Revealed Differences

Marco Castillo[a,*], Mikhail Freer[b]

[a]*Department of Economics, Texas A&M University, 3035 Allen Building, College Station, TX 77843, USA.*
[b]*Department of Economics, University of Leuven (KU Leuven), E. Sabbelaan 53, B-8500 Kortrijk, Belgium.*

Abstract

Two individuals are said to be *revealed different* if their joint decisions are more distant from rationality than either of their individual decisions taken separately. We show that the *revealed different* relation can be used to identify preference types and therefore to evaluate the heterogeneity of preferences in a completely nonparametric way. Using experimental data from a random sample of the Dutch population, we find that 1,182 individuals can be divided into 131 different preference types, men's preferences are more heterogeneous than and different from women's preferences. These results replicate in a different decision environment.

Keywords: rational behavior, heterogeneity of preferences
JEL: C38, D01, D12

*Corresponding author
Email addresses:* `marco.castillo@tamu.edu` (Marco Castillo), `mikhail.freer@kuleuven.be` (Mikhail Freer)

1. Introduction

We investigate the question of the heterogeneity of preferences from the point of view of revealed preference analysis. Specifically, we propose the *revealed different* relation to study differences in preferences in a completely nonparametric way. Two individuals are said to be *revealed different* if their joint decisions are more distant from rationality than either of their individual decisions taken separately. We can use the *revealed different* relation to answer the question of how heterogeneous preferences are within a group and which groups are more heterogeneous. Our approach can be extended to allow alternative and/or stricter versions of rationality, which allows us to evaluate the cost of different behavioral assumptions in detecting heterogeneity.

Our definition allows a comparison between individuals who themselves violate rationality. This is a necessary requirement since most individual data are not perfectly consistent with rationality. In our analysis, we use the preexisting level of rationality of the individuals being compared as a reference point. This construction allows us to define a *revealed different* relation which is complete and symmetric yet not transitive. We also discuss alternative approaches whereby two individuals are not comparable. The normalization we use is motivated by the observation that the joint decisions of two perfectly rational agents cannot be more rational than the decisions taken separately.

The *revealed different* relation can be used to evaluate the level of heterogeneity in a set of data. For instance, Pattanaik and Xu (2000) use the complementary *similarity* relation to characterize freedom of choice across sets of alternatives. This relation allows the comparison of sets of alternatives in terms of how many different options they provide. The authors suggest using the smallest partition of the similarity relation as a measure of freedom of choice. Because the *similarity* relation is complementary to the *revealed different* relation, the methodology we use to elicit the amount of types in the data corresponds to Pattanaik and Xu's (2000) methodology.

Preference types are defined by two conditions. First, no two members of the same preference type are *revealed different* from each other. Second, for any two different

preference types, there is at least one pair containing one person belonging to each type such that they are *revealed different* from each other. While the assignment of people to preference types is not unique, the cardinality of the smallest partition of people into preference types is always well-defined. We use this number as our measure of heterogeneity of preferences.

In order to evaluate the empirical content of the *revealed different* relation, some assumptions about how to measure distance from rationality are needed. In this paper we use three measures of distance from rationality: Afriat's (1973) critical cost efficiency index (CCEI), Echenique et al.'s (2011) money pump index (MPI) and Houtman and Maks (1985) index (HMI). The CCEI is appealing because it is monotone in the number of decisions a subject makes. The CCEI never decreases as more decisions are observed. Since the CCEI is bounded between 0 and 1, this implies that the CCEI will converge to its true value as more data are collected, which allows us to assess the extent to which our results are due to sampling error or true differences in preferences. Monotonicity is not a common property among measures of distance from rationality. The MPI and the HMI, are not monotone measures and provide a natural contrast to the CCEI.

Our approach to gathering data is to use a random sample of the Dutch population whose preferences regarding risky assets were elicited using experimental methods in Choi et al. (2014). In that study, participants were presented with 25 randomly generated budgets containing two assets that obtained with equal probability. By defining the *revealed different* relation based on the Generalized Axiom of Revealed Preferences (GARP), we find that the 1,182 individuals in the sample can be partitioned into 131 different preference types. The preference types have an average of 8.03 (s.d. 3.88) members, with the largest type containing 20 subjects and the smallest type containing a single individual. These results are comparable using the MPI to measure distance from rationality instead of the CCEI. In this case, the number of preference types is 101.

As discussed above, the *revealed different* relation can be used to assess differences in heterogeneity across populations and to test the cost for measured heterogeneity of additional behavioral assumptions on preferences. Regarding differences across

3

populations, we find evidence that men's preferences are more heterogeneous than women's preferences, i.e., there are preference types among men that are not found among women.[1] These results are robust. In an appendix we show that the reported gender differences in heterogeneity reproduce in a data set using different goods and a different population. This gives us confidence of the ability of nonparametrically identify differences across groups. We also show that the ability to detect differences across groups depends on the number of assumptions imposed on the preference relation. Imposing assumptions on preference relations not supported by the data significantly reduces the estimated number of types.

To measure the extent to which differences in individual preferences are due to differences in individual characteristics, we conduct a regression analysis of the *revealed different* relation as a function of the differences in demographic variables and individual fixed effects of the individuals being compared. We find that individual differences explain around 33 (max 53 percent) (according to logit regressions) percent of the variance of the *revealed different* relation. Including observable characteristics (sex, education, and income level) of the pair adds little to the proportion of the variance explained. This suggests that differences in preferences, regardless of being correlated, are largely driven by factors other than differences in individual characteristics.

We are not the first to suggest using revealed preference analysis to identify heterogeneity in preferences. For example, Gross (1995) proposed testing for GARP using cross-sectional data as a way to test for the commonality of preferences. As another example, Heufer (2014) offered a nonparametric way to classify people based on their level of risk aversion.[2] The work of Crawford and Pendakur (2013), who propose finding the smallest partition of people into preference types such that each

---

[1]Formally, there might be some types of preferences among women which are also unique. However, under the null hypothesis that both populations have the same types, the most conservative forecast is that there are some preference types among men which are not present among women.

[2]Note that this measure is not complete. The method may return people who are not comparable in terms of risk aversion. However, the measure can be completed by assuming that incomparable people are revealed different.

group's decisions do not violate GARP, is closest to our approach. Using observations of the demand for dairy in a cross-section of the Danish population, they find that there are up to 12 separate preference types in the population. An important difference between their approach and ours is that we explicitly address the possibility of behavior that is not perfectly rational. This is important since, in experimental settings, subjects routinely violate the weak axiom of revealed preferences (Sippel, 1997). It can also be shown that Crawford and Pendakur (2013) provide a lower bound for the degree of heterogeneity, because rational behavior is confounded with the choice sets' inability to reveal violations of rational behavior. Indeed, we use Crawford and Pendakur's (2013) method in the random sample of Dutch subjects by treating each separate budget as if it was a unique individual observation. This sample contains 29,550 (25×1182) different observations. We find that these 29,550 budgets can be grouped into only 11 types. Given that we observe that almost 40 percent of the population have a CCEI below 0.9, it is clear that some of the types must include budgets where the GARP cannot fail.[3]

Our research fits in with the larger literature on diversity (Weitzman, 1992; Pattanaik and Xu, 2000; Nehring and Puppe, 2002), which has tasked itself with axiomatizing the measurement of diversity. Weitzman (1992) and Nehring and Puppe (2002) take cardinal measure(s) of difference across individuals or alternatives as a starting point. While the *revealed different* relation can be used to construct a distance function, its appeal lies in the fact that it is an ordinal summary of differences. We consider this to be a necessary first step in exploring diversity as revealed by individual choices.

The remainder of this paper is organized as follows. Section 2 describes the revealed difference approach to heterogeneity. Section 3 describes the data we use and presents our results. Section 4 discusses alternative measures of heterogeneity, alternative axioms of rationality and alternative consumption contexts. Section 5 concludes the paper.

---

[3]Our approach takes advantage of a richer data set. Our method can be thought as a natural extension of Crawford and Pendakur's (2013) approach.

2. Preliminaries

In this section we provide the definitions that are necessary for formally defining the *revealed different* relation and showing when it can be used to elicit differences in the preferences of agents. First, we define the *revealed different* relation in general and provide the conditions under which two subjects's being *revealed different* shows that they do have different preferences. Second, we specify the definitions of rationality that we use in the analysis. Third, we describe the measures we use for the distance from rationality. Finally, we discuss how to classify people into preference types using the *revealed different* relation.

Let $E = \{x^i, p^i\}_{i=1}^N$ be a **consumption experiment**, where $x^i \in \mathbb{R}_+^L$ are consumption bundles chosen at prices $p^i \in \mathbb{R}_{++}^L$, and let $\mathcal{E}$ denote the set of all consumption experiments. Note that prices define the budget $B^i = \{x : \ p^i x \le p^i x^i\}$. Let us assume that every E is generated by a complete preference relation R over $\mathbb{R}_+^L$. Let $\rho : \mathcal{E} \to [0, 1]$ be a **distance from rationality**, where $\rho(E) = 0$ if and only if there is a complete and transitive relation that rationalizes E. We remark that the notion of rationality can include requirements other than completeness and transitivity of preferences. Let $D = \{E_1, E_2, \ldots, E_n\}, E_i \in \mathcal{E}$ be a set of consumption experiments.

