_i) = 1$.

We can think of Assumption 6 as placing a restriction on the relative sizes of the choice space $\mathcal{X}$ and the budget sets. Implicitly, we assume $\mathcal{X}$ is sufficiently large such that there exists a sampling distribution $F_{\mathcal{B}}$ satisfying the above conditions. The first part requires the support of $F_{\mathcal{B}}$ to be large enough that the chance of drawing the same budget set twice is effectively zero. In particular this rules out the case that, with sufficient observations, the researcher could observe all possible subsets of $\mathcal{X}$.[16] The second condition of Assumption 6 is more subtle and should be approached with caution in the context of discrete choice space $\mathcal{X}$.[17] It effectively says that, conditional on observing $(x_i, \mathcal{B}_i)$, there is zero probability of drawing a budget set j whose exterior intersects x_i. The second part automatically holds in Euclidean space whenever $\mathcal{B}$'s probability measure has a density, because the exterior of a budget set is always one dimension less than the maximal dimension of the budget set.

---

[14]There are different flavors of revealed preference relations that can be constructed from a finite dataset $\mathcal{D}_n$. This paper uses an observed direct revealed preference relation. More standard is the direct revealed preference relation R^0 which is defined as $x_i R^0 y$ if $y \in \mathcal{B}_i$. Whenever $y \in \mathcal{B}_i^{\downarrow\downarrow}$, we would write $x_i P^0 y$, which defines the *strict directly revealed preference relation*. Most existing revealed preference literature works with both P^0, and the transitive closure of R^0.

[15]An alternative to Assumption 6 is to pair Q_n^0 with a strict version, however this is a computationally cumbersome approach which we do not pursue in this paper.

[16]It is both a necessary and sufficient condition for $\mathcal{B}$'s probability measure to be non-atomic. A probability measure is *non-atomic* if no single element in the sample space $2^{\mathcal{X}}$ has probability greater than 0.

[17]For example, suppose choices were made from binary menus in L-dimensional Euclidean space with $\trianglerighteq = \geq$. If each menu consists of two random draws from $\{0, 1\}^L$, then the experiment does not satisfy Assumption 6. If however, each menu is two random draws $[0, 1]^L$, then the assumption does hold.

We now define a version of cyclical $\trianglerighteq$-consistency which can be checked with only the observed revealed preference relation Q_n^0.

Proposition 1. Under Assumption 6, acyclicity of Q_n^0 is equivalent to rationality of the observed choices $\mathcal{D}_n$.

Under Assumption 6, the event described in (2) occurs with probability zero. Coupled with the definition for $\trianglerighteq$-cyclicity and Theorem 1 from Nishimura, Ok, and Quah (2017) we are done.

1.4 Dispersion of the choice distribution and rationality

There are two sources of randomness in the choice data, which both contribute to stochasticity in the observed revealed preference relation. The first source is the sampling distribution of the budget sets. Then, conditional on the sample of budget sets, an additional source of randomness comes from the choice distribution. We would like to link the probability of observing a rationalizable dataset with the dispersion in the choice distribution. To do so, we first assume the true choices are rational. The following assumption replaces Assumption 1:

Assumption 7. There is some true preference relation $\succsim^*$, that extends $\trianglerighteq$, and for which $x_i^* = \max(\mathcal{B}_i, \succsim^*)$ for all observations i.

Under this assumption, any observed deviations from rationality must be driven by the introduction of error in the choices. In the case without dispersion, the researcher observes the true choices. The budget sets are sampled randomly from $\mathcal{X}$ so the revealed preference relation Q_n^0 is a random sample of $\succsim^*$.[18]

We must also require that non-zero dispersion in the choice distribution leads to some probability of observing a cycle. The following assumption strengthens Assumption 2.

Assumption 8. When the dispersion of the choice distribution is greater than zero, $\mathbb{P}_{x_i}(x_i \in \mathcal{B}_j^\downarrow \cap \mathcal{B}_i | \mathcal{B}_i, \mathcal{B}_j) > 0$ for any $i \neq j$ where $\mathcal{B}_i \cap \mathcal{B}_j^\downarrow \neq \emptyset$.

Conditional on $\mathcal{B}_i$, the choice distribution has large enough support that x_i falls inside some randomly drawn $\mathcal{B}_j$'s interior with positive probability.

We now have enough structure to present the main result of this section. Let $\mathbb{P}(\mathcal{D}_n$ is rational) be the probability of observing a rational dataset of size n. It is a joint probability of choices and budget sets.

---

[18]In the error-free case, the intersection $Q_n^0 \cap \succsim^*$ depends on n and the sampling properties of $F_{\mathcal{B}}$. For a related discussion on the transitive closures of samples of $\succsim^*$, see Chambers et al. (2018).

Proposition 2. Under Assumptions 1 to 8 the following are true:

1. If dispersion is exactly 0, then $\mathbb{P}(\mathcal{D}_n$ is rational$) = 1$
2. If dispersion is greater than 0, then $\mathbb{P}(\mathcal{D}_n$ is rational$) < 1$
3. As $n \to \infty$, if dispersion is greater than 0, $\mathbb{P}(\mathcal{D}_n$ is rational$) \to 0$.

Assumptions 1 and 2 link a zero dispersion to the true choice, which under Assumption 7 is rational. Assumptions 5 and 8 can be thought of as assumptions on falsifiability; they guarantee that a sample of budget sets can be drawn such that cycles are feasible, and that conditional on such a sample there is a probability of drawing choices that create cycles. Assumptions 3 and 4 are independence assumptions that are not crucial but are convenient for calculating joint probabilities. The full proof for Proposition 2 is in Appendix B.

Based on these results, the dispersion can be thought of as a measure for how rational is the data. It lends itself as well to this purpose from both a statistics and economic theory perspective as a "test" for rationality of the data.

There are existing indices of rationality, however they do not capture the dispersion. A discussion of these negative results is included in appendix A. I show that when errors are baked in to the choices, these measures do not appropriately separate out the error from rational choices. I prove that Afriat's (1975) efficiency index is not robust to outliers and converges to a limit that depends only on the support of the choice distribution. If two individuals have choices distributed with different dispersion but on the same support, their efficiency indices converge to the same value. I also show that Varian's (1990) index is not consistent, meaning it depends on the true choice, regardless of the dispersion. If two individuals have the same dispersion but different true choices, they will have different Varian's indices. I also show that when dispersion is positive, the Houtman-Maks (1985) index is biased and will underestimate the error.

1.5 Observed revealed preference relation as a random graph

Ideally we could say something more to link a higher probability for rationality with a lower dispersion. However, the observed revealed preference relation has certain dependencies that make such a statement difficult to prove without additional structure on the model. For the contaminated data model, which I present in section 2, it is the case that lower dispersion leads to higher probability of rationality.

We will now transition to thinking of the relation Q_n^0 as a directed random graph (or network). See Appendix D for an overview of the notation. The vertices are $\mathcal{V} = \{1, \ldots, n\}$, each representing a choice x_i in $\mathcal{D}_n$. An edge connects from k to i whenever $x_k Q_n^0 x_i$. Conditional on the budget set $\mathcal{B}_k$, the existence of this edge depends only on the location of x_i

with relation to the interior $\mathcal{B}_k^{\downarrow}$. Therefore, conditional on the collection of all n budget sets $\mathcal{B}$, any incoming edge to i depends only on the realization of x_i from its distribution. Thus, the edges connecting *to* a given vertex depend on each other, but from Assumption 4 they are independent of all other edges in the graph.

Observation 3. The observed revealed preference relation Q_n^0 is a directed random graph on n vertices. Conditional on the budget sets $\mathcal{B}_n$, for any $i, j, k \in \mathcal{V}$, the edges (k, i) and (j, i) are dependent.[19] Under Assumption 4, the edges (i, k) and (i, j) are independent.

We can store the observed revealed preference relation in terms of its adjacency matrix, **R**. Each element $r_{ki} = 1$ whenever $x_k Q_n^0 x_i$, and zero otherwise. The adjacency matrix **R** can also be written as a sequence of n columns, $[\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n]$. Observation 5 implies that the elements within each column r_{ji} and r_{ki} are dependent. However, under Assumption 4, the "observed revealed preferred to i" column $\mathbf{r}_i$ is independent of any other column $\mathbf{r}_k \in \mathbf{R}$.

2 Model

I apply the theoretical framework in Section 1 to the case when observed data is contaminated with error (Huber 1964; Horowitz and Manski 1995). In the contaminated data model, the researcher observes a randomly drawn *mistake* with probability π, or the true choice with probability $1 - \pi$. This model is suitable for implementation mistakes, or situations where the researcher thinks that error in the observed choice occurs sporadically. The model itself is agnostic on the source of the error, which could arise on the part of individuals making mistakes, or from idiosyncratic measurement errors on the part of the researcher.

This model nests any mistake distribution and any true choice. We maintain the assumption that the true choices are rational (Assumption 7), and assume only that the mistake distribution has sufficiently broad support to create cycles (in keeping with Assumption 8). The measure of dispersion π satisfies Assumption 2: when $\pi = 0$ the observation is the deterministic true choice (the classic rational model); when $\pi > 0$ the observed choice is random and selected according to an error distribution. In Section 2.1 I formally define the model and discuss its impact on the distribution of the observed revealed preference relation.

The parameter of interest is the dispersion π, which is easily interpreted as the *error*

---

[19]A small caveat to this dependency is the case where the budget sets $\mathcal{B}_i$ and $\mathcal{B}_k$ are such that the edge (k, i) either can never exist, or will always exist. In these cases the edge (k, i) is constant and is independent.

rate.[20] I present two identification results in section 2.2. The first result states that for any contaminating distribution of mistakes $\mathbf{R}$ is identified by the underlying true rational preferences and the error rate π (Proposition 4). The implication of this identification result is that when π is unknown, a single realization of $\mathbf{R}$ is not enough to recover $\mathbf{R}^*$. The second result uses the presence of 2-cycles between observations, which is captured in a graph whose adjacency matrix $\mathbf{S}$ is a transformation of $\mathbf{R}$.

Certain properties of the degrees of $\mathbf{S}$ will drive the estimation strategy in Section 3. I discuss the degree distribution of $\mathbf{S}$ in section 2.3 and some limiting properties in section 2.4.

2.1 Contaminated choices and revealed preference

We keep the assumption on the budget set sampling procedure from before (Assumptions 3, 5 and 6) and replace Assumptions 2 and 4 on the choice distribution with the following assumptions.

Assumption 9. The *contaminated choice* is

$$x_i = \begin{cases} x_i^* & \text{if } z_i = 0 \\ \mu_i & \text{if } z_i = 1. \end{cases} \tag{3}$$

The *mistake* is $\mu_i \sim_{iid} F_{\mu|\mathcal{B}_i}$ and the indicator of an error is $z_i \in \{0, 1\}$.

Assumption 10. The unobserved indicators of an error are $z_i \sim_{iid} Bernoulli(\pi)$

Assumption 11. The unknown error rate π is less than 0.5.

A contaminated choice is a random choice, drawn from a distribution with a point mass of at least $1 - \pi$ on the true choice, x^*. For clarity in exposition, a *mistake* is an element in the choice space μ_i, while an observation i is *an error* if $z_i = 1$. The set of *error observations* is $\mathcal{Z} = \{i : z_i = 1\}$, and the set of *error-free observations* is $\mathcal{Z}^c = \{k : z_k = 0\}$. A sufficient condition for conditional independence (Assumption 4) is that the contaminating mistakes are independent, $\mu_i \perp \mu_k$ for all $i \neq k$. We also maintain Assumption 7 which says that the true choice x^* is rational.

The predominant feature of the contaminated choice model is a point mass of $1 - \pi$ on the true choice. We will generally assume that $\pi \ll 0.5$, so the deterministic true choice

---

[20]Because π is a probability, it does not depend on the units of the choice space. This is in contrast with something like the standard deviation in a trembling hand model which is defined in units of the choice-space. This model therefore lends itself well to within-individual analysis, as the parameter π can be estimated from a number of different choice settings and compared across domains. It can also be used as a measure to compare rationality across individuals.

is emphasized in the model. This point mass is particularly useful since the model can have a discrete support. The results of this section can be applied to any discrete choice setting satisfying the assumptions outlined in section 1. This is a contribution over existing indices, which can be used only when choices are made from linear budget sets. Of course the contaminated choice model can also be used in continuous choice environments, which is what we demonstrate in this paper. Figure 1 presents three examples of contaminated choices. The first two panels show how different mistake distributions can contaminate the same true choice at the same error rate. The last panel shows how the model operates in a discrete choice setting, keeping the error rate fixed.

(i) Uniform　　　　　　　(ii) Truncated normal
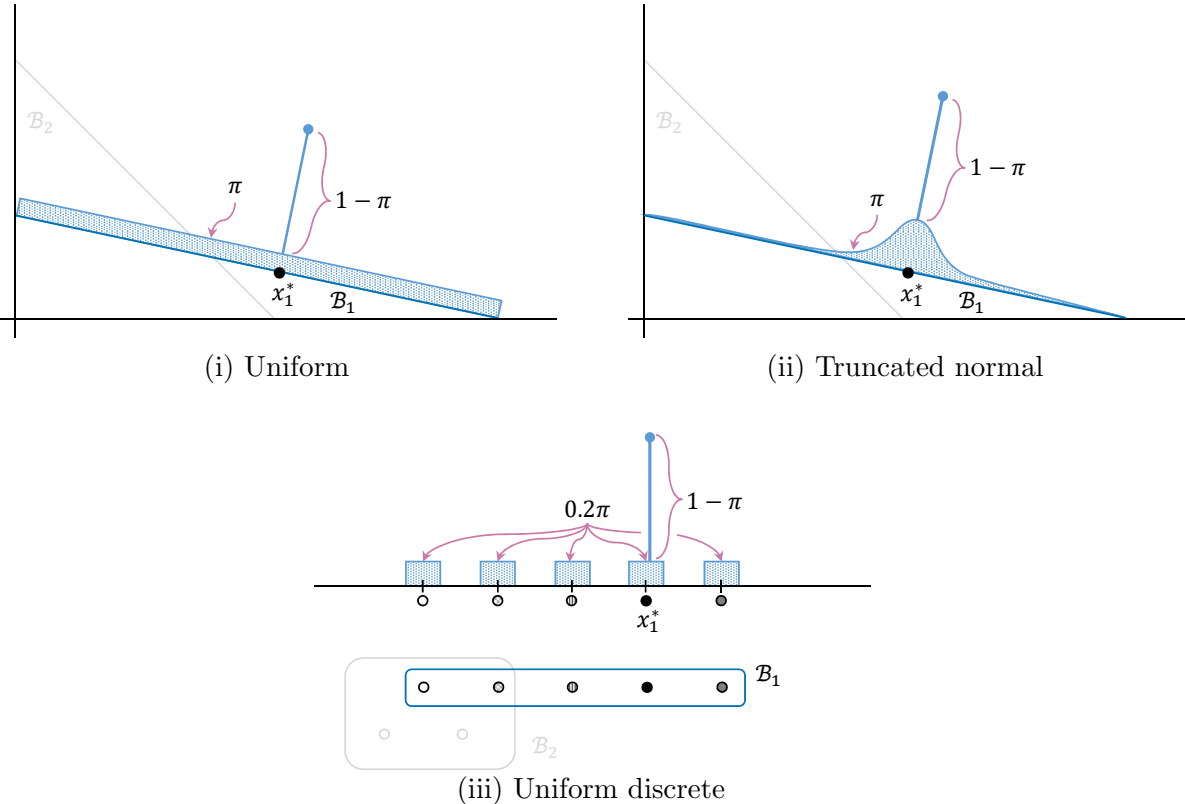
(iii) Uniform discrete

Figure 1: Examples of the contaminated choice probability density. The distinguishing feature of this distribution is a point mass of at least $1 - \pi$ on the location of the true choice. The remaining π density is distributed according to some mistake distribution whose support can be (i-ii) continuous, or (iii) discrete.

Recall from Section 1.5 that the observed revealed preference relation Q_n^0 has an adjacency matrix $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n]$ which is a sequence of n randomly contaminated $n \times 1$ columns. Each "observed revealed preferred to i" column $\mathbf{r}_i$, is either an error-free column $\mathbf{r}_i^*$ with probability $(1 - \pi)$, or an error column $\mathbf{e}_i$ with probability π.

16

$$\mathbf{r}_i = \begin{cases} \mathbf{e}_i & \text{if } z_i = 1 \\ \mathbf{r}_i^* & \text{if } z_i = 0 \end{cases} \tag{4}$$

The column $\mathbf{e}_i$ is derived directly from the mistakes μ_i, and its distribution is derived from $F_{\mu|\mathcal{B}_i}$. The probability of observing a given error column, conditional on the budget sets is an integral over the mistake distribution taken over regions of the budget set exterior. For a discussion of this discretization process, see Kitamura and Stoye (2018). Let $\mathbf{R}^*$ be a matrix of the n error-free columns and $\mathbf{E}$ be composed of the n randomly drawn error columns. It is helpful to think of $\mathbf{R}$ as a matrix composed of columns selected randomly from either $\mathbf{R}^*$ or $\mathbf{E}$ according to $\{z_i\}_{i=1}^n$.

