



The Money Pump as a Measure of Revealed Preference Violations: a Comment

Laurens Cherchye
CES, KULeuven

Bram De Rock
SBS-EM, ECARES, Université Libre de Bruxelles

Bart Smeulders
ORSTAT, KULeuven

Frits Spieksma
ORSTAT, KULeuven

ECARES working paper 2013-30

The Money Pump as a Measure of Revealed Preference Violations: A Comment^{*}

Laurens Cherchye[†], Bram De Rock[‡], Bart Smeulders[§],
and Frits C.R. Spieksma[¶]

Abstract

In a recent and insightful paper, Echenique et al. [4] proposed the Money Pump Index (MPI) as an intuitive measure to evaluate the severity of violations of consumer rationality (defined in terms of revealed preference axioms). For practical applications, they suggest using the Mean or Median MPI. In this note, we show that computing these Median and Mean MPI is computationally hard, which makes them impractical in the case of large datasets (including “scanner” datasets as the one used by Echenique et al.). To overcome this problem, we propose Maximum and Minimum MPIs as easy-to-apply alternatives. These MPIs preserve the intuition of the Median and Mean MPIs and can be computed efficiently (i.e. in polynomial time). We also show the practical usefulness of the Maximum and Minimum MPI through an application to the dataset of Echenique et al..

1 Introduction

Irrational behavior makes consumers vulnerable, as it allows arbitrageurs to “pump money” from them. In particular, arbitrageurs can extract money from irrational consumers by following an opposite purchasing strategy. In a recent and insightful contribution, Echenique et al. [4] operationalized this idea by proposing the Money Pump Index (MPI). These authors presented an intuitive MPI that is defined on the basis of revealed preference axioms characterizing rational consumer behavior (such as the Generalized Axiom of Revealed Preference (GARP), which we consider below).

The MPI concept provides an attractive solution to a frequently cited problem of standard revealed preference tests. For a given consumer, these tests are bound to produce a binary result: if the dataset with the consumer’s purchase observations satisfies the revealed preference axioms, then the consumer is said to act rational; alternatively, if the dataset violates the axioms, then the consumer is classified as irrational. However, datasets that fail such a binary test may actually be very close to consistency with rationality. It is therefore of interest to have information on the severity of observed violations. Based on the above money pump idea, the MPI measures the fraction of the budget that arbitrageurs can extract from an irrational consumer. An MPI value that is further away from zero then indicates a more severe violation of rationality (i.e. greater consumer vulnerability).¹

^{*}We thank Federico Echenique for usefull discussion.

[†]Center for Economic Studies, University of Leuven. Laurens Cherchye gratefully acknowledges financial support from the Research Fund K.U.Leuven through the grant STRT1/08/004.

[‡]Corresponidng author. ECARES and ECORE, Université Libre de Bruxelles, Avenue F.D. Roosevelt 50, CP 114-04, B-1050 Brussels, Belgium, e-mail: bderock@ulb.ac.be. Bram De Rock gratefully acknowledges the European Research Council (ERC) for his Starting Grant.

[§]ORSTAT, University of Leuven.

[¶]ORSTAT, University of Leuven.

¹We note that Afriat [1] introduced the first index to quantify the severity of violations of revealed preference axioms. This index was followed by a considerable number of other goodness-of-fit measures, with differing properties

This note is concerned with the practical computation of the MPI. As we will explain below, if a consumer violates rationality, then typically there will be multiple purchase observations implying such a violation. In principle, we can compute a money pump cost for each violation. This calls for an aggregate MPI that summarizes these money pump costs into a single metric. In their original contribution, Echenique et al. propose the mean and median money pump cost as such aggregate MPIs. Obviously, these Mean and Median MPIs have an intuitive interpretation in terms of the money lost by the consumer due to irrational behavior. A first contribution of this note is that we show that computing the Mean and Median MPIs is an NP-hard problem.² This result provides a formal statement of the fact that it is computationally challenging to compute these measures in practice, in particular for datasets with large numbers of observations.

Notable examples of such large datasets are household-level “scanner” datasets, which Echenique et al. [4] also considered in their empirical application. Scanner datasets contain information on household-level purchases collected at checkout scanners in supermarkets. They typically consist of multiple purchase observations for many households. Such large datasets are increasingly available, and Echenique et al. provide a particularly convincing case on the usefulness of their MPI concept in combination with scanner data. At this point, however, it is worth emphasizing that they also extensively discussed the computational complexity of the MPI for their own scanner dataset (see in particular their Remark 1 on p. 1207). To mitigate the computational burden, they therefore suggested as a practical method to compute approximations of the Mean and Median MPIs. Essentially, these approximations focus on violations of revealed preference axioms that involve only a small number of observations (see Section 5 for more details).

