



RUHR

ECONOMIC PAPERS

Jan Heufer
Per Hjertstrand

Consistent Subsets – Computationally Feasible Methods to Compute the Houtman-Maks-Index

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@wiso.uni-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel, Prof. Dr. Jochen Kluge
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #523

Responsible Editor: Wolfgang Leininger

All rights reserved. Bochum, Dortmund, Duisburg, Essen, Germany, 2014

ISSN 1864-4872 (online) – ISBN 978-3-86788-598-0

The working papers published in the Series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editors.

Ruhr Economic Papers #523

Jan Heufer and Per Hjertstrand

**Consistent Subsets – Computationally
Feasible Methods to Compute the
Houtman-Maks-Index**

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über:
<http://dnb.d-nb.de> abrufbar.

<http://dx.doi.org/10.4419/86788598>

ISSN 1864-4872 (online)

ISBN 978-3-86788-598-0

Jan Heufer and Per Hjerstrand¹

Consistent Subsets – Computationally Feasible Methods to Compute the Houtman-Maks-Index

Abstract

We provide two methods to compute the largest subset of a set of observations that is consistent with the Generalised Axiom of Revealed Preference. The algorithm provided by Houtman and Maks (1985) is not computationally feasible for larger data sets, while our methods are not limited in that respect. The first method is a variation of Gross and Kaiser's (1996) approximate algorithm and is only applicable for two-dimensional data sets, but it is very fast and easy to implement. The second method is a mixed-integer linear programming approach that is slightly more involved but still fast and not limited by the dimension of the data set.

JEL Classification: C14, D11, D12

Keywords: Demand theory; efficiency; nonparametric analysis; revealed preference; utility maximisation

December 2014

¹ Jan Heufer, TU Dortmund; Per Hjerstrand, Research Institute of Industrial Economics, Stockholm. – Access to data collected and previously used by Syngjoo Choi, Raymond Fisman, Douglas Gale, and Shachar Kariv is gratefully acknowledged. – All correspondence to: Jan Heufer, TU Dortmund University, Department of Economics and Social Science, 44221 Dortmund, Germany, e-mail: jan.heufer@tu-dortmund.de

1 INTRODUCTION

When consumer choice data violates utility maximising behaviour it is often desirable to know by “how much” the observed choices deviate from utility maximisation. Houtman and Maks (1985) proposed to measure the degree of inconsistency as the maximal number of observations in the observed sample consistent with rational choice. This measure (the HM-index) is calculated as the maximal subset of observations consistent with some revealed preference axiom. The method has the additional advantage that researchers can restrict further analysis of the data set to this maximal subset. This paper proposes two simple and fast methods to calculate the HM-index.

Choi, Fisman, Gale, and Kariv (2007) tested whether the choices of 93 subjects over 50 decision rounds were consistent with utility maximisation. Since most subjects were inconsistent, they computed the HM-index to obtain a measure of the degree of inconsistency. However, in doing so, they report that the HM-index is “*computationally intensive even for moderately large data sets*”. We apply our procedures to their data and find them to be very fast: the first procedure found a solution for every subject in at most 0.4 seconds, while the second found a solution in at most 3.3 seconds for every subject. Given the efficiency of our algorithms, researchers can use the procedures to run extensive Monte-Carlo simulations to approximate the test power based on the HM-index.

Our first method is a simple combinatorial algorithm based on Gross and Kaiser (1996). It is only applicable for two-dimensional data sets, but because experimental data sets are often two-dimensional, the method is useful for many purposes. As described above, it runs very fast and to our knowledge is the first efficient algorithm that does not require optimisation software. The second method is based on solving a mixed integer linear programming (MIP) problem. This method is not restricted by the dimension of the data set. As such, it is more general but slower than the first procedure. Implementations of the methods for Matlab® and Wolfram Mathematica® are available as supplementary material upon request.

Dean and Martin (2013) recently proposed a new measure of how close choice data is to satisfy utility maximisation. The algorithm used to implement this measure can also be used to calculate the HM-index, and like our second method it consists of solving a MIP problem. Thus, the two problems are similar in computational complexity. But while our MIP problem is deduced directly from the definition of the HM-index, Dean and Martin (2013)’s MIP problem is based on solving the so called “minimum set covering problem” which is shown to be equivalent to calculating the HM-index. In this respect, it is important to note the simplicity of our first algorithm which does not require using any optimisation packages.