2.2   Identification

For this subsection we condition on a collection of n budget sets, $\mathcal{B}_n$ which overlap enough that cycles are feasible. The observed revealed preference relation $\mathbf{R}$ is a semiparametric model whose distribution is governed by three objects: $\{\mathbf{R}^*, \pi, F_{\mu|\mathcal{B}}\}$. We are interested in recovering the parameter π while allowing $\mathbf{R}^*$ to be as flexible as possible. That end, we turn our attention to understanding the identification of this model, specifically under the assumptions outlined in Section 1.

The first identification result states that for any $F_{\mu|\mathcal{B}}$, $\mathbf{R}$ is identified by $\theta = \{\mathbf{R}^*, \pi\}$ (Proposition 4). The implication of this result is that when π is unknown, a single realization of $\mathbf{R}$ is not enough to recover $\mathbf{R}^*$. This result is true for any $\mathbf{R}^*$, including ones with cycles.

The second identification result uses the presence of 2-cycles between observations, which is captured in an matrix $\mathbf{S}$ and is a transformation of $\mathbf{R}$. The $\mathbf{S}$ has a structure that exposes the rationality of $\mathbf{R}^*$ (Observation 5). This structure drives the estimation strategy in Section 3.

For the first result, we use the moment of $\mathbf{R}$ because the elements within each column of $\mathbf{r}_i$ are dependent. The first moment is an $n \times n$ matrix whose elements are each in the range $[0,1]$

$$\mathbb{E}_{\theta, F_{\mu|\mathcal{B}}}[\mathbf{R}] = (1-\pi)\mathbf{R}^* + \pi\, \mathbb{E}_{\mu|\mathcal{B}}[\mathbf{E}].$$

Since $z_i \sim_{iid} Bernoulli(\pi)$ (Assumption 10), the moment of $\mathbf{R}$ is a mixture between the discrete $\mathbf{R}^*$ and an expectation over the random mistakes.[21] The identification relies on the fact that $\mathbf{R}^*$ has elements that are either 0 or 1, coupled with the following identification assumption on the mistakes μ_i. Each (i,j) element of $\mathbb{E}_{\mu|\mathcal{B}}[\mathbf{E}]$ is written $\bar{e}_{ij}$, and is the

---

[21]It is possible to generalize Assumption 10 to allow for a probability $\mathbb{P}(z_i = 1) = \pi_i$ as long as the probability $\pi_i \sim_{iid} F_\pi$. In this case the model is identified by the expected error rate $\mathbb{E}[\pi_i]$.

probability that $\mu_i \in \mathcal{B}_j^{\downarrow} \cap \mathcal{B}_i$ governed by $F_{\mu|\mathcal{B}_i}$. The assumption below eliminates the case where the distribution $F_{\mu|\mathcal{B}_i}$ is a point mass for every budget set $\mathcal{B}_i \in \mathbf{\mathcal{B}}_n$. We can think of this as an extension to the power assumption Assumption 8.

Assumption 12. Conditional on $\mathbf{\mathcal{B}}_n$, there is at least one (j, k) such that $\bar{e}_{jk} \notin \{0, 1\}$.

Proposition 4. Let $\widetilde{\theta} = \{\widetilde{\mathbf{R}}^*, \widetilde{\pi}; F_{\mu|\mathcal{B}}\}$ and $\theta = \{\mathbf{R}^*, \pi; F_{\mu|\mathcal{B}}\}$ where $F_{\mu|\mathcal{B}}$ is some fixed mistake distribution. Under Assumptions 9 to 12,
$$\mathbb{E}_\theta[\mathbf{R}] \neq \mathbb{E}_{\widetilde{\theta}}[\mathbf{R}] \text{ if and only if } \widetilde{\theta} \neq \theta.$$

A detailed proof is in Appendix B, and an illustrative example can be found in Appendix C. Importantly the proof of Proposition 4 does not use rationality of $\mathbf{R}^*$.

From Proposition 1, $\mathbf{R}^*$ is rational when it is acyclic. We know that $\mathbf{R}$ has an acyclic induced subgraph on the error-free observations $\mathcal{Z}^c$. There is also an induced subgraph on $\mathcal{Z}$ which is completely random according to $F_{\mu|\mathcal{B}_{\mathcal{Z}}}$. These insights are captured in Observation 5, which uses the graph of 2-cycles constructed from $\mathbf{R}$.[22] The adjacency matrix of this graph is $\mathbf{S}$. Each element $s_{ik} = r_{ik}r_{ki}$ indicates the existence of a cycle between i and k. By construction, $\mathbf{S}$ is a symmetric adjacency matrix and its graph is undirected. Figure 2 shows the adjacency matrices $\mathbf{R}$ and $\mathbf{S}$ for simulated data.

Observation 5. Define the count of 2-cycles between i and j as $s_{ji} = r_{ij}r_{ji}$ for any $i, j \in \{1, \ldots, n\}$. Conditional $\mathbf{\mathcal{B}}_n$ and the errors $\mathcal{Z}$, the following are true:

1. If $j, l \notin \mathcal{Z}$ then $s_{jl} = 0$.
2. If $k, i \in \mathcal{Z}$ then s_{ki} is identified by the marginal probabilities $\bar{e}_{ki}$ and $\bar{e}_{ki}$.

Observation 5 highlights two nice properties of the 2-cycle graph $\mathbf{S}$. The induced subgraph on the error-free vertices, $\mathbf{S}[\mathcal{Z}^c]$, has no edges, since the true choices are rational. Meanwhile, the induced subgraph on the errors, $\mathbf{S}[\mathcal{Z}]$ captures the way any two errors create a 2-cycle with each other. Neither subgraph depends on the underlying true preferences. As shown in the second panel of figure 2, the adjacency matrix $\mathbf{S}$ is a block matrix with one block of all 0 and another block with some random distribution.

The off-block elements of $\mathbf{S}$ are the 2-cycles that occur between an error-free choice and an error. All we know about $\mathbf{R}^*$ is that it is acyclic, which is not particularly informative. Take as fixed the edges drawn for the errors. If $\mathbf{R}^*$ has a total of K edges, then there are *at least* K cyclic graphs that are equal to $\mathbf{R}^*$ with one added edge. So the off-diagonal block, when $\mathbf{R}^*$ is acyclic could be identical, or with only a difference of one edge, when $\mathbf{R}^*$ is not

---

[22]In some settings, cycles of length K are more informative than 2-cycles. One interesting direction of future research is to define $\mathbf{S}$ for longer cycles. Each element s_{ij} counts the number of K-cycles between i and j and the first part of Observation 5 still holds.

acyclic. This is to say that we do not generally know, or have a good way of characterizing, how the off diagonal elements behave.
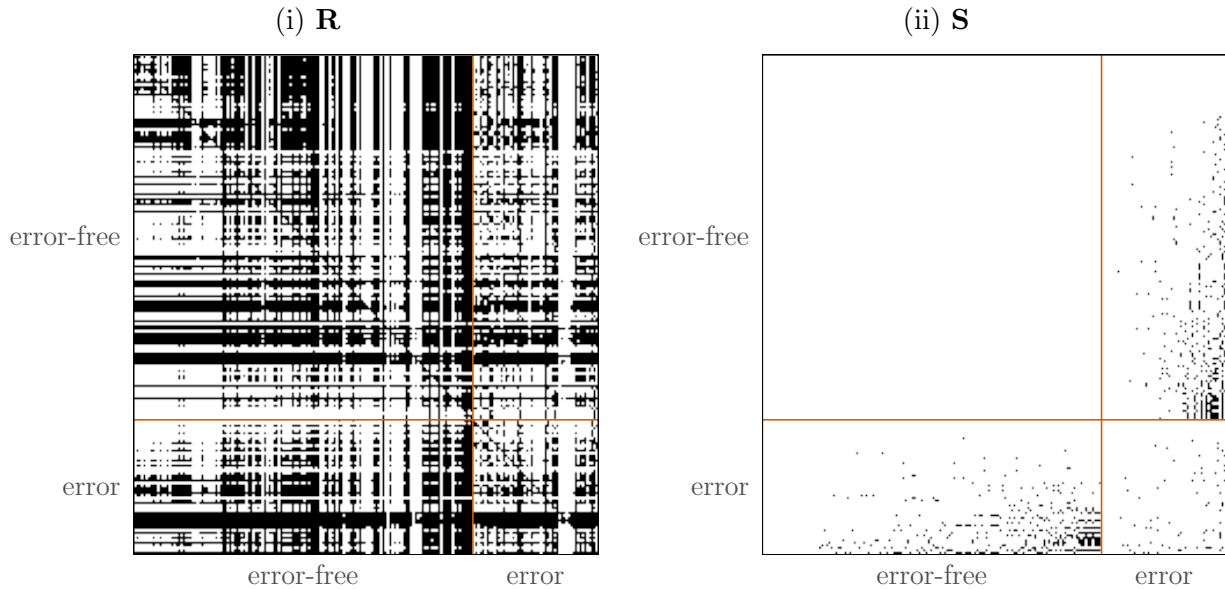

Figure 2: Simulated adjacency matrices for (i) the observed revealed preference relation and (ii) the graph of two-cycles. Here the observations have been ordered according to their latent classification $\mathcal{Z}$. The experiment consists of 200 budget lines on two continuous goods, selected randomly with intercepts between 10 and 100, with one at least 50. The error-free choices are perfect complements, and the error distribution is uniformly distributed on the exterior of budget set as in figure 1(i). The error rate $\pi = 30\%$.

2.3   The distribution of degrees on S

The structure produced by the contaminated choice model will be used in the classifier of Section 3.1 which will operate on the degree of **S**. Recall that the degree is $\bar{d}_i = \frac{1}{n} \sum_j s_{ji}$. There is a large empty induced subgraph on the error-free observations, so intuitively it might seem that the degrees for error-free observations should be smaller than those for the errors. After all, the error-free observations can only be involved in cycles with errors, while the errors can create cycles with both types of observations. In the simulated example from before, it seems the observations with high degree, those involved in many cycles, are all error observations. Figure 3 shows the histogram for degrees separated by type.

Formally defining the properties of the degree distribution of **S** with respect to the different types of observations is complicated by the unknown nature of $\mathbf{R}^*$. The two examples that follow illustrate how the degree distribution might change depending on the true preferences. Note that both examples 3 and 4 are ruled out by Assumption 5, but they provide
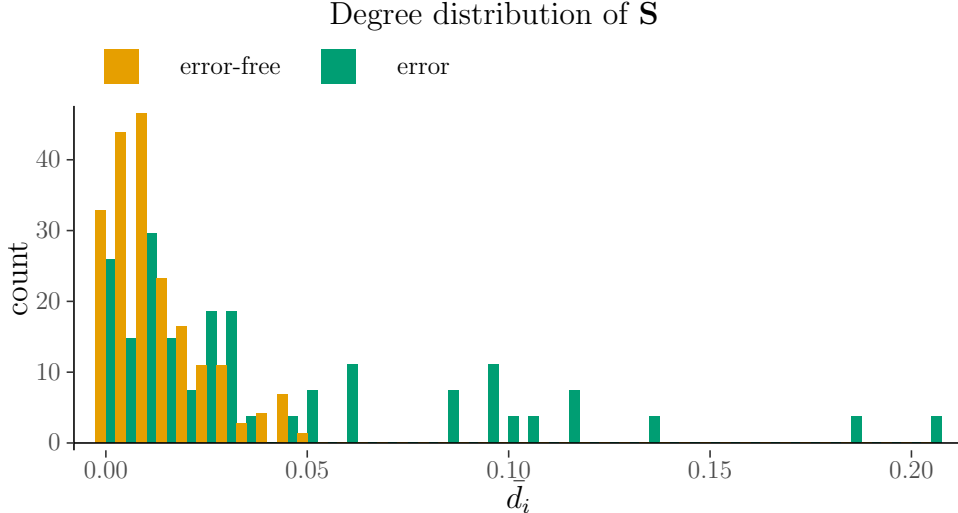
19

Figure 3: The degree distribution is notably different for error and error-free observations in **S**. The graph shows the degree distribution for the same 200 simulated observations as before, see caption of figure 2.

useful thought experiments.

Example 3. The true preference relation is devoid of information, so $\mathbf{R}^*$ is all 0. Any 2-cycles must necessarily involve two errors, and the off-diagonal blocks of **S** are all 0. The distribution of degrees has a point mass of at least $n - |\mathcal{Z}|$ on degree 0 from the error-free observations. the spread of the degree distribution is driven entirely by edges in the error induced subgraph $\mathbf{S}[\mathcal{Z}]$. Any observation with degree above zero must be to an error. ◀

Example 4. Suppose $\mathbf{R}^*$ is such that all errors are revealed preferred to all true choices, meaning $x_i^* \in \partial \mathcal{B}_j^\uparrow$ when $j \in Z$ and $i \in \mathcal{Z}^c$. Also suppose the mistakes μ_j are drawn in way that all the true choices are revealed preferred to the errors, so $\mu_j \in \partial \mathcal{B}_i^\uparrow \quad \forall j \in Z, \forall i \in \mathcal{Z}^c$. Every unique pair of error and error-free observations forms a 2-cycle and the off-diagonal blocks in **S** are all 1's. The error-free observations have degree exactly $\frac{|\mathcal{Z}|}{n}$, and the errors have degree at least $\frac{n - |\mathcal{Z}|}{n}$. When $|\mathcal{Z}| < \frac{n}{2}$ the degree distribution of **S** perfectly separates the error-free and the error distributions. ◀

One can think of other examples of random graphs that have variation in the degrees. For example the classic stochastic block model has edges drawn randomly and independently according to a fixed probability. The cross-type and same-type probabilities differ leading overlapping but non-identical degree distributions for the types. These models have the property of *degree concentration*; as the number of observations increases, the degree of each observation converges to the expected degree. Existing methods for detecting latent classifications in networks rely on degree concentration to separate out the types, however it is important to emphasize that neither **R** nor **S** have this property.

20

Another property that is commonly leveraged is independence (or weak dependence) in degrees. Unfortunately this too does not hold for $\mathbf{S}$. For any $j \neq i \neq k$, the edges $s_{ij} \not\perp s_{kj}$, due to the dependence of elements within a column in $\mathbf{R}$ (Observation 3). These two edges appear in the degrees $\bar{d}_i$ and $\bar{d}_k$ respectively, because $\mathbf{S}$ is symmetric. Thus $\bar{d}_i \not\perp \bar{d}_k$. Less critically for large enough n, both degrees $\bar{d}_i$ and $\bar{d}_k$ have the term s_{ik} in their sum. This goes to show that even if the observed choices are *iid*, the degree distribution of $\mathbf{S}$ has a complicated dependency structure.

The implication of this dependency is that the degree distributions of error and error-free observations overlap. We may be interested in how much the degree distributions overlap. Unfortunately, this depends on the true preferences $\mathbf{R}^*$ and the distribution of mistakes $F_{\mu|\mathcal{B}}$. Figure 4 has simulated data for three different underlying true preferences: (i) perfect substitutes, (ii) Cobbs-Douglas demand with $\alpha = 0.75$, and (iii) randomly generated choices that satisfy consistency. These were contaminated with the same errors, for the same 200 budgets, and with the same error $\mathcal{Z}$ classification. The panels on the left show the choices and the panels on the right show the degree distribution, this time stacked. Both the error and error-free degrees move when the true preferences are changed. This serves to highlight the dependency in $\mathbf{S}$.
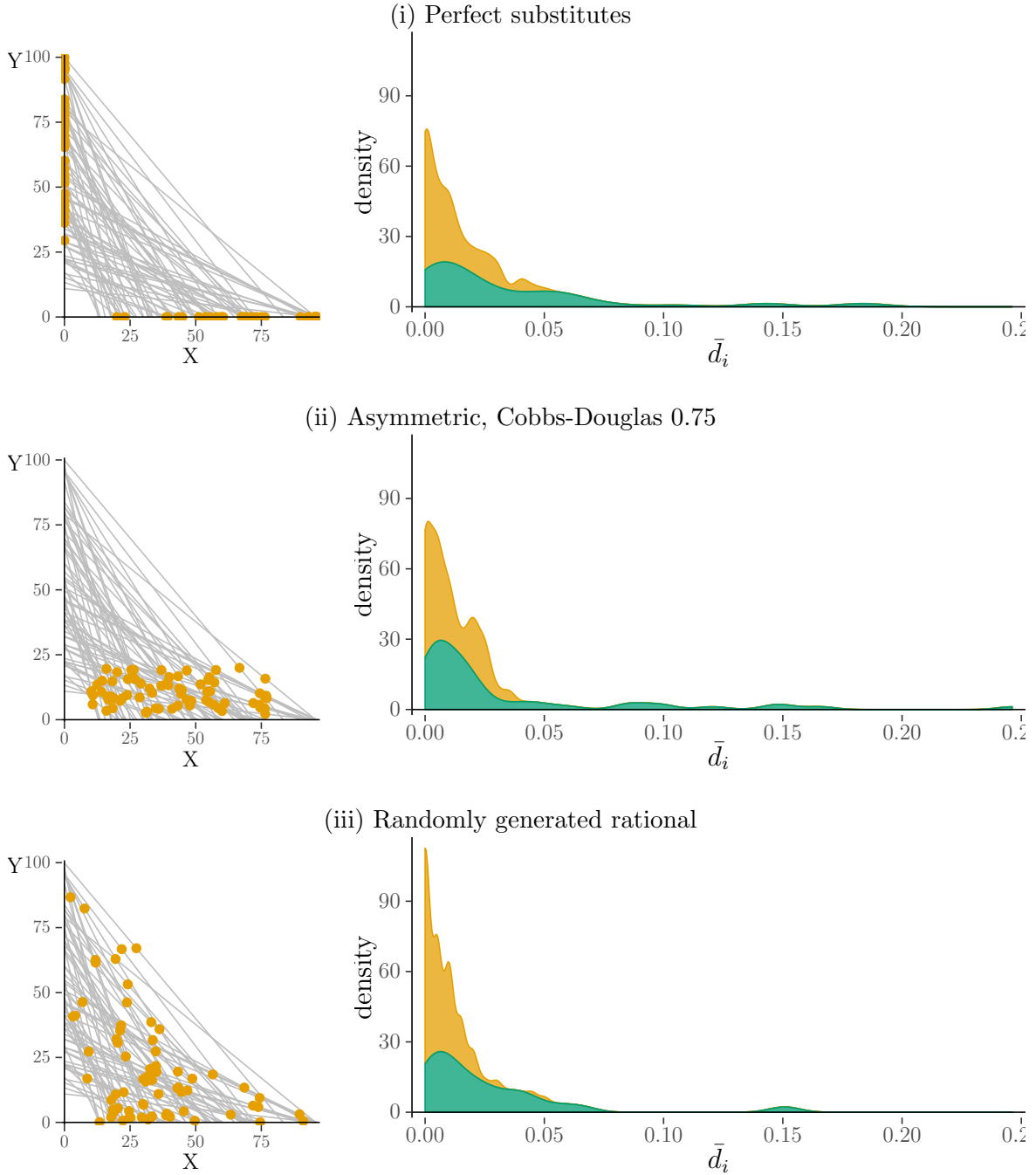
Figure 4: The degree distribution of **S** changes under different true preferences. Figure 4(i) has X and Y perfect substitutes. For figure 4(ii) choices were made using a Cobbs-Douglas utility function with $\alpha = 0.75$. A random iterative procedure generated the choices in figure 4(iii), which are in fact rational. The left-hand plots show the true choices. The right-hand plots are stacked degree density plots of error (green) and error-free (gold) degree. Across all three simulations the experiment $\mathcal{B}$, indicators $\{z_i\}$, and errors $\{\mu_i\}$ were kept the same; only the true choices $\{x_i^*\}$ were varied.

