The Stata Journal (2023)
23, Number 2, pp. 578–588 DOI: 10.1177/1536867X231175345

Consistent subsets: Computing the Houtman–Maks index in Stata

Marcos Demetry
Linnaeus University
Växjö, Sweden
marcos.demetry@lnu.se

Per Hjertstrand
Research Institute of Industrial Economics
Stockholm, Sweden
per.hjertstrand@ifn.se

Abstract. The Houtman–Maks index is a measure of the size of a violation of utility-maximizing (that is, rational) behavior. In this article, we introduce the command **hmindex**, which calculates the Houtman–Maks index for a dataset of prices and observed choices of a consumer. The command is illustrated with an empirical application.

Keywords: st0720, hmindex, Houtman–Maks index, revealed preference, WGARP, WARP

1 Introduction

The hypothesis that economic agents choose commodity bundles by maximizing a utility function subject to a budget constraint forms the core of neoclassical economics. Revealed preference is an efficient tool to test whether consumer-choice data satisfy utility-maximizing (that is, rational) behavior.[1] When these data violate utility maximization, it is often desirable to know the "size" of the violation. Houtman and Maks (1985) propose measuring the degree of inconsistency as the maximal number of observations in the observed sample consistent with rational choice. This measure (the Houtman–Maks [HM] index) is calculated as the maximal subset of observations consistent with some revealed preference axiom.[2]

In this article, we introduce the command **hmindex**, which calculates the HM index for a dataset of prices and observed choice quantities sampled from an individual consumer (or a cross-section of individuals). **hmindex** implements a combinatorial algorithm proposed by Gross and Kaiser (1996) to calculate the HM index for the weak generalized axiom of revealed preference (WGARP) and the weak axiom of revealed preference (WARP), which we refer to as the Gross–Kaiser (GK) algorithm.

When the data consist of two goods, WGARP (WARP) is a necessary and sufficient condition for the data to be (strictly) rationalized by a continuous, strictly increasing,

1. See Demetry, Hjertstrand, and Polisson (2022) for a brief introduction to empirical revealed preference theory.
2. There exist several other goodness-of-fit measures for revealed preference tests. One of the most prominent is the Afriat efficiency index, which is implemented in the command `aei`, documented in Demetry, Hjertstrand, and Polisson (2022). The HM index is a goodness-of-fit measure that is more disaggregated than the Afriat efficiency index because it gives a binary response to whether a specific observation is included in the maximal subset consistent with revealed preference.

and (strictly) concave utility function. Thus, in such cases the HM index gives the maximal subset of observations that is consistent with rational choice. For datasets with more than two goods, WGARP and WARP are only necessary but not sufficient conditions for rationality because they do not account for transitive binary relations. Thus, for such data the HM index gives an upper bound on the maximal subset of observations consistent with rationality. However, in many multidimensional datasets, violations of rationality also seem (almost invariably) to be WGARP violations. That is, higher-order insensitivities in the data detected by more stringent tests are often associated with pairwise inconsistencies that are detected by WGARP (Alston and Chalfant 1991). If so, the HM index would give the correct maximal subset consistent with rationality.

Our command allows the user to calculate the distribution of the HM index over uniformly random distributed data. We illustrate the command on experimental data collected by Choi et al. (2007).

2 The GK algorithm to calculate the HM index

Suppose there are T observations of the prices and quantities of $K \geq 2$ goods. At observation $t = 1, \ldots, T$, the prices and quantities are denoted by $\mathbf{p}^t = (p_1^t, \ldots, p_K^t)$ and $\mathbf{x}^t = (x_1^t, \ldots, x_K^t)$, respectively. We assume that all prices are strictly positive and that all quantities are nonnegative (that is, some but not all quantities at any given observation may be equal to zero). The dataset $(\mathbf{p}^t, \mathbf{x}^t)_{t=1,\ldots,T}$ usually describes a single consumer that is observed over time but can also describe a cross-section of consumers.[3,4]

