How Rational are your Choice Data?

Mark Dean

Department of Economics, Brown University, 64 Waterman Street, Providence, Rhode Island, 02912

Email address mark_dean@brown.edu

Daniel Martin

Department of Economics, New York University, 19 West 4th Street, New York, New York 10012.

Email address daniel.martin@nyu.edu

We are grateful to Shachar Kariv for providing the data and computer program used in Choi et al. [2006, 2007]; Martijn Houtman for providing the computer program used in Houtman [1995]; Ian Crawford, Hiroki Nishimura, Debraj Ray, Michael Richter, Joerg Stoye, Juan Carlos Suarez Serrato and participants in the NYU NRET seminar for helpful comments; Jesse Perla for programming advice; and Adam Sachs for research assistance.

How Rational are your Choice Data?

June 30, 2010

Abstract

We present an algorithm that finds the size of the largest subset of a choice data set that is consistent with acyclicality. Our algorithm is orders of magnitude more efficient than existing methods. It can be used to calculate measures of rationality, such as the Houtman-Maks index and the minimum multiple rationales, that have been previously impractical. We also develop a new measure of how close a data set is to rationality. Further, we demonstrate the efficiency of our algorithm on data from laboratory experiments [Choi et al. 2007] and household consumption data [Blundell, Browning and Crawford 2003, 2008]. In doing so, we provide previously unavailable rationality measures for these data sets.

1 Introduction

Much of economic theory is predicated on the assumption that people make choices as if they are maximizing a stable utility function. Choice data are consistent with this assumption if the preference information revealed by choice is *acyclic*.[1] If we observe choices from subsets

[1] Exactly what it is that reveals preference varies from theory to theory (see Varian [2006] for a review). The Strong Axiom of Revealed Preference (SARP) assumes that x is preferred to y if x is chosen when y was available. The Generalized Axiom of Revealed Preference (GARP) assumes that x is revealed preferred to y if x is chosen when y was available at a strictly lower price.

of some grand set Z, and the binary relation $\succ$ on Z represents the (strict) preferences revealed by those choices, acyclicality means that one cannot find a sequence z_1, z_2, ..., z_n in Z such that

$$z_1 \succ z_2 \succ ... \succ z_n \succ z_1.$$

Choice data that satisfy the acyclicality property are *rationalizable* as the outcome of utility maximization.[2]

Acyclicality is a stringent requirement for choice data to satisfy – for example, a single mistaken choice can lead to an entire data set being classified as 'irrational'. Unsurprisingly, in most data sets the acyclicality property is often violated. This is true both for data generated by laboratory experiments (see for example Choi, Gale, Fisman, and Kariv [2006, 2007]) and for consumption data from the field (see for example Famulari [1995]). The fact that acyclicality does not hold globally has led to theoretical and empirical work based on *which* subsets of a choice data set are consistent with acyclicality.[3]

Our paper contributes to this research in four ways:

1. We develop a new algorithm for calculating the size of the largest subset of a choice data set that is consistent with acyclicality. This algorithm is orders of magnitude more efficient than existing algorithms.

2. We show that this algorithm can be used to implement many existing tools for analyzing the rationalizability of choice data. The use of these tools has previously been severely limited by computational constraints.

3. We develop a new measure of how close a particular data set is to rationality that bridges the gap between previous measures based on the *number* of irrational choices and those based on the *severity* of violations of rationality.

[2] Assuming that the set of available alternatives is countable.

[3] See Koo [1963], Houtman and Maks [1985], Famulari [1995], Kalai, Rubenstein and Spiegler [2002], Blundell, Browning and Crawford [2003, 2008], Crawford and Pendakur [2008], Deb [2008], Apesteguia and Ballester [2009, 2010], Nobibon et al. [2009].

4. We use our algorithm to analyze experimental data from Choi, Gale, Fisman, and Kariv [2006, 2007] and household consumption data from Blundell, Browning and Crawford [2003, 2008]. We show that, in practice, our method allows analysis that was previously impossible, and in doing so we provide previously unavailable rationality measures for these data sets.

In section 2 we introduce our algorithm for finding the size of the largest subset of a choice data set that is consistent with acyclicality. We call this problem the *maximal acyclical set problem*, or MASP. While testing whether a set of preferences are acyclic is relatively easy,[4] MASP is NP-hard. This means that there is no known algorithm with solution times that are certain to only increase polynomially with the number of choices. The key to our approach is to show that MASP is equivalent to the minimum set covering problem (MSCP), which is well studied in the computer sciences and operations research literature. While MSCP is NP-hard in the strong sense, there are a wide variety of algorithms that are extremely efficient in solving it for practical cases and are included in standard 'solver' software packages (see Caprara, Toth, and Fischetti [2000]). For any choice data set, we can therefore translate the associated MASP into an equivalent MSCP, which can then be solved using one of these software packages.[5] Section 2 shows the equivalence between MASP and MSCP, and describes in detail the algorithm we use to employ MSCP solvers to solve MASP.[6]

As we show in section 3, the maximal acyclical subset plays a crucial role in several existing analytical tools. Prior to the development of our algorithm, computational costs meant that these tools were essentially unusable on reasonable sized data sets. These applications include:

1. *Measuring how close a data set is to rationality.* The size of the maximal acyclical subset has been proposed as such a measure of how close a set of choice data is to satisfying

[4] The Floyd-Warshall algorithm performs this test in a computationally efficient manner.

[5] Off the shelf algorithms for solving MSCP are included in many software packages that perform optimization (such as Matlab). More powerful solvers are available for free over the Internet (such as SCIP, GLPK and MINTO) or are available commercially (such as CPLEX).

[6] A usable version of our algorithm is available at www.————.com.

rationality (Koo [1963], Houtman and Maks [1985]).[7] Our algorithm increases the size of data set for which the size of maximal acyclical subset can practically be found, and allows for the benchmarking of this measure against simulated random choice data, as proposed by Bronars [1987].[8]

2. *Calculating the minimum number of rationales or 'types'.* Kalai, Rubenstein and Spiegler [2002] introduce the idea of using multiple 'rationales', or preference orderings to rationalize choice data.[9] Of particular interest is the minimum multiple rationales (MMR) for a data set, or the minimum number of preference orderings necessary to rationalize a data set. Calculating the MMR measure for a data set is also an NP-hard problem. Our algorithm provides an efficient way of calculating an upper bound for the MMR.

3. *Identifying choice sets on which to estimate demand correspondences and recover utility functions.* Some nonparametric methods of estimating demand correspondences and recovering utility functions (such as those used in Varian [1982, 1983]) are only valid when applied to data that satisfies acyclicality. A maximal acyclical subset is therefore the largest subset of data set to which such approaches can be applied, so it is a compelling choice for welfare analysis.[10]

The flexibility of our approach to solving MASP also allows us to develop a new measure of how close a particular data set is to rationality. Existing measures focus either on counting the number of violations of rationality exhibited by a subject (Houtman and Maks [1985]), or by calculating the severity of these violations (usually in monetary terms, such as in Afriat

[7] Equivalently, the smallest number of observations that must be removed in order for a data set to be acyclic.

[8] See Choi et al. [2006, 2007], Fisman et al. [2007] and Famulari [1995] for cases where computational constraints have been binding.

[9] A similar approach has been proposed for household consumption decisions, on the basis that there may be several types of households in a population (see Crawford and Pendakur [2008]) or because households consist of different members who may have different preferences (see Cherchye et al. [2007], Deb [2008], Nobibon et al. [2009]).

[10] This approach is pursued on the production side by Banker and Maindiratta [1988].

[1972] and Varian [1990]). In section 4, we propose a new measure that bridges the gap between these two approaches by modifying the Houtman and Maks (HM) index to allow for heterogeneity in the strength of revealed preference observations. Instead of finding the maximal acyclical subset, we find the 'minimum cost' way of constructing an acyclical set, with the cost of removing an observation related to the strength of the related revealed preference. This addresses an important criticism of the Houtman and Maks approach – that a data set can be classified as very irrational even though it contains a large number of acyclical choices caused by the subject being approximately indifferent between the available alternatives.

In section 5, we use our algorithm to analyze laboratory and field data. We begin by justifying the central claim of our paper – that our algorithm is *in practice* much better at calculating the size of the maximal acyclical set than existing methods. We use three different test environments: the choice data from the laboratory experiments of Choi et al. [2007], UK household consumption data used from Blundell, Browning and Crawford [2003, 2008], and simulated data based on the Choi et al. [2007] environment. We show that, across these environments, out algorithm is able to cope with data sets between 7 and 20 times greater than the next best alternative. As the complexity of these problems increases exponentially with the size of the data set, this means that our algorithm is orders of magnitude more powerful than those currently in use. Moreover, this gap will only increase as MSCP solvers improve in performance.[11]

Our algorithm allows us to calculate various measures of rationality which, until now, have been insolvable for many choice data sets of interest. In section 5 we also report the results of these analyses. We describe the Houtman and Maks (HM) scores for the data sets mentioned above, and how they compare to the random benchmarking. We also report results for our new measure of rationality and for the MMR. Notably, the number of rationales needed for any of our data sets is low, even in the case of random data. Finally, we report the correlations

[11]For ease of use, we selected the callable C library of GLPK as our solver. The performance of our algorithm could be substantially improved by using a stand-alone solver program. For more details, see appendix 1.3.

of various measures in our observed populations. On the experimental data of Choi et al. [2007] we find this correlation to be high – over 75% for all pairwise comparisons. We can; however, identify individuals who perform well according to one measure of rationality, but poorly relative to another.