2.4 Convergence of the degrees on the error subgraph

Insights from this section motivate the estimator presented in section 3.3. Recall from Observation 5 the induced subgraph on the errors is a random graph whose distribution is determined entirely by $F_{\mu|\mathcal{B}}$. We define the degrees on the errors for any i in $\mathcal{Z}$ as

$$\bar{d}_i^{\mathcal{Z}} = \frac{1}{|\mathcal{Z}|} \sum_{j \in \mathcal{Z}} s_{ji}.$$

This error degree is derived from two things. First, the realized mistake μ_i interacts with the sampled budget sets which are associated with the other errors. The second is the other realized mistakes interacting with i's budget set $\mathcal{B}_i$. There is high variance in $\bar{d}_i^{\mathcal{Z}}$ when the error subgraph is small, which can happen for small π or n. We therefore look at the limit of $\bar{d}_i^{\mathcal{Z}}$ as $|\mathcal{Z}|$ increases.

Conditional on a realized mistake μ_i and its budget set $\mathcal{B}_i$, as we increase the number of observations the error degree $\bar{d}_i^{\mathcal{Z}}$ converges to a *limiting degree D_i*. This is the probability that a realized μ_i is involved in a 2-cycle with another randomly drawn mistake from a randomly sampled budget set. Formally this is the joint probability of sampling some budget $\mathcal{B}_t$ such that $\mu_i \in \mathcal{B}_t$, and then drawing from that budget a mistake μ_t in the interior $\mathcal{B}_i^{\downarrow}$.

Proposition 6. Conditional on $(\mu_i, \mathcal{B}_i)$ the error degree $\bar{d}_i^{\mathcal{Z}}$ converges almost surely to D_i as $|Z| \to \infty$, where

$$D_i = \int \mathbb{P}_\mu(\mu_t \in \mathcal{B}_i^{\downarrow} | \mathcal{B}_t, \mathcal{B}_i) \mathbb{I}\{\mu_i \in \mathcal{B}_t^{\downarrow}\} dF_{\mathcal{B}}(\mathcal{B}_t). \tag{5}$$

Convergence follows from the strong law of large numbers, see Appendix B for a short proof.[23]

It is often the case that solving (5) analytically for D_i is computationally intractable, so instead we can simulate it. To do so sample a large number $m \gg n$ budge sets and from each of those draw a mistake according to $F_{\mu|\mathcal{B}}$; denote the set of simulated errors as $\mathcal{M}$. For an error $i \in \mathcal{Z}$ the *simulated limiting error degree* is $D_i^{\mathcal{M}}$. This is the number of 2-cycles formed between μ_i and the $\mathcal{M}$ simulated mistakes, normalized by m. Figure 5 shows how ten different error observations have error degrees $\bar{d}_i^{\mathcal{Z}}$ that converge to their simulated limiting degree $D_i^{\mathcal{M}}$ as the size of the induced subgraph $|\mathcal{Z}|$ increases. The same arguments in Proposition 6 apply to $D_i^{\mathcal{M}}$ which also converges almost surely to D_i conditional on $(\mu_i, \mathcal{B}i)$.

---

[23]Although it is not relevant to this paper, it is possible to use the central limit theorem to determine the rate of convergence.
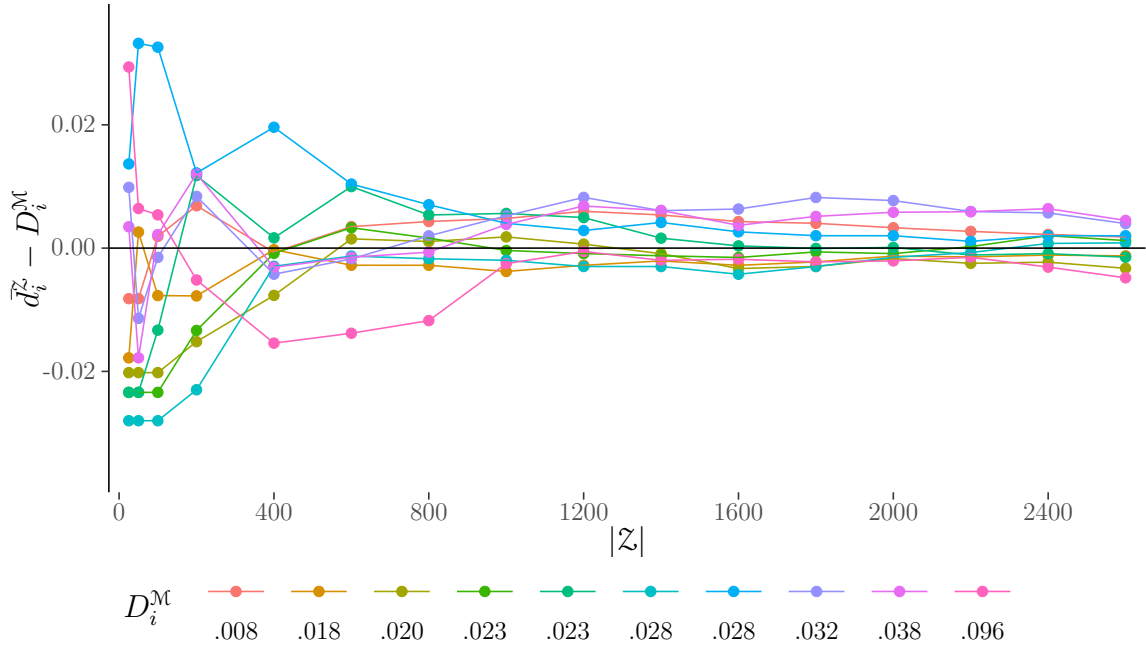
Figure 5: A sample of 10 uniformly drawn mistakes is observed. The size of the error subgraph $|\mathcal{Z}|$ is increased by uniformly drawing more mistakes and appending them to the dataset. The degree of the error subgraph $\bar{d}_i^{\mathcal{Z}}$, normalized by $|\mathcal{Z}|$, converges to i's limiting degree D_i conditional on $(\mu_i, \mathcal{B}i)$. We simulate the limiting degree for each observation by drawing $m = 5000$ mistakes and budget sets and counting the percent of 2-cycles formed between i's observed values $(\mu_i, \mathcal{B}i)$ and the $\mathcal{M}$ simulated mistakes.

Consider now the set of D_i for all errors $i \in \mathcal{Z}$. Sampling variation in the budget sets $\mathcal{B}_i$ and μ_i creates variation in $\{D_i\}_{i \in \mathcal{Z}}$. In fact, $\{D_i\}_{i \in \mathcal{Z}}$ is a sample from a distribution the limiting degrees which does not concentrate on its mean $\mathbb{E}[D_i]$. We can simulate this distribution using the simulated mistakes from before. If we construct the observed revealed preference relation on $\mathcal{M}$ then we can derive $\mathbf{M}$, the graph of 2-cycles for the m simulated mistakes. The degree on this graph is $D_j^{\mathcal{M}}$ for any simulated mistake $j \in \mathcal{M}$. The distribution of the limiting degrees is shown in Figure 6, where $m = 5000$.

The implication of these results is that the sample average of 2-cycles in the error-free subgraph will often be far from the mean of the limiting degree distribution. Furthermore, the sample degree $\bar{d}_i^{\mathcal{Z}}$ is also far from its own limit $D_i^{\mathcal{M}}$.
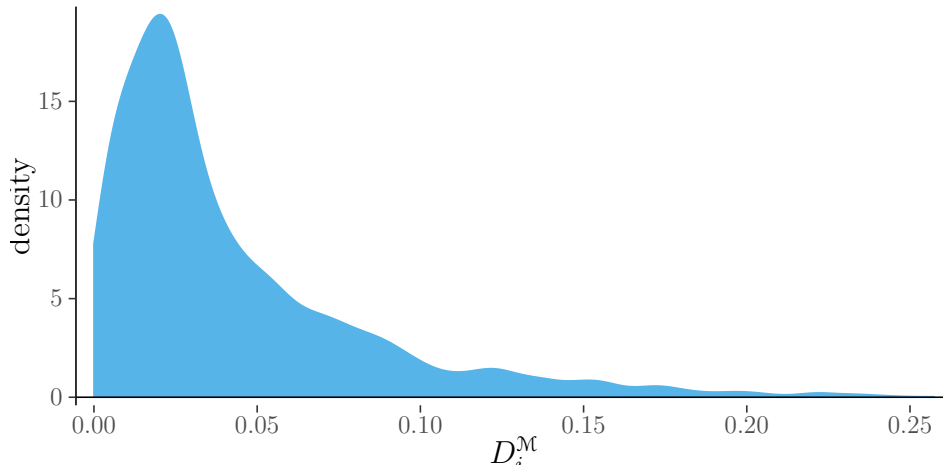
24

Figure 6: The distribution for the limiting degree D_i is simulated by drawing $m = 5000$ mistakes and budget sets. Using the observed revealed preference relation created for these $\mathcal{M}$ simulated mistakes, the degree of the graph on 2-cycles is $D_j^{\mathcal{M}}$ for each $j \in \mathcal{M}$. The variation in the distribution arises from variation in the realized values of $(\mu_j, \mathcal{B}j)$. For the set of errors in the observed data $\{D_i^{\mathcal{M}}\}_{i \in z}$ is an sample drawn from this distribution.

3   Estimation

The estimation problem we face is two-fold: classification and point-estimation. We would like to classify the observations as either errors or error-free, and to estimate the error rate π. Unfortunately, because of the overlap in degrees, it is not possible to fully differentiate all the errors.

I propose an estimation strategy that first separates out the high-degree observations, identifying them as errors. This is done using an iterative algorithm. If the error distribution is unknown to the researcher here is where we must stop. The output is a partial recovery of the observations made with error. We can also use the number of errors to estimate a lower bound on π.

If the researcher knows the distribution from which the errors are drawn, then we can simulate the limiting distribution for the degrees of the error subgraph. We can also simulate the limiting degrees for the specific observations that were classified as errors. Comparing the upper tail of the limiting distribution with the observed sample produces an estimate for the error rate. This procedure is unable to label the remaining errors because they are indistinguishable from the error-free observations.

3.1 Classification

To recover a subset of the error observations I propose algorithm A, based on a k-core peeling algorithm, in which the largest degree is iteratively removed. After each removal, the degrees for the remaining subgraph are recalculated, and the maximal degree is removed again. This continues until the maximal degree is zero, meaning we are left with an induced subgraph without 2-cycles. The observations that are removed are labeled as "errors". There are however, a number of low degree errors "disguised" as error-free which remain in the graph. As a result we have only recovered a subset of error observations.

Algorithm A: iterative removal of observation with largest degree

    input : an adjacency matrix $\mathbf{S}$ indicating 2-cycles
    output: a set $\widehat{\mathcal{Z}}_A$ of indices categorized as errors

1 Initialize $\widehat{\mathcal{Z}}^c = \{1, \ldots, n\}$
2 **for** $t = 0 \ldots n$ **do**
3   Find $\mathbf{S}^{(t)}$ the induced subgraph on $\widehat{\mathcal{Z}}^c$
4   Calculate the degrees for all $i \in \widehat{\mathcal{Z}}^c$ using $\mathbf{S}^{(t)}$, normalized by $n - t$
5   Find k, the index of the maximal degree
6   **if** *degree at k is 0* **then**
7   **return** $\widehat{\mathcal{Z}} \equiv \{1, \ldots, n\} \setminus \widehat{\mathcal{Z}}^c$
8   **else**
9   Remove $\{k\}$ from the set $\widehat{\mathcal{Z}}^c$

To understand why algorithm A works, it is helpful to consider a simplified model that mimics the characteristic of the random revealed preference relation. A formal analysis of the performance of algorithm A in this simplified model is part of a forthcoming project.

Example 5. As before there are n vertices classified as either $\mathcal{Z}$ or $\mathcal{Z}^c$, where the number of vertices in $\mathcal{Z}$ is much less than half of n. For each vertex in $\mathcal{Z}$ draw a degree $\bar{d}_i$ from a degree distribution F_d. Suppose that F_d allows for low degrees; figure 6 is an example. Match the $n\bar{d}_i$ edges for each vertex i in $\mathcal{Z}$ randomly to different vertices in the graph. The vertices in $\mathcal{Z}^c$ do not have any edges drawn for them directly, rather they are only connected through edges matched to them from $\mathcal{Z}$. So by construction there is an empty induced subgraph on $\mathcal{Z}^c$. Edges between two vertices in $\mathcal{Z}$ are matched iteratively to ensure the degree $\bar{d}_i$ for each vertex is retained. The result of this process is a graph with dependent edges, and overlapping degree distributions for vertices in $\mathcal{Z}^c$ and $\mathcal{Z}$. The limiting degree distribution on the subgraph $\mathcal{Z}$ is, by construction, F_d which does not converge to its mean. The model

for random revealed preferences shares these features although it has more dependencies and is less analytically clean.[24]

To see how algorithm A works in the simplified model, first observe that the maximal degree of the graph will alway be a vertex in $\mathcal{Z}$. The degree of a vertex in $\mathcal{Z}^c$ is a sum over $|\mathcal{Z}|$ elements, while the degree in $\mathcal{Z}$ is a sum over $|\mathcal{Z}^c|$ *and* $|\mathcal{Z}|$ elements. Because the edges are uniformly matched, it is very unlikely that a vertex in $\mathcal{Z}^c$ has a high degree which can be confused for a vertex in $\mathcal{Z}$. Put in terms of the adjacency matrix: the degrees of $\mathcal{Z}$ are uniformly distributed along each *column* so the degrees of $\mathcal{Z}^c$ are uniformly distributed along each *row*. Since degrees are column sums, for there to be a high-degree vertex in $\mathcal{Z}^c$, the edges from a number of vertices in $\mathcal{Z}$ would need to bunch together.

In the first step of algorithm A, removing the maximal-degree vertex removes all its edges. Since this includes edges connected to vertices in $\mathcal{Z}^c$, those vertices will also have a lower degree. In general, many high-degree vertices in the graph will belong to $\mathcal{Z}$, although the exact proportion depends on the degree distribution F_d. As we continue to remove high-degree vertices, we shrink the number of vertices from $\mathcal{Z}$ in the remaining subgraph, and thereby lower its overall maximum degree. Eventually, vertices in $\mathcal{Z}$ with low degree, and whose edges are mostly connected to other vertices in $\mathcal{Z}^c$ will not be distinguishable from vertices in $\mathcal{Z}^c$. This will result in a small number of removed vertices that are from $\mathcal{Z}^c$. It will also result in a number of vertices from $\mathcal{Z}$ remaining in the subgraph the end of algorithm A. An analytical proof of the above discussion is not possible for our original model, since $\mathbf{R}^*$ is unknown. ◄

To show how this operates in practice, we run algorithm A on the simulated data from before. The initial degree distribution can be seen in figure 3, while figure 7 shows the degree distribution at subsequent iterations of the algorithm. For this dataset, the algorithm ends after 39 observations have been removed. Each iteration results in a lower maximal degree the remaining graph. Importantly, the difference between the error observations' maximal degree and the error-free observations' maximal degree also shrinks, which is what leads to a small amount of misclassification.

