CHICAGO JOURNALS

The Money Pump as a Measure of Revealed Preference Violations: A Comment
Author(s): Bart Smeulders, Laurens Cherchye, Bram De Rock, and Frits C. R. Spieksma
Source: *Journal of Political Economy*, Vol. 121, No. 6 (December 2013), pp. 1248-1258
Published by: The University of Chicago Press
Stable URL: http://www.jstor.org/stable/10.1086/674077
Accessed: 03/02/2014 04:28

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Journal of Political Economy.*

http://www.jstor.org

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

Comment

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

Bart Smeulders

University of Leuven

Laurens Cherchye

University of Leuven

Bram De Rock

Université Libre de Bruxelles

Frits C. R. Spieksma

University of Leuven

I. 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, Lee, and Shum (2011) 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, which we consider below).

We thank the editor Phil Reny and Federico Echenique, Sangmok Lee, and Matthew Shum for useful discussion. Laurens Cherchye gratefully acknowledges financial support from the Research Fund KU Leuven through the grant STRT1/08/004. Bram De Rock gratefully acknowledges the European Research Council for his Starting Grant. Frits Spieksma's research has been partially funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Office.

[*Journal of Political Economy,* 2013, vol. 121, no. 6]
© 2013 by The University of Chicago. All rights reserved. 0022-3808/2013/12106-0001$10.00

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

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 data set with the consumer's purchase observations satisfies the revealed preference axioms, then the consumer is said to act rationally; alternatively, if the data set violates the axioms, then the consumer is classified as irrational. However, data sets 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. On the basis of 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]

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. (2011) 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 as a result of 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 data sets with large numbers of observations.

Notable examples of such large data sets are household-level "scanner" data sets, which Echenique et al. also considered in their empirical application. Scanner data sets contain information on household-level purchases collected at checkout scanners in supermarkets. They typically consist of multiple purchase observations for many households. Such large data sets 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 data set (see in particular their

[1] We note that Afriat (1967) 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 in regard to outliers, the number of violations, complexity of calculation, etc. See Echenique et al. (2011) for a discussion on the relation between these measures and the MPI.

[2] We refer to Garey and Johnson (1979) for an introduction to the theory of NP-completeness.

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

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 Sec. V 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) data sets. We also indicate how such computation can proceed in practice. Next, we use Echenique et al.'s data set 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 II sets the stage by introducing the generalized axiom of revealed preference (GARP), which forms the basis for the MPI. Section III then introduces the MPI concept and the associated notions of mean and median MPIs. Section IV contains our core results: it introduces maximum and minimum MPIs and defines the computational complexity of the different MPIs that we consider. Section V shows the practical usefulness of our results through an application to the scanner data set of Echenique et al. Section VI, finally, presents conclusions.

II. Generalized Axiom of Revealed Preference

Suppose that we have a data set, $S = \{(p_i, q_i) | i = 1, \ldots, n\}$, 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 $i = 1, \ldots, n$.

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*

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

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.

DEFINITION 1 (GARP). A data set S satisfies GARP if for each pair of bundles q_i, q_j ($i, j = 1, \ldots, n$, with $i \neq j$), the following holds: if $q_i R_0 q_j$, 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 who maximizes his or her preferences given the budget constraint. Varian (1982) (based on Afriat [1967]) formalized this result by showing that a data set 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 data set; see, for example, Varian (1982) for more discussion. The MPI 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 (1985); see Smeulders et al. (forthcoming) for a discussion.

III. Money Pump Index

As explained by Echenique et al. (2011), 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 purchasing 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

$$p_i(q_i - q_j) + p_j(q_j - q_i). \tag{1}$$

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

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

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

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

To be able to make meaningful comparisons between GARP violations involving different sequences, the 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,

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

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

If a data set for a given consumer violates GARP, then there are typically several sequences of observations that are involved in a violation. Therefore, Echenique et al. introduce the mean and median MPIs 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.

IV. Complexity Results

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

To demonstrate that the mean and median MPIs are computationally hard, we derive a reduction from the problem #cycle. Arora and Barak (2009, chap. 9) 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 in general (unless $P = NP$).

THEOREM 1. The mean MPI and the median MPI cannot be calculated in polynomial time unless $P = NP$.

The Appendix contains the proofs of our results. In these proofs, the number of goods, N, is not bounded by a constant. Hence, it remains an open question whether or not polynomial time algorithms exist in the case of a fixed number of goods. Recent results by Deb and Pai (2013) show that, given a large number of observations compared to the num-

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

ber of goods, some structure will appear in the preference relations. It is possible that this structure can be exploited to find polynomial time algorithms in these cases. Given our focus, we decided not to explore this further in this note.

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 time required to compute the maximum MPI and the minimum MPI is polynomial in the number of observations.

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 (1979) showed that one can compute the minimum cycle ratio in polynomial time, this reduction proves our result.

V. Empirical Application: Deterministic Test Results

We next compute the newly proposed maximum and minimum MPIs for the data set reported in Echenique et al. (2011). This data set contains 494 households (i.e., 494 consumers), with 26 purchase 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, Magnanti, and Orlin (1993) 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
DESCRIPTIVE STATISTICS FOR ALTERNATIVE MPIS

	Minimum MPI	Maximum MPI	Range	Approximated Mean MPI	Approximated Median MPI
Average	.0341	.0936	.0595	.0610	.0591
Standard deviation	.0287	.0616	.0603	.0359	.0369
Number of zeros	0	0	74	0	0
Minimum	.0002	.0048	.0000	.0048	.0048
First quartile	.0154	.0489	.0115	.0355	.0322
Median	.0268	.0797	.0443	.0543	.0497
Third quartile	.0429	.1274	.0867	.0809	.0791
Maximum	.2782	.4010	.3350	.2782	.2782

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

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 that 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 percent, while the average minimum MPI amounts to 3.41 percent. 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 percent 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 fewer than 74 households; that is, 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, that is, 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.