2 Solving the Maximal Acyclical Set Problem

In this section we formally define MASP, and describe our algorithm for solving the problem.[12] In doing so, we present the key theoretical result in the paper. A MASP can be solved by identifying and solving a related minimum set covering problem – a class of problems which is well studied within the field of operations research. This is both good news and bad news for solving incidences of MASP. The bad news is that MSCP (and so MASP) are *NP-hard.* Appendix 1.2 discusses this concept in more detail, but the implication is that there is no known method that can guarantee solution times will increase only as a polynomial function of the number of inputs to MASP. In other words, one cannot guarantee that solution times will not get very large, very quickly as the size of the input data grows. However, the good news is that researchers in operations research have developed a number of tools that *in practice* solve MSCP quickly and exactly for many data sets. The equivalence we show between MSCP and MASP means that these techniques can be adopted wholesale for solving MASP.

2.1 Definitions

MASP is the problem of finding the size of the largest subset of a set of choice data that generates acyclical revealed preference relations. The primitives of the problem are a grand set of *alternatives* Z, a set of *observations* X and a *relation function* $D : X \rightarrow 2^{Z \times Z}$ that characterizes a set of binary relations on Z generated by each observation in X. We call the

[12] An implementation of this algorithm is available at www.———-.com/software/.

triple $\{Z, X, D\}$ a *data set*.[13]

As an example, consider the case of a laboratory experiment in which we observe a subject making choices from subsets of Z. Furthermore, assume that we are prepared to say that the chosen object in any set is strictly preferred to all the other available alternatives.[14] In this case we could think of each observation in X as consisting of a tuple (z, A) with $z \in A$ and $A \in 2^Z/\emptyset$., implying that alternative z has been observed as being chosen from the set of alternatives A. The function $D : X \rightarrow 2^{Z \times Z}$ would then be defined as the revealed preference relations generated by X:

$$\begin{aligned} \forall\ (z, A) &\in X, \\ D(z, A) &= \{(z, y) \mid y \in A/\{z\}\} \end{aligned}$$

We denote by the binary relation $\succ_x \subset Z \times Z$ the relations generated by the observation $x \in X$, so that $\succ_x = D(x)$. For any $B \subset X$, we define the binary relation $\succ^B$ on Z as

$$z \succ^B w \text{ if, for some } x \in B,\ z \succ_x w.$$

For an arbitrary binary relation $\succ$, a *cycle* refers to a set of alternatives $z_1, z_2, ..., z_n \in Z$ such that $z_1 \succ z_2 \succ .... \succ z_n \succ z_1$. We say that a set of observations $B \subset X$ is *acyclic* if the binary relation $\succ^B$ generated by B contains no cycles. Thus, we define MASP as the problem of finding the size of the largest subset $B \subset X$ such that the resulting binary relation $\succ^B$ is acyclic.

[13] We assume that X is finite. In this case, we can solve MASP whatever the cardinality of Z. Moreover, acyclicality is enough to guarantee that choices can be rationalized by utility maximization even if Z is uncountable. This is because we can concentrate on the (finite) set of objects $\overline{Z}$ that are chosen in any observation X. If the data is acyclic, we can generate a utility function $u : \overline{Z} \rightarrow \mathbb{R}$ that rationalizes choice between these objects. All remaining alternatives can be assumed to have a utility equal to $\min_{z \in \overline{Z}} u(z) - 1$.

[14] Note that this assumption is *not* central to our methodology, which is flexible enough to cope with almost any definition of revealed preference. For example, if we are observing choices from budget sets, we could say that x is revealed preferred to y only if x was chosen when y was available at strictly lower cost (per the Generalized Axiom of Revealed Preference). The precise assumption is defined by the nature of the mapping D.

Definition 1 *The maximal acyclical set problem (MASP) for a data set* $\{Z, X, D\}$ *is the problem of finding the size of a set* $B \subset X$ *such that*

$$\text{(i) } B \text{ is acyclic}$$

$$\text{(ii) if } B' \subset X \text{ and } |B'| > |B| \text{, then } B' \text{ is not acyclic}$$

In other words, MASP is the problem of finding the size of the largest acyclical subsets of X. Note that the maximal acyclical set may not be unique.

Next, we define a MSCP. In order to explain the idea behind this class of problems, we illustrate it with the following example:

Example 1 *Imagine you are setting up a cellphone network and need to buy rights to bandwidth in all 50 states. However, bandwidth is being sold in packages of different states (e.g. package 1 includes Alabama, Rhode Island and Wyoming, package 2 includes South Dakota, Minnesota and Wyoming and so on). Each package has a particular cost. The problem you face as a cellphone provider is: "What collection of packages should I buy to ensure some bandwidth in all 50 states at the lowest possible cost?" In other words, what is the minimum cost way of covering all 50 states?*

A formal statement of this class of problems is as follows:[15]

Definition 2 *Let* S *be a (finite) set ,* $\Theta \subset 2^S$ *be a collection of subsets of* S *and* $k : S \rightarrow \mathbb{R}$ *be a cost function which attaches a cost to each element of* S*. A covering of* Θ *is a subset* $T \subset S$ *such that* $\theta \cap T \neq \varnothing \ \forall \ \theta \in \Theta$*. In other words, every set in* Θ *contains at least one element of* T*. A minimum set covering problem (MSCP) is the problem of finding the minimum cost of covering of* Θ*, or*

$$\min_{T \in 2^S} \sum_{s \in T} k(s)$$

$$\text{subject to } \theta \cap T \neq \varnothing \ \forall \ \theta \in \Theta$$

[15] Note that some people call the problem stated in this way as the 'minimum hitting problem'. However, Ausiello et al. [1980] show that this is equivalent to other statements of the minimum set covering problem.

Again, note that the minimum covering set may not be unique.

In the bandwidth example above, we can let S be the set of packages and Θ be a collection of 50 sets, one for each state, containing the packages which cover each state (e.g. if Alabama was covered by packages 1, 7 and 9 then $\theta_1 = \{1, 7, 9\}$, if Alaska was covered by packages 3, 14, 19 and 23 then $\theta_2 = \{3, 14, 19, 23\}$ and so on). k would contain information on the cost of each package.

2.2 Equivalence of MASP and MSCP

In order to show the equivalence of MASP and MSCP, we need to formalize the concept of the set of cycles generated by a data set.

Definition 3 *A* ***cycle*** *generated by a data set* $\{X, Z, D\}$ *consists of a non-repeating sequence* $z_1, ..., z_n$ *in* Z *and a sequence* $x_1, ..., x_n$ *in* X *such that*

$$z_1 \succ_{x_1} \ldots \succ_{x_{n-1}} z_n \succ_{x_n} z_1$$

Let C denote the set of all cycles generated by X.

We will say that observation $x \in X$ *breaks* a cycle $c \in C$ if x appears in the sequence $x_1, ..., x_n$. Note that if a subset B of X breaks all cycles in C, then the complement of that subset X/B is acyclic.

Next, we will define the components of a *complimentary* MSCP for a particular MASP in the following way.

Definition 4 *For the MASP associated with a data set* $\{X, Z, D\}$*, we define the complimentary minimum set covering problem by the following elements* $\bar{S}$*,* $\bar{\Theta}$ *and* $\bar{k}$*:*

1. $\bar{S} = X$

2. $\bar{k}(x) = 1 \ \forall \ x \in S$

3. $\bar{\Theta} = \{\overline{\theta}(c) \mid c \in C\}$, *where* $\overline{\theta}(c) = \{x \in X \mid x \text{ breaks } c\}$

With this structure, the problem of finding the size of the smallest subset of X which breaks all cycles in C is the same as finding the cost of the minimum covering of Θ. Further, a smallest subset of X which breaks all cycles in C is the complement of a largest subset of X that is acyclic.

Theorem 1 *For the MASP associated with a data set* $\{X, Z, D\}$*, if some number* s *is the solution to the complimentary MSCP* $(\bar{S}, \bar{\Theta}, \bar{k})$*, then* $|X| - s$ *is the solution to the MASP.*

Proof. See appendix 1.1. ■

Thus any MASP can be solved by solving the equivalent MSCP. While the MSCP is NP-hard, these problems have been studied exhaustively in the operations research literature because they can be applied to many real world situations, such as train scheduling and city planning. As a result, algorithms have been developed to solve or approximate solutions to MSCP quickly for larger and larger data sets. In appendix 1.3 we describe in detail the steps required to solve a MASP using this equivalence. We also provide links to software that automates this process.

3 Economic Applications of the Maximal Acyclical Set Problem

MASP is of interest because of its relation to the classic assumptions of economic rationality. Acyclicality is a necessary and sufficient property for a binary relation to have a complete extension that is a linear order.[16] This means that a binary relation that is acyclic can be thought of as being derived from a complete *preference relation*.[17] Such a binary relation can be represented (in the case of countable number of objects) by a utility function. MASP

[16] A linear order is a binary relation that is complete, transitive, antisymmetric and reflexive.

[17] A relation that is complete, transitive and reflexive.

therefore identifies the size of the largest subset of a choice data set that is consistent with utility maximization.

In this section we discuss three economic applications of MASP.