2.1. The Revealed Different Relation

In this section we make explicit the meaning that two consumption experiments are revealed different. Additionally, we show the conditions under which two consumption experiments will be revealed different only if they are generated by two different underlying preference relations.

Two consumption experiments are *revealed different* if jointly they are further from rationality than each of them taken separately:

$$RD_\varepsilon(E_r, E_s) = \begin{cases} 1, \text{ if } \rho(E_r \cup E_s) - \max\{\rho(E_r); \rho(E_s)\} > \varepsilon \\ 0, \ otherwise. \end{cases}$$

Let us illustrate the mechanics of the RD_ε using a simple example. Assume we have two consumption experiments, each one of them consisting of only one choice:

$E_1 = \{((1, .75), (.25, 1))\}$ and $E_2 = \{((2, .5), (1, .5))\}$. Each of the experiments is trivially rational, however, the joint experiment violates the Weak Axiom of Revealed Preferences. That is, the joint decisions are more distant to rationality than individual choices.

Even if all choices are generated by the same underlying preferences, it is possible to observe that $\rho(E_r \cup E_s) > 0$ for choices r and s due to the fact that we only observe an estimate of the "underlying" distance from rationality. For instance, two experiments in the previous paragraph could be generated by the same (not fully rational) preference relation. For that reason, we introduce a precision parameter (ε) such that $\rho(E_r \cup E_s) - \max\{\rho(E_r); \rho(E_s)\} > \varepsilon$ as a way to reduce the likelihood of false positives. Note that the precision of this estimate depends on the desired power of the test, i.e. the number of observations per person.

Observation 1. *For every $E_r, E_s \in \mathcal{E}$, $RD_\varepsilon(E_r, E_s)$ satisfies the following properties:*

– $RD_\varepsilon(E_r, E_r) = 0$

– $RD_\varepsilon(E_r, E_s) \geq 0$

– $RD_\varepsilon(E_r, E_s) = RD_\varepsilon(E_s, E_r).$

Observation 1 shows that RD_ε is a distance function.[4] This is important since Weitzman (1992) shows that if differences between species can be represented by a distance function, it is always possible to construct a value function that represents the costs of losing a species. Note that in the current paper we do not address the issue of cardinal measures of diversity. However, Observation 1 shows that it is possible to use RD_ε to construct a cardinal measure of diversity.

Let us now show that there is always such measurement error such that RD_ε returns one only if two consumption experiments are generated by different preference relations. This is important, because it shows that for any data set there is a RD_ε

---

[4]We use the distance function in the sense introduced by Weitzman (1992), that is for every x, y, the following are true if d is a distance function $d(x, y) \geq 0$; $d(x, x) = 0$; $d(x, y) = d(y, x)$.

that do not produce false positives. Let $\mathcal{E}(R) \subseteq \mathcal{E}$ denote the space of consumption experiments generated by R.

Observation 2. *For every $E_r, E_s \in \mathcal{E}$, there is a ε such that $RD_\varepsilon(E_r, E_s) = 1$ only if E_r and E_s are generated by different preference relations.*

Proof. Let R_r be the complete preference relation. Let E_r be a finite consumption experiment generated by R_r, i.e. $E_r \in \mathcal{E}(R_r)$. Let $\varepsilon_r = \sup_{E_r \subset E \in \mathcal{E}(R_r)} |\rho(E) - \rho(E_r)|$. Since ρ is bounded, there is a finite supremum $\varepsilon_r \leq 1$. Similarly we can define ε_s. Let $\varepsilon = \max\{\varepsilon_r, \varepsilon_s\}$. Assume on the contrary that E_r and E_s are generated by the same preference relation and $RD_\varepsilon(E_r, E_s) = 1$. This implies that $\rho(E_r \cup E_s) - \min\{\rho(E_r), \rho(E_s)\} > \varepsilon$. Without loss of generality assume that $\rho(E_r) \leq \rho(E_s)$, then $\rho(E_r \cup E_s) - \rho(E_r) > \varepsilon_r$. This implies a contradiction, because $\varepsilon_r \geq \rho(E_r \cup E_s) - \rho(E_r)$ (implied by $E_r \cup E_s \in \mathcal{E}(R_r)$). $\qquad\square$

Observation 2 shows that there always is a precision level ε such that two consumption experiments are revealed different only if they are generated by different preference relations. Note that the observation establishes that preferences are revealed different by the data, only if they are generated by different preference relations; it does not claim the reverse. This implies that RD_ε tends to underestimate heterogeneity.

A distance function ρ is **monotone** if $\rho(E)$ increases as new budgets are added to the consumption experiment E, i.e. $E \subseteq E'$ implies $\rho(E) \leq \rho(E')$. Let $\rho(R)$ be the distance from rationality of the preference relation, that is the distance from rationality of one of all possible consumption experiments. The following observation allows us to use RD_0 instead of RD_ε:

Observation 3. *If ρ is monotone and for every $E \in D$, $\rho(E) = \rho(R)$, then $RD_0(E_r, E_s) = 1$ only if E_r and E_s are generated by different preference relations.*

The proof of this is straightforward and, therefore, it is omitted. We have to assume that the observed distance from rationality is equal to the underlying preference relation's distance from rationality. This assumption implies that the consumption experiment is powerful enough to elicit the real distance from rationality. However,

even under this assumption this cannot be true unless we assume distance from rationality to be monotone function.

2.2. Axioms of Revealed Preferences

Because the *revealed different* relation is defined with respect to a distance from rationality and distance from rationality in turn depends on the definition of rationality, we now state the axioms that specify the definitions of rationality that we use for the analysis.

For a given $e \in [0,1]$, define the revealed preference relation as $x^i R(e) x^j$ if $p^i x^j \leq e p^i x^i$ and the strict preference relation as $x^i P(e) x^j$ if $p^i x^j < e p^i x^i$. Let $T(e)$ denote the transitive closure of $R(e)$.

Definition 1. *A consumption experiment $E = \{x^i, p^i\}_{i=1}^N$ satisfies the* **Generalized Axiom of Revealed Preference (GARP)** *if $x^i T(1) x^j$ implies not $x^j P(1) x^i$.*

Afriat (1967) showed that a consumption experiment satisfies GARP[5] if and only if there exists a nonsatiated and continuous utility function that represents $R(1)$. Hence, GARP is a necessary and sufficient condition for the existence of a complete and transitive preference relation that rationalizes the consumption experiment. This justifies using GARP as our main definition of rationality.

2.3. Measures of Rationality

Recall that to define RD_ε, we need to specify the distance from rationality. We introduce the following measures of rationality: the Critical Cost Effieciency Index (CCEI) (Afriat, 1973), the Money Pump Index (MPI) (Echenique et al., 2011) and Houtman-Maks Index (HMI) (Houtman and Maks, 1985).

Definition 2. *The Critical Cost to Efficiency Index* *(CCEI) is the maximum $e \in [0,1]$ such that $R(e)$ is consistent with GARP.*

---

[5]GARP is introduced by Varian (1982), while Afriat (1967) refers to an equivalent condition called cyclical consistency.

The value 1 - CCEI can be interpreted as the share of income a person is willing to waste to behave irrationally. Note that a subject who is rational has a CCEI of 1. Importantly, Afriat (1973) showed that e is the CCEI of a consumption experiment if and only if there exists a nonsatiated and continuous utility function that represents $R(e)$. In this setup, $\rho = 1 - e$, and therefore larger values of ρ mean greater distances from rationality. The CCEI is monotone in the sense that it cannot increase as budgets are added to the consumption experiment. Hence, the corresponding distance from rationality (ρ) is monotone (but increasing) as well.

The second distance function we use is the MPI. Let $x^{i_1}, x^{i_2}, \ldots, x^{i_k}$ be the sequence that violates GARP, i.e., $x^{i_j} R(1) x^{i_{j+1}}$ for any $j \in \{1, \ldots, k-1\}$ and $x^{i_k} P(1) x^{i_1}$. The *money pump* of the sequence is defined as $mp = \sum_{j=1:k} p^{i_j}(x^{i_j} - x^{i_{j+1}})$, where $i_{k+1} = i_1$. The *relative money pump* is defined as the money pump divided by the total income in the chain ($rmp = \frac{\sum_{j=1:k} p^{i_j}(x^{i_j} - x^{i_{j+1}})}{\sum_{j=1:k} p^{i_j} x^{i_j}}$).

Definition 3. *The* ***Money Pump Index*** *(MPI) is the median (or the mean) relative money pump over all sequences that violate GARP.*

The third index is HMI, originally it is defined as the maximum subset of data that is consisted with GARP. However, for the sake of our method we need to redefine it as the ratio of the maximum subset of data that satisfies GARP to the size of consumption experiment.

Definition 4. *The* **Houtman-Maks Index** *(HMI) is defined as follows:*

$$HMI = \frac{\max_{E' \subseteq E} \{|E'| : E' \text{ satsisfies } GARP\}}{|E|}$$

Note that of the above mentioned indices, only the CCEI is monotone. Therefore, it is the only one to which we can formally apply RD_0 (see Observation 3). Moreover, both the MPI and HMI are not feasible to compute (NP) for the general case ($L \geq 3$). We, therefore, use the average money pump methodology from Smeulders et al. (2013) to approximate the MPI. For the case of HMI we follow Gross and Kaiser

10

(1996) since we have linear budgets.[6]

2.4. How should heterogeneity be measured?

We now discuss how to measure heterogeneity using the *revealed different* relation.