VI. Conclusion

We have shown that the mean and median MPIs originally proposed by Echenique et al. (2011) are generally difficult to compute, which makes them impractical in the case of large data sets (including scanner data sets). 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 MPIs. We also demonstrated the practical usefulness of these maximum and minimum MPIs through an application to the scanner data set that 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.

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

Appendix

Proof of Theorem 1

Proof for the Mean MPI

We first explain the concept of a so-called graph representation of a data set S, denoted by $G(S)$. Given a data set 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 data sets consist of $n + 2$ observations. In fact, observations $1,\ 2,\ \ldots,\ n$ 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 $i \neq j$ ($\epsilon < 1/2n$) 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 data set 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 $p_{n+1} = (2,\ 1,\ 1,\ \ldots,\ 1)$, $p_{n+2} = (1,\ 2,\ 1,\ \ldots,\ 1)$, $q_{n+1} = (3,\ 2,\ 2,\ \ldots,\ 2)$, and $q_{n+2} = (2,\ 3,\ 2,\ \ldots,\ 2)$. Notice that no observation $\{1,\ 2,\ \ldots,\ n\}$ 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 $2/(2n + 4)$. Let us write $\mathrm{MPI}(C)$ for the value of the MPI corresponding to a cycle C in the graph representation of the data set S. Then, the mean MPI of data set S can be written as

$$\begin{aligned} \mathrm{MPI} &= \frac{\sum_{C \in G(S_1)} \mathrm{MPI}(C)}{\#\mathrm{cycle} + 1} \\ = \mathrm{MPI} &= \frac{\sum_{C \in G([1,2,\ldots,n])} \mathrm{MPI}(C) + [2/(2n + 4)]}{\#\mathrm{cycle} + 1} \end{aligned} \tag{A1}$$

(where $G([1,\ 2,\ \ldots,\ n])$ stands for the graph representation of data set S, restricted to observations $1,\ 2,\ \ldots,\ n$).

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

$$\begin{aligned} \mathrm{MPI} &= \frac{\sum_{C \in G(S_1)} \mathrm{MPI}(C)}{\#\mathrm{cycle} + 1} \\ = \mathrm{MPI} &= \frac{\sum_{C \in G([1,2,\ldots,n])} \mathrm{MPI}(C) + [2/(2n + 6)]}{\#\mathrm{cycle} + 1}. \end{aligned} \tag{A2}$$

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

Now suppose that we have a polynomial time algorithm for finding the mean MPI of a data set. Then we can find the mean MPI for S_1 and S_2, compute the difference, and with the knowledge that this difference is

$$\frac{[2/(2n+6)] - [1/(2n+4)]}{\#\text{cycle} + 1},$$

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 data set 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 $j \neq i$ ($\epsilon < 1/[2n^2]$), $(p_i)_i = 1$. 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 $1 + [2(n-1)/2n^2]$. Similarly, if the arc (i, j) exists, then $1 - (1/n) + (3/2n^2)$ is a lower bound on $p_i q_i - p_i q_j$. From combining these two bounds, it follows that $(n^2 - n + 3/2)/(n^2 + n - 1) > 0.5$ obtains a lower bound on the minimum MPI of data set D^-.

We now add two more observations and two 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 $p_{n+1} = (\epsilon, \ldots, \epsilon, 1, 0.5)$, $p_{n+1} = (\epsilon, \ldots, \epsilon, 0.5, 1)$, $q_{n+1} = (0, \ldots, 0, 1, 0)$, and $q_{n+1} = (0, \ldots, 0, 0, 1)$. 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 data set 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 data set 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 MPIs 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 data sets 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 fewer 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

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

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 that x arcs must be added to D to form D^{+}. Then find $\max_k[f(k) < x]$ and add $x/f(k)$ 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 MPIs 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. 1993). This is done by first creating a graph corresponding to the consumption data. More precisely, we build the graph G corresponding to the data set S. For every observation $i = 1, \ldots, n$ in the data set, 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 (1979) 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

Afriat, S. N. 1967. "The Construction of Utility Functions from Expenditure Data." *Internat. Econ. Rev.* 8 (1): 67–77.

[3] It 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 $(p_i)_j = \epsilon$.

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin. 1993. *Network Flows: Theory, Algorithms, and Applications.* Englewood Cliffs, NJ: Prentice-Hall.

Arora, S., and B. Barak. 2009. *Computational Complexity: A Modern Approach.* Cambridge: Cambridge Univ. Press.

Deb, R., and M. Pai. 2013. "The Geometry of Revealed Preference." Working paper, Univ. Toronto.

Echenique, F., S. Lee, and M. Shum. 2011. "The Money Pump as a Measure of Revealed Preference Violations." *J.P.E.* 119 (6): 1201–23.

Garey, M. R., and D. S. Johnson. 1979. *Computers and Intractability.* San Francisco: Freeman.

Houtman, M., and J. Maks. 1985. "Determining All Maximal Data Subsets Consistent with Revealed Preference." *Kwantitatieve Methoden* 19:89–104.

Megiddo, N. 1979. "Combinatorial Optimization with Rational Objective Functions." *Math. Operations Res.* 4 (4): 414–24.

Smeulders, B., F. C. R. Spieksma, L. Cherchye, and B. De Rock. Forthcoming. "Goodness of Fit Measures for Revealed Preference Tests: Complexity Results and Algorithms." *ACM Trans. Econ. and Computation.*

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

This content downloaded from 134.58.253.57 on Mon, 3 Feb 2014 04:28:26 AM
All use subject to JSTOR Terms and Conditions