Because of the computational difficulties associated with the Mean and Median MPIs, our second contribution is that we propose the Maximum and Minimum MPIs as easy-to-apply alternatives. The Maximum MPI gives the percentage of money lost in the most severe violation of rationality, while the Minimum MPI does the same for the least severe violation. Clearly, these measures preserve the intuition underlying the Mean and Median MPIs. In particular, they figure as natural bounds on the amounts of money that an arbitrageur can extract from irrational consumers.

Importantly, our newly proposed Maximum and Minimum MPIs have clear practical usefulness. We show that the Maximum and Minimum MPIs can be computed efficiently (i.e. in polynomial time), which makes them easily applicable to large (e.g. scanner) datasets. We also indicate how such computation can proceed in practice. Next, we use the dataset of Echenique et al. [4] to demonstrate the application of the Maximum and Minimum MPIs. Here, our particular focus is on assessing the performance of these measures relative to the Mean and Median MPIs. In addition, we show that comparing the values of the Maximum and Minimum MPIs can reveal interesting information to the empirical analyst. This makes us believe that our results may contribute to the further dissemination of the intuitive MPI concept in empirical analyses of (ir)rational consumer behavior.

The rest of this note unfolds as follows. Section 2 sets the stage by introducing the Generalized Axiom of Revealed Preference (GARP), which forms the basis for the MPI. Section 3 then introduces the MPI concept and the associated notions of Mean, Median, Maximum and Minimum MPI. Section 4 contains our core results and defines the computational complexity of the different MPIs that we consider. Section 5 shows the practical usefulness of our results through an application to the scanner dataset of Echenique et al.. Section 6, finally, concludes.

in regards to outliers, the number of violations, complexity of calculation, etc. See Echenique et al. [4] for a discussion on the relation between these measures and the MPI.

²We refer to Garey and Johnson [5] for an introduction into the theory of NP-completeness.

2 Generalized Axiom of Revealed Preference

Suppose that we have a dataset, $S = \{(p_i, q_i) \mid i = 1, \dots, n\}$, of n observed purchases by a consumer. That is, we observe N -dimensional vectors of prices $p_i \in \mathbb{R}_{++}^N$ and quantities $q_i \in \mathbb{R}_+^N$ for every observation $i = 1, \dots, n$.

To explain the concept of revealed preferences, we need to consider two observations i and j . If $p_i q_i \geq p_i q_j$, then bundle q_i is *directly revealed preferred* to bundle q_j , since at prices p_i both bundles are available but q_i is chosen. This is expressed by writing $q_i R_0 q_j$. The transitive closure of R_0 is denoted by R and is called the indirect revealed preference relation. If $p_i q_i > p_i q_j$, we say that bundle q_i is *strictly directly revealed preferred* to bundle q_j , which is denoted by $q_i P_0 q_j$.

We can now define the Generalized Axiom of Revealed Preference (GARP).

Definition 1. (GARP): A data set S satisfies GARP if for each pair of bundles, q_i, q_j , ($i, j = 1, \dots, n$, with $i \neq j$), the following holds: if $q_i R q_j$ then it is not the case that $q_j P_0 q_i$.

In words, GARP states that it cannot be that bundle q_i is preferred over bundle q_j and that at the same time bundle q_i costs at prices p_j strictly less than bundle q_j . Intuitively this does not comply with the basic notion of a rational consumer that maximizes his/her preferences given the budget constraint. Varian [8] (based on Afriat [1]) formalized this result by showing that a dataset S satisfies GARP if and only if it is “rationalizable” by a well-behaved utility function. This means that each observed consumer purchase can be represented as maximizing this utility function subject to the observed budget (which is assumed to be equal to the observed expenditures). Given this, consistency with GARP provides a natural rationality condition for a consumer dataset; see, e.g., Varian [8] for more discussion. The money pump index is an example of a goodness-of-fit measure: it gives an indication of the severity of the violation of GARP. Other goodness-of-fit measures exist: we mention Afriat’s index, Varian’s index and an index proposed by Houtman and Maks, see Smeulders et al. [7] for a discussion.