---

[24]In the literature on random graphs, this is type of graph is called a *configuration model* CITE.
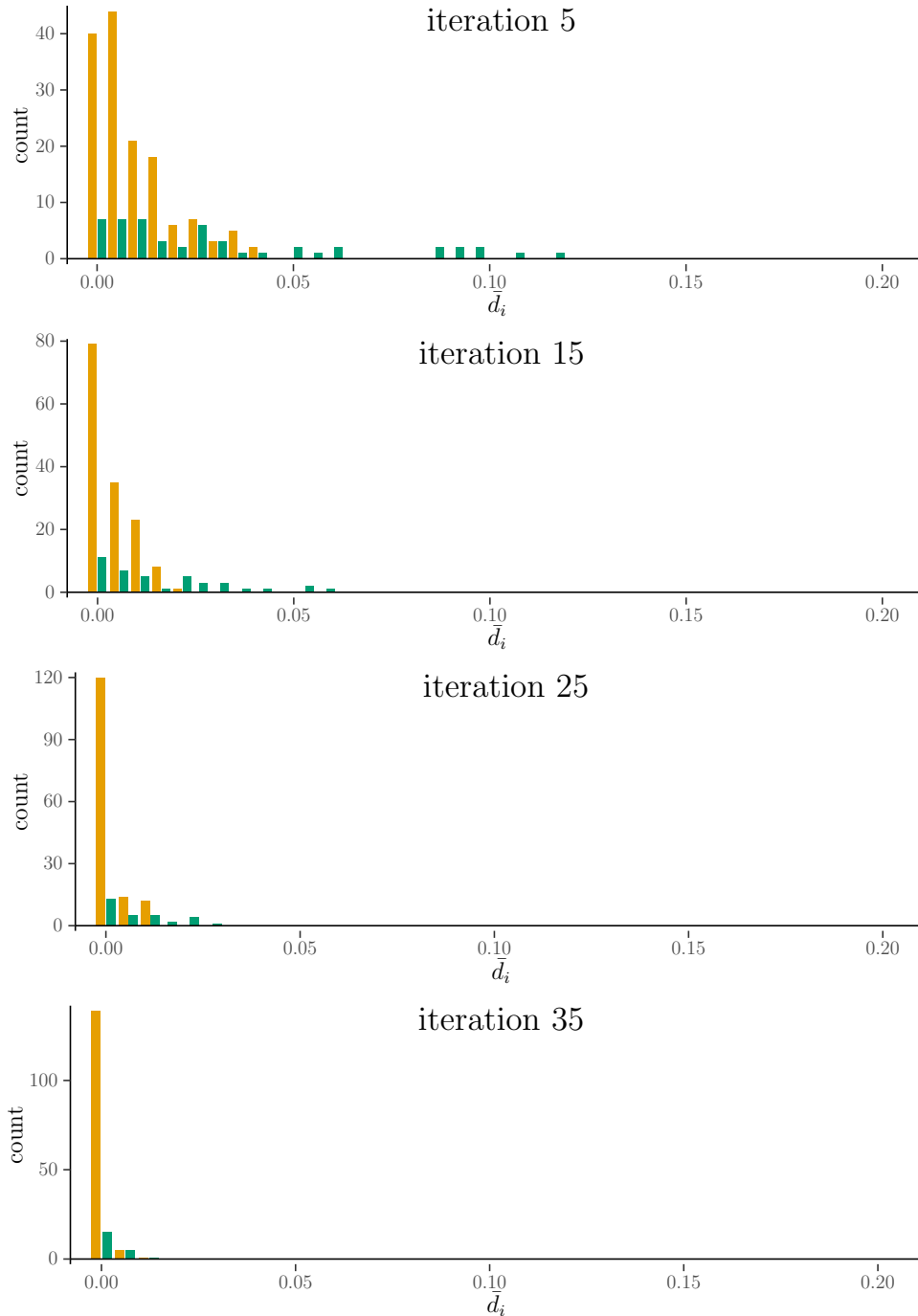
Figure 7: The degree distribution at different iterations of algorithm A. A total of 37 observations are removed until all remaining degrees are 0. Note the changing y-axis which indicates more observations with lower degrees as the algorithm progressively removes high-degree observations.

For this particular example, 39 observations are removed by the algorithm and classified as errors. Of these, 36 are *true positives* (correctly classified errors), and the remaining 3 are *false positives* since they are in actually error-free. The remaining 161 observations were not removed by the algorithm, but 18 of them are *false negatives* since they are errors.

$$precision = \frac{\text{\# true positives}}{\text{\# true positives} + \text{\# false positives}}$$

$$recall = \frac{\text{\# true positives}}{\text{\# true positives} + \text{\# false negatives}}$$

There are two statistics we use to assess the performance of the algorithm: precision and recall, defined above. The *precision* measures how many of the classified errors are actually errors. So conditional on being in the set $\widehat{\mathcal{Z}}_A$, how likely is an observation to be an error with $z = 1$. The *recall* tells us how many of the errors we were able to recover. Recall tells us how likely an error is to be in $\widehat{\mathcal{Z}}_A$. The higher both numbers are, the better the algorithm's performance.

Table 1 measures the precision and recall of the classifier under a variety of parametric specifications. Both the mean and standard deviation are reported for 100 replications of the simulations. In general the precision is higher and the recall is lower. Differences in the true preferences do not affect the algorithm's performance. Randomly generated true preferences are somewhat of an exception. These are choices selected from a uniform distribution on the budget set in an iterative way so they satisfy GARP. It is therefore not surprising that the algorithm is less able to identify uniform mistakes from this type of true choice, since they are quite similar.

When we change the contaminating error distribution we see a change in the recall but not the precision. When the mistake is likely to be indistinguishable from the true choice, the recall drops. This can happen through a restriction on the distribution's support, or when it has a lower dispersion around the true choice. In the column FOSD, mistakes are drawn uniformly from the region of the exterior below the 45-degree line (on the cheaper asset). Since true choices rarely fall in this range, the mistakes are easily picked up and the recall is better than when mistakes are drawn uniformly from the full budget set exterior. When mistakes are drawn normally around the true choice with standard deviation of 15 (or 15% of the budget set exterior), recall is even lower. The mistakes are so similar to the true choices that they are indistinguishable.

Performance of Algorithm A Classification in Simulations

| | *True Preferences* | | | |
	substitute	complement	asymmetric	random
precision	90.55	90.48	90.99	83.96
	(5.84)	(5.33)	(4.6)	(6.92)
recall	62.21	62.69	56	50.1
	(6.21)	(6.66)	(6.02)	(5.89)

| | *Error Distribution* | | | |
	Uniform	FOSD	$N(\sigma^2 = 15)$	$N(\sigma_i^2 = 15\%)$
precision	90.48	95.77	90.04	87.19
	(5.33)	(3.03)	(5.14)	(5.71)
recall	62.69	88.31	59.1	55.55
	(6.66)	(4.31)	(6.59)	(6.68)

| | *Sample Size n* | | | | | |
	25	50	100	200	400	800
precision	74.27	81.64	86.36	90.48	92.29	94.48
	(29.76)	(15.7)	(9.05)	(5.33)	(3.18)	(1.94)
recall	27.6	41.31	52.87	62.69	72.6	79.6
	(18.01)	(13.56)	(9.39)	(6.66)	(4.29)	(2.4)

| | *Error Rate π* | | | | |
	$\pi = 0.05$	$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$	$\pi = 0.4$
precision	90.87	91.91	90.56	90.48	86.83
	(12.28)	(7.42)	(6.61)	(5.33)	(5.77)
recall	66.54	66.06	66.03	62.69	59.79
	(16.45)	(10.83)	(7.87)	(6.66)	(5.36)

Table 1: The mean precision and mean recall from 100 replications are reported in units of percent. The standard deviation across replications is below in parentheses. Unless otherwise stated, in each replication the sample size is $n = 200$, true preferences are perfect complements, the error rate π is 0.3, and the contaminating error is uniformly distributed on the exterior of the budget set. Budget lines are generated randomly as in Choi et al. (2007a)

The discrete nature of the classification means the number of errors $|\mathcal{Z}|$ can cause small-sample variability in performance. The number of errors is small when $n = 200$ and the error rate π is small, or when $\pi = 0.3$ and the sample size n is small. Increasing the overall sample size is the best way to improve both precision and recall. Keeping the sample size fixed, the higher the error rate, the more errors there are with lower degree, so there are more opportunities for the algorithm to mis-classify an error and an error-free observation, resulting in lower precision and recall. However higher error rate π also yields, in expectation, a larger sized subgraph $\mathcal{Z}$ which is why the variability drops.

3.2 Set estimation with unknown distribution of mistakes

The number of errors $|\mathcal{Z}|$ is a sufficient statistic to estimate π. Recall that the z_i are Bernoulli random variables, and the sum of n Bernoulli random variables is a $Binomial(n, \pi)$ random variable. So the sample statistic $|\mathcal{Z}|/n$ converges to π as the sample size n increases. If algorithm A were a perfect classifier, then $\widehat{\pi}_A \equiv |\widehat{\mathcal{Z}}_A|/n$ could estimate π. Even with mis-classification, if the true positives canceled out the true negatives, $\widehat{\pi}_A$ could still estimate π consistently. Unfortunately, the simulation results in table 1 show that recall is systematically lower than precision, even as the sample size increases. This means there are more false negatives than false positives, especially at the small sample sizes we often observe. So the researcher cannot with confidence use $\widehat{\pi}_A$ as an estimator for π.

This systematic difference suggests that $|\widehat{\mathcal{Z}}_A|$ can estimate a lower bound for $|\mathcal{Z}|$. In every simulation run so far $|\mathcal{Z}| \geq |\widehat{\mathcal{Z}}_A|$ holds. An analytic proof of this result is beyond the scope of this paper. Instead we can gain some insight as to why this might be true with a small divergence. Consider the smallest set of observations $\mathcal{A}$ which, when removed, results in an empty subgraph. The size of $\mathcal{A}$ will always be a lower bound for $|\mathcal{Z}|$ since $\mathbf{S}[\mathcal{Z}^c]$ is empty. So either $\mathcal{Z}$ is the minimizing set, or there is a smaller set of observations which result in an empty subgraph when removed. If some errors are not involved in any cycles, then the minimizing set will be smaller than $|\mathcal{Z}|$. The problem of finding the minimal set $\mathcal{A}$ is known as the minimal vertex cover problem, and is famously NP-hard. To get around this computational complexity, various approximation algorithms have been proposed over the years, including Algorithm 2 in Gross and Kaiser (1996) which is equivalent to algorithm A with an added step removing non-unique maxima. It is therefore not entirely surprising that

algorithm A yields a lower bound for $|\mathcal{Z}|$ since $|\widehat{\mathcal{Z}}_A|$ is a reasonable approximation of $\mathcal{A}$.[25]

As a lower bound, $\widehat{\pi}_A$ gets closer to π as the sample size increases. Table 1 shows the lower bound estimate $\widehat{\pi}_A$ and the bias from $|\mathcal{Z}|/n$ in simulations. When the error rate increases, the lower bound also increases although the bias, as a fraction of π increases slightly.

<div align="center">Performance of $\widehat{\pi}_A$ as a Lower Bound for π in Simulations</div>

	Sample Size n					
	25	50	100	200	400	800
lower bound	0.112	0.145	0.185	0.207	0.236	0.254
	(0.06)	(0.05)	(0.04)	(0.03)	(0.02)	(0.01)
bias	0.193	0.142	0.117	0.092	0.064	0.047
	(0.08)	(0.05)	(0.03)	(0.02)	(0.01)	(0.01)

	Error Rate π				
	$\pi = 0.05$	$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$	$\pi = 0.4$
lower bound	0.036	0.072	0.145	0.207	0.275
	(0.01)	(0.02)	(0.03)	(0.03)	(0.03)
bias	0.013	0.028	0.054	0.092	0.124
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)

Table 2: The mean lower bound $\widehat{\pi}_A$ and mean bias $|\mathcal{Z}|/n - \widehat{\pi}_A$ from 100 replications. The standard deviation across replications is below in parentheses. Unless otherwise stated, in each replication the sample size is $n = 200$, true preferences are perfect complements, the error rate π is 0.3, and the contaminating error is uniformly distributed on the exterior of the budget set. Budget lines are generated randomly as in Choi et al. (2007a)

3.3   Point estimation with a known distribution of mistakes

The problem of correcting for the bias in $\widehat{\pi}_A$ is a matter of recovering how many errors are missing from the initial classification $\widehat{\mathcal{Z}}_A$. If the distribution of mistakes $F_{\mu|\mathcal{B}_i}$ is known, then for any observed choice and budget set $(x_i, \mathcal{B}_i)$ we can derive the expected frequency of 2-cycles that i is involved in with other errors, conditional on the i being an error. This value is defined in Section 2.4, and is i's limiting degree on the error-subgraph. We calculate it using a simulated graph $\mathbf{M}$ of 2-cycles between m errors. For any error $j \in \mathcal{Z}$, the simulated

---

[25]It is possible to construct a counter-example graph in which $|\widehat{\mathcal{Z}}_A| > |\mathcal{Z}|$, meaning algorithm A does not result in a lower bound. Consider for example figure 8 in Gross and Kaiser (1996) with $\mathcal{Z} = \{a, b, c, d\}$. It is my conjecture that the degree distributions observed in practice make such an instance unlikely, based on the results of the simulations in this paper (and Comment 1 in GK(1996)). The concerned researcher can estimate a more conservative lower bound using a different approximating algorithm for $|\mathcal{A}|$, such the one provided by Halevy et al. (2018).

limiting degree $D_j^{\mathcal{M}}$ is an independent random draw from a distribution, which captures the joint variation in $F_{\mathcal{B}}$ and $F_{\mu|\mathcal{B}_j}$. Figure 8 shows the simulated limiting degree density, and a histogram-density of the *iid* sample of limiting degrees $\{D_i^{\mathcal{M}}\}_{i\in\mathcal{Z}}$ for the errors in $\mathcal{Z}$.
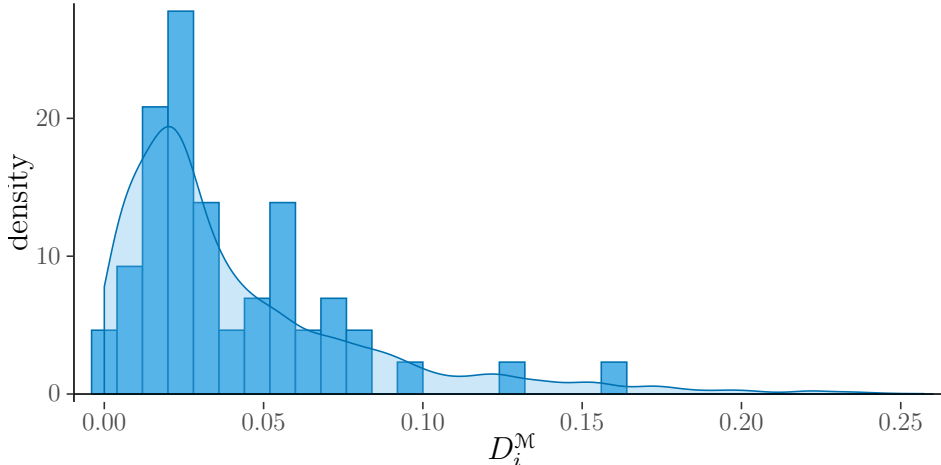

Figure 8: Histogram of the simulated limiting degrees for the sample of 54 errors in $\mathcal{Z}$ scaled as a density. Overlaid on top is the simulated density of $D_i^{\mathcal{M}}$.

We do not observe $\mathcal{Z}$, but we do have a candidate classification $\widehat{\mathcal{Z}}_A$ from Algorithm A. The estimation strategy compares the sample of simulated limiting degrees $\{D_i^{\mathcal{M}}\}_{i\in\widehat{\mathcal{Z}}_A}$ against the distribution with the goal of estimating *how many are missing*. This problem is related to a missing data problem, however we do not want to estimate the degrees of the missing observations only their count. To that end we use what we know about distribution and what we know about the missing observations.

By its design Algorithm A has trouble finding errors with low limiting degrees $D_i^{\mathcal{M}}$. Figure 9 shows the order in which Algorithm A classifies each observation as an error. The vertical ticks along the top represent errors that were not included in $\widehat{\mathcal{Z}}_A$ (false negatives); these tend to have lower limiting degree values. We can therefore think of the set $\{D_i^{\mathcal{M}}\}_{i\in\widehat{\mathcal{Z}}_A}$ as if it is a censored sample of $\{D_j^{\mathcal{M}}\}_{j\in\mathcal{Z}}$. The presence of error-free observations in $\widehat{\mathcal{Z}}_A$ (false-positives) cause the estimate of π to be biased down.