For any pair of observations (t, s), we say that $\mathbf{x}^t$ is directly revealed in preference to $\mathbf{x}^s$, written $\mathbf{x}^t R^D \mathbf{x}^s$, if $\mathbf{p}^t \cdot \mathbf{x}^t \geq \mathbf{p}^t \cdot \mathbf{x}^s$. This means that $\mathbf{x}^t$ is chosen even though the cost of the bundle $\mathbf{x}^s$ (at prices $\mathbf{p}^t$) does not exceed $\mathbf{p}^t \cdot \mathbf{x}^t$. For all $\mathbf{x} \in \mathbb{R}_+^K$ and any $t = 1, \ldots, T$ such that $\mathbf{p}^t \cdot \mathbf{x}^t \geq \mathbf{p}^t \cdot \mathbf{x}$, the data $(\mathbf{p}^t, \mathbf{x}^t)_{t=1,\ldots,T}$ are rationalized by a utility function u if $u(\mathbf{x}^t) \geq u(\mathbf{x})$ and strictly rationalized if $u(\mathbf{x}^t) > u(\mathbf{x})$ whenever $\mathbf{x} \neq \mathbf{x}^t$.

A dataset $(\mathbf{p}^t, \mathbf{x}^t)_{t=1,\ldots,T}$ satisfies the WGARP if $\mathbf{x}^t R^D \mathbf{x}^s$ implies $\mathbf{p}^s \cdot \mathbf{x}^s \leq \mathbf{p}^s \cdot \mathbf{x}^t$.

When the data consist of only two goods, it is well known that WGARP is a necessary and sufficient condition for a dataset to be rationalized by a continuous, strictly increasing, and concave utility function (see Banerjee and Murphy [2006] and Varian [1982]).

3. If the data form a cross-section, then the problem of calculating the maximal subset of the data consistent with revealed preference can be thought of as a problem of finding the maximal number of consumers that share a common utility function (that is, the same preference).
4. The following brief discussion of rationalizability, WGARP, and WARP draws on Demetry, Hjertstrand, and Polisson (2022).

When the data consist of more than two goods, Aguiar, Hjertstrand, and Serrano (2020) show that WGARP (WARP) is a necessary and sufficient condition for a dataset to be (strictly) rationalized by a continuous, strictly increasing, piecewise (strictly) concave, and skew-symmetric preference function.[5]

A dataset $(\mathbf{p}^t, \mathbf{x}^t)_{t=1,\dots,T}$ satisfies the WARP if $\mathbf{x}^t R^D \mathbf{x}^s$ implies $\mathbf{p}^s \cdot \mathbf{x}^s < \mathbf{p}^s \cdot \mathbf{x}^t$ whenever $\mathbf{x}^t \neq \mathbf{x}^s$.

In the two-dimensional case, WARP is a necessary and sufficient condition for a dataset to be strictly rationalized by a continuous, strictly increasing, and strictly concave utility function (see Rose [1958] and Matzkin and Richter [1991]). Economically, the difference between WGARP and WARP is that WGARP allows for indifference between bundles. In other words, while WGARP accommodates demand correspondences, WARP is consistent only with demand functions.

The HM index is defined as the maximal subset of observations from the data $(\mathbf{p}^t, \mathbf{x}^t)_{t=1,\dots,T}$ such that WGARP or WARP holds.[6] All observations that are removed from the data form the violator set (VS), while the remaining observations satisfying WGARP or WARP form the consistent set (CS). The value of the HM index is usually presented as either 1) the number of observations in CS, that is, the maximal number of observations satisfying WGARP or WARP, or 2) the fraction of observations in CS, that is, the number of observations in CS divided by the total number of observations in the data.