3 Money Pump Index

As explained by Echenique et al. [4], if GARP is violated, a money pump cost (MPC) can be calculated for every violation. This MPC is the amount of money an arbitrageur could gain from the consumer by following an appropriate buying strategy. More precisely, suppose that we have two observations i and j for which $p_i q_i \geq p_i q_j$ and $p_j q_j > p_j q_i$. This implies a violation of GARP that involves the observations i and j . Then, the arbitrageur can make money by buying bundle q_i at prices p_j and reselling it at p_i , and by buying q_j at prices p_i and reselling it at p_j . The total profit following from these transactions gives the corresponding MPC, which equals

$$p_i(q_i - q_j) + p_j(q_j - q_i). \quad (1)$$

Generalizing this argument, we can compute the MPC associated with a GARP violation involving a sequence of observations i_1, i_2, \dots, i_k as follows

$$MPC = \sum_{j=1}^k p_{i_j}(q_{i_j} - q_{i_{j+1}}), \quad (2)$$

with $q_{i_{k+1}} = q_{i_1}$.

To make meaningful comparisons between GARP violations involving different sequences, the Money Pump Index (MPI) of a violation is calculated by dividing the associated MPC by the total budget of the observations that are involved in the violation. That is

$$MPI = \frac{\sum_{j=1}^k p_{i_j} (q_{i_j} - q_{i_{j+1}})}{\sum_{j=1}^k p_{i_j} q_{i_j}}, \quad (3)$$

with again $q_{i_{k+1}} = q_{i_1}$.

If a dataset for a given consumer violates GARP, then there are typically several sequences of observations that are involved in a violation. Therefore, Echenique et al. [4] introduce the Mean and Median MPI of consumers as measures of consumer irrationality. More precisely, each violation gives rise to an MPI value (as defined in (3)). The Mean MPI is then defined as the mean of these MPI values, while the Median MPI equals the median of these values. These measures indeed have an intuitive meaning as quantifying the severity of consumer irrationality.

4 Complexity results

In their original contribution, Echenique et al. already argued that computing the Mean and Median MPI is a challenging task and, therefore, they propose to approximate these MPIs in practical applications (see also Section 5). In what follows, we will formally state that computing the Mean and Median MPI is indeed an NP-hard problem.

To demonstrate that the Mean and Median MPI are computationally hard, we derive a reduction from the problem #Cycle. Arora and Barak [3] showed that a polynomial time algorithm for #Cycle implies that $P = NP$. Thus, our reduction implies that computing the Mean and Median MPIs is NP-hard, meaning that there cannot exist a polynomial time algorithm for computing them (unless $P=NP$). The Appendix contains the proofs of our results.

Theorem 1. *The Mean MPI and the Median MPI cannot be calculated in polynomial time, unless $P = NP$.*

We suggest using the Maximum and Minimum MPIs as easy-to-apply alternative measures of irrationality. These measures are calculated as, respectively, the maximum and minimum MPI values defined over all violations. Interestingly, we can prove that these Maximum and Minimum MPIs can be computed in polynomial time, which makes them particularly attractive from an empirical point of view.

Theorem 2. *The Maximum MPI and the Minimum MPI can be calculated in polynomial time.*

We prove this theorem by reducing the problem of computing the Maximum and Minimum MPIs to the Minimum Cycle Ratio problem (see the Appendix). Since Megiddo [6] showed that one can compute the Minimum Cycle Ratio in polynomial time, this reduction proves our result.

5 Empirical application: deterministic test results

We next compute the newly proposed Maximum and Minimum MPIs for the dataset reported in Echenique et al. [4]. This dataset contains 494 households (i.e. 494 consumers), with 26 purchase

Table 1: Descriptive statistics for alternative MPIs

	Minimum MPI	Maximum MPI	Range	Approximated Mean MPI	Approximated Median MPI
Average	0.0341	0.0936	0.0595	0.0610	0.0591
Standard Deviation	0.0287	0.0616	0.0603	0.0359	0.0369
Number of zeros	0	0	74	0	0
Minimum	0.0002	0.0048	0.0000	0.0048	0.0048
First quartile	0.0154	0.0489	0.0115	0.0355	0.0322
Median	0.0268	0.0797	0.0443	0.0543	0.0497
Third quartile	0.0429	0.1274	0.0867	0.0809	0.0791
Maximum	0.2782	0.4010	0.3350	0.2782	0.2782

observations per household. Out of these 494 households, there are 396 that violate GARP. The numbers reported in Table 1 pertain to this subset of households.