We focus our attention on right tails of the sample and distribution. For any threshold θ define the number of observations in $\widehat{\mathcal{Z}}_A$ that have $D_i^{\mathcal{M}}$ above the threshold as

$$h_{\widehat{\mathcal{Z}}_A}(\theta) = \sum_{i\in\widehat{\mathcal{Z}}_A} \mathbb{I}\{D_i^{\mathcal{M}} \geq \theta\}. \tag{6}$$

This *count* is the shaded region to the right of θ in the top panel of Figure 10. In the bottom panel the area above θ shows the expected *proportion* of observations in a random
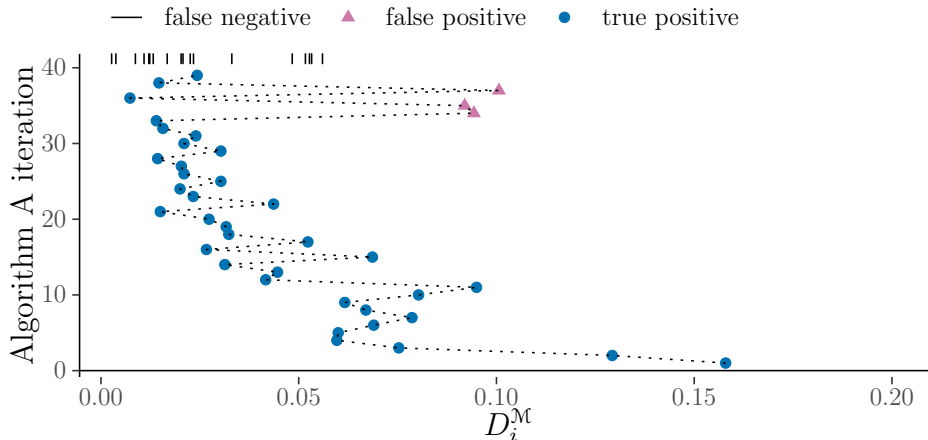
Figure 9: Comparing the limiting degree $D_i^{\mathcal{M}}$ to the iteration of Algorithm A at which i is classified as an error belonging to $\widehat{\mathcal{Z}}_A$. Later iterations remove low degrees and begin to introduce false positives (error-free observations misclassified as errors). False negatives (errors that are missing from $\widehat{\mathcal{Z}}_A$) generally have low degree. The set $\{D_i^{\mathcal{M}}\}_{i \in \widehat{\mathcal{Z}}_A}$ is a randomly censored sample of $\{D_j^{\mathcal{M}}\}_{j \in \mathcal{Z}}$.

sample that have $D_i^{\mathcal{M}}$ above θ. This probability is the complementary cumulative distribution function (CCDF), which we denote as $G(\theta)$.[26]

For a sufficiently high threshold, there is no censoring and the observed count in $\widehat{\mathcal{Z}}_A$ is exactly equal to the count in the uncensored sample. In this range of limiting degrees, normalizing by the true size $|\mathcal{Z}|$ of the sample returns the empirical CCDF. For lower thresholds, the censoring kicks in, and the count we observe is smaller than what would have been observed without censoring. For a sufficiently low threshold (often 0) dividing by the true size yields the $|\widehat{\mathcal{Z}}_A|/|\mathcal{Z}|$.

The function $h_{\widehat{\mathcal{Z}}_A}(\theta)/z$ converges to a censored version of the limiting CCDF. Different types of censoring lead to different shapes of $\bar{G}$. We will assume there is some fixed censoring threshold, below which Algorithm A censors. The CCDF is then $G(\theta)$ until that threshold, after which it reaches its maximum value of $|\widehat{\mathcal{Z}}_A|/|\mathcal{Z}|$. From Figure 10 we know that the type of censoring we observe is not fixed, but rather random. This strong assumption introduces bias in our estimates, but given that we do not know the exact nature of the censoring, it yields reasonable estimation results.

We replace $|\mathcal{Z}|$ with a candidate sample size k. Our estimation will find the $k \in \{|\widehat{\mathcal{Z}}_A| \ldots {}^n/_2\}$ that minimizes the loss between the empirical censored CCDF, $h_{\widehat{\mathcal{Z}}_A}(\theta)/k$, and the expected censored CCDF $\bar{G}(\theta; k) = \min\{G(\theta), |\widehat{\mathcal{Z}}_A|/k\}$. Figure 11 shows the empirical

---

[26]The CCDF is also referred to as the tail distribution or exeedance. As its name suggests, it is the complement of the more common cumulative distribution function (CDF).
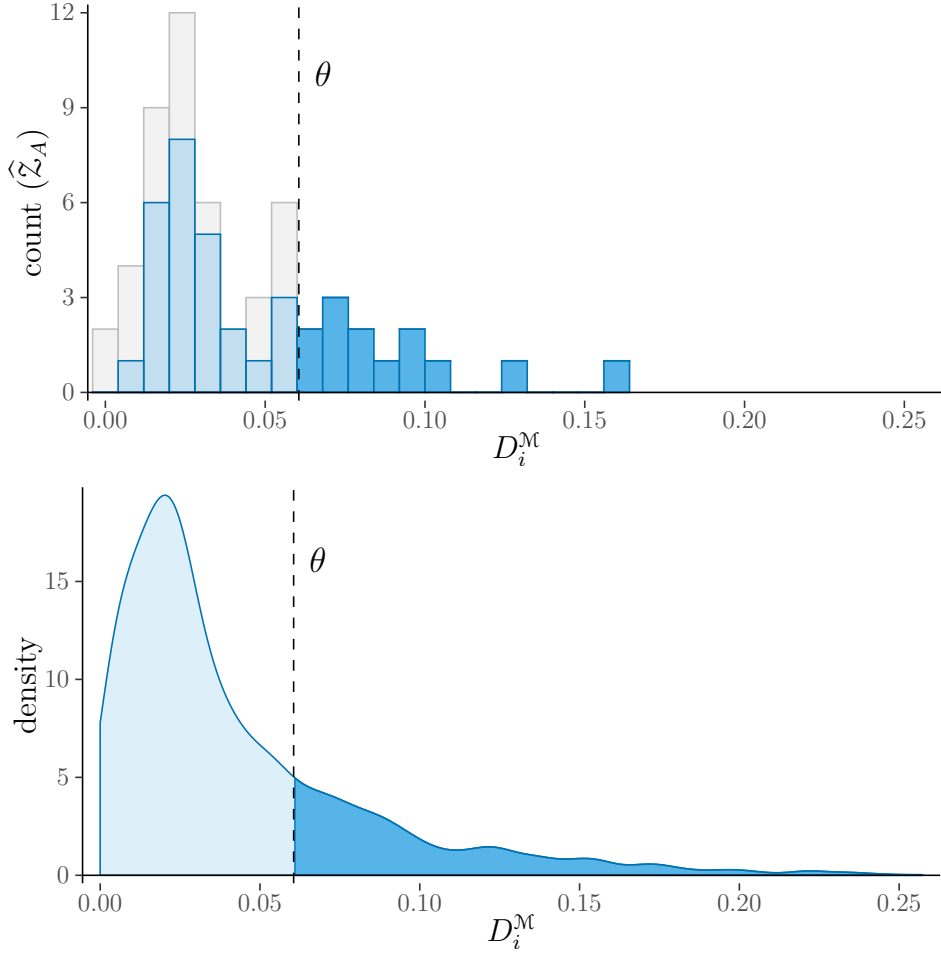
Figure 10: Comparison between the tail above a threshold θ in the observed censored sample (top) and limiting density (bottom). The shaded density above θ is $G(\theta)$, and the shaded count above θ is $h_{\widehat{\mathcal{Z}}_A}(\theta)$. The gray bars in the bottom panel show the uncensored sample $\mathcal{Z}$. When θ is low the censoring kicks in and the censored sample count is less than $h_{\mathcal{Z}}(\theta)$.

and expected censored CCDFs at different values of k. Each point corresponds to a different threshold θ; higher thresholds are on the right and result in lower values of both the CCDFs. At the leftmost point, $\theta = 0$, so $\bar{G}(0)$ is the normalized sample count $|\widehat{\mathcal{Z}}_A|/k$, which shrinks as k increases from $|\widehat{\mathcal{Z}}_A|$. The top panel shows the minimal possible value of $k = |\widehat{\mathcal{Z}}_A|$, and the bottom panel is the largest possible value $k = 0.5n$. In between is the graph for $k = |\mathcal{Z}|$ the true value, and $\widehat{k}_B$, the estimated value that minimizes the loss.

We use the least absolute deviation loss function. For a sequence of thresholds Θ, the LAD finds k that minimizes the sum of the absolute value of deviations for each $\theta \in \Theta$. Each deviation is defined as the difference between the empirical censored CCDF at θ and
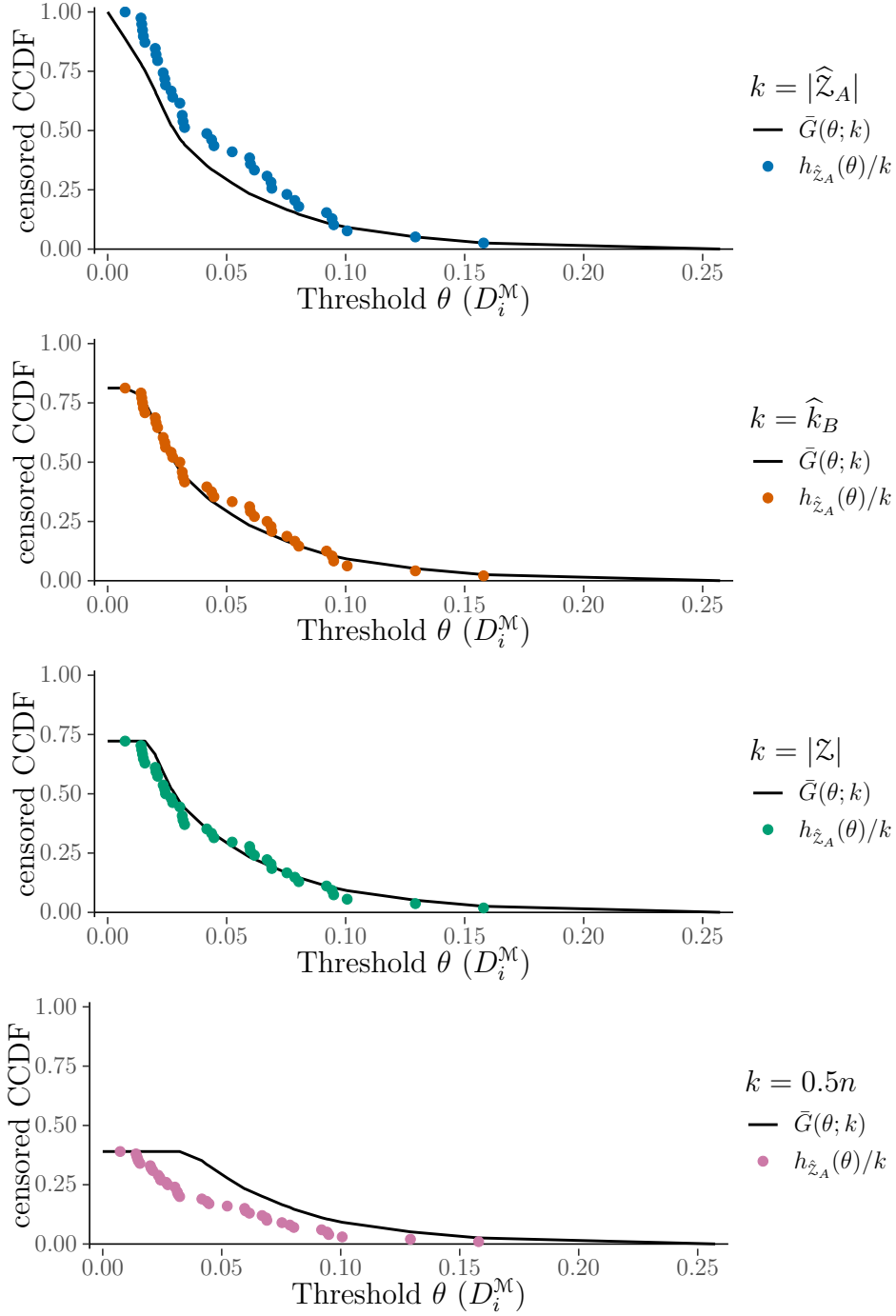
Figure 11: The empirical and expected censored complementary cumulative distribution functions (CCDF) at different candidate sample sizes. The empirical censored CCDF $\frac{h_{\widehat{\mathcal{Z}}_A}(\theta)}{k}$ counts how many observations in the censored sample are above the threshold θ and normalizes it by k. The expected censored CCDF is $\bar{G}(\theta; k) = \min\{G(\theta), |\widehat{z}_A|/k\}$. The top panels shows the minimal possible $k = |\widehat{\mathcal{Z}}_A|$, while the bottom panel shows the maximum value $0.5n$. The second panel shows the estimated $k = \widehat{k}_B$ and the third panel shows the true sample size $|\mathcal{Z}|$.

the expected censored CCDF.

$$\widehat{k}_B = \underset{k \in \{|\widehat{\mathcal{Z}}_A|,...,^n/_2\}}{\arg\min} \frac{1}{\sqrt{|\Theta|}} \sum_{\theta \in \Theta} \left| \frac{h_{\widehat{\mathcal{Z}}_A}(\theta)}{k} - \bar{G}(\theta; k) \right| \tag{7}$$

Typically Θ is the order statistics of $\{\mathcal{D}i\}_{i \in \widehat{\mathcal{Z}}_A}$. However, when these limiting degrees are derived through simulation, the values of $D_i^{\mathcal{M}}$ are discrete. For this reason, we use the non-repeating values in the same set. The normalization constant $^1/\sqrt{|\Theta|}$ ensures efficient convergence of the estimator.[27]

Figure 12 shows the minimized function of the estimator at different values of k section 3.3. The function has a unique global minimum over the range of possible k, although in this instance the estimated $\widehat{k}_B = 48$ is less than the true $|\mathcal{Z}| = 54$.
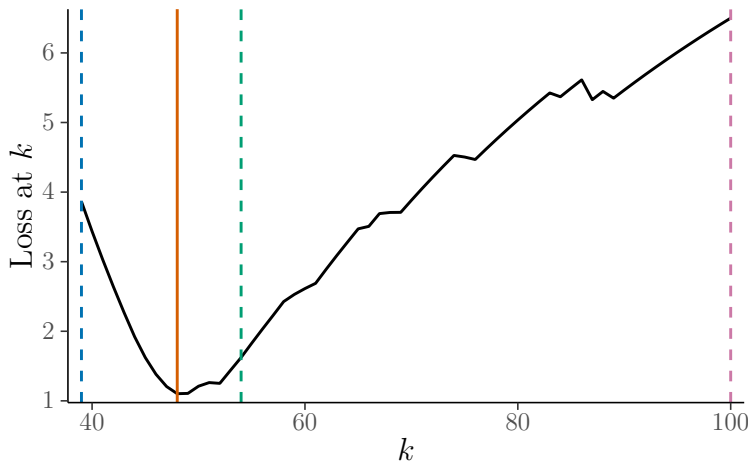

Figure 12: The least absolute deviations loss for different candidate values of k. The leftmost dashed line corresponds to $k = |\widehat{\mathcal{Z}}_A|$, and the rightmost one is $k = 0.5n$. The solid line in the middle corresponds to the estimated $\widehat{k}_B$ which minimizes the loss. The dashed line in the middle corresponds to the truce size of the uncensored sample $|\mathcal{Z}|$.

Table 3 shows the results of monte carlo simulations for the estimator from Algorithm B. We do not see a significant difference in performance for different true preferences or different error rates. The performance for different error distributions changes owing to the bias introduced by the the censoring assumption. Note that we cannot calculate truncated normals if they are centered around the true choices. Importantly, increasing the sample size does not change the performance of the estimator; the bias remains fairly constant.

---

[27]Proof forthcoming.

Performance of Algorithm B Estimation in Simulations

	True Preferences			
	substitute	complement	asymmetric	random
estimate	0.24	0.266	0.228	0.254
	(0.05)	(0.04)	(0.05)	(0.05)
bias	-0.199	-0.115	-0.239	-0.152
	(0.18)	(0.15)	(0.15)	(0.17)

	Error Distribution			
	Uniform	FOSD	$N(\sigma^2 = 15)$	$N(\sigma_i^2 = 15\%)$
estimate	0.266	0.312	NA	NA
	(0.04)	(0.04)	(NA)	(NA)
bias	-0.115	0.04	NA	NA
	(0.15)	(0.14)	(NA)	(NA)

	Sample Size n					
	25	50	100	200	400	800
estimate	0.266	0.26	0.278	0.266	0.269	0.279
	(0.15)	(0.12)	(0.08)	(0.04)	(0.04)	(0.02)
bias	-0.113	-0.134	-0.072	-0.115	-0.103	-0.071
	(0.49)	(0.41)	(0.28)	(0.15)	(0.14)	(0.07)

	Error Rate π				
	$\pi = 0.05$	$\pi = 0.1$	$\pi = 0.2$	$\pi = 0.3$	$\pi = 0.4$
estimate	0.049	0.09	0.186	0.266	0.353
	(0.02)	(0.03)	(0.04)	(0.04)	(0.05)
bias	-0.023	-0.098	-0.068	-0.115	-0.118
	(0.47)	(0.28)	(0.2)	(0.15)	(0.13)

Table 3: The estimate and bias averaged over 100 replications are reported. The standard deviation across replications is below in parentheses. Unless otherwise stated, in each replication the sample size is $n = 200$, true preferences are perfect complements, the error rate π is 0.3, and the contaminating error is uniformly distributed on the exterior of the budget set. The simulating graph is of size $m = 1500$. Budget lines are generated randomly as in Choi et al. (2007a)

Algorithm B: Estimation of $\widehat{\pi}_B$

Simulate Limiting Error Degree Distribution

    inputs : experiment generating function, error distribution

1   **for** $j = 1 \ldots m$ *(very large)* **do**

2   Generate an experiment $\mathcal{B}_j$

3   Simulate a mistake μ_j from $\mathcal{B}_j$

4   Create the graph $\mathbf{M}$, indicating 2-cycles in the simulated errors $\mathcal{M}$

5   Calculate degrees for each j using $\mathbf{M}$, normalized by m

6   **return** limiting degree distribution$\{D_j^{\mathcal{M}}\}_{j\in\mathcal{M}}$; simulated mistakes $\{(\mu j, \mathcal{B}_j)\}$

Simulate Limiting Degrees for $\widehat{\mathcal{Z}}_A$

    inputs : observed choices in $\widehat{\mathcal{Z}}_A$, and the simulated errors

7   **for** i *in* $\widehat{\mathcal{Z}}_A$ **do**

8   Find how many 2-cycles i forms in $(x_i, \mathcal{B}_i) \cup \{(\mu_j, \mathcal{B}_j)\}_{j\in\mathcal{M}}$

9   save the error degree $D_i^{\mathcal{M}}$, normalized by $m+1$

10   **return** the sample of error degrees $\{D_i^{\mathcal{M}}\}_{i\in\widehat{\mathcal{Z}}_A}$

Derive CCDF

    Define Θ, the unique values of $\{D_i^{\mathcal{M}}\}_{i\in\widehat{\mathcal{Z}}_A}$

    for $\theta \in \Theta$ **do**

        find $G(\theta) = \frac{1}{m}\sum_{j=1}^{m} \mathbb{I}\{D_j^{\mathcal{M}} \geq \theta\}$

        find $h_{\widehat{\mathcal{Z}}_A}(\theta) = \sum_{i\in\widehat{\mathcal{Z}}_A} \mathbb{I}\{D_i^{\mathcal{M}} \geq \theta\}$

    return the CCDF $G(\cdot)$ and the empirical count $h_{\widehat{\mathcal{Z}}_A}(\cdot)$

Estimate $\widehat{\pi}$

    for $k \in \{|\widehat{\mathcal{Z}}_A| \ldots {}^n/_2\}$ **do**

        find the loss at k per section 3.3

    Find $\widehat{k}$ the argument that minimizes the loss **return** the estimated $\pi = \frac{\widehat{k}}{n}$.