To calculate the HM index, Gross and Kaiser (1996) use a graph-theoretic approach.[7] Every observation is interpreted as a node of a graph. If observations s and t form a violation of WGARP, then the nodes for s and t are adjacent. The degree of a node t, $\text{degr}(t)$, is the number of nodes to which it is adjacent. Define At as the set of nodes adjacent to node t and $1At$ as the set of nodes that are adjacent to t with degree 1. The GK algorithm consists of two parts:

- **Step 1**. Whenever $\text{degr}(t) = \max_{s=1,\dots,T} \text{degr}(s)$ and $\text{degr}(r) < \text{degr}(t)$ for all $r \in At$, then remove observation t.
- **Step 2**. Whenever $\text{degr}(t) = \text{degr}(i) = \max_{s=1,\dots,T} \text{degr}(s)$ and $i \in At$, then a) if $1At = \varnothing$, remove observation t; b) if $1Ai = \varnothing$, remove observation i; and c) if $1At = 1Ai = \varnothing$, remove either t or i.

Steps 1 and 2 are repeated sequentially until no additional observation is removed. Heufer and Hjertstrand (2015) suggested the GK algorithm for WGARP. Gross and Kaiser

5. See Aguiar, Hjertstrand, and Serrano (2020) for the definition of a preference function and the relevant properties pertaining to preference functions.
6. Strictly speaking, the HM index is defined as the maximal subset of observations such that some revealed preference axiom holds. But because we are concerned only with WGARP or WARP, we consider only the more narrow definition stated here.
7. The standard way of calculating the HM index in empirical applications has been to iteratively delete observations and test for revealed preference. However, even for relatively small datasets, this can be very computationally demanding and sometimes even practically unfeasible.

(1996) originally suggested the algorithm for WARP, which requires only redefining adjacency. If two nodes, t and s, are defined as adjacent whenever t and s form a violation of WARP, then the GK algorithm will provide the set of indices consistent with WARP.

Gross and Kaiser (1996) point out that the algorithm may remove additional observations, in which case it produces only a lower bound and not an exact solution to the HM index. However, they also argue, based on experimental evidence, that this should occur very rarely.

As a benchmark to rational choice, empirical applications of the HM index using experimental data sometimes compare the distribution of the HM index across all subjects with the distribution of the HM index under the assumption of uniformly random consumption behavior. `hmindex` allows the user to calculate the empirical distribution of the HM index under this type of irrational consumer behavior by drawing uniformly random data, as explained in section 2.4 of Demetry, Hjertstrand, and Polisson (2022), and calculating the HM index for every simulated dataset. The distribution can then be visualized by plotting the values of the HM index in a kernel density plot (see figure 2 in section 4 for such an application to experimental data).

3 hmindex

3.1 Syntax

`hmindex, price(`*mname*`) quantity(`*mname*`)` [`axiom(`*axiom*`) distribution simulations(`#`) seed(`#`)`]

3.2 Options

`price(`*mname*`)` specifies a $T \times K$ price matrix, where each row corresponds to an observation t and the columns correspond to the goods. All prices are required to be strictly positive. If any of the elements in the price matrix are nonpositive (or if the price and quantity matrices have different dimensions), the command returns an error message. `price()` is required.

`quantity(`*mname*`)` specifies a $T \times K$ quantity matrix, where each row corresponds to an observation t and the columns correspond to the goods. All quantities are required to be nonnegative. Some (but not all) quantities at a given observation may be equal to zero. If the quantity matrix violates these conditions (or if the price and quantity matrices have different dimensions), the command returns an error message. `quantity()` is required.

`axiom(`*axiom*`)` specifies the axiom that the user would like to apply. The default is `axiom(WGARP)`. The user can apply WARP by specifying `axiom(WARP)`. The user may also apply both axioms simultaneously by specifying `axiom(all)`.

`distribution` specifies whether the user would like to calculate the empirical distribution of the HM index under uniformly random consumption behavior. This produces an output table with the mean (`Mean`), standard deviation (`Std. Dev.`), minimum (`Min`), first quartile (`Q1`), median (`Median`), third quartile (`Q3`), and maximum (`Max`) over all simulated data. The simulated uniformly random data are calculated as explained in sections 2.4 and 3.3 of Demetry, Hjertstrand, and Polisson (2022). The user can set the number of simulations and the random seed in the simulation of the uniformly random data (see the next two options).