Pattanaik and Xu (2000) proposed a measure of freedom of choice based on the "similarity" of the alternatives available in different choice sets. The *revealed similar* and *revealed different* relations are complements of each other. According to Pattanaik and Xu (2000), the diversity of a set is determined by the smallest partition of alternatives into homogeneous disjoint subsets. A set is said to be homogeneous if none of its elements are different from each other (or, alternatively, any two elements of the set are similar). Pattanaik and Xu's (2000) definition of diversity has a natural appeal for classifying people into preference types.

We remark that given this definition, there might be two individuals who are not *revealed different* from each other but are nevertheless assigned to different preference types. This can occur if one of these individuals, a, is not *revealed different* from any other individual, while the other, b, is *revealed different* from a third individual, c. In this case, b and c belong to different preference types, and a can freely be assigned to either b's preference type or c's preference type. This makes it clear that measurement of the heterogeneity of preferences is akin to the well-studied problem of graph coloring in graph theory (see Agnarsson and Greenlaw (2006)). We will now provide the necessary definitions to show the equivalence of the two problems and to explain how we group preferences into types.

A graph is a tuple $G = (V, A)$ in which V is the set of vertices and $A \subseteq \{V, V\}$, where $\{V, V\}$ is the set of all 2-element subsets of V. Vertices $u, v \in V$ are said to be **adjacent** if $\{u, v\} \in A$. Note that we can define the graph that corresponds to the revealed difference relation as follows: Let G be the set of individuals $(i, j, \ldots)$, and let $\{i, j\} \in A$ if and only if $RD_\varepsilon(i, j) = 1$. A **coloring** is a function $c : V \to \mathbb{N}$, such that if $\{u, v\} \in A(G)$ for some $u, v \in V$, then $c(u) \neq c(v)$. This simply means that none of the two adjacent vertices can have same color. A graph G is said to be

---

[6]Note that this works only for two-good case. For the higher dimensional problem one can apply the Mixed Integer Programming method from Heufer and Hjertstrand (2015).

k-colorable, if there is a coloring function $c : V \rightarrow \{1, 2, \ldots, k\}$ – the graph can be colored using no more than k-colors.

Definition 5. *A number $\chi(G)$ is said to be **chromatic number** of the graph if G is $\chi(G)$-colorable and there is no $k < \chi(G)$ such that G is k-colorable.*

The chromatic number of a graph is the minimum number of colors required for a graph coloring – in our context this would specify the minimum amount of types to cover all the population. Importantly, the chromatic number is unique, and therefore the measure of diversity is always well-defined. Note that coloring in the context of *revealed differences* implies, that there is no couple of consumption experiments which are revealed different and belong to the same type. As we already mentioned this has a direct connection to the Pattanaik and Xu (2000) measure of the freedom of choice – let graph be defined over alternatives and two vertices being adjacent if alternatives are not similar. Then, the minimum amount of disjoint homogeneous sets necessary to cover all space of alternatives would be equal to the chromatic number of this graph.

It is important to note that while the chromatic number is unique, the actual coloring of a graph need not be. This can be shown using a simple example. Assume there is a vertex that is not connected to any other vertex (an isolated vertex); then this can be colored with an arbitrary color. Hence, an isolated vertex can belong to any type, and therefore, coloring is not unique. Thus, while we can answer the question of how many preference types there are, we cannot (uniquely) determine the composition of the types.

3. Results

In this section, we provide an empirical analysis of heterogeneity using the *revealed different* relation. We first discuss the data we use. Second, we determine ε for RD_ε according to Observation 2. Third, we use RD_ε to determine the number of preference types in the data.

3.1. Data

We employ the experimental data that uses contingent assets. Let $i = 1, \ldots, L$ denote the states of the world. Then $\pi \in [0;1]^L$ such that $\sum_{i=1}^{L} \pi_i = 1$ denote the probability distribution over the states of the world, where π_i is the probability that state of the world i happens. In this case x is the vector of investments in every state of the world given prices p and x_i is paid to the subject if i-th state of the world is realized.

The analysis is based on the following data sets:

– Choi et al. (2007). The participants are 93 undergraduates and staff members from the University of California – Berkeley. Each subject faced 50 different budgets with two risky assets that obtained with probabilities $\frac{1}{2}, \frac{1}{3}$, or $\frac{2}{3}$. The prices of the assets were assigned randomly.

– Choi et al. (2014). The participants are a random sample of 1182 Dutch adults. Each participant faced 25 budgets with two risky assets that obtained with equal probability. The prices of the assets were assigned randomly.

3.2. Testing Assumptions

For the *revealed different* relation to elicit real differences in preferences we have to make two assumptions: (i) distance from rationality converges to some "true" distance from rationality, and (ii) observed distances from rationality are "close enough" to the "real" distance from rationality.

While assumption (i) has to be taken for granted, assumption (ii) can be tested. Recall that the data in Choi et al. (2014) has 25 budgets per subject, and for the unions of two consumption experiments we obtain 50 budgets per pair. Therefore, using Choi et al.'s (2007) data (which has 50 budgets per subject), we can estimate the changes in the distance from rationality resulting from increasing the number of budget from 25 to 50 (ε). Note however, that these data sets sample different populations. Choi et al. (2007) is an experiment with undergraduate students from

Berkeley, while Choi et al. (2014) takes the sample of Dutch adults.[7]

This is necessary to control for possible false positive results in which two consumption experiments generated by the same underlying preference relation appear to be *revealed different.*

For this purpose, we bootstrap 100 samples of 25 budgets for each subject and measure the distance from rationality using these sub-experiments. Then we compare the distances from rationality obtained from the bootstrapped sub-experiments using 25 budgets to the distances from rationality obtained from the consumption experiment with 50 budgets. Since the consumption experiments are generated using the data from the same subjects, they are generated by the same underlying preference relation. Hence, the distance from rationality in the experiment with 50 budgets minus the distance from rationality in the experiment with 25 budgets characterizes the value of ε that is needed to eliminate the possibility of a subject being *revealed different* from him- or herself. Figure 1 shows the distribution of changes in distance from rationality between 25 and 50 budgets for all indices.

[Figure 1 about here.]

We construct RD_ε relations with ε equal to 0, to the median change in the distance from rationality, and to the 95th percentile of the empirical distribution of the change in the distance from rationality. Note that Figure 1 shows that the indices behave differently. In particular, the CCEI is, by definition, monotone and stabilizes at quite small levels of ε. While the MPI is not monotone, the changes are quite small. This is due to the fact that the method we use estimates the median MPI for all cycles and hence for a large enough number of budgets adding more data will not shift the median if preferences are similar. Finally, the HMI is not monotone and it does not stabilize. It is, therefore, difficult to perform the proposed analysis using HMI since we would have to account for quite a large ε. This would significantly reduce the ability of our method to find out the differences if they are present.

---

[7]Choi et al. (2014) show that the distribution of the CCEI is larger in the Berkeley sample than in the Dutch population. That is subjects in the Dutch sample are further from rationality.

[Table 1 about here.]

Table 1 shows the values of ε we use in the analysis. To simplify notation, we use RD_0 to denote the relation that does not allow for error, RD_M to denote the relation that uses the median change in distance from rationality, and RD_{95CI} to denote the relation that uses the 95th percentile of the empirical distribution of the change in distance from rationality.

3.3. The Revealed Different Relation

This section presents basic information about the *revealed different* relation. The richness of the *revealed different* relation can be appreciated through Figure 2. It shows the density of the relation for a random sample of 100 subjects and an example of the coloring of the RD relation.

[Figure 2 about here.]

Table 2 presents the basic statistics for the *revealed different* relation. The average number of people a subject is different from is 695 (s.d. 237), with a minimum of 20 and a maximum of 1176 (numbers provided use RD_0 according to CCEI).

[Table 2 about here.]

We want to answer the following questions with our analysis: first, whether heterogeneity in terms of preferences is present in the data, and how many types there are; second, whether heterogeneity can be explained by some observable characteristics (demographic data) or is mostly idiosyncratic; and third, how our method should be used to reveal differences across groups – in particular, differences by gender. Note that the method we construct allows us not only to answer the question of whether groups have different types of preferences, but also to compare the groups on the basis of how heterogeneous they are.

As mentioned above, we use the data from Choi et al. (2014) for this purpose. Table 3 provides descriptive statistics for the sample. The sample has a majority of men, people who are older than 50 years of age, people who are salaried, and people who are in a stable relationship.

[Table 3 about here.]

15

3.3.1. Are Preferences Different?

Figure 3 shows the number of different preference types as a function of the population size. These estimates are obtained by random sampling subsets of individuals and calculating the chromatic numbers for those sets based on the RD_ε relation. The RD_ε relation is defined in terms of the CCEI, MPI and HMI. Panel (a) shows the results using the CCEI, Panel (b) shows the results using the MPI and Panel (c) shows the results using HMI. Solid lines are the (mean) number of types for the corresponding group, and dashed lines show the 95% confidence intervals for the number of types given a population size. However, due to the tightness of confidence intervals in this case they are hardly visible. Henceforth, we use 400 bootstrapped samples per population size. In both cases, the higher we set ε, the fewer the types that are revealed.