3.1 Measures of Rationality

Acyclicality offers a very stark measure of whether a particular data set is consistent with rationality: either acyclicality is satisfied or it is not. There is no concept of how 'close' a particular data set is to satisfying this condition. In practice, almost all real life choice data exhibit *some* violation of acyclicality, meaning that such a stark measure is not particularly useful in comparing behavior across individuals or decision making circumstances.[18] Instead, it is useful to have some measure as to whether one data set is closer to rationality than another under according to some definition. Such a measure could be used, for example, to test the class of 'coherent arbitrariness' models discussed by Ariely et al. [2003] and Rubinstein and Salant [2008], which propose that choices may appear more rational (or coherent) for choice problems that are framed the same way, than between choices framed in different ways.

One such measure of rationality was proposed by Houtman and Maks [1985]: the size of the largest subset of choice observations that satisfy acyclicality (henceforth the HM index).

Imagine a choice experiment in which subject A exhibits the following behavior:

$$\begin{aligned}
C_A(\{x,y\}) &= \{x\} \\
C_A(\{x,y,z\}) &= \{z\} \\
C_A(\{x,z\}) &= \{z\} \\
C_A(\{y,z\}) &= \{y\} \\
C_A(\{x,y,w\}) &= \{w\}
\end{aligned}$$

If we assume that choice is synonymous with (strict) revealed preference, then these data are

[18] For example, Choi et al. [2007] report that 83% of subjects violated GARP and Cherchye et al. [2008] find that 55% of households fail the (unitary) test of GARP.

not consistent with acyclicality, as x is revealed preferred to y, while y is revealed preferred to z, which is in turn revealed preferred to x. However, if one were to remove the observation $C_A(\{y,z\}) = \{y\}$, then the resulting system would be consistent with acyclicality.

Now imagine that subject B exhibits the following behavior:

$$\begin{aligned} C_B(\{x,y\}) &= \{x\} \\ C_B(\{x,y,z\}) &= \{z\} \\ C_B(\{x,z\}) &= \{z\} \\ C_B(\{y,z\}) &= \{y\} \\ C_B(\{x,y,w\}) &= \{y\} \end{aligned}$$

This data set is also not consistent with acyclicality. However, in this case one would have to remove *two* observations before subject B's choices are consistent with acyclicality. In this sense, subject B could be described as *less rational* than subject A. This, in essence, is the meaning of the HM index: The HM index of subject A is 4, while for subject B it is 3.

While this measure is not without its problems (see section 4 for a discussion), it has the advantage of being applicable to wide variety of data sets: the HM index can in principle be applied to choices from any arbitrary choice set. For example, the HM index can be used to test for consistency in choices from dynamic and discrete environments with foods and colors (Echenique, Wilson, and Yariv [2009]) or numeric values (Caplin, Dean, and Martin [2010]). In contrast, the rationality measures of Afriat [1972] and Varian [1991] are only applicable to data obtained by observing choices derived from different budget sets.

Formally, we define the HM index using the concept of the maximal acyclical set defined in section 2.1 above.

Definition 5 *The HM index for a data set $\{Z,X,D\}$ is a number M such that $M = |A|$, where A is a maximal acyclical set of that data set.*[19]

It is obvious from this definition that the problem of calculating the HM index is identical

[19] The HM index can be normalized by dividing it by the total number of observations in a data set.

to MASP. This means both that calculating the index is NP-hard and that the algorithm developed in section 2 can be applied directly to calculating it.

One issue with rationality measures such as the HM index is that it can be hard to interpret what a particular value tells us about the underlying data. For example, consider a data set in which we observe choices from two disjoint choice sets. In this case the (normalized) HM index will be 1 for *any* observed pattern of choice. In other words, such a data set offers no meaningful test of rationality. One way to address this shortcoming is to compare the values of our chosen index to the distribution of values we would see under some alternative 'null hypothesis' for behavior. Such a comparison allows one to determine whether observed behavior shows more, less or similar levels of rationality than the null hypothesis.

One popular benchmark is to compare index values to those that one would expect to see under random choice – in each choice set individuals have an equal chance of choosing any object in the choice set.[20] Although random choice is a relatively weak null hypothesis, it is applicable to almost any choice setting. The role of random choice in determining the statistical power of rationality measures is discussed by Bronars [1987] and is applied to Selten's measure of predictive success by Beatty and Crawford [2010].[21]

One advantage of our algorithm is that it makes this type of benchmarking feasible. In almost all practical cases, benchmarking against random choice has previously proved too computationally intensive (see Choi et al. [2007]).

Once we have generated a benchmark, the next question is how to compare the experimental data to this benchmark. For a joint test of all subjects, one can compare the *distribution* of the index scores in the data with the *distribution* of index scores generated under the null hypothesis using some nonparametric measure of the difference between distributions (such as the Kolmogorov-Smirnoff test). In the case of a single observation, one can simply read off the percentile of the simulated data in which that observation falls. Another intuitive

[20] Or, in the case of budget sets, an equal chance of choosing any object on the budget line.

[21] Alternatively, we could generate a distribution of possible index values for a given choice environment using a more plausible error model or decision rule. For example, see Choi et al. [2006] and Andreoni and Harbaugh [2006].

measure is to subtract the average simulated score from an actual score – in the style of Selten's measure of predictive success [1991]. For the HM index, the resulting number would represent the fraction of a data set consistent with rationality over and above what could be explained by random choice. We refer to this measure, which combines elements of predictive power and statistical power tests, as the 'Selten-Bronars' measure.

3.2 Minimal Multiple Rationales

If a data set is cyclical, it cannot be explained perfectly by a single preference order, or rationale. However, it may be that agents have different rationales for different states. For example, the relative ranking of an umbrella and a bicycle may differ depending on whether it is raining or not. If we do not observe these different states, then the resulting choices may appear irrational. This notion was captured by Kalai, Rubinstein, and Spiegler [2002], who introduced the concept of *rationalization by multiple rationales.* A rationale is a preference ordering, and a data set is rationalized by a collection of rationales if all observations are explicable as the maximization of one of the rationales. Thus, choice data that can be rationalized by n rationales can be thought of as being generated by an individual who at any time is in one of n different 'states' and in each state has a different set of preferences. Such an approach has also been applied to the analysis of household-level data, to determine if household choices can be rationalized as preference maximization by one of the members of the household (Deb [2008], Nobibon et al. [2009]) and to determine if households are heterogeneous (Crawford and Pendakur [2008]).

Formally, we define the concept of rationalizing by multiple rationales as follows:

Definition 6 *For a data set $\{Z, X, D\}$, a rationalization by multiple rationales is a set of complete preference relations*[22] *R on Z such that, for every $x \in X$, there exists $\trianglerighteq \in R$ such that $\trianglerighteq$ is an extension of $D(x)$.*

In other words, a rationalization by multiple rationales is a collection of preference rela-

[22]i.e. complete, transitive and reflexive binary relations.

tions such that each observation $x \in X$ can be explained by one of these rationales.

Definition 7 *A minimal multiple rationales (MMR) is a rationalization by multiple rationales such that no smaller collection of rationales could rationalize the data.*

Appendix 1.4 provides examples that illustrate the difference between the concept of MMR and the HM index.

The problem of calculating the MMR for a data set is equivalent to the problem of finding a partition of the observation set X such that:

1. The observations in each set of the partition are acyclic.
2. Any partition of X that is composed of fewer divisions has at least one division that is not acyclic.

Solving this problem for incomplete data has been shown to be NP-hard by both Apesteguia and Ballester [2010] and Deb [2008]. These papers draw on graph theory to suggest methods for finding or providing approximations to the size of the MMR. Crawford and Pendakur [2008] present a method for placing a lower and upper bound on the size of the MMR by iteratively testing randomly generated partitions, and Cherchye et al. [2007] present the corresponding IP conditions for an analogous problem.

The techniques that we have developed for solving MASP offer a new way of calculating an upper bound on the size of the minimal MMR. By iteratively applying MASP, we can find a partition of the observations such that each set in the partition is acyclic. While this is not guaranteed to be the *minimal* partition, in practice this technique (a form of 'greedy' algorithm) provides a better approximation then the methods of Deb [2008], as we show in section 5.

A formal description of the algorithm is as follows:

Algorithm 1 *The following algorithm calculates an upper bound for the size of the minimal MMR. Let $X_1 = X$, the original set of observations, and let $i = 1$.*

1. *Solve MASP for* Z, X_i *and* $D : X_i \to 2^{Z \times Z}$

2. *Select a maximal acyclical set from the above solution, and set* P_i *equal to that subset of* X_i

3. *If* $P_i = X_i$ *continue, otherwise set* $X_{i+1} = X_i / P_i$ *and return to step 1*

4. *Let* $P = \{P_1, ..., P_I\}$

The collection of sets P is a partition of X and each P_i is acyclic. Thus, the cardinality of P is an upper bound on the size of the minimal MMR.

Clearly, if this algorithm returns a value of 1 or 2, then it has found the size of the minimum MMR for the data set. Thus, it is sufficient to answer the question of whether a household of size 2 can rationalize the data (see Nobibon et al. [2009]). However, this algorithm is not guaranteed to find the exact size of the minimal MMR for more than 2 partitions.