`simulations(#)` specifies the number of repetitions of the simulated uniformly random datasets. The default is `simulations(1000)`. This option is useful only in combination with the `distribution` option.

`seed(#)` specifies the random seed in the generation of uniformly random datasets (see sections 2.3 and 3.3 of Demetry, Hjertstrand, and Polisson [2022] for a detailed explanation). The default is `seed(12345)`. This option is useful only in combination with the `distribution` option.

3.3 Stored results

`hmindex` stores the following in `r()`:

Scalars	
`r(OBS)`	number of observations
`r(GOODS)`	number of goods
`r(HM_NUM_`*axiom*`)`	maximal number of observations that satisfy the axiom; given by `#HM` in the output
`r(HM_FRAC_`*axiom*`)`	maximal fraction of observations that satisfy the axiom; given by `%HM` in the output (calculated as maximal number of observations satisfying the axiom divided by total number of observations)
`r(SIM)`	number of simulated uniformly random datasets (if option `distribution` is specified)
Macros	
`r(AXIOM)`	axiom or axioms being tested
Matrices	
`r(INDICATOR_`*axiom*`)`	T-dimensional binary array indicating whether the observation is in the CS (`1`) or the VS (`0`)
`r(OBSDROP_`*axiom*`)`	list of observations in the VS, that is, which observations are dropped from the dataset
`r(CS_price_`*axiom*`)`	price matrix in the CS for the specified axiom, that is, the price data corresponding to the goods in CS
`r(CS_quantity_`*axiom*`)`	quantity matrix in the CS for the specified axiom, that is, the quantity data corresponding to the goods in CS
`r(VS_price_`*axiom*`)`	price matrix in the VS for the specified axiom, that is, the price data corresponding to the goods in VS
`r(VS_quantity_`*axiom*`)`	quantity matrix in the VS for the specified axiom, that is, the quantity data corresponding to the goods in VS
`r(SUMSTATS_`*axiom*`)`	summary statistics for random data: number and fraction of observations in CS
`r(SIMRESULTS_`*axiom*`)`	number and fraction of observations in CS for every simulated uniformly random dataset

3.4 Examples

The following two examples illustrate **hmindex** using a dataset of 20 observations on the prices and quantities of two goods. The consumed quantities of goods 1 and 2 are x1 and x2, which form the quantity matrix X = (x1, x2). The corresponding price matrix with prices p1 and p2 of the two goods is P = (p1, p2). The first example runs **hmindex** using its default options, that is, for WGARP.

```
. use hmindex_example_data
. mkmat x1 x2, matrix(X)
. mkmat p1 p2, matrix(P)
. hmindex, price(P) quantity(X) axiom(wgarp)
    Number of obs          =        20
    Number of goods        =         2
```

Axiom	#HM	%HM
WGARP	15	.75

We see that the HM index is 0.75 (or 15 in absolute terms), which implies that 25% of the observations would have to be removed from the original dataset for the data to be rationalizable by WGARP. The second example runs **hmindex** for WARP, with the **distribution** option using 1,000 simulations.

```
. hmindex, price(P) quantity(X) axiom(warp) distribution simulations(1000)
    Number of obs          =        20
    Number of goods        =         2
    Simulations            =      1000
```

Axiom	#HM	%HM
WARP	15	.75

Summary statistics for simulations:

WARP	#HM	%HM
Mean	15.809	.79045
Std. Dev.	1.510232	.0755116
Min	11	.55
Q1	15	.75
Median	16	.8
Q3	17	.85
Max	20	1

The first output table shows that applying WARP gives the same value of the HM index as applying WGARP. The second table shows that if consumption behavior was uniformly random—which in this case is a proxy for irrational behavior—then the HM index is slightly higher with a median value of 16 over all simulations. This would suggest that the consumption behavior displayed through the choices in the data is slightly more irrational than uniformly random consumption behavior.