We observe that the (mean) number of preference types in a population of 1000 individuals is 90 according to the CCEI, 77 according to the MPI and 166 according to HMI. In the population at large (1,182 people), we find that there are 131 types if we use the CCEI to define RD_0, there are 116 types if we use the MPI to define RD_0 and 248 types if we use HMI to define RD_0. If we restrict the definition of revealed difference to RD_M, then the number of types is 90 for the CCEI, 116 for the MPI and 176 for HMI. If we restrict the definition of revealed difference to RD_{95CI}, then the number of types is 35 for the CCEI, 17 for the MPI and 5 for HMI. This shows that there is a large degree of heterogeneity in the data regardless of the measure used for distance from rationality. Note that there are fewer types according to the MPI compared to the CCEI. This can be explained by the fact that the MPI is not a monotone measure of distance from rationality. For the case of the HMI, the number of types is dramatically reduced. This is due to the large measurement error associated with this measure.

[Figure 3 about here.]

Figure 3 suggests that the total number of groups in the population is finite. We do not observe a linear relationship between population size and the number of groups. Given that the minimum size of a group is one, we would expect that

the number of groups eventually converges. Note, however, that the answer to the question of how many types there are is complicated by measurement error. Figure 3 demonstrates that making the *revealed different* relation stricter leads to a reduction in the number of types in the population. One implication is that heterogeneity is likely to be underestimated due to the fact that all consumption experiments are finite and therefore unable to detect some differences (see Observation 2). More importantly, erroneous assumptions regarding rationality are likely to reduce the measured level of heterogeneity. We will discuss this issue more extensively below.

Figure 1 shows a significant portion of the population undergoing very large changes in their CCEIs as we move from 25 to 50 budgets. This suggests that estimates of ε based on means will likely be biased upwards. That is, the ε needed to statistically reject the null hypothesis of no difference in preferences is likely overestimated. Figure 3 illustrates this. The estimated number of groups based on the median change in CCEI is much larger than the estimates based on the 95% percentile of the change in CCEI. However, using this very strict threshold, we still find a large number of types in the population. We can then conclude that there is a high degree of heterogeneity in the data.

3.3.2. What Can Explain Heterogeneity?

To investigate whether the observed level of heterogeneity in the data is due to observed individual characteristics, we conduct regression analysis on the *revealed different* relation on the characteristics of the individuals in a pair. Table 4 presents a linear probability model on a variable that equals 1 if individuals i and j are *revealed different* and 0 if they are not. The main statistic of interest is the portion of the variance of the *revealed different* relation that can be explained by differences in individual characteristics. Given that there are 1,182 subjects, we have 1,397,124 such comparisons. All the regressions cluster errors at the i level and all regressions include fixed effects on the i and j subject.

The first four columns in Table 4 show estimates for the case in which the *revealed different* relation is calculated using Afriat's CCEI. The first column shows a regression in which only fixed effects for each member of the pair are included. Fixed

effects capture any individual differences across individuals. They capture about 33 percent of the variance in the *revealed different* relation. This means that two-thirds of the variance in the relation cannot be explained by individual differences. The next regressions then include characteristics of the pair. We observe in the second column that adding characteristics of the pair adds little to the proportion of the variance explained. This finding holds regardless of how strict the definition of the *revealed different* relation is. This suggests that there other factors that are particular to the pair that explain the differences in preferences. The fifth and sixth column reproduce the analysis using the MPI. The last three columns present the corresponding regressions using the HMI. We note that the results using the HMI are unstable, contradicting results using either the CCEI or the MPI. This is likely due to the fact that the HMI does not stabilize as more information is added and suggests caution in using measures of distance from rationality that are not monotone.

Table 4 allows us to take a look at differences in preferences across and within subpopulations. We observe that a pair of males are more likely to be different than a mixed gender pair and that a pair of females are less likely to be different than a mixed gender pair. This result is consistent for all ε for the RD_ε computed for CCEI and MPI. Regarding age differences, we do not observe a consistent pattern across definitions and assumptions. Finally, differences across other characteristics are relatively small and not always consistent. Table 5 reproduces these results using logit regressions instead of linear regressions. This demonstrate that the results are not due to parametric assumptions made in the regression analysis.

[Table 4 about here.]

[Table 5 about here.]

3.3.3. Which Groups Are Different?

We have already shown evidence regarding the amount of heterogeneity within and between subpopulations. We now investigate whether there are differences across groups and whether there are groups with more diverse preferences.

To answer the first question, we check whether the number of groups in the population as a whole is larger than the number of groups in its constituent subpopulations. We do this because for an additional group to appear in the population as a whole, there must exist a member of one of the subpopulations who is *revealed different* than some person in each of the groups of the other subpopulations. That is, there must be a person in one subpopulation who cannot be included in any of the existing groups in other subpopulations. To answer the second question, we simply look at the number of groups of different subpopulations. Populations that are more heterogeneous are expected to be divided into more groups.

Since the number of groups tends to be larger in larger populations, we need to study the distribution of types across populations holding constant the sample size of individuals used in the estimation. This can be done by taking random samples of equal size from the population of interest (males, females, mixed gender). We present results for samples sizes ranging from 50 to 500.

[Figure 4 about here.]

Figure 4 shows the number of types for males, females, and the population at large. The graphs are presented by the measure of rationality we use (CCEI, MPI or HMI) and the value of ε. Solid lines are the (mean) number of types for the corresponding group and dashed lines show the 95% confidence intervals for the number of types given a population size.

Figure 4 shows two main findings. First, males and females have different preferences. We can conclude this from the fact that the number of types in the population at large is larger than the number of types in the two mono-gender samples (54.6 for the mixed gender group, 48.5 for the males-only group, and 45.8 for the females-only group). In other words, there are types of preferences that are specific to males and females. Second, males are more heterogeneous than females. We can infer this from the fact that the number of types in the males-only group is greater than the number of types in the females-only group (48.5 for the males-only group and 45.8 for the

19

females-only group).[8]

We verify the existence of gender differences in risk preferences by comparing their propensity to deviate from asset allocations that give equal payments in all of the states of the world. Figure 5 shows the average absolute deviation from equal payments for male and female subjects.[9] The average deviation of males is significantly different from that of females (0.1979 vs. 0.1763, t-test = 3.1138, p-value = 0.0019). This comparison is valid since budgets were randomly assigned to subjects. Any systematic difference in asset allocation must be due to difference in risk attitudes. While reassuring, this exercise also illustrates an advantage of our method. The observed difference in asset allocations might be due to the existence of different preference types among male and female subjects or to the fact that some preference types are over-represented in these populations. Our method shows that there are different preference types in these populations. Note that this result cannot be found using the HMI$\times RD_{95CI}$. In this case, the measurement error is too large to detect differences across populations.

[Figure 5 about here.]

4. Discussion

4.1. Robustness

We complement our analysis by exploring two questions: (1) How does the *revealed difference* relation change as stricter versions of rationality are considered? and (2) Are the differences we obtain dependent on the particular context of contingent assets? To answer the first question, we analyze the *revealed difference* relation with the added assumption that subjects respect first-order and second-order stochastic dominance (FOSD, SOSD). This extension is natural since most models of behavior

---

[8]The numbers presented in this paragraph are for RD_0 based on the CCEI at the group size 500. However, the difference is robust to different values of ε and to using the MPI instead of the CCEI.

[9]The formula is $\frac{1}{K}\sum_{i=1}^{i=K}|\frac{x}{x+y} - \frac{1}{2}|$ where x and y is the allocations in each one of the assets and K is the number of budgets faced by the subject.

under risk, neoclassical or not, assume FOSD. This analysis is presented in detail in the Appendix. To answer the second question, we analyze the data in Castillo and Freer (2016) using *revealed difference* relation (see Appendix). These data used budgets defined over goods, not contingent assets.[10]

Our analysis produce two clear results. First, the *revealed difference* relation is sensitive to the validity of the assumptions on rationality. Assumptions like FOSD, which is not supported in these data, hampers our ability to detect heterogeneity across groups. This exercise shows that the measurement of the heterogeneity of preferences is intimately related to the assumptions made about behavior. Wrong, parametric or non-parametric assumptions about preferences will lead to an underestimation of the number of preference types. This also illustrates the general fact that the further are agents from rationality the harder it is to identify the heterogeneity of preferences.[11] The second results is that gender difference in heterogeneity we detected are present in the context of goods as well. This is remarkable given that the data set is much smaller (63 subjects in total).

4.2. Alternative ways to group people

Recently, Crawford and Pendakur (2013) have proposed using revealed preference methods to determine the number of different preference types in a population. The intuition behind their approach is similar to the intuition behind ours, with the difference that they use cross-sectional data with one observation per person. A potential shortcoming of using cross-sectional data to group people into preference types is that the method cannot distinguish between being mutually consistent and not being comparable. That is, a person whose budget sets are contained in the budget sets of another person cannot be revealed different from this person without invoking stronger preference assumptions such as homotheticity. This could generate a serious downward bias on the amount of heterogeneity whenever the number of

---

[10]The goods are: cash, Barnes and Noble gift card, and Mason Money. The last one is the internal monetary system which can be used for any cafes and restaurants on campus.