To compute our results for the Maximum and Minimum MPIs, we implemented an algorithm described in Ahuja et al. [2] for solving the Minimum Cycle Ratio problem. This algorithm is very quick in practice: we needed only a few seconds to compute the results for all 494 households.

Table 1 presents summary statistics on the different MPIs under consideration. Let us first consider our findings for the maximum and minimum MPIs. As indicated in the Introduction, we believe these results reveal interesting information, as they tell about maximum and minimum amounts of money that an arbitrageur can extract from irrational consumers. We find that the average Maximum MPI equals 9.35%, while the average Minimum MPI amounts to 3.41%. However, the corresponding standard deviations also reveal that these numbers hide quite some variation across households. Next, we observe that the range between the Maximum and Minimum MPIs is on average 5.95%, and that this range also varies quite substantially across households. In this respect, however, it is also worth noting that the range turns out to be zero for no less than 74 households, i.e. for about one-fifth of our total sample (with 396 households) we obtain that the Maximum MPI exactly equals the Minimum MPI.

As a final base of comparison, we compare our results to the ones reported by Echenique et al.. As indicated above, these authors recognized the complex nature of the Mean and Median MPIs and therefore resorted to computing approximations of these MPIs in their empirical application. In particular, they approximated the Mean and Median MPIs by focusing on short violations only, i.e. violations consisting of at most four observations. Table 1 reports the associated descriptive statistics. When comparing Echenique et al.’s results for the Mean and Median MPIs to the ones for our Maximum and Minimum MPIs, we conclude that, in many cases, these last two “extreme” MPIs spread symmetrically around the first two “central” MPIs. This suggests that the average of the Maximum and Minimum MPIs may actually provide relevant information. In particular, these numbers can be used to obtain a good estimate of the Mean and Median MPIs.

6 Conclusion

We have shown that the Mean and Median MPIs originally proposed by Echenique et al. [4] are generally difficult to compute, which makes them impractical in the case of large datasets (including scanner datasets). As alternatives, we therefore proposed the Maximum and Minimum MPIs. These MPIs can be computed efficiently (i.e. in polynomial time) and preserve the attractive interpretation of the Mean and Medium MPI. We also demonstrated the practical usefulness of these Maximum and Minimum MPIs through an application to the scanner dataset which Echenique et al. also studied. We hope that our results will contribute to the further dissemination of the intuitive MPI concept in empirical analyses of (ir)rational consumer behavior.

Appendix

Proof of Theorem 1

Proof for the Mean MPI. We first explain the concept of a so-called graph representation of a dataset S , denoted by $G(S)$. Given a dataset S , $G(S)$ arises by having a node in $G(S)$ for each observation in S , and having an arc between nodes i and j whenever $p_i q_i \geq p_i q_j$.

Consider an instance of the problem $\#Cycle$, that is we have a directed graph $G = (V, A)$, $|V| = n$ with the question: how many directed cycles exist in G ? We will answer the question by computing the Mean MPI of two specially constructed sets of consumer data S_1 and S_2 . Both datasets consist of $n + 2$ observations. In fact, observations $1, 2, \dots, n$ are identical for S_1 and S_2 and can be described as follows.

For each vertex $i \in V$, we construct a price vector p_i with $(p_i)_j = \epsilon$ for $i \neq j$ ($\epsilon < \frac{1}{2n}$) and $(p_i)_i = 1$. For every vertex i we create a quantity vector q_i with $(q_i)_i = 1$, $(q_i)_j = 0$ if there is an arc from j to i in G (for $i \neq j$), and $q_j^i = 2$ if there is no arc (again for $i \neq j$). Observe that an arc in G corresponds to an arc in the graph representation of the dataset consisting of these n observations, and vice versa. We now finish the description of S_1 , by specifying observations $n + 1$ and $n + 2$ as follows. Let $p_{n+1} = (2, 1, 1, \dots, 1)$, $p_{n+2} = (1, 2, 1, \dots, 1)$, $q_{n+1} = (3, 2, 2, \dots, 2)$ and $q_{n+2} = (2, 3, 2, \dots, 2)$. Notice that no observation $\{1, 2, \dots, n\}$ is preferred over observation $n + 1$ and $n + 2$. Further notice that observation $n + 1$ is preferred over $n + 2$, and vice versa. Hence, the number of cycles in $G(S_1)$ is $1 +$ the number of cycles in G , or $1 + \#Cycle$ for short. In particular, we can easily verify that the MPI (see(3)) of the additional cycle equals $\frac{2}{2n+4}$. Let us write $MPI(C)$ for the value of the MPI corresponding to a cycle C in the graph representation of the dataset S . Then, the mean MPI of dataset S can be written as:

$$MPI = \frac{\sum_{C \in G(S_1)} MPI(C)}{\#Cycle + 1} = MPI = \frac{\sum_{C \in G([1, 2, \dots, n])} MPI(C) + \frac{2}{2n+4}}{\#Cycle + 1} \quad (4)$$

(Where $G([1, 2, \dots, n])$ stands for the graph representation of dataset S , restricted to observations $1, 2, \dots, n$.)