---

4 Empirics

Now that we have presented the methods, we can apply them to a number of datasets. As of the writing of this draft I have results for a dataset of individual choices over risky assets collected in the lab (Choi, Fisman, Gale, and Kariv 2007a). This is the benchmark dataset which can be used to compare existing indices against results from the estimation procedure described in section 3.

4.1 Choi, Fisman, Gale, and Kariv (2007a)

In their paper on preferences over uncertainty, Choi et al. collect a rich dataset in the lab of 47 individuals' choices. During the experiment each individual participated in 50 decision rounds. In each round the participant allocated tokens between two arrow securities, X and Y, each of which had an equal probability of paying out. The allocation was constrained by a linear budget set, which was generated randomly for each round.

At the end of the experiment one of the rounds was chosen, and the winning arrow security paid out to the participant at a rate of 2 tokens for every dollar. The budget lines were chosen to have intercepts between 10 and 100 tokens, with at least one between 50 and 100.[28] The variation in slopes captures the trade-off between efficiency and risk; investing everything in the cheapest security maximizes expected payout; investing equal amounts in both goods minimizes variance in payout. The steeper the slope, the higher the maximum expected payout. Since the arrow securities had an equal probability of paying out, the labels of the arrow securities were arbitrary. In additional to the original 50 observations, the X and Y labels were swapped and added to the choice dataset, making a total of 100 observations. Details of how this was done can be found in the replication code.

The top panel of figure 13 shows the result of the estimation procedure described in previous sections. The estimation of $\widehat{\pi}_B$ assumes a uniform error distribution on the choices. The bottom panel shows estimates of Afriat's CCEI for the subjects in the experiment (left). The second distribution is what wold have been observed if the choices were drawn completely uniformly random from the budget set (e.g. with $\pi = 1$). In general we find that most people have error rates between 5% and 14 %.

---

[28]Some prices were recalculated to correct rounding errors in the raw data file.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	St. Dev.
$\widehat{\pi}^A$	0.00	0.04	0.08	0.09	0.13	0.42	0.08
$\widehat{\pi}^B$	0.00	0.05	0.10	0.12	0.14	0.50	0.11

Table 4: Summary statistics for the distribution of error rates estimtaed from the Choi et al. (2007b) data.
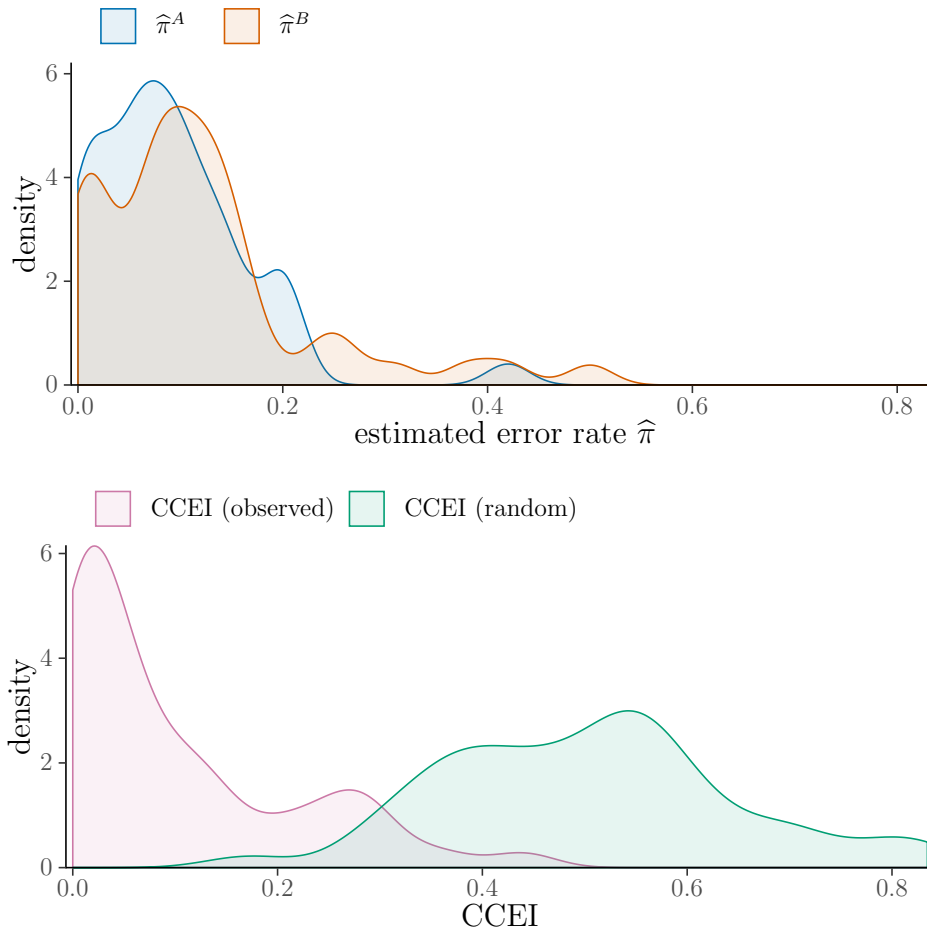

Figure 13: The top panel shows the distribution of estimated error rates across the 47 individuals in Choi et al. (2007a). $\widehat{\pi}_A$ is the lower bound from the classifier and $\widehat{\pi}_B$ is the result of the estimator when mistakes are drawn uniformly from the budget set exterior. The bottom panel shows the distribution on the same individuals for Afriat's CCEI (left). The distribution on the bottom right is what would have been observed in terms of the CCEI had the error rate been 100%.

Summary and Concluding Remarks

In this paper I address the question of how to separate preferences from error when the researcher observes a single individual make choices from a sequence of overlapping but non-identical menus. I develop the observed revealed preference relation, which incorporate error in choices and sampling variation.

Acyclicity of this relation is equivalent to rationality under certain conditions on the sampling distribution of the budget sets. I nest the model of classical (deterministic) rational choice through a dispersion parameter in the choice distribution; when dispersion is zero the researcher always observes the true choice which is rational. When dispersion is non-zero, I prove that the probability of observing a rational dataset shrinks to zero with more observations. I show that the observed revealed preference relation, by nature of its construction, is a random graph that has a specific dependency structure.

I use the contaminated data model to impose more structure on this graph. This turns the nonparametric revealed preference relation into a semi-parametric model. I prove that when the distribution of mistakes is known, this model is identified by the true preferences and a parameter of error rate π. This π is related to the Houtman-Maks index, which is a lower-bound estimate. I also show that under models such as this, where choices are observed with error, Varian's Index is not robust to differences in preferences, and Afriat's index converges to an uninformative lower bound.

In the case of the contaminated data model, I provide a classifier that partially recalls the errors. I also provide an estimator of π that can correct for the missing errors, as long as the researcher knows the contaminating error distribution. Taking these methods to a benchmark dataset of choices over risky assets collected in the lab, I find that most individuals have an error rate of between 5 and 14%.

There are of course limitations to these methods. Many of the theoretical results rely on independence of the choice distribution and budget set sampling distribution. They also take a philosophical stance that the errors affect the choices but not the budget sets. This assumption is reasonable in the context of lab experiments, but should be applied cautiously when using observational datasets. There are also some theoretical gaps in this approach. More investigation is required to prove exactly how and when the classifier works. This would also help the main limitation of the estimator, namely its bias.

As long as the assumptions on the choice space, budget sampling distribution, and choice distribution or met, this method can be applied to a variety of large datasets measuring individual choices from different menus of alternatives. Unlike other approaches, the methods of this paper can be applied to discrete and continuous menus of alternatives, opening up the

potential applications of empirical revealed preference techniques. The classifier is especially unique in its ability to recover (partially) the errors. These can be correlated with other observables at the time of the decision, or used to prune the dataset from noise. The estimators are also interesting in their own right; unlike existing indices they are easily interpretable and can be compared both across individuals and across choice settings. Further work is needed to develop statistical properties for these approaches.

More broadly, this project speaks to a deep connection between the theory of revealed preferences, random graphs, and models of choice with error. The dependency structure inherent in the observed revealed preference relation carries over as more observations are collected. With enough observations, it may be possible to construct a limiting graph (or graphon) on these observations. An interesting line of research naturally follows, asking how we can we more fully connect between the limiting behavior of an ordering and a utility function.

References

Abaluck, J. and A. Adams (2018). What do Consumers Consider Before they Choose? Identification from Asymmetric Demand Responses.

Afriat, S. N. (1967). The Construction of Utility Functions from Expenditure Data. *International Economic Review 8*(1), 67–77.

Afriat, S. N. (1972). Efficiency Estimation of Production Functions. *International Economic Review 13*(3), 568–598.

Aguiar, V. H. and N. Kashaev (2018). Stochastic Revealed Preferences with Measurement Error. (April).

Apesteguia, J. and M. A. Ballester (2015). A Measure of Rationality and Welfare. *Journal of Political EconomyJournal of Political Economy 123*(6), 1278–1310.

Arrow, K. J. (1959). Rational Choice Functions and Orderings. *Economica 26*(102), 121–127.

Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and Newman–Girvan and other modularities. *Proceedings of the National Academy of Sciences 106*(50), 21068–21073.

Blundell, R. W., M. Browning, and I. A. Crawford (2003). Nonparametric Engel Curves and Revealed Preference. *Econometrica 71*(1), 205–240.

Chambers, C. P. and F. Echenique (2016). *Revealed Preference Theory.* Cambridge University Press.

Chambers, C. P., F. Echenique, and N. S. Lambert (2018). Preference Identification.

Channarond, A., J.-J. Daudin, and S. Robin (2012). Classification and estimation in the Stochastic Blockmodel based on the empirical degrees. *Electronic Journal of Statistics 6*, 2574–2601.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007a). Consistency and Heterogeneity of Individual Behavior under Uncertainty. *American Economic Review 97*(5), 1921–1938.

Choi, S., R. Fisman, D. Gale, and S. Kariv (2007b). Revealing Preferences Graphically: An Old Method Gets a New Tool Kit. *American Economic Review 97*(2), 153–158.

Dean, M. and D. Martin (2016). Measuring Rationality With the Minimum Cost of Revealed Preference Violations. *The Review of Economics and Statistics 98*(3), 524–534.

Diewert, W. E. (1973). Afriat and Revealed Preference Theory. *The Review of Economic Studies 40*(3), 419–425.

Echenique, F., S. Lee, and M. Shum (2011). The Money Pump as a Measure of Revealed Preference Violations. *Journal of Political Economy 119*(6), 1201–1223.

Epstein, L. G. and A. J. Yatchew (1985). Non-parametric Hypothesis Testing Procedures and Applications to Demand Analysis. *Journal of Econometrics 30*(1-2), 149–169.

Fishkind, D. E., D. L. Sussman, M. Tang, J. T. Vogelstein, and C. E. Priebe (2013). Consistent Adjacency-Spectral Partitioning for the Stochastic Block Model when the Model Parameters are Unknown. *SIAM Journal of Matrix Analysis and Applications 34*(1), 23–39.

Fleissig, A. R. and G. A. Whitney (2005). Testing for the Significance of Violations of Afriat's Inequalities. *Journal of Business & Economic Statistics 23*(3), 355–362.

Gale, D. (1960). A Note on Revealed Preference. *Economica 27*(108), 348–354.

Gross, J. (1995). Testing Data for Consistency with Revealed Preference. *The Review of Economics and Statistics 77*(4), 701–710.

Gross, J. and D. Kaiser (1996). Two Simple Algorithms for Generating Subset of Data Consistent with WARP and Other Binary Relations. *Journal of Business & Economic Statistics 14*(2), 251–255.

Halevy, Y., D. Persitz, and L. Zrill (2018). Parametric Recoverability of Preferences. *Journal of Political Economy 126*(4), 1558–1593.

Hanoch, G. and M. Rothschild (1972). Testing the Assumptions of Production Theory: A Nonparametric Approach. *Journal of Political Economy 80*(2), 256–275.

Hey, J. D. and C. Orme (1994). Investigating Generalizations of Expected Utility Theory Using Experimental Data. *Econometrica 62*(6), 1291–1326.

Hjertstrand, P. (2013). A Simple Method to Account for Measurement Errors in Revealed Preference Tests.

Horowitz, J. L. and C. F. Manski (1995). Identification and Robustness with Contaminated and Corrupted Data. *Econometrica 63*(2), 281–302.

Houthakker, H. S. (1950). Revealed Preference and the Utility Function. *Economica 17*(66), 159–174.

Houtman, M. and J. Maks (1985). Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve methoden 19*, 89–104.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics 35*(1), 73–101.

Kitamura, Y. and J. Stoye (2018). Nonparametric Analysis of Random Utility Models. *Econometrica 86*(6), 1883–1909.

Manzini, P. and M. Mariotti (2014). Stochastic Choice and Consideration Sets. *Econometrica 82*(3), 1153–1176.

McFadden, D. L. (2005). Revealed stochastic preference: A synthesis. *Economic Theory 26*(2), 245–264.

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

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

Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics 39*(4), 1878–1915.

Samuelson, P. A. (1938). A Note on the Pure Theory of Consumer's Behavior. *Economica 5*(17), 61–71.

Sen, A. K. (1971). Choice Functions and Revealed Preference. *The Review of Economic Studies 38*(3), 307–317.

Snijders, T. A. B. and K. Nowicki (1997). Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure. *Journal of Classification 14*, 75–100.

Uzawa, H. (1956). Note on Preference and Axioms of Choice. *Annals of the Institute of Statistical Mathematics 8*(1), 35–40.

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

Varian, H. R. (1985). Non-parametric analysis of optimizing behavior with measurement error. *Journal of Econometrics 30*(1-2), 445–458.

Varian, H. R. (1990). Goodness-of-Fit in Optimizing Models. *Journal of Econometrics 46*(1), 125–140.

Zhao, Y., E. Levina, and J. Zhu (2012). Consistency of Community Detection in Networks Under Degree-Corrected Stochastic Block Models. *The Annals of Statistics 40*(4), 2266–2292.

A   Negative Results on Existing Goodness-of-Fit Indices

A number of existing indices seek to measure "how far" a given dataset is from satisfying rationality. When choices are drawn from a distribution with dispersion around some rational true choice, it becomes clear that these indices suffer from a number of undesirable properties. The following section applies only to the case of choices from linear budget sets in $\mathbb{R}_+^L$, since the most commonly used indices are defined only on this space. As before, the choices have any distributional form $\mathbf{x}_i \sim F(\cdot; \theta | \mathcal{B}_i)$ where θ are the parameters for the distribution. The budget sets are defined by normalized price vectors such that $\mathcal{B}_i = \{\mathbf{x} \in \mathbb{R}_+^L : \mathbf{p}_i \cdot \mathbf{x} \leq 1\}$. We will refer only to the price vectors $\mathbf{p}_i$ moving forward.

Define $\mathbf{v}$ to be a vector of length n, where each element $v_i \in [0, 1]$ captures the amount by which each budget set i is shifted inwards. We can define $R_{\mathbf{v}}^0$, the perturbed revealed preference relation, as $\mathbf{x}_i R_{\mathbf{v}}^0 \mathbf{x}_j$ whenever $v_i \mathbf{p}_i \cdot \mathbf{x}_i \geq \mathbf{p}_i \cdot \mathbf{x}_j$. When the inequality is strict we say that $\mathbf{x}_i P_{\mathbf{v}}^0 \mathbf{x}_j$.