[11]The same idea is included in the economic significance index by Heufer (2014), especially in the empirical analysis using the index.

goods a person considers is small.

We test whether Crawford and Pendakur's (2013) estimate of heterogeneity might be biased downwards when the set of goods is small. To do this, we consider each budget in our data as if it were a unique individual observation and calculate the number of types using Crawford and Pendakur's (2013) method. We find that a total sample of $29,550$ (25×1182) can be classified into 11 different types according to their method. Note, however, that these results might not be reliable due to the fact that many subjects are not perfectly rational in contradiction with the method's assumptions. However, the method requires restricting the analysis to situations that make decisions across individuals comparable.

5. Conclusions

We use the revealed preference approach to analyze the heterogeneity of preferences. We define two individuals to be *revealed different* if their joint decisions are further from rationality than each of their individual decisions taken separately. The method is useful for analyzing the structure of preferences and assessing the effect of stricter definitions of rationality on the measured heterogeneity of preferences. Using a random sample of Dutch adults, we find that men possess more heterogeneous preferences than women. We also find that stricter definitions of rationality decrease the level of heterogeneity observed in the data. This suggests there is a trade-off between precision in estimates and revealed heterogeneity in preferences in the parametric approach. We show that our results are robust to using alternative definitions of distance from rationality and robust across different contexts. Our method shows that revealed preference analysis can be a useful tool for analyzing differences in preferences across individuals.

Appendix

A: Additional Axioms

We show that adding more unfeasible assumptions is costly in terms of heterogeneity. In the case of contingent assets, every element of $x \in \mathbb{R}_+^L$ corresponds to

the payoff in every state of the world. Let $\pi \in [0;1]^L$ such that $\sum_{i=1}^{L} \pi_i = 1$ denote the probability distribution over the states of the world, where π_i is the probability that state of the world i happens. Denote by $F_x(\xi)$ the cumulative distribution function that corresponds to the lottery x – the probability that x would pay no more than ξ. For every $x, y \in \mathbb{R}_+^L$ let $\xi^i \in \{x_1, \ldots, x_L\} \cup \{y_1, \ldots, y_L\}$ and assume that all ξ^i are ordered in the increasing order. Now we can define First Order Stochastic Dominance and Second Order Stochastic Dominance relations by

$$x \geq_{FSD} y \text{ if } F_x(\xi^i) \leq F_y(\xi^i) \text{ for all } \xi^i$$

and

$$x \geq_{SSD} y \text{ if } \sum_{i=1}^{l} F_x(\xi^i)[\xi^{i+1} - \xi^i] \leq \sum_{i=1}^{l} F_y(\xi^i)[\xi^{i+1} - \xi^i] \text{ for all } l < n$$

A preference relation R satisfies First Order Stochastic Dominance (FOSD) if $\geq_{FSD} \subseteq R$ and Second Order Stochastic Dominance (SOSD) if $\geq_{SSD} \subseteq R$. Note that $\geq_{FSD} \subseteq \geq_{SSD}$ therefore every preference relation that satisfies SOSD satisfies FOSD as well. Further we present the revealed preference axioms that guarantee existence of complete, transitive preference relation that satisfies FOSD (FSD-GARP) and that guarantees existence of complete, transitive preference relation that satisfies SOSD (SSD-GARP). These axioms has been introduced by Heufer (2014) who also shows that they are equivalent to the existence of corresponding complete preference relation.

Denote by $R_{FSD}(e) = R(e) \cup \geq_{FSD}$ and by $T_{FSD}(e)$ denote the transitive closure of $R_{FSD}(e)$. Denote by

$$P_{FSD}(e) = \left\{ (x,y) : x T_{FSD}(e) z (P(e) \cup \geq_{FSD}) z' T_{FSD}(e) y \text{ for some } z, z' \in R_+^L \right\}$$

Similar definitions can be introduced for SOSD we will refer to them just changing the subscript. Let $\sigma_l(x)$ be the l-th permutation of x and by $L!$ the factorial of L.

Denote by

$$\sigma(E) = \{y : y = \sigma_l(x^i) \text{ for some } i = 1, \ldots, n; \ l = 1, \ldots, L!\}$$

Definition 6. *A consumption experiment $E = \{x^i, p^i\}_{i=1}^N$ satisfies the* **Generalized Axiom of Revealed Preference and First Order Stochastic Dominant (FSD-GARP)** *if $s^i T_{FSD}(1) s^j$ implies not $s^j P_{FSD}(1) s^i$ for every $s^i, s^j \in \sigma(E)$.*

Further we introduce the similar condition for SSD-GARP. Let $x \sim_E y$ if the expected payoffs of x and y are equal, i.e. $\pi x = \pi y$. Note that in the case of FSD-GARP it is necessary to check GARP for every permutation of the chosen points, for the SSD-GARP we need to check GARP for the limit points of the set of points that weakly SOSD at least one of the chosen point given that they have same expected value. Further we formally define this set of points.

$$M(x, \{i_1\}) = \left\{ y \in \mathbb{R}_+^L : \ y = \underset{y'(\geq_{SSD} \cap \sim_E)x}{\arg\max} \ y'_{i_1} \right\}$$

Now take a sequence of indices $\{i_j\}_{j=1}^n, n \leq L - 1, 1 \leq i_j \leq L$, then

$$M(x, \{i_1\}_{j=1}^n) = \left\{ y \in \mathbb{R}_+^L : \ y = \underset{y' \in M(x, \{i_1\}_{j=1}^{n-1})}{\arg\max} \ y'_{i_n} \right\}$$

Let $\hat{M}(x)$ denotes the union for all $M(x, \{i_1\}_{j=1}^n)$ for every permutation of indices from 1 to L and

$$\tau(E) = \{y \in \mathbb{R}_+^L : y \in \hat{M}(x^i) \text{ for some } i = 1, \ldots, n\}$$

that is that $\hat{M}(x^i)$ for every chosen point.

Definition 7. *A consumption experiment $E = \{x^i, p^i\}_{i=1}^N$ satisfies the* **Generalized Axiom of Revealed Preference and Second Order Stochastic Dominant (SSD-GARP)** *if $s^i T_{SSD}(1) s^j$ implies not $s^j P_{SSD}(1) s^i$ for every $s^i, s^j \in \tau(E)$.*

We do similar calibration procedure to determine the correct median and confidence interval ε for the FSD-GARP and SSD-GARP. However, in this case we use only the 47 subjects in Choi et al. (2007) symmetric treatment ($\pi = (\frac{1}{2}, \frac{1}{2})$). Table 6 presents the measurement errors obtained from the calibration procedure. We note that these estimates imply a more restrictive notion of heterogeneity.

[Table 6 about here.]

[Table 7 about here.]

Table 7 presents the summary statistics for RD_0, RD_M and RD_{95CI} using GARP, FSD-GARP and SSD-GARP. Figure 6 presents the results for the amount of types obtained using GARP, FSD-GARP and SSD-GARP. The amount of types obtained using GARP, FSD-GARP and SSD-GARP for the whole sample (1182 people) are 131, 85 and 30. We see that additional restriction on the definition of rationality lead to a reduction in the measured amount of hetorogeneity.

[Figure 6 about here.]

Figure 7 presents the results for males and females. Using FSD-GARP, i.e. consistency with first-order stochastic dominance, we cannot reject that males and females are equally heterogeneous. The same holds if we use SSD-GARP. Note that this simple exercise shows us that adding an unfeasible assumption is costly in terms of measuring heterogeneity, since the different preferences would be hidden by the large error terms.

[Figure 7 about here.]

B: Preferences over Consumption Goods

We use the data from Castillo and Freer (2016) who conduct an experiment using real consumption goods. In the experiment, subjects allocate an endowment among three goods: Cash, Barnes and Noble gift card and Mason Money. Barnes and Noble is book distributor and stands for necessities since textbooks are a required

expenditure for students. Mason Money is George Mason internal monetary system that can be used at any on-campus restaurant, therefore, it stands for food spending. The commodities are chosen as to minimize the transaction costs of consumption. The unit of measurement of each commodity is $1. Subjects are asked to allocated a 100 tokens between the above described goods facing prices that are denominated in tokens per dollar. Each subject faces 30 decision problems, one of which is chosen at random to be implemented. The experiment was conducted with 64 George Mason undergraduates of whom 35 are female.

Since we do not have an additional experiment to calibrate the measurement error in the RD relation, we restrict the analysis to the RD_ε relation. Due to the robustness of the results using the CCEI only, we only present the results using this index. Table 8 presents the amount of types found in our experiments.

[Table 8 about here.]

We use the same methodology as in the paper to construct measures of number of types per sample size. Since the total number of men is 28, we estimate these numbers for populations ranging from 5 to 25.

[Figure 8 about here.]

Figure 8 presents the results. The results are similar to those using the data from Choi et al. (2014). Males and females have different preferences and males have more preference types than females.

References

Afriat, S.N., 1967. The construction of utility functions from expenditure data. International Economic Review 8, 67–77.

Afriat, S.N., 1973. On a system of inequalities in demand analysis: an extension of the classical method. International Economic Review , 460–472.