Now, we finish the description of dataset S_2 by specifying observations $n + 1$ and $n + 2$ as follows: $p_{n+1} = (2, 1, 1, \dots, 1)$, $p_{n+2} = (1, 2, 1, \dots, 1)$, $q_{n+1} = (4, 2, 2, \dots, 2)$ and $q_{n+2} = (2, 4, 2, \dots, 2)$. As in dataset S_1 there is one additional cycle between nodes $n + 1$ and $n + 2$, which has MPI equal to $\frac{2}{2n+6}$. Thus, the mean MPI of dataset S_2 can be written as.

$$MPI = \frac{\sum_{C \in G(S_1)} MPI(C)}{\#Cycle + 1} = MPI = \frac{\sum_{C \in G([1,2,\dots,n])} MPI(C) + \frac{2}{2n+6}}{\#Cycle + 1} \quad (5)$$

Now suppose we have a polynomial time algorithm for finding the mean MPI of a dataset, then we can find the mean MPI for S_1 and S_2 , compute the difference and with the knowledge that this difference is $\frac{\frac{2}{2n+6} - \frac{1}{2n+4}}{\#cycle+1}$ find $\#Cycle$. This implies that we would have a polynomial time algorithm for solving $\#Cycle$, which in turn implies $P = NP$.

Proof for the Median MPI. Consider an instance of the problem $\#Cycle$. We will solve this problem by computing the median MPI of a polynomial number ($n \log(n)$) of specially constructed sets of consumer data. First, number the vertices of G from 1 to n . We will then construct a dataset D^- with n observations and n goods as follows. For every vertex $i \leq n$ we construct a price vector p_i with $(p_i)_j = \epsilon$ for $j \neq i$ ($\epsilon < \frac{1}{2n^2}$), $(p_i)_i = 1$. For every vertex i we create a quantity vector q_i with $(q_i)_i = 1$, $(q_i)_j = 0$ if there is an arc from j to i in G (for $i \neq j$), and $(q_i)_j = 2$ if there is no arc (again for $i \neq j$). It can be easily checked that the graph representation of D^- has the same set of arcs as the original graph G . It follows that both have the same number of cycles. Given the construction, it is easy to see that the upper bound on the budgets is $1 + \frac{2(n-1)}{2n^2}$ and $1 - \frac{1}{n} + \frac{1}{n^2}$ a lower bound on $p_i q_i - p_i q_j$ the arc (i, j) exists. It follows that $\frac{n^2 - (n-1)}{n^2 - (n+1)} > 0.5$ is a lower bound on the minimum MPI of dataset D^- .

We now add 2 more observations and 2 goods to D^- , creating D . For every observation $i < n+1$, $(p_i)_{n+1} = (p_i)_{n+2} = 2$ and $(q_i)_{n+1} = (q_i)_{n+2} = 0$. Set $p_{n+1} = (\epsilon, \dots, \epsilon, 1, 0.5)$, $p_{n+2} = (\epsilon, \dots, \epsilon, 0.5, 1)$, $q_{n+1} = (0, \dots, 0, 1, 0)$ and $q_{n+2} = (0, \dots, 0, 0, 1)$. It is clear that $n+1$ and $n+2$ are preferred over every other observation and that no observation $i < n$ is preferred over either $n+1$ or $n+2$. In this way, one more violation is added, with an MPI of 0.5. It follows that the minimum MPI of dataset D has a value of 0.5 and that there is one unique violation that has this MPI.

Now, consider that we add additional goods and observations to the dataset D , creating D^+ . For these new goods and observations, the prices and quantities are so that all existing violations remain with the same MPI, a known number of new violations are created, and the MPI of these new violations are smaller than the minimum MPI of violations in D . It is clear that if the Median MPI of D^+ is equal to the minimum MPI of D , then the number of new violations created in D^+ is equal to the number of violations in D and thus one more than the number of cycles in G . We will now show that we can efficiently add new goods and observations to D to create a known number of extra violations in D^+ , and that creating a polynomial number of datasets D^+ is sufficient for finding a D^+ for which the median MPI is equal to the minimum MPI of D .