2 PRELIMINARIES

The commodity space is \mathbb{R}_+^L and the price space is \mathbb{R}_{++}^L , where $L \geq 2$ is the number of different commodities. A budget set is defined as $B^i = B(\mathbf{p}^i, w^i) = \{\mathbf{x} \in \mathbb{R}_+^L : \mathbf{p}^i \mathbf{x}^i \leq 1\}$, where $\mathbf{p}^i = (p_1^i, \dots, p_L^i)' \in \mathbb{R}_{++}^L$ is the price vector and income is normalised to 1. We assume that $\mathbf{p}^i \mathbf{x}^i = 1$; the only observables of the model are N budgets and the corresponding consumer demand. As price vectors characterise budgets, the entire set of N observations is denoted $\Omega = \{(\mathbf{x}^i, \mathbf{p}^i)\}_{i=1}^N$.

The bundle \mathbf{x}^i is *directly revealed preferred* to a bundle \mathbf{x} , written $\mathbf{x}^i R^0 \mathbf{x}$, if $\mathbf{p}^i \mathbf{x}^i \geq \mathbf{p}^i \mathbf{x}$; it is *strictly directly revealed preferred* to \mathbf{x} , written $\mathbf{x}^i P^0 \mathbf{x}$, if $\mathbf{p}^i \mathbf{x}^i > \mathbf{p}^i \mathbf{x}$; it is *revealed preferred* to \mathbf{x} , written $\mathbf{x}^i R \mathbf{x}$, if R is the transitive closure of R^0 , that is, if there exists a sequence $\mathbf{x}^j, \dots, \mathbf{x}^k$, such that $\mathbf{x}^i R^0 \mathbf{x}^j R^0 \dots \mathbf{x}^k R^0 \mathbf{x}$. The bundle \mathbf{x}^i is *strictly revealed preferred* to \mathbf{x} , written $\mathbf{x}^i P \mathbf{x}$, if $\mathbf{x}^i R \mathbf{x}^j P^0 \mathbf{x}^k R \mathbf{x}$ for some $j, k = 1, \dots, N$.

Axiom (Samuelson 1938) *A set of observations Ω satisfies the Weak Axiom of Revealed Preference (WARP) if for all $i, j = 1, \dots, N$, it holds that $[\text{not } \mathbf{x}^i R^0 \mathbf{x}^j]$ whenever $\mathbf{x}^j R \mathbf{x}^i$ and $\mathbf{x}^i \neq \mathbf{x}^j$.*

Axiom (Varian 1982) *A set of observations Ω satisfies the Generalised Axiom of Revealed Preference (GARP) if for all $i, j = 1, \dots, N$, it holds that $[\text{not } \mathbf{x}^i P^0 \mathbf{x}^j]$ whenever $\mathbf{x}^j R \mathbf{x}^i$.*

Axiom (Banerjee and Murphy 2006) *A set of observations Ω satisfies the Weak GARP (WGARP) if for all $i, j = 1, \dots, N$, it holds that $[\text{not } \mathbf{x}^i P^0 \mathbf{x}^j]$ whenever $\mathbf{x}^j R^0 \mathbf{x}^i$.*

Varian (1982) showed that GARP is a necessary and sufficient condition for the existence of a continuous, monotonic, and concave utility function to rationalise Ω . Banerjee and Murphy (2006) showed that for $L = 2$, GARP and WGARP are equivalent.

Houtman and Maks (1985) introduced the HM-index to measure the degree of inconsistency of GARP. To formally define the HM-index, let $\mathbf{v} = (v^1, \dots, v^N)$ be a vector of binary variables (i.e., $v^i \in \{0, 1\}$ for all $i = 1, \dots, N$), and define the relation $R^0(\mathbf{v})$ as $\mathbf{x}^i R^0(\mathbf{v}) \mathbf{x}^j$ if $v^i \mathbf{p}^i \mathbf{x}^i \geq \mathbf{p}^j \mathbf{x}$, and let $R(\mathbf{v})$ be the transitive closure of $R^0(\mathbf{v})$; furthermore, let $P^0(\mathbf{v}^i)$ if $v^i \mathbf{p}^i \mathbf{x}^i > \mathbf{p}^i \mathbf{x}$.