Agnarsson, G., Greenlaw, R., 2006. Graph Theory: Modeling, applications, and algorithms. Prentice-Hall, Inc.

Castillo, M., Freer, M., 2016. A revealed preference test of quasi-linear preferences. Working Paper .

Choi, S., Fisman, R., Gale, D., Kariv, S., 2007. Consistency and heterogeneity of individual behavior under uncertainty. American Economic Review 97, 1921–1938.

Choi, S., Kariv, S., Mueller, W., Silverman, D., 2014. Who Is (More) Rational? American Economic Review 104, 1518–1550.

Crawford, I., Pendakur, K., 2013. How many types are there? Economic Journal 123, 77–95.

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

Gross, J., 1995. Testing data for consistency with revealed preference. Review of Economics and Statistics 77, 701–710.

Gross, J., Kaiser, D., 1996. Two simple algorithms for generating a subset of data consistent with warp and other binary relations. Journal of Business & Economic Statistics 14, 251–255.

Heufer, J., 2014. Nonparametric comparative revealed risk aversion. Journal of Economic Theory 153, 569–616.

Heufer, J., Hjertstrand, P., 2015. Consistent subsets: Computationally feasible methods to compute the houtman–maks-index. Economics Letters 128, 87–89.

Houtman, M., Maks, J., 1985. Determining all maximal data subsets consistent with revealed preference. Kwantitatieve Methoden 19, 89–104.

Nehring, K., Puppe, C., 2002. A theory of diversity. Econometrica 70, 1155–1198.

Pattanaik, P.K., Xu, Y., 2000. On diversity and freedom of choice. Mathematical Social Sciences 40, 123–130.

Sippel, R., 1997. Experiment on the pure theory of consumer's behaviour. Economic Journal 107, 1431–1444.

Smeulders, B., Cherchye, L., Spieksma, F.C., De Rock, B., 2013. The money pump as a measure of revealed preference violations: A comment. Journal of Political Economy 121, 1248–1258.

Varian, H.R., 1982. The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society , 945–973.

Weitzman, M.L., 1992. On diversity. Quarterly Journal of Economics 107, 363–405.

Figures

(a) Critical Cost Efficiency Index

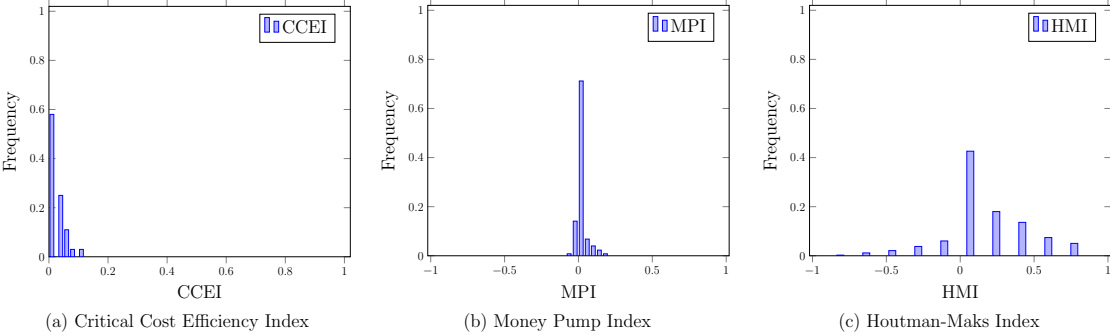(b) Money Pump Index

(c) Houtman-Maks Index

Figure 1: Relative change in distance from rationality (using Choi et al. (2007) data): changing the number of budgets from 25 to 50.
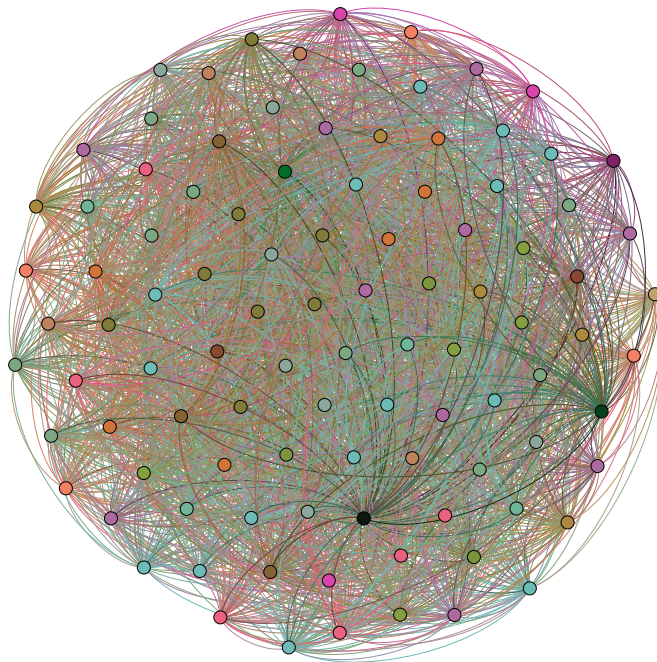
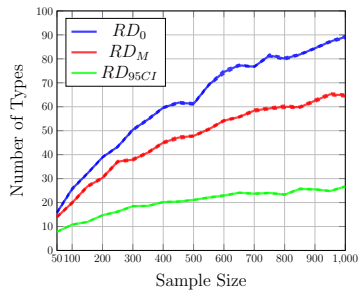Figure 2: Graph of the *revealed different* relation
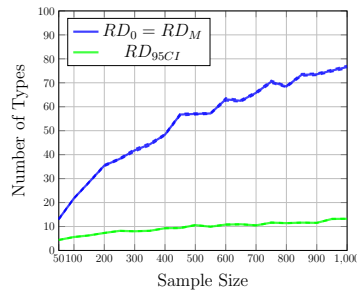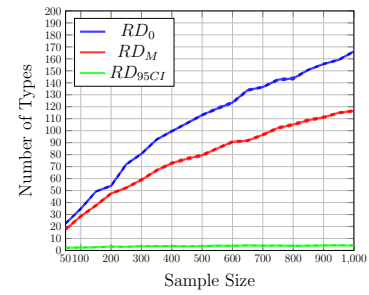
(a) Critical Cost Efficiency Index   (b) Money Pump Index   (c) Houtman-Maks Index

Figure 3: Number of types in the *revealed different* relation for different measures of distance from rationality and levels of ε

Figure 4: Number of types in the *revealed different* relation for mixed gender, males-only and females-only sub-samples

32
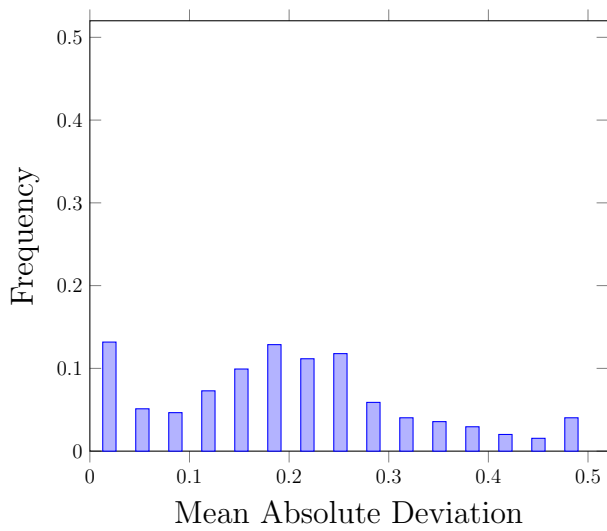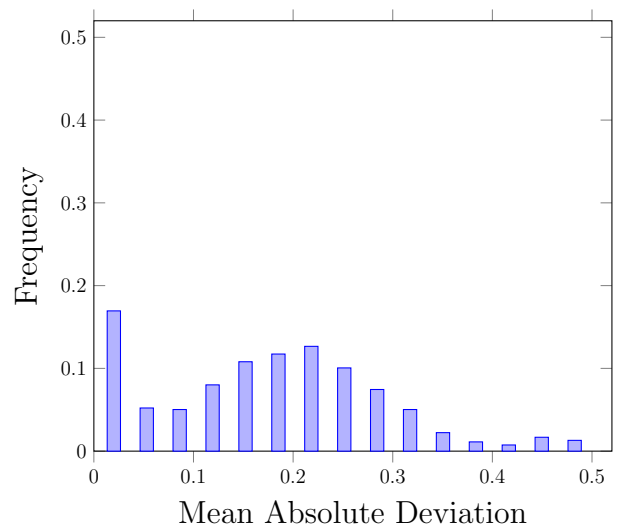
(a) Males

(b) Females

Figure 5: Distributions of mean absolute deviations from 50-50 allocation.
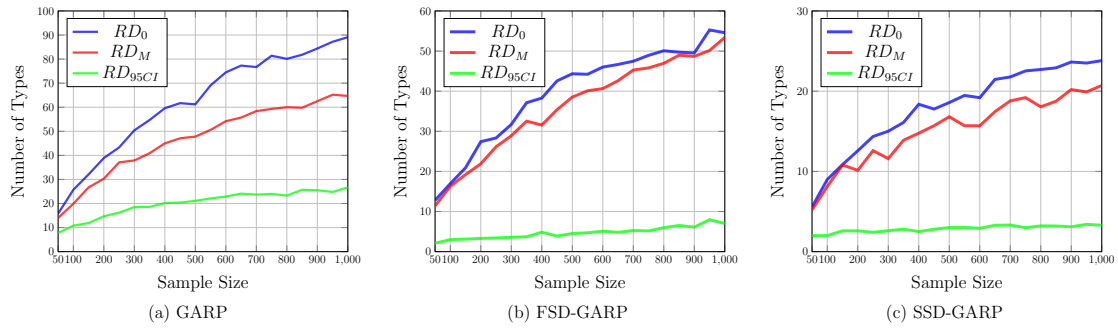
Figure 6: Number in *revealed different* relation for GARP, FSD-GARP and SSD-GARP and different levels of ε