First, we notice that G has less than $O(n \times n!) < O(n^{n+1})$ cycles. A binary search over this number can be done in $O(\log(n^{n+1})) = O((n+1) \log(n))$ time. At each step in this binary search, a component C is added to D so that $D \cup C = D^+$. This component is created as follows. Let $f(k)$ be the number of cycles in a fully connected digraph. Now assume x arcs must be added to D to form D^+ , then find $\max_k (f(k) < x)$ and add $\left\lfloor \frac{x}{f(k)} \right\rfloor$ subcomponents of k observations to C . Set the prices and quantities so that all observations within one subcomponent are preferred

to all other observations in that group, and so that the MPI of these violations are smaller than the minimum MPI of D and so that no cycles that include observations of multiple subcomponent exist.³ It is easy to see that for a given x we can efficiently find the groups to be added and, as $(2k + 1) \times f(k) > f(k + 1)$, the number of subcomponents is polynomial.

In conclusion, the #Cycle problem for a graph G can be solved by calculating the Median MPI of at most $O((n + 1) \log(n))$ graphs, which can be constructed in polynomial time and have a size that is polynomial in the size of the graph G . As such, a polynomial time algorithm for the Median MPI would mean a polynomial time algorithm for #Cycle, which implies $P = NP$.

Proof of Theorem 2

Proof for the Maximum MPI. We prove this theorem by reducing the problem of computing the Maximum MPI to the Minimum Cycle Ratio problem (see Ahuja et al. [2]). This is done by first creating a graph corresponding to the consumption data. More precisely, we build the graph G , corresponding to the dataset S . For every observation $i = 1, \dots, n$ in the dataset, there is a vertex i in the graph. If and only if $q_i R_0 q_j$, there is a directed arc $(i, j) \in E$. Each arc has a weight $w(i, j)$ and a budget $b(i, j)$. We set the weight of each arc, $w(i, j) := p_i q_j - p_i q_i$ and $b(i, j) := p_i q_i$. Notice that for a given cycle in the graph, the ratio of the sum of the weights of the arcs over the sum of budgets of the arcs is the MPI of the corresponding GARP violation. It is clear that the cycle with the minimum cycle ratio in this graph corresponds to the violation with the maximum MPI. Since constructing the graph is possible in $O(n^2)$ time, and Megiddo [6] showed that computing the Minimum Ratio Cycle has a time complexity of $O(n^3 \log n)$, the theorem follows.

Proof for the Minimum MPI. The proof of this result is directly analogous to the one above. For compactness, we do not include it here.

References

- [1] S.N. Afriat. The construction of utility functions from expenditure data. *International Economic Review*, 8(1):67–77, 1967.
- [2] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, 1993.
- [3] S. Arora and B. Barak. *Computational complexity: a modern approach*, pages 171–186. Cambridge University Press, 2009. Chapter 9: Complexity of Counting.
- [4] F. Echenique, S. Lee, and M. Shum. The money pump as a measure of revealed preference violations. *Journal of Political Economy*, 119(6):1201–1223, 2011.
- [5] M.R. Garey and D.S. Johnson. *Computers and intractability*, volume 174. Freeman San Francisco, CA, 1979.

³It can be easily seen that this is possible by adding as many goods as observations and ranking the subcomponents. For every observation i added, set $(q_i)_i = 1$, $(q_i)_j = 0$ otherwise. Set $(p_i)_i = 1$, $(p_i)_j = 0.75$ if j is associated with an observation in the same subcomponent. If j is associated with an observation in a higher subcomponent $(p_i)_j = 2$. Finally, if j is associated with an observation in a lower subcomponent or D , set $(p_i)_j = \epsilon$.

- [6] N. Megiddo. Combinatorial optimization with rational objective functions. *Mathematics of Operations Research*, 4(4):414–424, 1979.
- [7] B. Smeulders, L. Cherchye, B. De Rock, and F. Spieksma. Goodness of fit measures for revealed preference tests: Complexity results and algorithms. *Transactions on Economics and Computation*, forthcoming, 2013.
- [8] H.R. Varian. The nonparametric approach to demand analysis. *Econometrica*, 50(4):945–973, 1982.