3.3 Estimating Demand Correspondences and Recovering Utility Functions

To determine the size of the maximal acyclical subsets, our algorithm finds a single maximal acyclical subset. This subset is a natural choice for conducting welfare analyses that require acyclicity because it contains the largest possible amount of acyclical revealed preference information.

For example, the nonparametric method of placing bounds on demand correspondences developed by Varian [1982] requires acyclic data because observed choices must satisfy GARP for the correspondence to be well-defined. Because individual consumer choices are often cyclical, economists have used aggregate data to perform these estimations (Varian [1982]). Another approach is to generate bundles that are acyclical for a subset of years by estimating Engel curves for each year (Blundell, Browning and Crawford [2003]). Our algorithm allows for a new approach, which is to perform these types of analyses on one or more maximal acyclical subsets (Martin [2010], Suarez Serrato [2010]).

Importantly, there may be a multiplicity of maximal acyclical sets for a choice data set. It may be sufficient to present a distribution of analytical results, where each maximal acyclical set contributes a single data point, but if a unique result is desired, a single result can be taken from a location in the distribution or generated by a summary statistic over the distribution. Alternatively, a single maximal acyclical set can be identified with desirable properties (such as demographic balance) or maximal acyclical sets can be aggregated in some way. For example, if we take the intersection of all maximal acyclical sets, then we have an "unambiguous choice relation" in the style of Bernheim and Rangel [2008].

4 A Modification of the HM Index

Apart from computational complexity, the HM index has other weaknesses as a measure of rationality, as discussed in Choi et al. [2006]. We therefore introduce an alternative measure which deals with two common criticisms of the HM index.

The first of these problems can be illustrated by the following example. Consider first choice data from subject C, which consists of the following observations:

$$\begin{aligned} C_C(\{x,y\}) &= \{x\} \\ C_C(\{z,y\}) &= \{y\} \\ C_C(\{w,y\}) &= \{y\} \\ C_C(\{x,y,z,w\}) &= \{y\} \end{aligned}$$

While choice data from subject D has the following observations:

$$\begin{aligned} C_D(\{x,y\}) &= \{x\} \\ C_D(\{z,y\}) &= \{z\} \\ C_D(\{w,y\}) &= \{w\} \\ C_D(\{x,y,z,w\}) &= \{y\} \end{aligned}$$

Both subjects have an HM index of 3. However, there is a clear sense in which we might think of subject D as being 'less rational' than C. In C, the only irrationality is caused by x being chosen from the set $\{x, y\}$, while y is chosen over x from $\{x, y, z, w\}$. In other words, there is only one piece of conflicting revealed preference information. However, in D the problem is caused by y being chosen from $\{x, y, z, w\}$, even though it is not chosen when y is paired individually with x, z or w. Thus we have three pieces of conflicting revealed preference information: According to the pairwise choices, subject D chose the *worst* alternative from $\{x, y, z, w\}$. In this sense, the selection of y from $\{x, y, z, w\}$ is a bigger departure from the other choices by D than the selection of y from $\{x, y, z, w\}$ in C.

A second problem with the HM index is that it does not contain any notion of the 'severity' of a particular violation of acyclicality. This is most obvious in the case in which the observed choices are over bundles of commodities from different budget sets. Consider the following choice behavior for hypothetical subjects E and F from budget sets in a commodity space that contains two goods (a and b):

- Budget set 1 : income is 10, price of good A is 2, price of good B is 2
  - E buys 1 unit of good A and 4 units of good B
  - F buys 2 units of good A and 3 units of good B
- Budget set 2 : income is 10, price of good A is 3, price of good B is 1
  - E buys 3 unit of good A and 1 unit of good B
  - F buys 3 unit of good A and 1 units of good B

Figure 1 illustrates the choices of these two subjects.

FIGURE 1 ABOUT HERE

Both of these consumers violate acyclicality, as in both cases the bundle bought in budget set 2 was available in budget set 1, and vice versa. However, the 'cost'[23] of the acyclicality

[23] Here cost can be thought of as a potential 'money pump' or as 'wasted' income.

violation for subject E was higher than that for subject F. For E, the bundle chosen from budget set 1 was available at a cost of 7 from budget set 2, while for consumer F, the bundle chosen from set 1 was available at a cost of 9. Thus the 'cost' of the acyclicality violation for E is 3, while for F it is only 1. Yet both subjects would have the same HM index. Cost measures of this type are the basis of another group of tests of rationality, including those of Afriat [1972] and Varian [1991].

More generally, a researcher may have some form of external metric on how different are a pair of objects or how strong is a particular revealed preference relation. A desirable property in a measure of the rationality of a particular data set is that it can take into account differences in the 'strength' of a particular revealed preference relation – punishing cycles that involve only 'strong' relations more than cycles that involve only 'weak' relations. The HM index has no way of incorporating this information into its measure of irrationality.

In the above example, a natural metric for the strength of preference exhibited when x is chosen over y could be the cost difference between x and y measured in terms of some denominator good. In other cases we might have some intuition that two objects are similar (a weak relation that is easy to break), or that they are very different (a strong relation that is hard to break).

In order to address these two shortcomings in the HM index we propose a modified measure. This modified index differs from the standard HM index in two key respects:

1. As the basic unit of analysis, we use each revealed preference relation, rather than each observed choice.

2. We allow for the varying 'costs' of removing different revealed preference relations, depending on an external metric for the strength of each preference.

To understand the impact of the first change, consider the case of object y being chosen from a choice set $\{x, y, z, w\}$. The HM index treats this as a single observation, and asks only whether or not this observation needs to be removed in order to guarantee acyclicality. Our modified index treats this as three separate pieces of information: that y is preferred to x, that y is preferred to z and that y is preferred to w. By doing so, it would treat the choices of subject D as less rational than those of subject C.

The second change allows for the measure to be modified to account for the strength of the revealed preference relations. Formally, the primitives of our modified index are a data set $\{Z, X, D\}$ and a weighting function $w : D(X) \to \mathbb{R}$. The weighting function carries information on the strength of different revealed preference relations. Note that we do not propose any particular metric – the weighting function forms part of the input to the measure. However, as we discussed above, in the case of choice of budget sets, one natural measure would be the cost difference between the chosen and unchosen bundle. Such a weighting function would bridge the gap between rationality measures that only count the *number* of violations of rationality (Famulari's [1995] measure and the HM index) and those that look only at the *cost* of such violations (Afriat [1972] and Varian [1991]).[24]

Given these primitives, the modified HM index we propose is defined as follows:

Definition 8 *For a data set $\{Z, X, D\}$ and a weighting function $w : D(X) \to \mathbb{R}$, the modified HM index is defined as*

$$W = \min_{B \subset D(X)} \sum_{a \in B} w(a)$$

such that $D(X)/B$ is acyclic.

In other words, the modified HM index is the minimum cost way to create an acyclical set of preference relations according to the weighting function w.

Note that it is only the relationship between MASP and MSCP that makes our modified index easy to calculate. Because MSCP has a cost function embedded in it, the steps outlined in section 2 can be extended to this measure, allowing us to use IP solvers to calculate it as well. Alternative algorithms, such as Houtman [1995] cannot be used to calculate the modified index because they require that all nodes and edges have the same cost.

[24] For an example of this implementation, see section 5.2.

4.0.1 Relation to Apesteguia and Ballester [2010]

In a recent paper, Apesteguia and Ballester [2010] suggest a set of desirable properties for measures of rationality. Their basic unit of analysis is an index that indicates how close a *particular* preference relation is to rationalizing a data set, whereas the HM measure indicates how close *any* preference relation is to rationalizing a data set. They consider 5 desirable axioms for such an index to have, and provide a representation theorem for measures that satisfy these axioms.

The HM index can be altered to consider a single reference preference relation, which will allow us discuss their axioms in relation to this measure. For a data set $\{Z, X, D\}$ and preference relation P on Z, we can define the following index:

$$I(P, \{Z, X, D\}) = |x \in X \text{ such that } P \text{ is an extension of } D(x)|.$$

$I(P, \{Z, X, D\})$ counts the number of observations in X which are consistent with the preference relation P. We can think of the HM index as maximizing the value of $I(P, \{Z, X, D\})$ across all preference relations on Z. An equivalent index can be defined for the modified HM index.

Having transformed the HM index to the framework of Apesteguia and Ballester [2010], we can ask whether or not $I(P, \{Z, X, D\})$ satisfies their 5 properties. In fact, for both the HM and modified HM indices, four of the five properties are satisfied: it is invariant to permutations, it has full value only if P rationalizes all observations, it obeys a version of independence[25] and a version of the Archimedean axiom. However, neither satisfy what Apesteguia and Ballester [2010] term the Composition condition. This requires that, if for two sets A,$B \in Z$, such that everything in A is preferred to everything in B according to P, $I(P, \{Z, X, D\})$ has to take the same value for the following two data sets:

1. x is chosen from A and y is chosen from $B \cup \{x\}$

[25] If, for two sets of observations B and B', the $I(P)$ index is higher for B than for B' then this is also true for $B \cup B''$ and $B' \cup B''$ for any additional observations B''.

2. y is chosen from $A \cup B$

Our measures could violate this axiom because the HM approach could find two violations in the first data set and just one in the second data set and the modified HM approach could find more revealed preference violations in the second data set than the first.

In fact, our algorithm cannot be used to find the preference relation that minimizes rationality violations for indices of this form. For such indices, the cost of deleting a particular observation depends on the preference relation being considered. For our algorithm to work, the cost of deleting a particular observation must be predetermined.