Definition $(GARP_{\mathbf{v}})$. A dataset $\mathcal{D}_n$ satisfies $GARP_{\mathbf{v}}$ if there does not exist a sequence $k \in (1), ...(K)$ such that $\mathbf{x}_{(1)} R_{\mathbf{v}}^0 \mathbf{x}_{(2)}$, $\mathbf{x}_{(2)} R_{\mathbf{v}}^0 \mathbf{x}_{(3)}$,..., $\mathbf{x}_{(K-1)} R_{\mathbf{v}}^0 \mathbf{x}_{(K)}$ and $\mathbf{x}_{(K)} R_{\mathbf{v}}^0 \mathbf{x}_{(1)}$, where *at least one* of $R_{\mathbf{v}}^0$ is strict $P_{\mathbf{v}}^0$.

The standard definition of GARP (the generalized axiom of revealed preferences) is equivalent to $GARP_{\mathbf{1}}$ and cyclical $\unrhd$-consistency on the Euclidean space with $\geq$ ordering. As in Halevy et al. (2018), the most common existing indices can be written as solving for the maximum of a function of $\mathbf{v}$ subject to the data satisfying $GARP_{\mathbf{v}}$. We turn our attention to three such indices, Afriat's CCEI (Afriat 1967), Varian's Index (Varian 1983), and Houtman-Maks (Houtman and Maks 1985).

All proofs are in appendix B.

A.1   Afriat's Index

Afriat's index measures the largest uniform shift necessary such that all observations satisfy $GARP_{\mathbf{v}}$.

Definition (Afriat's Index). $I_A(\mathcal{D}_n) = \max_{v \in [0,1]} v$ such that $\mathcal{D}_n$ satisfies $GARP_{\mathbf{v}}$, where $\mathbf{v}$ is a vector of v, of length n.

We are interested in the limiting behavior of Afriat's index. The first thing to notice is that additional observations to the dataset can only decrease or maintain the index's value.

Lemma 7. Consider a sequence of datasets $\{\mathcal{D}_n\}_n$ where each $\mathcal{D}_n \subset \mathcal{D}_{n+1}$. Afriat's index is weakly decreasing in this sequence, $I_A(\mathcal{D}_n) \geq I_A(\mathcal{D}_{n+1})$.

Applying the monotone convergence theorem, we could show that for any sequence of datasets (incrementally increasing in observations), Afriat's index converges to a limit. But we are interested in the nature of this limit when the sequence of datasets is drawn randomly.

It is possible to imagine a very unlucky draw from the budget sampling and choice distributions which creates a large violation that requires a very small v to break all the cycles. Because Afriat's index is a maximum, it is not robust to these potentially rare events. The worst of these events can be calculated using only the *support* of $F(\cdot, \theta|\mathcal{B})$ and $F_\mathcal{B}$.[29] Holding $F_\mathcal{B}$ fixed, consider two different distributions for choice F. Even if they have very different dispersion, as long as they are distributed on the same support F, their Afriat's index will converge to the same value. This implies that Afriat's index does not tell us anything meaningful beyond the support of the distribution F. At the limit, the Afriat's index from any choice distribution with equal support will converge to the same value as if the choices were selected uniformly on the same support.[30]

Proposition 8. Suppose $\mathbf{x}_i \sim_{iid} F(\cdot, \theta|\mathcal{B}_i)$, where F is a generic distribution function with closed support $S_F(\mathcal{B})$. Let the budget sets be generated *iid* according to a distribution with closed support $S_\mathcal{B}$. Then there exists a $\Lambda < 1$ such that $I_A(\mathcal{D}_n) \to \Lambda$ as $n \to \infty$ almost surely, where Λ is a function only of the supports S_F and $S_\mathcal{B}$.

For the proofs in this section we use the fact that each observation $(\mathcal{B}_i, x_i)$ in $\mathcal{D}_n$ is conditionally independent and identically distributed. This means that each observation is interchangeable, and the order of observations does not matter.

A.2   Varian's Index

Unlike Afriat's index, Varian's index finds a value v_i for each observation and maximizes an aggregation of all the v_i.

Definition (Varian's Index). $I_V(\mathcal{D}_n) = \max_{\mathbf{v} \in [0,1]^n} f(\mathbf{v})$ such that $\mathcal{D}_n$ satisfies $GARP_\mathbf{v}$, where $f(\cdot)$ is a normalized aggregator function such as $\frac{1}{n} \sum_{k=1}^n v_k$.

Varian's index is normalized by n and so it does not suffer from the same sensitivity to outliers as Afriat's.[31] However, Varian's index does rely on the interaction between the choice

---

[29]By support we mean essential support- the largest closed set such that every open set in the complement has probability zero.

[30]Equal support of distributions is a sufficient but not necessary condition for the limit to be the same. In fact, as long as one "end" of the support is identical the limiting value will be the same.

[31]One common critique of Varian's index is that for moderate sized datasets it is impossible to calculate exactly, only approximations exist.

distribution's central tendency and the budget sets' data generating function. So two choice distributions can have the same spread (amount of irationality) but different centrality and they will converge to different values of Varian's index.

Example 6. Suppose we have two observations, and for ease of exposition suppose there are two goods $L = 2$. Fix the two budget sets defined by $\mathbf{p}$ and $\mathbf{q}$ and let the price vectors be normalized so $p_1 x_1 + p_2 x_2 = 1$ and $q_1 y_1 + q_2 y_2 = 1$. Define the intersection between these two as $\mathbf{w}$, not on the 45-degree line. We observe a dataset $\{(\mathbf{p}, \mathbf{x}), (\mathbf{q}, \mathbf{y})\}$. If the choices are such that both $\mathbf{py} < 1$ and $\mathbf{qx} < 1$ hold, then there is a violation of GARP and $I_V(\mathbf{x}, \mathbf{y}) = \max\{\mathbf{py}, \mathbf{qx}\}$. Since choices will always be on the budget exterior set we can rewrite the index in terms of x_1 and y_1. $I_V(x_1, y_1) = \max\{\left(p_1 y_1 + \frac{p_2}{q_2}(1 - q_1 y_1)\right), \left(q_1 x_1 + \frac{q_2}{p_2}(1 - p_1 x_1)\right)\}$

Fixing the price vectors, consider the choices as drawn from a contaminated choice distribution. Let there be a point mass of probability $1 - \pi$ on the true choices, with a probability mass π distributed on the budget set exterior $\partial\mathcal{B}_p^\uparrow = \{\mathbf{x} : \mathbf{p} \cdot \mathbf{x} = 1\}$ and $\partial\mathcal{B}_q^\uparrow = \{\mathbf{y} : \mathbf{q} \cdot \mathbf{y} = 1\}$.
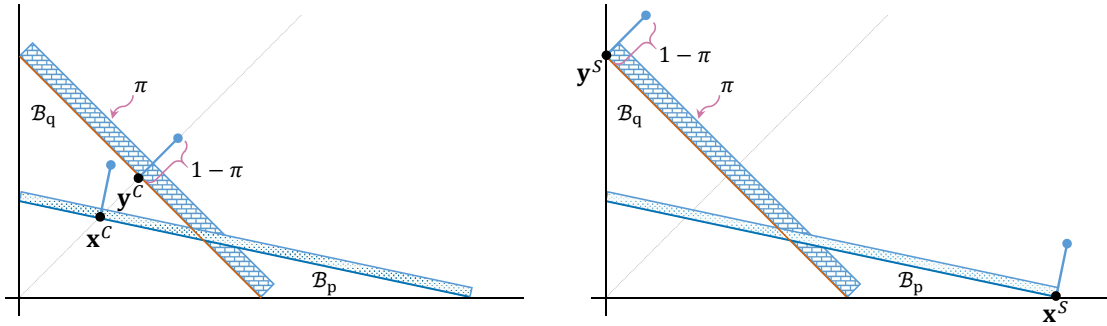
Consider two different underlying preferences: the individual C prefers equal amounts of both goods (perfect complements), and the individual S prefers only the cheapest good (perfect substitutes). The figure illustrates the distribution of choices under each of these two "true" preferences. If we were to compare the expected value of I_V, conditioning on $\mathbf{p}$ and $\mathbf{q}$, the difference $\mathbb{E}_S[I_V(x_1, y_1)|\mathbf{p}, \mathbf{q}] - \mathbb{E}_C[I_V(x_1, y_1)|\mathbf{p}, \mathbf{q}]$ is not zero. In fact, it depends in magnitude on the underlying preferences C relative to the intersection point $\mathbf{w}$.

The expected value of S is written out below. The first term is the probability that "there are no GARP violations"; the index in this event is 1. The second term integrates the index for all combinations of choices that do result in a GARP violation, multiplied by the probability of such a combination occurring.
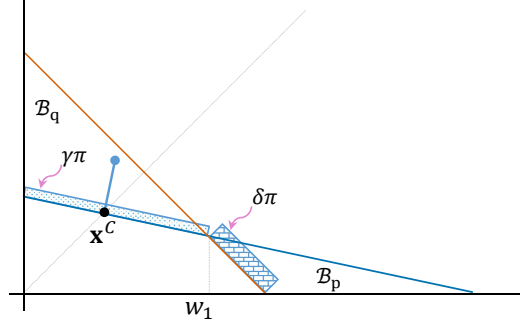
$$\mathbb{E}_S[I_V(x_1, y_1)|\mathbf{p}, \mathbf{q}] = \left(1 - \delta\gamma\pi^2\right) + \int\limits_{x_1=0}^{w_1} \int\limits_{y_1=w_1}^{1/q_1} \delta\gamma\pi^2 I_V(x_1, y_1) dy_1 dx_1$$

The relevant difference between to two distributions is that under C, the location of the point mass on $\mathbf{x}^C$ is in the interior of the budget set defined by $\mathbf{p}$. In this case, the probability of "no GARP violations" decreases by $\delta\pi(1-\pi)$, and the expected value when there is a violation has an additional term owing to the point mass. With some algebra we have

$$\mathbb{E}_S[I_V(x_1, y_1)|\mathbf{p}, \mathbf{q}] - \mathbb{E}_C[I_V(x_1, y_1)|\mathbf{p}, \mathbf{q}] = \delta\pi(1-\pi)\left(1 - \int\limits_{y_1=w_1}^{1/q_1} I_V(x_1^C, y_1) dy_1\right)$$

Since the integral is over points that are involved in a GARP violation, the integral will be less than 1, so the difference in expected values is strictly positive. Furthermore, the magnitude of the difference depends explicitly on the location of $\mathbf{x}^C$ relative to w_1. ◀

Generally, the expected value of Varian's index, conditional on the collection of price vectors $\mathcal{P}_n$, is

$$\mathbb{E}_F[I_V(\mathcal{D}_n)|\mathcal{P}_n] = \int\limits_{\partial\mathcal{B}_1^\uparrow} \int\limits_{\partial\mathcal{B}_2^\uparrow} \cdots \int\limits_{\partial\mathcal{B}_n^\uparrow} I_V(\{\mathbf{x}_i\}_{i=1}^n, \mathcal{P}_n) dF\mathbf{x}_n \dots dF\mathbf{x}_2 dF\mathbf{x}_1.$$

Consider an arbitrary observation i, which for ease of notation we will take to be the first observation $i = 1$. The exterior of i's budget can be divided into subsets $\mathcal{X}_1^k$, where each subset is overlapped by a different collection of budget sets. Holding the distribution of all other observations $2, \dots, n$ fixed, and applying the law of iterated expectation, the expected value of Varian's index is

$$\mathbb{E}_F[I_V(\mathcal{D}_n)|\mathcal{P}_n] = \sum_k \mathbb{P}_{F_1}(\mathbf{x}_1 \in \mathcal{X}_1^k) \int_{\mathcal{X}_1^k} \tilde{I}_V(\mathbf{x}_1, \mathcal{P}_n) dF_{|\mathcal{X}_1^k}\mathbf{x}_1.$$

The function $F_{|\mathcal{X}_1^k}$ is the distribution of $\mathbf{x}_1$ conditional on being in subset k. The function $\tilde{I}_V$ is the conditional expectation over all other budget sets

$$\tilde{I}_V(\mathbf{x}_1, \mathcal{P}_n) = \int_{\partial\mathcal{B}_2^\uparrow} \cdots \int_{\partial\mathcal{B}_n^\uparrow} I_V(\{\mathbf{x}_1 \cup \mathbf{x}_i\}_{i=2}^n, \mathcal{P}_n) dF\mathbf{x}_n \ldots dF\mathbf{x}_2.$$

Assuming the prices are normalized, the value of Varian's index is always $\mathbf{p}_i \cdot \mathbf{x}_j \in [0, 1]$ for some i and j. This means that for all $\mathbf{x}_1 \in \mathcal{X}_1^k$, $\tilde{I}_V(\mathbf{x}_1, \mathcal{P}_n)$ is either constant in $\mathbf{x}_1$, or a linear combination of the elements of $\mathbf{x}_1$. By definition, all $\mathbf{x}_1 \in \mathcal{X}_1^k$ will have the same function value $\tilde{I}_V(\mathbf{x}_1, \mathcal{P}_n)$, meaning it will be integrable over the region $\mathcal{X}_1^k$.

Now consider two different distributions of choice, F and G, that are otherwise the same except for on budget set 1, where they have the same measure of dispersion, but different centrality. Then there is some k such that $\mathbb{P}_{F_1}(\mathbf{x}_1 \in \mathcal{X}_1^k) > \mathbb{P}_{G_1}(\mathbf{x}_1 \in \mathcal{X}_1^k)$ and there is some j where the sign is reversed.

$$\mathbb{E}_F[I_V(\mathcal{D}_n)|\mathcal{P}_n] - \mathbb{E}_G[I_V(\mathcal{D}_n)|\mathcal{P}_n] = \sum_k \mathbb{P}_{F_1}(\mathbf{x}_1 \in \mathcal{X}_1^k) \int_{\mathcal{X}_1^k} \tilde{I}_V(\mathbf{x}_1, \mathcal{P}_n) dF_{|\mathcal{X}_1^k}\mathbf{x}_1 -$$

$$\mathbb{P}_{G_1}(\mathbf{x}_1 \in \mathcal{X}_1^k) \int_{\mathcal{X}_1^k} \tilde{I}_V(\mathbf{x}_1, \mathcal{P}_n) dG_{|\mathcal{X}_1^k}\mathbf{x}_1$$

A.3   Houtman-Maks Index

The Houtman-Maks Index finds the fewest number of observations that need to be removed in order to leave an acyclic graph. Under Assumption 5 there is some probability of cycles between two errors in which case the Houtman-Maks will be a lower-bound for the amount of error in a contaminated choice model. See discussion of estimation with an unknown error distribution in Section 3.2.

B   Proofs

Proposition 2

Proof. 1. If dispersion is 0, then $\mathbb{P}(\mathcal{D}_n$ is rational$) = 1$.

From Assumption 2 if dispersion is 0, the observed choice is the deterministic true choice. From Assumption 7, the true choice is rationalizable by some true preference relation $\succsim^*$ which extends $\rhd$. By construction Q_n^0 satisfies $\rhd$-consistency and we can apply Theorem 1 from Nishimura, Ok, and Quah (2017) to get rationality.

2. If dispersion is greater than 0, then $\mathbb{P}(\mathcal{D}_n$ is rational$) < 1$.

Consider the event of generating a sequence of $k \geq 2$ budget sets $\boldsymbol{\mathcal{B}}^k = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ for which the following hold: $\mathcal{B}_1 \cap \mathcal{B}_2^\downarrow \notin \{\emptyset, \mathcal{B}_1\}, \ldots, \mathcal{B}_k \cap \mathcal{B}_1^\downarrow \notin \{\emptyset, \mathcal{B}_k\}$. Call the probability of such a sequence $\mathbb{P}_{\boldsymbol{\mathcal{B}}}(\boldsymbol{\mathcal{B}}^k)$. From Assumption 5 we have that $\mathbb{P}_{\boldsymbol{\mathcal{B}}}(\boldsymbol{\mathcal{B}}^k) > 0$. Define the conditional probability $\mathbb{P}_{x_i}(x_i \in \mathcal{B}_i \cap \mathcal{B}_j^\downarrow | \mathcal{B}_i, \mathcal{B}_j) = p_{ij}$. From Assumption 8, conditional on the k budget sets, $p_{ij} \in (0,1)$ for all $i, j \in \{1, \ldots, k\}$.

By conditional independence of the observed choices (Assumption 4), the probability of a cycle in Q_n^0 on $\boldsymbol{\mathcal{B}}^k$ is a product of probabilities, so $\mathbb{P}(Q_n^0$ has cycle on $\boldsymbol{\mathcal{B}}^k | \boldsymbol{\mathcal{B}}^k) = p_{12} \cdots p_{k1}$. By the law of iterated probability, and independence of the budget sets (Assumption 3), $\mathbb{P}(Q_n^0$ has cycle on $\boldsymbol{\mathcal{B}}^k) = p_{12} \cdots p_{k1} \cdot \mathbb{P}_{\boldsymbol{\mathcal{B}}}(\boldsymbol{\mathcal{B}}^k)$. Since all component probabilities are greater than 0, and a cycle on $\boldsymbol{\mathcal{B}}^k$ is just one type of many, $\mathbb{P}(Q_n^0$ not acyclic$) > 0$. Using Assumption 6 we can apply Proposition 1, and get that $\mathbb{P}(\mathcal{D}_n$ is rational$) < 1$.