4 Empirical illustration

We applied `hmindex` to experimental data collected by Choi et al. (2007). These data consist of portfolio choice allocations in a two-dimensional setting from 93 experimental subjects over 50 decision rounds; that is, $T = 50$. Each subject split his or her budget between two Arrow–Debreu securities, with each security paying 1 token if the corresponding state was realized and 0 tokens otherwise. The experiment consisted of two treatments. In the first (symmetric) treatment, with 47 subjects, each state of the world occurred with probability 1/2, which was objectively known to the subjects. The second (asymmetric) treatment was applied to 46 subjects, who faced states occurring with objective probabilities 1/3 and 2/3. All state prices were randomly chosen and varied across all decision rounds and subjects.

Choi et al. (2007) reported the HM index for all but six subjects, whom they were unable to find an optimal solution for. Using `hmindex`, we calculated the HM index for WGARP for all 93 subjects, including the 6 unreported subjects in Choi et al. (2007). For every subject, the following code calculates (and saves to the matrix `results`) the actual HM index and the mean, minimum, and maximum HM indices over all simulations in the uniformly random data.

```
matrix results = J(93,4,0)
matrix colnames results = "HM" "Mean" "Min" "Max"

forvalues subject = 1/93 {

      quietly hmindex, price(P`subject') quantity(X`subject') ///
         distribution simulations(1000)

      matrix sumstats = r(SUMSTATS_WGARP)

      matrix results[`subject', 1] = r(HM_NUM_WGARP)          /* Actual HM */
      matrix results[`subject', 2] = sumstats[1, 1]           /* Mean */
      matrix results[`subject', 3] = sumstats[3, 1]           /* Min */
      matrix results[`subject', 4] = sumstats[7, 1]           /* Max */

}
```

Figure 1 presents these results.[8] The gray-colored triangle markers represent the calculated HM index for each subject. The connected intervals give the maximum and minimum values of the HM index calculated from the simulated uniformly random data, while the black dots give the mean HM index calculated over all simulations. We find

8. The help file accompanying `hmindex` contains the code to generate figure 1.

that 77 subjects have an HM index of 45 or higher (that is, %HM $\geq$ 0.9), 14 subjects have values between 40 and 44 (that is, $0.8 \leq$ %HM < 0.9), and 2 subjects have values below 40 (that is, %HM < 0.8). Moreover, of all 93 subjects, the HM indices from 6 of them are equal to or lower than the maximum HM indices calculated from the simulated uniformly random datasets, which may cast doubt on whether the actual choices from these subjects are rational.