5 Data Analysis

5.1 Measuring the Performance of the MSCP-Based Algorithm

The central claim of our paper is that our algorithm based on MSCP is, in practice, much better than existing methods of solving MASP. In this section we demonstrate this claim. In order to do so, we compare our algorithm to two existing methods for solving MASP. Our two comparators are the methodology applied by Choi et al. [2007] (henceforth CGFK) and Houtman [1995].[26]

A priori there is reason to expect our algorithm will outperform our two comparators. CGFK use an algorithm that examines separately each strongly connected component[27] of a directed graph to find the minimum number of removals for a data set to satisfy acyclicality. However, after a few initial steps, their algorithm employs brute force to remove observations one by one, which is typically very computationally expensive in data sets with dense strongly connected components. Alternately, Houtman [1995] extends the algorithm developed by Guardabassi [1971], which has the advantage that it does not require cycles to be found in

[26] We do not compare ourselves to the algorithm developed by Gross and Kaiser [1996], as it only applies to undirected graphs.

[27] The largest set of nodes such that every node can be reached by a path of directed edges from every other node in the set.

the initial step. While this approach works very well for small data sets, Smith and Walford [1975] report that Guardabassi's algorithm is "practical only for small graphs (circuits)."

We use three different environments to benchmark our data. The first is the experimental data of CGFK. The second is simulated data set based on the characteristics of CGFK, which we use to 'destruction test' the three algorithms. The third is consumer choice data from the British Family Expenditure Survey, as used in Blundell, Browning and Crawford [2003, 2008].

When comparing the performance of different algorithms, we use a standard set of computing resources. All programs were run on a desktop computer with a 2.8 GHz dual core processor, 8 GB of RAM and the Windows 7 operating system. The IP solver we use is the callable C library of GLPK.

5.1.1 Environment 1: The Experimental Data of Choi, Gale, Fisman, and Kariv [2007]

The first data we use to test our algorithm comes from the experimental data reported in CGFK. In these experiments, subjects make repeated choices between risky alternatives using a novel graphical interface. Subjects make choices over bundles of units of asset x and units of asset y, with their choice set described by a randomly generated budget line. The state of the world is uncertain, and in one state of the world, a unit of asset x pays \$0.50 while asset y pays nothing, and in the other state of the world, the reverse is true. Each subject makes 50 choices, and CGFK collect data for 93 subjects. The authors selected the number of choices as the number necessary to provide a reasonably powerful test of rationality.

In order to use the technology we have so far introduced in this paper to examine the data, we have to identify for each subject the set of choice objects Z, the set of observations X and the set of revealed preference relations $D : X \to 2^{Z \times Z}$. Initially, it seems that identifying the set of alternatives is problematic, as each budget set contains an infinite number of alternatives, making the size of set Z infinite. However, we can simplify this problem by noting that the only objects that can be involved in a violation of irrationality are those that have at one time been chosen. Thus, without any loss of generality, for any individual subject we can set Z to be equal to the set of bundles that they have chosen from *some* budget set, meaning that $|Z| \leq 50$. Then X simply contains the 50 choice observations for that subject, and $D(z, A) = \{(z, y) \mid y \in A / \{z\}\}$ for $\forall$ $(z, A) \in X$.

TABLE 1 ABOUT HERE

CGFK attempt to calculate the HM index for all 93 of their subjects. However, they report being unable to do so for 6 of their subjects. It is these 6 'problem cases' that we focus on here. Table 1 reports the time taken to find solve MASP for these subjects.[28] Unsurprisingly, the CGFK algorithm struggles – for 4 subjects it fails to solve MASP within an hour,[29] while for the other two subjects it solves the problem within about 120 seconds. However, both Houtman's algorithm and our MSCP-based algorithm solve MASP for these subjects almost instantaneously.

5.1.2 Environment 2: Simulated Choice Data

In order to differentiate between the power of our algorithm and that of Houtman, we use simulated choice data of increasing size to find the functional limit of each. In order to do so, we create an increasing number of hypothetical choice sets (budget sets) using the same method of the CGFK. We then simulate the choices of a 'random' subject – one that chooses a point on each budget constraint at random. Such subjects are likely to create more irrationality than actual subjects, and therefore provide a harder test for the algorithms. Additionally, random subjects are of independent interest as the form the basis of the benchmarking measures discussed in section 3.1.

FIGURE 2 ABOUT HERE

[28] CGFK also report results for an algorithm that approximates a solution to MASP. However, our results show that this approximation is a poor one, so we do not report its performance here.

[29] The algorithm converged in about 2 hrs for one subject, 13 hrs for another subject, and did not converge within 48 hrs for the remaining two subjects.

Figure 2 reports how long it takes each algorithm to solve MASP for 100 simulated subjects with difference choice set sizes. As mentioned above, these 100 simulations could generate similar algorithm run times as 100 very irrational experimental subjects or as a small benchmark set. The figure clearly shows the set size at which each algorithm breaks down. The CFGK algorithm breaks down in data sets of about 30 choices (this explains why they cannot benchmark their results against random choice for their 50 choice data sets). Houtman's algorithm breaks down for data sets that contain about 65 choices (below the size of other experimental data sets, such as Suarez Serrato [2010]). Our MSCP-based algorithm begins to break down at choice sets of approximately 350. As the complexity of the problem roughly increases in a exponential way with the size of the choice set, this demonstrates the considerable power of the MSCP-based algorithm relative to that of Houtman [1995].

5.1.3 Environment 3: Household Consumption Data

The third data set on which we test our algorithm is the household consumption data set used in Blundell Browning and Crawford [2008] (henceforth BBC). The authors use 25 years of data from the (annual) British Family Expenditure Survey[30] to gather household level consumption data, and the annual Retail Price Index to construct price measures. For each household, consumption is aggregated into three categories: food, other nondurables and expenses. Prices for each of these categories are then calculated using a retail price index, which is the same for all households in a given year. Thus, each household is identified by a 6-tuple describing their consumption of the three good types, and a price for each of these goods.[31] Note that if we assume that strict revealed preference only occurs when one bundle is chosen over another that is strictly cheaper, then acyclicality can only revealed in a data sets that contain budget lines that cross (i.e. intersect at one and only one point). This implies that violations of acyclicality require observations from households with different price vectors, and so different years.

In order to compare the performance of our algorithm to that of Houtman, we take data

[30] A survey that reports consumption data for repeated cross sections of British households.

[31] For details of how these data are constructed, see Blundell, Browning and Crawford [2008].

from 1997-1999, the final three years of the period studied by BBC. For each of these years, we consider households in the 50th-75th percentile in terms of income, giving a total of 1090 observed choices. We focus on households from the same income quartile because they are more likely to have budget lines that cross, increasing the probability of a meaningful test of acyclicality. From these 1090 observed choices, we draw subsets of increasing size, on which we test the ability of our algorithm and Houtman's algorithm to solve MASP. The results are shown in figure 3.

FIGURE 3 ABOUT HERE

Houtman's algorithm breaks down in choice sets of size of 210 or more. In contrast, our MSCP-based algorithm can solve MASP for the entire data set (1090 observations) in roughly 5 seconds. In fact, our algorithm can solve MASP for all observations from the years 1997-1999, for a total of 4,354 observations. In this setting, our algorithm can handle a data set over 20 times larger than the next best algorithm. This differs from the result in environment 2 because the structure of the resulting graph is quite different.

While our computer program fails when a fourth year of data is added, it is not due to the speed of our algorithm. For 1996-1999, the size of the preference graph that the data generates exceeds the available memory in 32 bit computers because although the number of observations is just 5,830, the number of revealed preference relations is 16,990,374.

5.1.4 Calculating Minimal Multiple Rationales

The final way in which we test our algorithm is to examine its ability to approximate the MMR. Here we are interested not in the speed of the algorithm (the computational limits will be similar to its ability to solve MASP), but instead its accuracy. As we discuss in section 3.2, we can use our MASP solver to provide an upper bound to the size of the MMR. We compare this upper bound to that provided by the algorithm proposed by Deb [2008].[32]

[32] We are aware of another approximation technique developed by Crawford and Pendakur [2008]. As yet we have not compared our performance to that of their algorithm.

In both experimental and household consumption data we find our algorithm to provide a tighter upper bound than the algorithm of Deb. In the CGFK data, our approach is strictly better than Deb's in 11 of 93 subjects, and is never strictly worse. In the case of the BBC data, we approximate the MMR for 25 sets of 1,000 observations taken randomly from 1990-1999, and our approach is strictly better than Deb's for 20 sets, and again, is never strictly worse.[33]

5.2 Measuring Rationality

In this section we discuss briefly what the various analyses described in section 3 have to say about our two data sets: the experimental data of CGFK and the household consumption data of BBC. Previously it has not been possible to perform these analyses, so it is interesting to know what these tests actually have to say. We begin by reporting the HM index for each of the data sets and benchmarking these against random choice in the manner of Bronars [1987] and Selten [1991]. Next, we implement our modified version of the HM measure, and discuss how it compares to the standard measure. Finally, we describe our estimates for the MMR.

5.2.1 The HM Index

Figure 4 shows the distribution of the HM index in the 93 subjects in the CGFK data, compared to the distribution generated from 93,000 simulated subjects. For each subject, we simulated 1,000 random choices from the exact budget sets that the subject faced, and then we pooled these simulations.