3. As $n \to \infty$, if dispersion is greater than 0, then $\mathbb{P}(\mathcal{D}_n$ is rational$) \to 0$.

Without loss of generality, suppose $n = 2K$ for some integer K. We can partition any dataset of size n into K partitions of datasets with 2 observations, which we index with k as $\mathcal{D}_n = \bigcup_{k=1}^K \mathcal{D}_2^k$. We construct an observed revealed preference relation for the k-th dataset of two observations, which we denote Q_2^k. Note that if Q_n^0 is acyclic then it has no cycles of length 2, so we can bound the probability from above $\mathbb{P}(Q_n^0$ acyclic$) \leq \mathbb{P}(Q_n^0$ has no cycles of length 2$)$. If there are no 2-cycles in Q_n^0 then every Q_2^k is acyclic. Let $a_2 \equiv \mathbb{P}(Q_2^0$ is acyclic$)$. Then by independence we have that

$$\mathbb{P}(Q_n^0 \text{ acyclic}) \leq (a_2)^K. \tag{8}$$

From item 2. above, when dispersion is not zero, $a_2 < 1$. Taking the limit as $n \to \infty$, we get that $(a_2)^K \to 0$, so we have that $\mathbb{P}(Q_n^0$ acyclic$) \to 0$. Applying Proposition 1, we have that $\mathbb{P}(\mathcal{D}_n$ is rational$) \to 0$. $\qquad \square$

Proposition 4

Proof. Take $\theta = (\mathbf{R}^*, \pi)$ and $\widetilde{\theta} = (\widetilde{\mathbf{R}}^*, \widetilde{\pi})$. The contrapositive of "only if" follows from the definition of the first moment: if $\theta = \widetilde{\theta}$ then $\mathbb{E}_{\theta, F_{\mu|\mathcal{B}}}[\mathbf{R}] = \mathbb{E}_{\widetilde{\theta}, F_{\mu|\mathcal{B}}}[\mathbf{R}]$.

We prove "if" by contradiction. Suppose $\theta \neq \widetilde{\theta}$ and $\mathbb{E}_{\theta,F_{\mu|\mathcal{B}}}[\mathbf{R}] = \mathbb{E}_{\widetilde{\theta},F_{\mu|\mathcal{B}}}[\mathbf{R}]$. Assumptions 9 and 10 and the equality of expectations imply

$$(1 - \pi)\mathbf{R}^* + \pi\,\mathbb{E}_{\mu|\mathcal{B}}[\mathbf{E}] = (1 - \widetilde{\pi})\widetilde{\mathbf{R}}^* + \widetilde{\pi}\,\mathbb{E}_{\mu|\mathcal{B}}[\mathbf{E}]. \tag{9}$$

(i) $\underline{\mathbf{R}^* = \widetilde{\mathbf{R}}^*}$. In this case, (9) implies $(\pi - \widetilde{\pi})\mathbf{R}^* = (\pi - \widetilde{\pi})\mathbb{E}_{\mu|\mathcal{B}}[\mathbf{E}]$. Since $\theta \neq \widetilde{\theta}$ but $\mathbf{R}^* = \widetilde{\mathbf{R}}^*$, it must be that $\pi \neq \widetilde{\pi}$. Since $\pi - \widetilde{\pi}$ is therefore not equal to 0, we have that $\mathbf{R}^* = \mathbb{E}_{\mu|\mathcal{B}}[\mathbf{E}]$. From Assumption 12, there is a pair (k, l) such that $\bar{e}_{kl} \notin \{0, 1\}$. Since r_{kl}^* is an element of $\{0, 1\}$, then $r_{kl}^* \neq \bar{e}_{kl}$, and we have a contradiction.

(ii) $\underline{\mathbf{R}^* \neq \widetilde{\mathbf{R}}^*}$. Then $\exists i, j$ such that $r_{ij}^* \neq \widetilde{r}_{ij}^*$. Note that this is not necessarily the same (k, l) from above. Equation (9) implies that

$$(1 - \pi)r_{ij}^* + \pi\,\bar{e}_{ij} = (1 - \widetilde{\pi})\widetilde{r}_{ij}^* + \widetilde{\pi}\,\bar{e}_{ij}.$$

Both $\mathbf{R}^*$ and $\widetilde{\mathbf{R}}^*$ are $\{0, 1\}$-valued matrices. Suppose without loss of generality that $r_{ij}^* = 0$ and therefore that $\widetilde{r}_{ij}^* = 1$. Then the above equation implies that $\pi\,\bar{e}_{ij} = 1 - \widetilde{\pi}(1 - \bar{e}_{ij})$. If the marginal probability $\bar{e}_{ij} = 1$, then $\pi = 1$, which contradicts Assumption 11. If $\bar{e}_{ij} = 0$, then $\widetilde{\pi} = 1$ which is also a contradiction. Otherwise, $\bar{e}_{ij} \in (0, 1)$ and

$$\widetilde{\pi} = \frac{1 - \pi\,\bar{e}_{ij}}{1 - \bar{e}_{ij}}.$$

Since $\pi, \widetilde{\pi}$ are both probabilities, the equality above implies that $\pi = \widetilde{\pi} = 1$ which is a contradiction. $\qquad\square$

Proposition 6

Proof. The error degree $\bar{d}_i^{\mathcal{Z}}$ is an average of $|\mathcal{Z}|$ indicators s_{ji} of a 2-cycle between i and j for any $j \in \mathcal{Z}$. Conditional on a realized pair $(\mu_i, \mathcal{B}_i)$, $s_{ji} \perp s_{ki}$ for any $j \neq k \neq i \in \mathcal{Z}$. Also conditional on $(\mu_i, \mathcal{B}_i)$, the probability of an indicator s_{ji} being 1 is exactly D_i. We can apply the strong law of large numbers to get almost sure convergence of $\bar{d}_i^{\mathcal{Z}}$ to D_i as $|\mathcal{Z}| \to \infty$. $\qquad\square$

Lemma 7

Proof. Consider the dataset of size n where each choice was made from a linear budget set defined by a price vector $\mathbf{p}$; $\mathcal{D}_n = \{(\mathbf{p}_i, \mathbf{x}_i)\}_{i=1}^n$. Let $v^* \in \arg\max_{v \in [0,1]} v$ s.t. $\mathcal{D}_n$ satisfies $GARP_{\mathbf{v}}$. Now suppose we have another observation, yielding $\mathcal{D}_{n+1} = \{\mathcal{D}_n; (\mathbf{p}_{n+1}, \mathbf{x}_{n+1})\}$. Let $w^* \in \arg\max_{w \in [0,1]} w$ s.t. $\mathcal{D}_{n+1}$ satisfies $GARP_{\mathbf{w}}$. Consider $\mathbf{w}'$, a vector of n elements each with value w^*. The original dataset $\mathcal{D}_n$ satisfies $GARP_{\mathbf{w}'}$. Thus, by definition of v^* as a maximizer, it must be the case that $v^* \geq w^*$. This means that $I_A(\mathcal{D}_n) \geq I_A(\mathcal{D}_{n+1})$. □

Proposition 8

Proof. Define $S_F(\mathbf{p})$ as the support of the choice distribution F conditional on a budget set defined by $\mathbf{p}$. Throughout this proof we will use m as a fixed sample size and n as the sample size which will increase to infinity. Given the set of prices vectors supported by $S_{\mathcal{B}}$, pick the combination of prices and choices that yield a dataset $\mathcal{D}_m^*$ of size m with the lowest possible index value which we call λ_m:

$$\lambda_m = \min_{\{\mathbf{p}_i \in S_{\mathcal{B}}\}_{i=1}^m} \min_{\{\mathbf{y}_i \in S_F(\mathbf{p}_i)\}_{i=1}^m} I_A\left(\{\mathbf{y}_i, \mathbf{p}_i\}\right)$$

For any sample size m the maximum cycle is of length m, so it is possible that different sample sizes have different minima. We will first show that the sequence of values of λ_m is weakly decreasing and convergent. We will then use this convergence to prove almost sure convergence of $I_A(\mathcal{D}_n)$ to the same limit. Note that λ_m is a function only of the supports S_F and S_B.

(I) $\underline{\exists \Lambda \in [0, 1) \text{ such that } \lambda_n \to \Lambda}$. Consider the dataset $\mathcal{D}_m^*$ that is the minimizer yielding λ_m, and the minimizer $\mathcal{D}_{m+1}^*$ yielding λ_{m+1}. If $\mathcal{D}_m^* \subset \mathcal{D}_{m+1}^*$ then $\lambda_{m+1} \leq \lambda_m$ by Lemma 7. If $\mathcal{D}_m^* \not\subset \mathcal{D}_{m+1}^*$ then it must be the case that $\lambda_{m+1} \leq \lambda_m$. Else one could construct a dataset $\mathcal{D}_{m+1}' = \mathcal{D}_m^* \cup (\mathbf{p}_i, \mathbf{x}_i)$ where $(\mathbf{p}_i, \mathbf{x}_i) \in \mathcal{D}_m^*$. By construction $I_A(\mathcal{D}_{m+1}') = \lambda_m$, so $\lambda_m < \lambda_{m+1}$ violates $\mathcal{D}_{m+1}^*$ being a minimizer. Therefore λ_n is monotonically decreasing.

By the monotone convergence theorem, there exists a $\Lambda \in [0, 1]$ such that $\lambda_n \to \Lambda$. By Assumptions 5 and 8, there is a large enough n such that $\lambda_n < 1$ so $\Lambda \in [0, 1)$. Since λ_n is a function only of the supports S_F and S_B, Λ is also a function only of the supports.

(II) $\underline{\Lambda \leq \liminf_{n \to \infty} I_A(\mathcal{D}_n)}$. By construction, $\lambda_n \leq I_A(\mathcal{D}_n)$ for all n and any $\mathcal{D}_n$. Therefore $\lim_{n \to \infty} \lambda_n \leq \liminf_{n \to \infty} I_A(\mathcal{D}_n)$. From convergence of λ_n, $\lim_{n \to \infty} \lambda_n = \Lambda$.

(III) $\underline{\limsup_{n\to\infty} I_A(\mathcal{D}_n) \leq \Lambda \text{ almost surely.}}$ Fix some m. For any choice of a sequence of dataset $\mathcal{D}_m$, with iid drawn observations, let the probability of the index falling $\varepsilon > 0$ away from the minimum λ_m be

$$\mathbb{P}\left(I_A(\mathcal{D}_m) \geq \lambda_m + \varepsilon\right) = 1 - \alpha \tag{10}$$

where α is some function of ε.

Now consider $n = Km$ where K is an integer much larger than 1. Let $\mathcal{D}_m^k$ be one of K randomly selected (without replacement) subsets of $\mathcal{D}_n$ of size m. The value $v_m = \min_k I_A(\mathcal{D}_m^k)$ will break cycles of length at most m. Either $\mathcal{D}_n$ satisfies $GARP_{v_m}$ or there are additional cycles in $\mathcal{D}_n$ that require a lower value of v to break, meaning that by Lemma 7, $I_A(\mathcal{D}_n) \leq \min_k I_A(\mathcal{D}_m^k)$. By definition, $\mathbb{P}\left(\min_k I_A(\mathcal{D}_m^k) \geq b\right) = \mathbb{P}\left(\forall k, I_A(\mathcal{D}_m^k) \geq b\right)$. The set of $\mathcal{D}_m^k$ can be thought of as k independent draws of size m, since observations are *iid* (Assumptions 3 and 4) and each dataset k is sampled randomly without replacement. So the probability that a dataset of size n falls ε away from λ_m can be bounded from above by $\left(\mathbb{P}(I_A(\mathcal{D}_m) \geq \lambda_m + \varepsilon)\right)^K$.

$$\mathbb{P}\left(I_A(\mathcal{D}_n) \geq \lambda_m + \varepsilon\right) \leq (1 - \alpha)^K$$

This means that summing over all possible sample sizes

$$\sum_{n=2}^{\infty} \mathbb{P}\left(I_A(\mathcal{D}_n) \geq \lambda_m + \varepsilon\right) \leq \sum_{n=2}^{\infty} (1 - \alpha)^{\lfloor n/m \rfloor}$$

The sum on the right hand side is finite, so we apply the Borel-Cantelli lemma. The event where '$\{I_A(\mathcal{D}_n) \geq \lambda_m + \varepsilon\}$ occurs finitely many times' thus has probability of 1. This means that with a large enough sample size n_0, for any $n \geq n_0$ it will be the case $I_A(\mathcal{D}_n) \leq \lambda_m + \varepsilon$. This applies for any $\varepsilon_n = \frac{1}{n}$, of which there are countably many. Taking the limit in n, the sequence ε_n converges to zero and we have

$$\limsup_{n\to\infty} I_A(\mathcal{D}_n) \leq \lambda_m \qquad a.s.$$

Then by (I), if we take the limit as $m \to \infty$, since m only appears on the RHS,

$$\limsup_{n\to\infty} I_A(\mathcal{D}_n) \leq \lim_{m\to\infty} \lambda_m = \Lambda \qquad a.s.$$

$\underline{\text{(II) and (III) imply that } I_A(\mathcal{D}_n) \xrightarrow{a.s.} \Lambda.}$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

C   Identification Example

This section contains an example with $n = 2$ that illustrates how the distribution of $\mathbf{R}$ is identified by $\theta = (\mathbf{R}^*, \pi)$.

Example 7. Consider the case of $n = 2$. The deterministic $\mathbf{R}^*$ belongs to a feasible set

$$\mathcal{F} = \left\{ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

The marginal probabilities of the error columns $\mathbf{e}_i$ are derived from a known error distribution. They are constants and denoted $\mathbb{P}_{\mu_1}(e_{21} = 1) = a$, $\mathbb{P}_{\mu_2}(e_{12}) = b$. The expected value of the completely random graph $\mathbf{E}$ is

$$\mathbb{E}[\mathbf{E}] = \begin{bmatrix} 1 & b \\ a & 1 \end{bmatrix}.$$

Conditional on $\theta = (\mathbf{R}^*, \pi)$, applying Assumptions 9 and 10 gives the first moment of $\mathbf{R}$.

$$\mathbb{E}_\theta[\mathbf{R}] = (1 - \pi) \begin{bmatrix} 1 & r_{12}^* \\ r_{21}^* & 1 \end{bmatrix} + \pi \begin{bmatrix} 1 & b \\ a & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & (1 - \pi)r_{12}^* + \pi b_{12} \\ (1 - \pi)r_{21}^* + \pi b_{21} & 1 \end{bmatrix}.$$

Now take any $\widetilde{\theta} = (\widetilde{\mathbf{R}}^*, \widetilde{\pi})$ such that $\theta \neq \widetilde{\theta}$. What conditions are necessary for $\mathbb{E}_\theta[\mathbf{R}] = \mathbb{E}_{\widetilde{\theta}}[\mathbf{R}]$? The two cases outlined in the proof are: (i) $\widetilde{\mathbf{R}}^* = \mathbf{R}^*$, and (ii) $\widetilde{\mathbf{R}}^* \neq \mathbf{R}^*$. Working through both of these will result in the cases that are eliminated by Assumptions 11 and 12. Note that in this example Assumption 12 implies that at least one of a or b are in $(0, 1)$.

D   Review of Graph Notation

A *graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a collection of vertices $\mathcal{V}$ and edges $\mathcal{E}$. In a directed graph an *edge* is an ordered pair (k, i), where $k, i \in \mathcal{V}$. For convenience we label the *vertices* with integers $\{1, \ldots, N\}$ where $N = |\mathcal{V}|$. The graph $\mathcal{G}$ can also be represented as an *adjacency matrix* $\mathbf{A}$ where an element a_{ki} is equal to 1 whenever the edge (k, i) belongs to the graph. If the edge (k, i) doesn't belong to the graph then adjacency matrix stores $a_{ki} = 0$. In an undirected graph, the edge pair is unordered and $\mathbf{A}$ is a symmetric adjacency matrix.

An *induced subgraph* of $\mathcal{G}$ on a set of vertices $\mathcal{W}$ is denoted $\mathcal{G}[\mathcal{W}]$. It is the graph on vertices $\mathcal{W}$ with edges $\{(k, i) \in \mathcal{E} : k, i \in \mathcal{W}\}$. The induced subgraph excludes any edges that connect to vertices outside of $\mathcal{W}$. If $\mathcal{G}$'s adjacency matrix is $\mathbf{A}$, then $\mathbf{A}[\mathcal{W}]$ is the adjacency matrix of the induced subgraph which includes only the rows and columns corresponding to vertices $\mathcal{W}$.

In an undirected graph, the (normalized) *degree* of a vertex i is the number of edges to which it is connected, normalized by the number of vertices in the graph. This is usually referred to as the normalized degree, but since we only use the normalized version in this paper it will simplify exposition to refer to it simply as the degree. It can be calculated by taking the sum on the i-th column of the adjacency matrix and dividing by N.

$$\bar{d}_i = \frac{1}{N} \sum_{k=1}^{N} a_{ki}.$$

Note that the degree $\bar{d}_i$ is always a number in $[0, 1]$. The degree on the induced subgraph of $\mathcal{W}$ is called the *degree on* $\mathcal{W}$. It is denoted for any $i \in \mathcal{W}$ by

$$\bar{d}_i^{\mathcal{W}} = \frac{1}{|\mathcal{W}|} \sum_{k \in \mathcal{W}} a_{ki}.$$

A graph is *acyclic* if there does not exist a sequence of edges $\{(1), (2), \ldots, (k)\}$. Acyclicity of the adjacency matrix $\mathbf{A}$ means there is no sequence of indices $\{(1), (2), \ldots, (k)\}$ such that $a_{(1)(2)} = 1; a_{(2)(3)} = 1; \ldots; a_{(k)(1)} = 1$. Any induced subgraph of an acyclic graph is also acyclic.