Axiom *A set of observations Ω satisfies GARP(\mathbf{v}) for some $\mathbf{v} \in \{0, 1\}^N$ if for all $i, j = 1, \dots, N$, it holds that $[\text{not } \mathbf{x}^i P^0(\mathbf{v}^i) \mathbf{x}^j]$ whenever $\mathbf{x}^j R(\mathbf{v}) \mathbf{x}^i$.*

Definition *The Houtman-Maks (HM) index is the maximal fraction of non-zero elements in the binary vector \mathbf{v} such that GARP(\mathbf{v}) holds.*

Thus, the HM-index is the solution to the problem

$$HM = \max_{\mathbf{v}} \sum_{i=1}^N \frac{v^i}{N} \text{ such that GARP}(\mathbf{v}) \text{ holds and } \mathbf{v} \in \{0, 1\}^N. \quad (1)$$

3 THE TWO-DIMENSIONAL CASE

As Houtman and Maks (1985), Gross and Kaiser (1996) took a graph-theoretic approach. Every observation is interpreted as a node of a graph. In their definition, if observations i and j form a violation of WARP, then the nodes for i and j are adjacent. The degree of a node i , $\text{degr}(i)$, is the number of nodes to which it is adjacent. Define A_i as the set of nodes adjacent to node i , and $1A_i$ as the set of nodes which are adjacent to i with degree 1.

The algorithm consists of two parts. First, whenever $\text{degr}(i) = \max_{j \in \{1, \dots, N\}} \text{degr}(j)$ and $\text{degr}(k) < \text{degr}(i)$ for all $k \in A_i$, remove i . Repeat this step until no index is removed anymore. Second, whenever $\text{degr}(i) = \text{degr}(h) = \max_{j \in \{1, \dots, N\}} \text{degr}(j)$ and $h \in A_i$, then (1) if $1A_i \neq \emptyset$, remove i , (2) if $1A_h \neq \emptyset$, remove h , (3) if $1A_i = 1A_h = \emptyset$, remove either i or h . Again, repeat this step until no index is removed anymore. All nodes not removed in this process belong to the set of indices consistent with WARP. Gross and Kaiser (1996) point out that there is a special case in which the algorithm will fail to provide a maximal subset. However, they argue that this case is extremely rare, and in any case, the algorithm provides a lower bound.

The algorithm is very efficient and easy to implement and therefore suitable for practical purposes. It can easily be adapted for WGARP by simply redefining adjacency.¹ If two nodes i and j are defined as adjacent

¹This has already been noted by Heufer (2014) who did not go into the details of the approach.

whenever i and j form a violation of WGARP , then the same algorithm will provide the set of indices consistent with WGARP . With Banerjee and Murphy's (2006) result, we then have a computationally efficient method to compute the maximal subset of indices which are consistent with GARP in the two-dimensional case.

The same method can also be applied to compute a homothetic HM -index for homotheticity in the two-dimensional case as Heufer (2013) showed that a pairwise version of Varian's (1983) homothetic axiom is sufficient in that case.

4 A MIXED INTEGER LINEAR PROGRAMMING APPROACH

A direct way to calculate the HM -index in the multi-dimensional setting (i.e., $L > 2$) is to numerically solve (1). However, this problem may become difficult to solve because of the complexity in implementing the $\text{GARP}(\mathbf{v})$ -constraint. Instead we suggest to reformulate the problem as a simple mixed integer programming (MIP) problem by replacing $\text{GARP}(\mathbf{v})$ with an equivalent condition.

Theorem 1 *For any $\mathbf{v} \in \{0, 1\}^N$, the following conditions are equivalent:*

1. *the set of observations Ω satisfies $\text{GARP}(\mathbf{v})$;*
2. *there exist numbers $U^i \in [0, 1]$ and $\psi^{ij} \in \{0, 1\}$ such that*

$$U^i - U^j < \psi^{ij}, \tag{2a}$$

$$(\psi^{ij} - 1) \leq U^i - U^j, \tag{2b}$$

$$v^i \mathbf{p}^i \mathbf{x}^i - \mathbf{p}^i \mathbf{x}^j < \psi^{ij} A^i, \tag{2c}$$

$$(\psi^{ij} - 1) A^j \leq \mathbf{p}^j \mathbf{x}^i - v^j \mathbf{p}^j \mathbf{x}^j, \tag{2d}$$

for all $i, j = 1, \dots, N$, where A^i is a number greater than $\mathbf{p}^i \mathbf{x}^i$.