FIGURE 4 ABOUT HERE

The data from the actual subjects is clearly significantly 'more rational' than that of the simulated subjects. Of the actual subjects, 14 required no removals, and the modal size of

[33] Our approximation is only one rationale tighter in all cases, but the approximate number of rationales needed to rationalize both sets is relatively small.

the maximal acyclical set (MAS) was 47. In contrast, in the simulated data, there are no observations that require no removals, and the modal size of the MAS was 34. The difference between the two sets is confirmed by a Kolmogorov-Smirnoff test ($p < 0.001$).

Despite the fact that the population of subjects is clearly different from random choosers, there are outliers who score poorly on the HM index. The Selten-Bronars measure can help us to identify which subjects did poorly given the budget sets they faced. On the extreme, if budget sets never intersect, then random choice will generate an HM index of 1, so a subject facing the same budget lines will always have a Selten-Bronars score of 0. On the other hand, if budget lines intersect a lot and so random data produces an average HM index of 0.25, then a seemly low HM index of 0.75 could actually produce a high Selten-Bronars score of 0.50 (see section 3.1 for more details)

For this data, the mode Selten-Bronars score was 0.255, which indicates that over 25 percent more of the data set is acyclical than with random choice. Clearly, the modal subject was much closer to rationality than random choice. At the 5th percentile, the score was still 0.1313, but four subjects had Selten-Bronars scores below 0.1 – the minimum was just 0.01802, which means it is almost indistinguishable from random choice.

Tables 2 and 3 show HM indices and associated Selten-Bronars measures (based on 100 simulations) for the household consumption data of BBC for the years 1997 to 1999. Table 2 breaks down the aggregate data by income percentile. Table 3 breaks it down by geographic grouping.

TABLE 2 ABOUT HERE

TABLE3 ABOUT HERE

As table 2 shows, the average HM index across the four quartiles is very close to the overall HM index for all 4,354 subjects pooled together. This indicates that there are few cycles that overlap the different income quartiles, which makes sense given that budget lines are less likely to cross at very different income levels. This result implies that we could separately determine the HM index for each income quartile if a data set proved too large for our algorithm to process or solve. On the other hand, table 3 shows that the average HM index across all regions is much higher than the overall HM index, which indicates that inconsistencies are generated when pooling data across regions. One way to interpret this result is that preferences show more differences across than within regions.

The Selten-Bronars measure provides contrast to the absolute HM index scores. In general, the Selten-Bronars scores indicate that only about 4-6 percent more of the actual data is acyclical than random choice data. Notably, the income groups that appear the most consistent by their HM index alone are the least consistent when we consider how far they are from random choice. The differences are even more varied among regions. For the two regions with the highest HM index scores (0.92), one (Wales) has one of the highest Selten-Bronars scores (0.06), while the other (Southwest) has a one of the lowest (0.03). Meanwhile, the Southeast region, which includes London, has a very low consistency score by the HM index alone (0.81), but one of the highest Selten-Bronars scores (0.06). This indicates that even though there were more cycles in household purchases in London, there were a lot more crossing budget constraints as well.

5.2.2 The Modified HM Index

In section 4 we proposed a modification of the HM index. This modification allowed for the removal of some preference relations to be more 'costly' than others, based on a weighting function that indicated the strength of that relation. Here we calculate this modified index for the CFGK data based on a particular weighting function. The weighting function we use is in the spirit of the rationality measures developed by Afriat [1972] and Varian [1991]. The strength of a preference relation is based on the cost difference between the chosen and not chosen object. For example, if bundle x is chosen when y is available for \$2 less, we give this revealed preference relation a higher weight than if y was available for \$1 less.

Formally, let $x \in X$ be the observation that bundle a was chosen from budget set with price vector p_a, when bundle b was available for a lower total cost. Then

$$w(D(x)) = \frac{p_a a}{p_a b}.$$

As a result, this modified HM index is the smallest possible total efficiency loss from cycles present in the data. Unlike the other efficiency measures, we treat each of the violations that an observation generates as separate efficiency losses.

Figure 5 shows the distribution of our modified index for the 93 subjects in the CGFK data, compared to the distribution generated from 93,000 simulated random subjects.

FIGURE 5 ABOUT HERE

Once again, the data from the actual subjects is clearly significantly 'more rational' than that of the simulated subjects. A Kolmogorov-Smirnoff test ($p < 0.001$) confirms that the distributions are quite different.

TABLE 4 ABOUT HERE

A natural question is how this modified measure compares to existing measures of rationality. Table 4 reports the Spearman Rank Correlation for four measures of rationality, calculated on the subjects in CFGK. These 4 measures are the standard HM index, our modified HM index, the Varian measure[34] and the Afriat measure.[35] Overall, the correlations are quite high – above 0.75 for all measures. Additionally it seems our modified measure is 'inbetween' the purely counting based measure of the HM index and the purely cost based measures of Afriat and Varian, as it has a closer correlation to each type than the other type does.

FIGURE 6 ABOUT HERE

[34] The maximum of the minimum reduction of each budget constraint needed to remove all cycles generated by the corresponding choice.

[35] The minimum reduction of every budget constraint in order to remove all cycles.

What types of subjects appear rational according to the standard HM index but not the modified one, and vice versa? Subject 512 is ranked as the 82nd least rational subject under the HM index but only the 43rd least rational under the modified index. In contrast, subject 603 is only the 40th least rational subject under the HM index but the 91st least rational subject under the modified index. As shown in figure 6, this is due to the fact that violations for subject 512 were very slight, while the violations for subject 603 involve large efficiency losses. Some of these losses stem from subject 603 selecting at the ends of the budget constraint.

5.2.3 Minimal Multiple Rationales

Figure 7 shows our estimated MMRs for the CFGK data, benchmarked against random choices

FIGURE 7 ABOUT HERE

Again, it is clear that the behavior of the CFGK subjects is a long way from their random counterparts (Kolmogorov-Smirnoff $p < 0.001$). Interestingly, the number of rationales needed to explain the CFGK data is low – no subject requires more than 3 rationales. Even in the simulated data with random choices, no subject requires more than 6 rationales. For the BBC data, when we look at 25 sets of 1,000 observations taken randomly from 1990-1999, the number of rationales needed is between 4 and 5 (out of a maximum of 10).

6 Conclusion

In this paper we have introduced a new method for finding the size of the maximal acyclical subset of a set of choice data. This method is orders of magnitude more powerful than existing methods, and opens up important new areas of analyses that can be applied to choice data, both from experiments and from household expenditure surveys.

One critique of the rationality measures which we report is that they are somewhat ad hoc, and do not admit the possibility of formal statistical tests. To some extent we sympathize with this claim. While we believe that these measures offer intuitive criteria for whether a particular data set is 'close to' satisfying rationality (particularly when augmented by tools such as the Selten-Bronars measure), it is clear that a formal testing procedure would be a significant improvement. We see the development of such a test as important future work, and hope that the increased analytical power provided by our algorithm will aid such an effort.

References

Afriat, S. (1972) "Efficiency Estimates of Production Functions," *International Economic Review*, 8, 568-598.

Andreoni, J. and W. Harbaugh (2006) "Power Indices for Revealed Preference Tests," Unpublished.

Apesteguia, J., and M. Ballester (2010) "A Measure of Rationality and Welfare," Unpublished.

Apesteguia, J., and M. Ballester (2009) "The Computational Complexity of Rationalizing Behavior," *Journal of Mathematical Economics*, forthcoming.

Ausiello, G., Marchetti-Spaccamela, A., and M. Protasi (1980) "Toward a Unified Approach for the Classification of NP-Complete Optimization Problems," *Theoretical Computer Science*, 12, 83-96.

Banker, R. D., and A. Maindiratta (1988) "Nonparametric Analysis of Technical and Allocative Efficiencies in Production," *Econometrica*, 56, 1315–1332.

Beatty, T. and I. Crawford (2010) "How Demanding is the Revealed Preference Approach to Demand?," Unpublished.

Bernheim, B. D., and A. Rangel (2008) "Choice-Theoretic Foundations for Behavioral

Welfare Economics." In *The Foundations of Positive and Normative Economics*, ed. Andrew Caplin and Andrew Schotter. New York: Oxford University Press.

Blundell, R., M. Browning, and I. Crawford (2003) "Nonparametric Engel Curves and Revealed Preference," *Econometrica*, 71, 205-240.

Blundell, R., M. Browning, and I. Crawford (2008) "Best Nonparametric Bounds on Demand Responses," *Econometrica*, 76, 1227-1262.

Bronars, S. (1987) "The Power of Nonparametric Tests of Preference Maximization," *Econometrica*, 55, 693-698.

Caplin, A., Dean, M., and D. Martin (2010) "Search and Satisficing," Unpublished.

Caprara, A., Toth, P., and M. Fischetti (2000) "Algorithms for the Set Covering Problem," *Annals of Operations Research*, 98, 352-371.

Cherchye, L., De Rock, B., Sabbe, J., and F. Vermeulen (2008) "Nonparametric Tests of Collectively Rational Consumption Behavior: An Integer Programming Procedure," *Journal of Econometrics*, 147, 258-265.

Choi, S., Gale, D., Fisman, R., and S. Kariv (2006) "Substantive and Procedural Rationality in Decisions under Uncertainty," Unpublished.