34
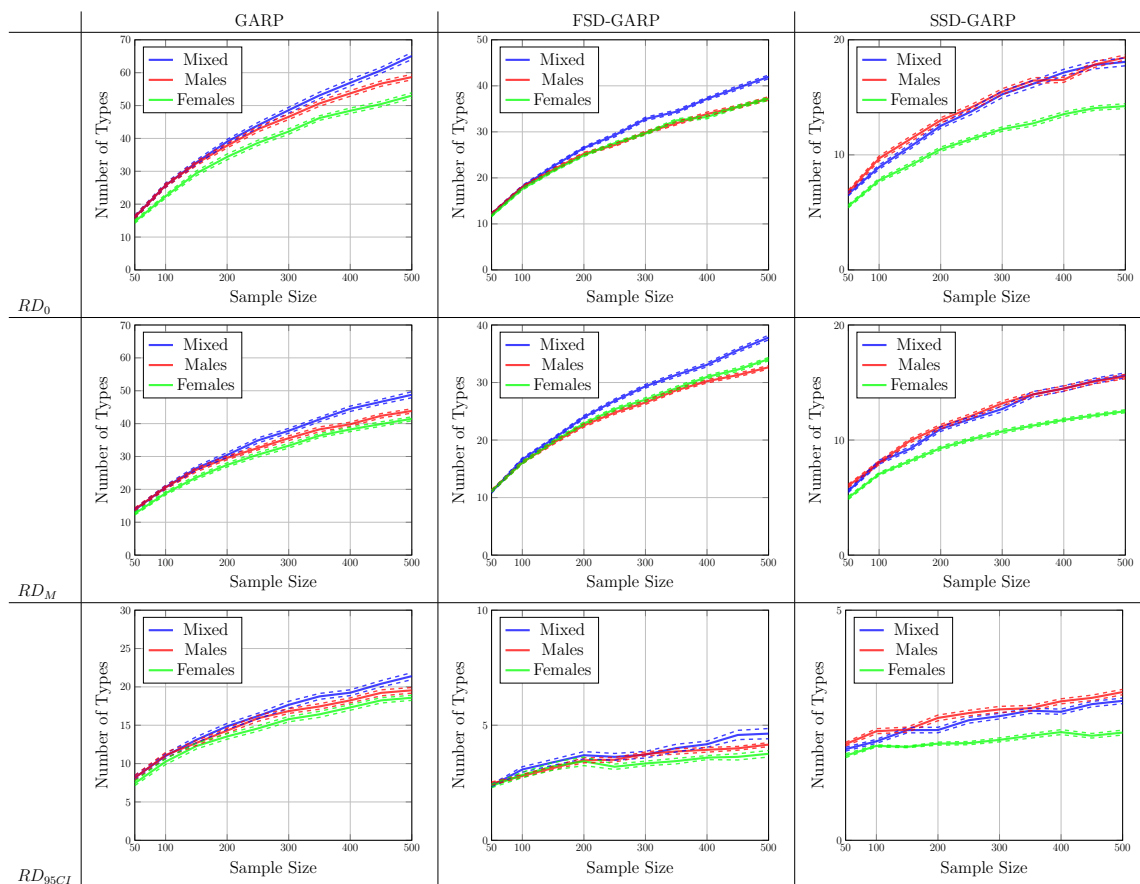
Figure 7: Number of types in the *revealed different* relation for mixed gender, males-only and females-only sub-samples

35

(a) RD_0 (b) RD_M (c) RD_{95CI}

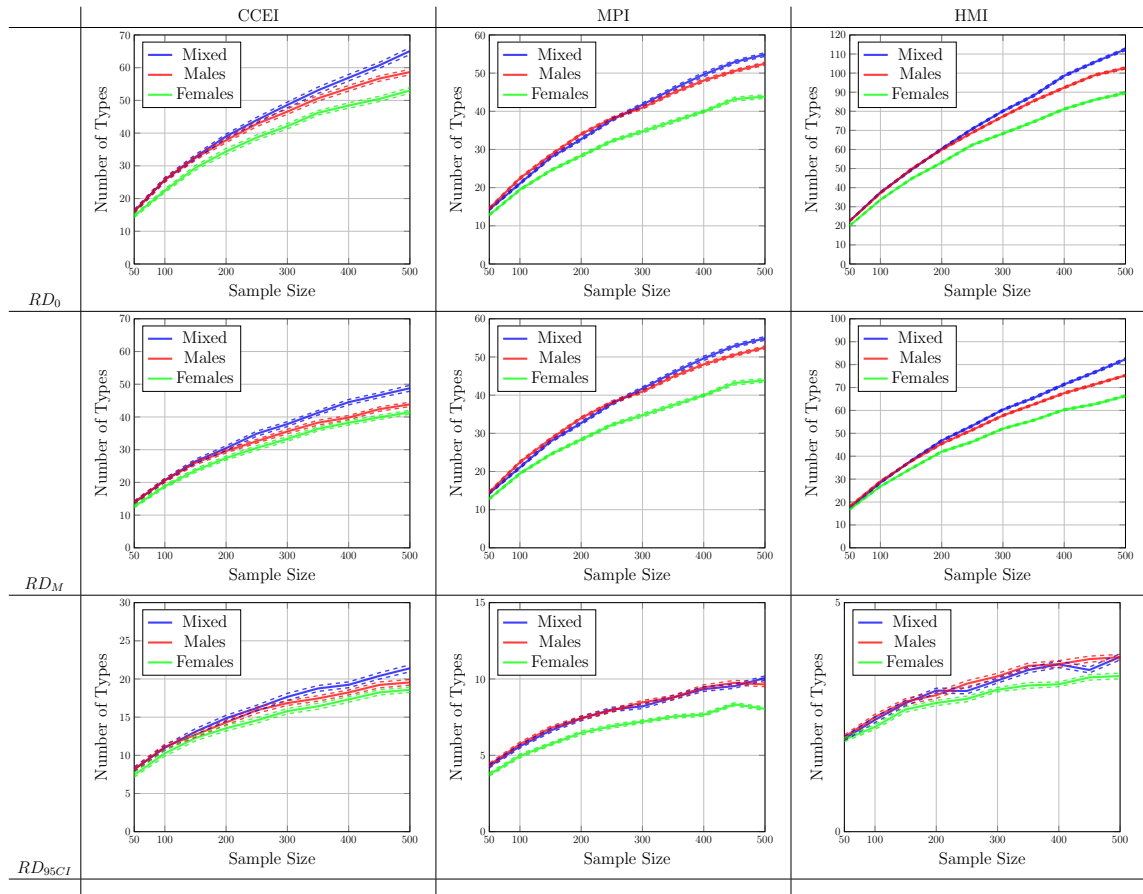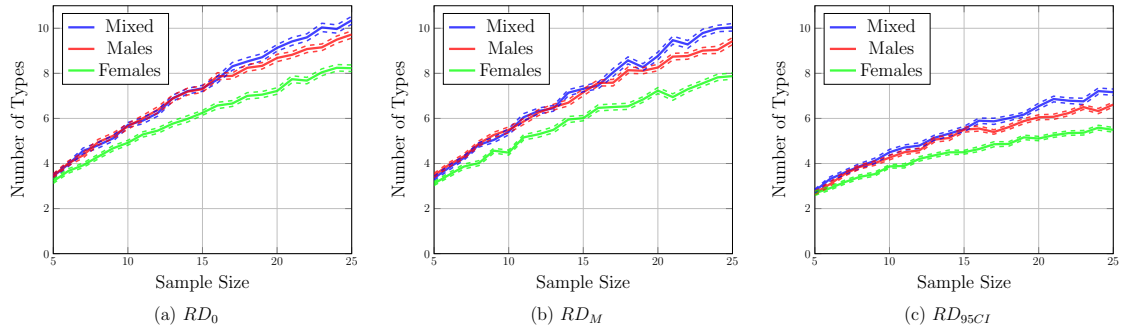Figure 8: Number of types in the *revealed different* relation for mixed gender, males-only and females-only sub-samples in Castillo and Freer (2016) data

Tables

Table 1: Median and 95th percentile of change in distance from rationality ε for *revealed different* relation.