This theorem states that $\text{GARP}(\mathbf{v})$ is equivalent to the set of linear inequalities (2a)-(2d).² We suggest to replace $\text{GARP}(\mathbf{v})$ with the linear inequalities (2a)-(2d) in (1) and calculate the HM -index by solving the following mixed-integer programming problem:

$$\text{HM} = \max_{\mathbf{v}, U^i, \psi^{ij}} \sum_{i=1}^N \frac{v^i}{N} \text{ subject to (2a)-(2d) and } \mathbf{v} \in \{0, 1\}^N. \tag{3}$$

This problem gives an exact and global solution (because every local solution to a MIP problem is a global solution), and there exist efficient algorithms for solving such MIP problems in practice (branch and bound, cutting plane, etc.).

5 EMPIRICAL APPLICATION AND CONCLUDING REMARKS

We applied our methods to data from Choi et al. (2007). This data consists of portfolio choice allocations in a two-dimensional setting (i.e., $L = 2$) from 93 experimental subjects over 50 decision rounds (i.e., $N = 50$). Choi et al. (2007) reported the HM -index for all but six subjects which they were unable to find an optimal solution for. We calculated the HM -index for every subject including the six unreported subjects (detailed

²The proof of Theorem 1 follows (with some simple modifications) from Cherchye, Demuynck, De Rock, and Hjertstrand (2014) who prove a similar theorem in the context of weakly separable utility maximisation.

results for each subject is available as supplementary material upon request). We found the method based on the Gross-Kaiser algorithm to be very fast: it found a solution for every subject in at most 0.4 seconds.³ Solving the MIP problem (3) is more involved and took a maximum of 3.3 seconds for every subject, which, in our view, can be considered fast enough given the relatively large N .⁴

This paper introduced two simple and efficient algorithms for computing the Houtman-Maks-index. The first algorithm is applicable for the two-dimensional setting and does not require any optimisation software. The second algorithm is based on solving a mixed-integer programming problem and can be applied to any dimensional setting. Both of these algorithms can be modified to calculate the HM-index for other revealed preference axioms, such as those for homotheticity.

REFERENCES

- Banerjee, S. and Murphy, J. H. (2006): "A Simplified Test for Preference Rationality of Two-Commodity Choice," *Experimental Economics* 9(9):67–75.
- Cherchye, L., Demuynck, T., De Rock, B., and Hjertstrand, P. (2014): "Revealed Preference Tests for Weak Separability: An Integer Programming Approach," *Journal of Econometrics* forthcoming.
- Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007): "Consistency and Heterogeneity of Individual Behavior under Uncertainty," *American Economic Review* 97(5):1921–1938.
- Dean, M. and Martin, D. (2013): "Measuring Rationality with the Minimum Cost of Revealed Preference Violations," Working paper.
- Gross, J. and Kaiser, D. (1996): "Two Simple Algorithms for Generating a Subset of Data Consistent With WARP and Other Binary Relations," *Journal of Business and Economic Statistics* 14(2):251–255.
- Heufer, J. (2013): "Testing Revealed Preferences for Homotheticity with Two-Good Experiments," *Experimental Economics* 16(1):114–124.
- (2014): "Nonparametric Comparative Revealed Risk Aversion," *Journal of Economic Theory* 153:569–616.
- Houtman, M. and Maks, J. (1985): "Determining all Maximal Data Subsets Consistent with Revealed Preference," *Kwantitatieve Methoden* 19:89–104.
- Samuelson, P. A. (1938): "A Note on the Pure Theory of Consumer's Behavior," *Economica* 5(17):61–71.
- Varian, H. R. (1982): "The Nonparametric Approach to Demand Analysis," *Econometrica* 50(4):945–972.
- (1983): "Non-parametric Tests of Consumer Behaviour," *Review of Economic Studies* 50(1):99–110.

³The mean running time over all subjects was 0.09 seconds with a variance of 0.004 seconds.

⁴The mean running time for the MIP solution over all subjects was 1.42 seconds with a variance of 0.26 seconds.