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

Crawford, I., and K. Pendakur (2008) "How Many Types Are There? A Simple, Theory-Consistent Approach to Unobserved Heterogeneity." Unpublished.

Deb, R. (2008) "An Efficient Nonparametric Test of the Collective Household Model," Unpublished.

Fisman, R., Kariv, S., and D. Markovits (2007) "Individual Preferences for Giving," *American Economic Review*, 97(2), 153-158.

Echenique, F., A. Wilson, and L. Yariv (2009) "Clearinghouses For Two-Sided Matching:

An Experimental Study," Caltech SS Working Paper 1315.

Garey, M. R. and D. S. Johnson (1979) "Computers and Intractability: A Guide to the Theory of NP-Completeness," New York: W.H. Freeman.

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

Guardabassi, G. (1971) "A Note on Minimal Essential Sets," *IEEE Transactions on Circuit Theory*, 19, 557-560.

Houtman, M. (1995) "Nonparametric Consumer and Producer Analysis," Thesis Rijksuniversiteit Limburg Maastricht.

Houtman, M., and J. A. H. Maks (1985) "Determining all Maximal Data Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden*, 19, 89-104.

Johnson, D. (1975) "Finding all the Elementary Circuits of a Directed Graph," *SIAM Journal of Computing*, 4, 77-84.

Kalai, G., A. Rubinstein, and R. Spiegler (2002) "Rationalizing Choice Functions by Multiple Rationales," *Econometrica*, 70, 2481-2488.

Koo, A. (1963) "An Empirical Test of Revealed Preference Theory," *Econometrica*, 31(4), 646-664

Nobibon, T., Cherchye, B., De Rock, J., Sabbe, J., and F. Spieksma (2009) "Heuristics for Deciding Collectively Rational Consumption Behavior," Unpublished.

Martin, D. (2010) "Cost-of-Living Changes for Core Consumer Groups," Unpublished.

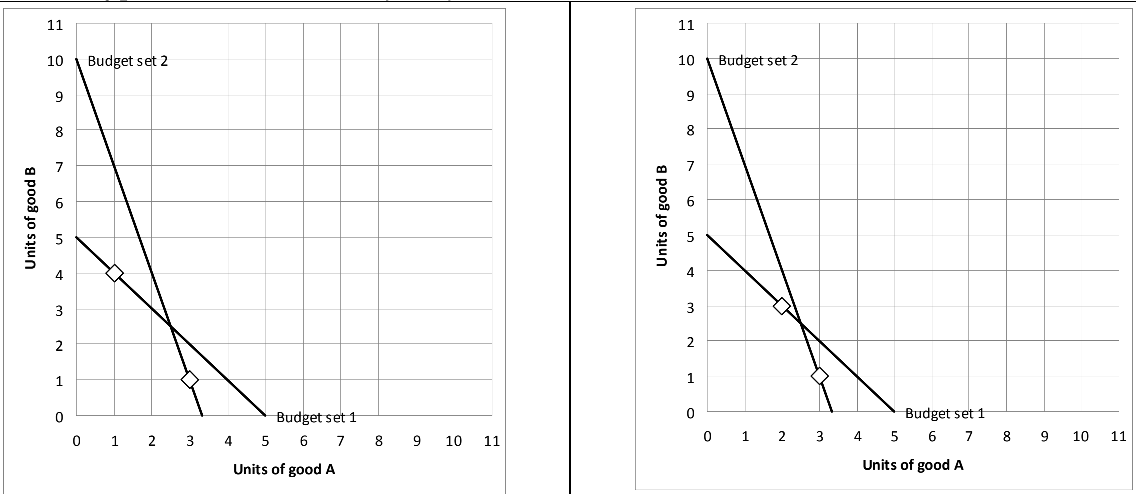
Smith, G., and R. Walford (1975) "The Identification of a Minimal Feedback Vertex Set of a Directed Graph," *IEEE Transactions on Circuits and Systems*, 22, 9-15.

Suarez Serrato, J. C. "An Experimental Exploration of Economic Behavior on Kinked Budget Sets," Unpublished.

Varian, H. (1991) "Goodness-of-Fit for Revealed Preference Tests," University of Michigan CREST Working Paper # 13.

Varian, H. (2006) "Revealed Preference," In *Samuelsonian Economics and the Twenty-First Century.*

Figure 1: Hypothetical Choices of Subjects H and J



Subject E	*Subject F*

Figure 2: Run Times for 100 Simulated Subjects

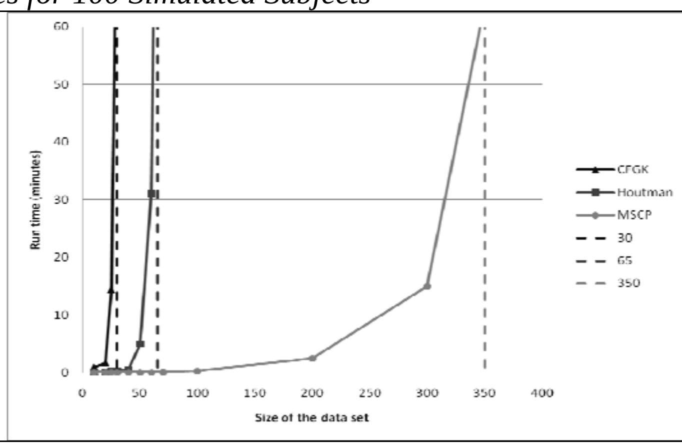


Figure 3: Run Times for Consumer Expenditure Data

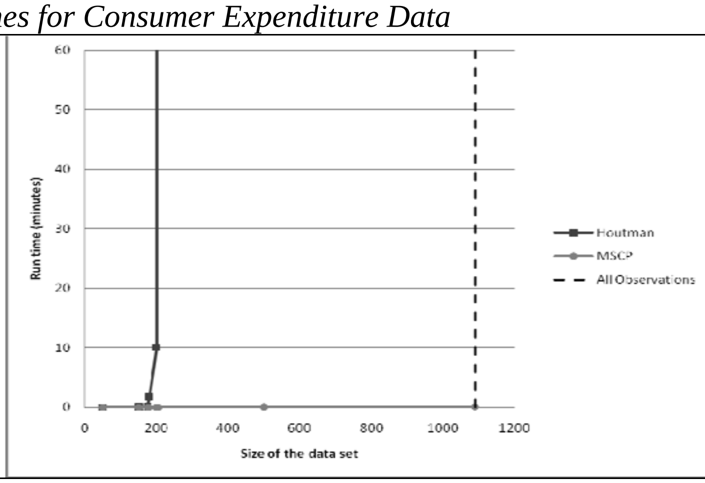


Figure 4: HM Index Scores for Actual Subjects (from CFGK) and 93,000 Simulations

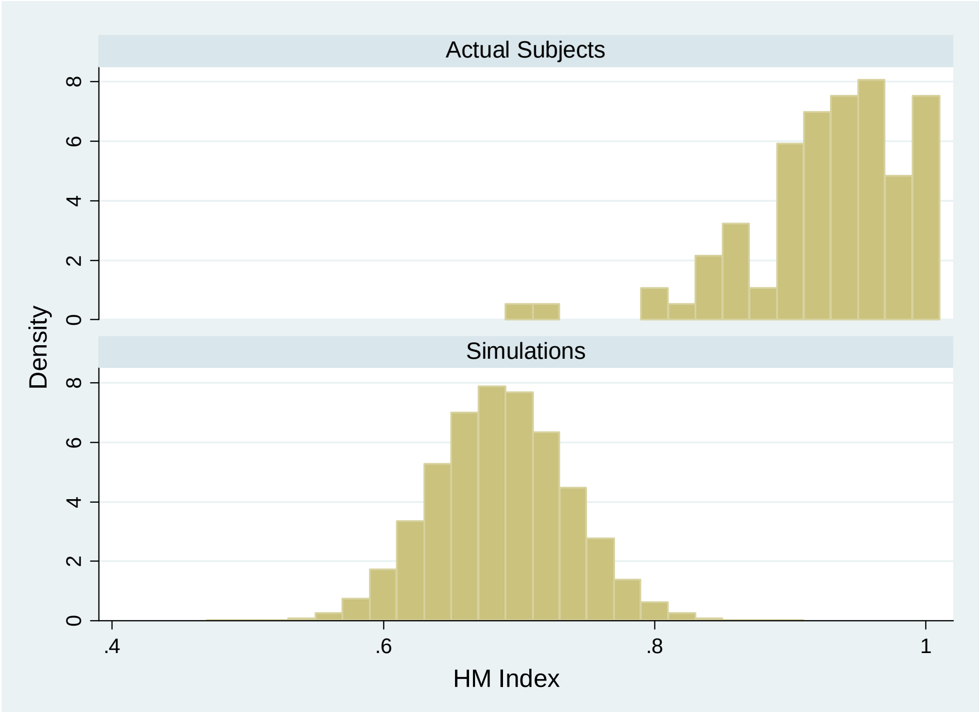


Figure 5: Modified HM Index Scores for Actual Subjects (from CFGK) and 93,000 Simulations