	Median	95th percentile
CCEI	.01	.07
MPI	0	.1
HMI	.1	.7

Table 2: Summary statistics for the probability of being *revealed different*

CCEI

Type of RD	Mean	SD	Min	Max
RD_0	0.5879	0.2006	0.1690	0.9949
RD_M	0.5402	0.2014	0.0093	0.9873
RD_{95CI}	0.3283	0.1988	0.0000	0.9501

MPI

Type of RD	Mean	SD	Min	Max
$RD_0 = RD_M$	0.5282	0.2044	0.0093	0.9882
RD_{95CI}	0.1047	0.1296	0.0000	0.7022

HMI

Type of RD	Mean	SD	Min	Max
RD_0	0.7168	0.1716	0.0059	0.9509
RD_M	0.5645	0.2165	0.0000	0.8790
RD_{95CI}	0.0048	0.0115	0.0000	0.1701

Table 3: Descriptive statistics of the sample of Dutch population from Choi et al. (2014).

	Mean	Std. Dev.
Females	0.454	0.498
Age :		
16-34	0.185	0.389
35-49	0.261	0.440
50-64	0.356	0.479
65+	0.197	0.397
Education:		
Low	0.336	0.472
Medium	0.297	0.457
High	0.364	0.481
Occupation:		
Paid work	0.531	0.499
Housework	0.116	0.320
Retired	0.209	0.407
Others	0.144	0.351
Household Composition:		
Partner	0.809	0.393
Number of children	0.843	1.126
Income :		
0-2499	0.228	0.419
2500-3499	0.255	0.436
3500-4999	0.292	0.455
5000+	0.225	

Table 4: Explaining differences: $y_{i,j} = 1$ if i is revealed different than j, 0 otherwise

		GARP×CCEI			GARP×MPI		GARP×HMI		
VARIABLES	RD_0	RD_0	RD_M	RD_{95CI}	$RD_0 = RD_M$	RD_{95CI}	RD_0	RD_M	RD_{95CI}
Female−Female		-0.580***	-0.274***	-0.340***	-0.391***	-0.058***	0.296***	0.322***	-0.009***
		[0.015]	[0.013]	[0.014]	[0.015]	[0.007]	[0.014]	[0.014]	[0.003]
Male−Male		0.572***	0.271***	0.337***	0.386***	0.056***	-0.299***	-0.326***	0.008***
		[0.015]	[0.013]	[0.014]	[0.015]	[0.007]	[0.014]	[0.014]	[0.003]
Age > 49 & Age > 49		0.540***	0.238***	0.325***	0.666***	0.408***	-0.473***	-0.580***	0.001***
		[0.015]	[0.012]	[0.014]	[0.014]	[0.014]	[0.019]	[0.019]	[0.000]
Age ≤ 49 & Age ≤ 49		-0.509***	-0.226***	-0.310***	-0.636***	-0.399***	0.507***	0.613***	0.001***
		[0.015]	[0.012]	[0.014]	[0.014]	[0.014]	[0.019]	[0.019]	[0.000]
One-level diff. in educ.[+]		-0.005**	0.000	-0.002	-0.006***	0.000	-0.003	-0.003	0.000
		[0.002]	[0.001]	[0.002]	[0.002]	[0.001]	[0.002]	[0.002]	[0.000]
Two-level diff. in educ.[+]		-0.015***	-0.001	-0.005**	-0.018***	-0.001	-0.011***	-0.011***	0.001*
		[0.004]	[0.002]	[0.003]	[0.004]	[0.002]	[0.004]	[0.004]	[0.000]
One-level diff. in income[+]		0.002	0.001	0.002	-0.001	-0.001	0.003***	0.002	0.000
		[0.002]	[0.001]	[0.001]	[0.002]	[0.001]	[0.001]	[0.001]	[0.000]
Two-level diff. in income[+]		0.003	0.002	0.003	-0.001	-0.001	0.006***	0.004**	0.000*
		[0.002]	[0.002]	[0.002]	[0.003]	[0.002]	[0.002]	[0.002]	[0.000]
Three-level diff. in income[+]		0.000	-0.001	0.000	-0.008*	-0.006**	0.007***	0.004	0.001***
		[0.004]	[0.003]	[0.003]	[0.004]	[0.003]	[0.003]	[0.003]	[0.000]
Constant	0.712***	0.755***	0.271***	0.381***	0.681***	0.358***	0.271***	0.101***	0.003
	[0.014]	[0.012]	[0.010]	[0.012]	[0.014]	[0.005]	[0.013]	[0.013]	[0.003]
i's fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
j's fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,397,124	1,397,124	1,397,124	1,397,124	1,397,124	1,397,124	1,397,124	1,397,124	1,397,124
R-squared	0.332	0.332	0.370	0.358	0.335	0.358	0.290	0.381	0.055

Robust standard errors in brackets, clustered at the individual level
*** p<0.01, ** p<0.05, * p<0.10
[+] We use variables as described in Table 3. For instance, "three-level difference in income" equals 1 if one subject has an income in the bracket [0-2499] and the other subject has an income in the bracket [5000+]

Table 5: Explaining differences (logits): $y_{i,j} = 1$ if i is revealed different than j, 0 otherwise

VARIABLES	RD_0	GARP×CCEI			GARP×MPI		GARP×HMI		
		RD_0	RD_M	RD_{95CI}	$RD_0 = RD_M$	RD_{95CI}	RD_0	RD_M	RD_{95CI}
Female-Female		-3.626***	-4.883***	-3.989***	-2.984***	-4.987***	2.103***	2.044***	-0.630
		[0.123]	[0.332]	[0.204]	[0.128]	[1.043]	[0.103]	[0.091]	[0.519]
Male-Male		3.580***	4.865***	3.967***	2.959***	4.924***	-2.125***	-2.071***	0.551
		[0.122]	[0.332]	[0.204]	[0.128]	[1.040]	[0.103]	[0.092]	[0.519]
Age > 49 & Age > 49		3.395***	4.633***	3.905***	4.522***	8.974***	-2.947***	-3.485***	-0.234
		[0.118]	[0.334]	[0.206]	[0.142]	[1.070]	[0.139]	[0.143]	[0.677]
Age ≤ 49 & Age ≤ 49		-3.210***	-4.529***	-3.798***	-4.356***	-8.843***	3.178***	3.694***	0.374
		[0.119]	[0.334]	[0.207]	[0.145]	[1.077]	[0.140]	[0.144]	[0.680]
One-level diff. in educ.[+]		-0.034**	0.005	-0.014	-0.038***	0.021	-0.019	-0.017	0.066
		[0.014]	[0.012]	[0.011]	[0.015]	[0.023]	[0.014]	[0.012]	[0.052]
Two-level diff. in educ.[+]		-0.096***	-0.002	-0.039*	-0.108***	0.029	-0.072***	-0.064***	0.155*
		[0.025]	[0.021]	[0.020]	[0.026]	[0.040]	[0.026]	[0.023]	[0.082]
One-level diff. in income[+]		0.018*	0.017	0.019**	-0.003	0.014	0.026***	0.016**	0.056
		[0.009]	[0.010]	[0.009]	[0.010]	[0.016]	[0.008]	[0.007]	[0.045]
Two-level diff. in income[+]		0.019	0.028*	0.026*	-0.005	0.053**	0.046***	0.037***	0.114*
		[0.014]	[0.015]	[0.014]	[0.015]	[0.027]	[0.011]	[0.011]	[0.068]
Three-level diff. in income[+]		0.012	0.010	0.016	-0.040	0.011	0.059***	0.046***	0.250***
		[0.024]	[0.023]	[0.022]	[0.026]	[0.045]	[0.017]	[0.016]	[0.094]
Constant	1.036***	1.292***	-1.042***	-0.449***	1.017***	-0.715***	-1.413***	-2.371***	-5.486***
	[0.078]	[0.075]	[0.070]	[0.066]	[0.070]	[0.130]	[0.065]	[0.064]	[0.353]
i's fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
j's fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,397,124	1,397,124	1,382,976	1,387,684	1,397,124	1,373,584	1,397,124	1,322,500	233,289
Pseudo R-squared	0.299	0.299	0.377	0.339	0.301	0.529	0.260	0.304	0.236

Robust standard errors in brackets, clustered at the individual level

*** p<0.01, ** p<0.05, * p<0.10

[+] We use variables as described in Table 3. For instance, "three-level difference in income" equals 1 if one subject has an income in the bracket [0-2499] and the other subject has an income in the bracket [5000+]

Table 6: Measurement error for FSD-GARP and SSD-GARP

	Median	95th percentile
GARP	.01	.07
FSD-GARP	.004	.19
SSD-GARP	.008	.22

Table 7: Summary statistics for the probability of being *revealed different*

	GARP			
Type of RD	Mean	SD	Min	Max
RD_0	0.5879	0.2006	0.1690	0.9949
RD_M	0.5402	0.2014	0.0093	0.9873
RD_{95CI}	0.3283	0.1988	0.0000	0.9501
	FSD-GARP			
Type of RD	Mean	SD	Min	Max
RD_0	0.4759	0.1892	0.0034	0.9239
RD_M	0.4531	0.1866	0.0017	0.9196
RD_{95CI}	0.0391	0.0631	0.0000	0.3968
	SSD-GARP			
Type of RD	Mean	SD	Min	Max
RD_0	0.1609	0.1485	0.0000	0.6244
RD_M	0.1453	0.1413	0.0000	0.6091
RD_{95CI}	0.0145	0.0354	0.0000	0.2614

Table 8: Amount of types in Castillo and Freer (2016) data

	Total	Males	Females
RD_0	24	15	14
RD_M	23	14	13
RD_{95CI}	13	9	8