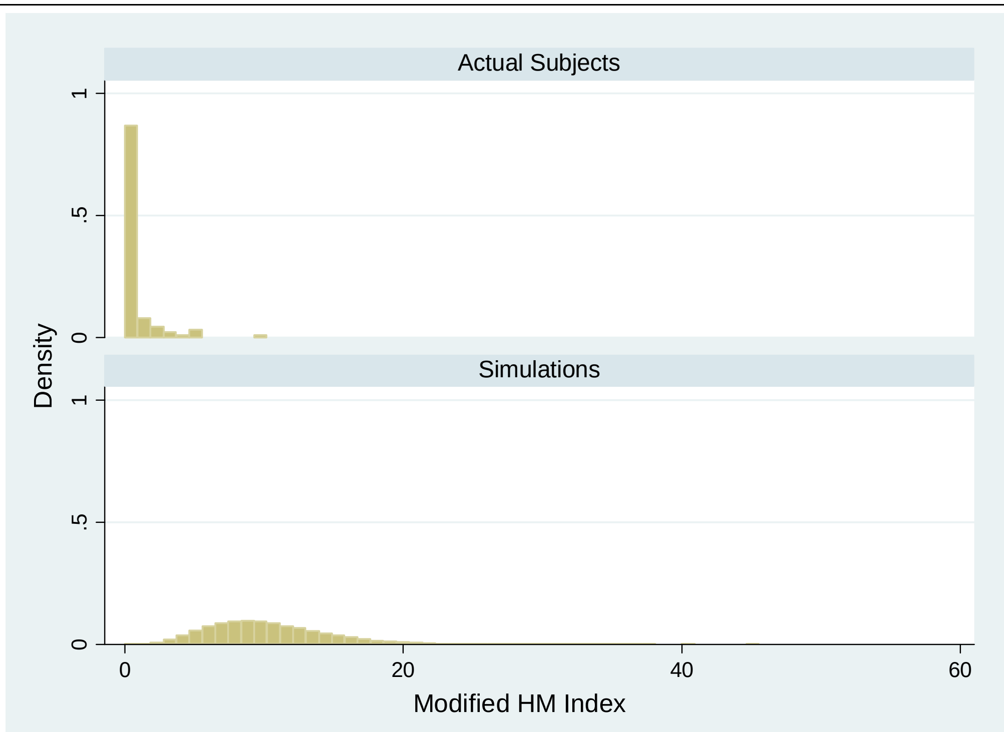
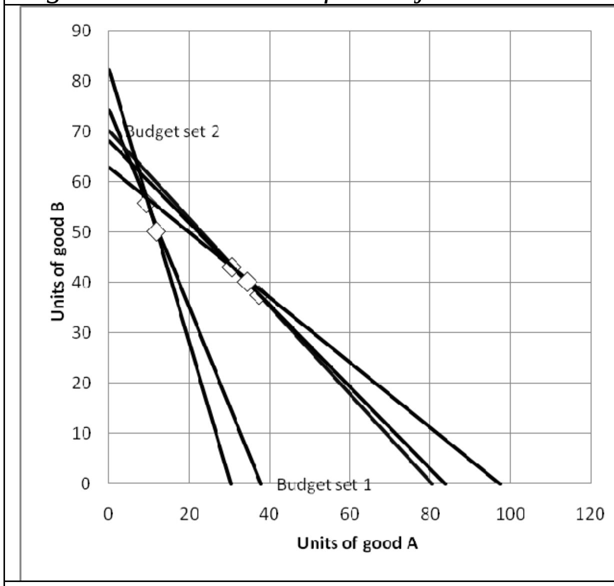
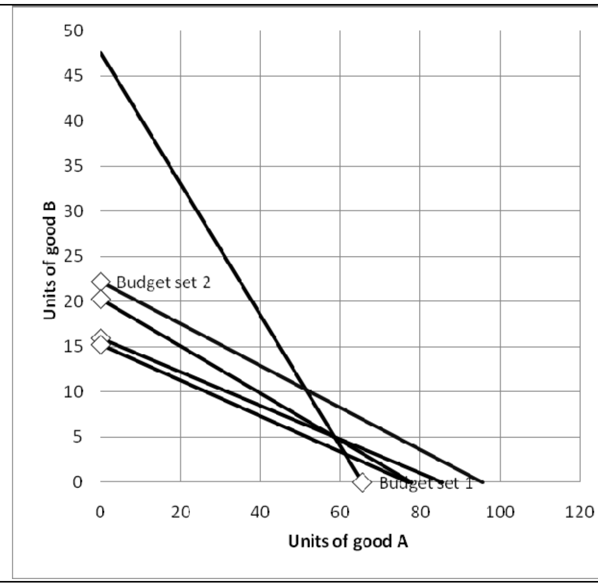


Figure 6: Five Choices for Subjects 512 and 603

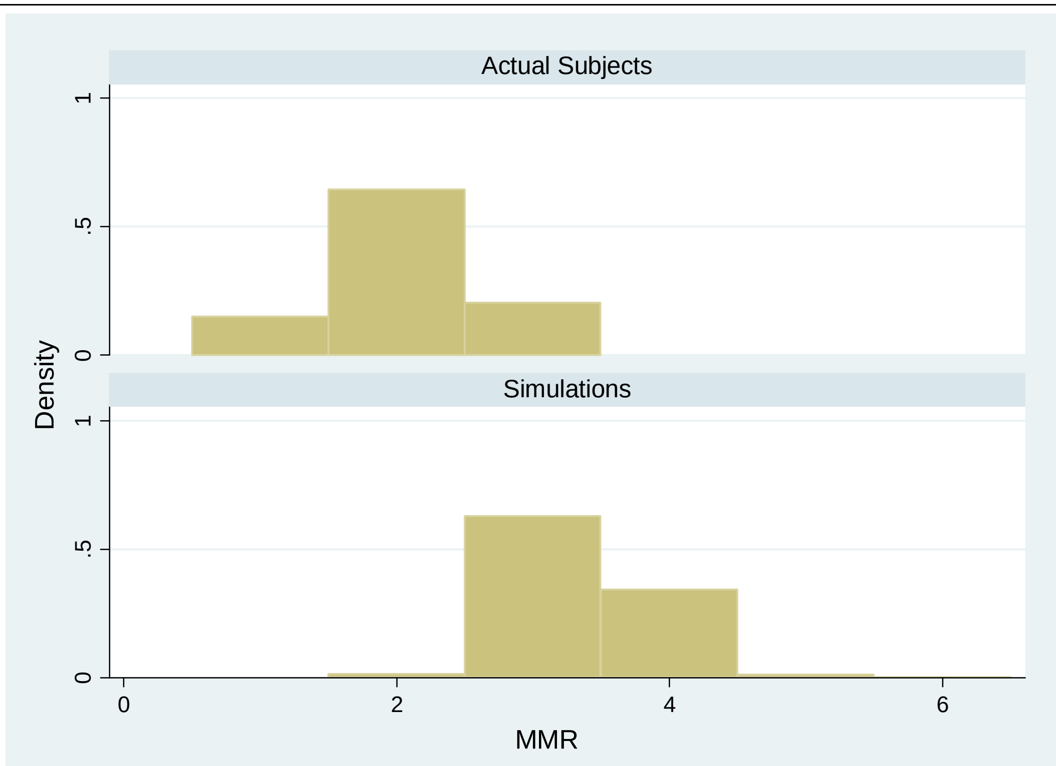


Subject 512



Subject 603

Figure 7: Minimal Multiple Rationales Estimates for Actual Subjects (from CFGK) and 93,000 Simulations



Tables

Table 1

Subject	CFGK	Houtman	MSCP
211	> 1 hour	< 1 sec	< 0.1 sec
324	120 sec	< 1 sec	< 0.1 sec
325	> 1 hour	< 1 sec	< 0.1 sec
406	> 1 hour	< 1 sec	< 0.1 sec
504	90 sec	< 1 sec	< 0.1 sec
608	> 1 hour	< 1 sec	< 0.1 sec

Run time for each algorithm for the 6 'problem subjects' from CFGK

Table 2

Income Percentile	**Obs**	**Removals**	**HM Index**	**Selten-Bronars**
1-25	1089	243	0.78	0.04
26-50	1089	284	0.74	0.06
51-75	1090	286	0.74	0.06
76-100	1089	240	0.78	0.05
Overall	**4354**	**1150**	**0.74**	**0.05**
Average			*0.75*	*0.05*

Rationality measures for household consumption data from BBC for years 1997-1999, broken down by income percentile

Table 3				
Geographical Grouping	**Obs**	**Removals**	**HM Index**	**Selten-Bronars**
East Anglia	169	16	0.91	0.04
East Midlands	309	35	0.89	0.03
London	373	46	0.88	0.06
North Ireland	297	29	0.90	0.04
North	197	19	0.90	0.02
North West	436	62	0.86	0.04
Scotland	324	34	0.90	0.05
South East	856	161	0.81	0.06
South West	418	34	0.92	0.03
Wales	224	19	0.92	0.06
West Midlands	380	42	0.89	0.02
Yorkshire	367	38	0.90	0.04
Overall	**4354**	**1150**	**0.74**	**0.05**
Average			*0.89*	*0.04*

Rationality measures for household consumption data from BBC for years 1997-1999, broken down by geographic region

Table 4				
Measure	HM	Modified HM	Varian	Afriat
HM	1	0.82	0.75	0.78
Modified HM		1	0.95	0.95
Varian			1	0.91
Afriat				1

Spearman Rank Correlation between rationality measures for CFGK data