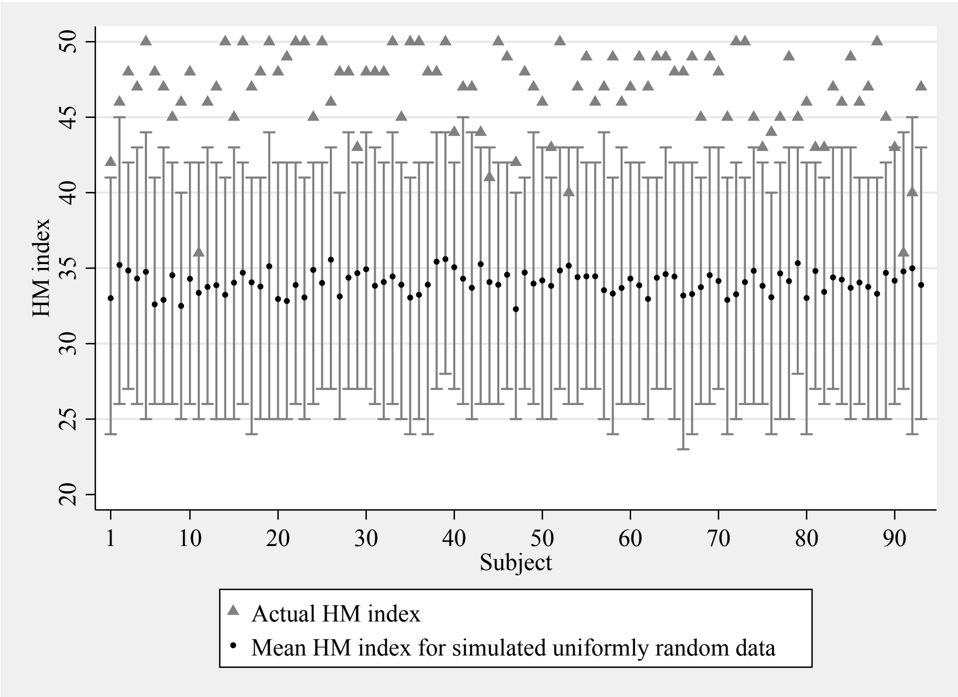

Figure 1. Calculated HM indices from actual choices (gray-colored triangle markers) and from simulated uniformly random data. The connected intervals give the maximum and minimum values of the HM index calculated from the simulated data, while the black dots refer to the mean HM index over the simulated data.

In figure 2, we plot the kernel cumulative distribution functions (left) and probability distribution functions (right) of the HM indices calculated from the subjects' actual choices (93 observations, solid line) and the HM indices from the simulated uniformly random data ($93 \times 1000 = 93000$ observations, dashed line). Overall, these results show that the actual choices are more consistent with rational choice than uniformly random choices.

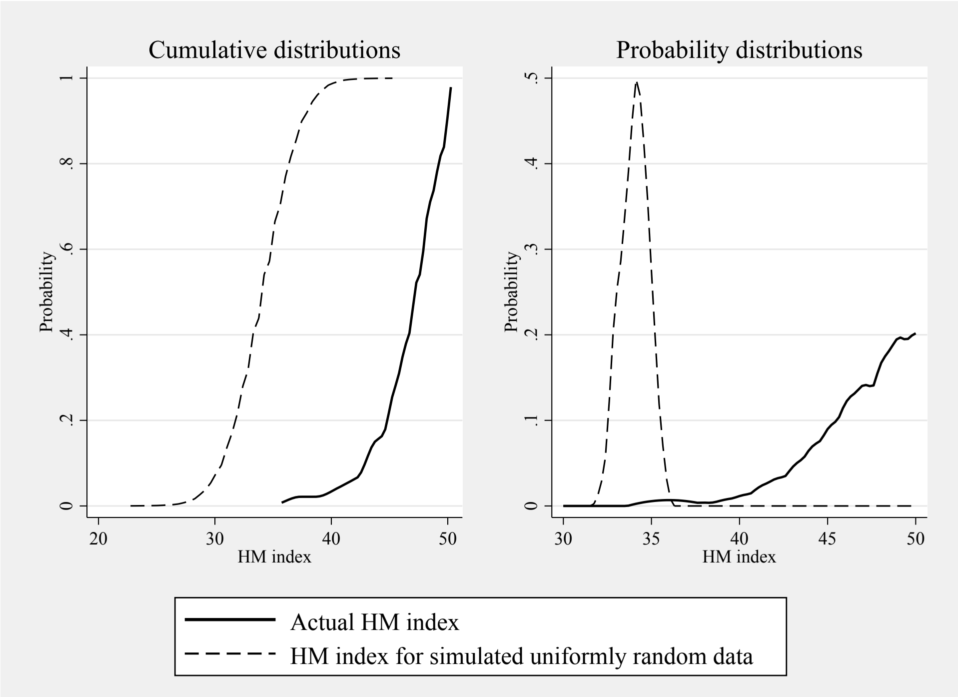

Figure 2. Left: Kernel cumulative distribution functions of HM indices from actual choices (solid line) and from simulated uniformly random data (dashed line) over all subjects. The kernel cumulative distribution functions were generated using the `kdensity` package in Stata. Right: Kernel probability densities of HM indices from actual choices (solid line) and from simulated uniformly random data (dashed line) over all subjects. The kernel probability density functions were generated using the `kdens` package in Stata with the `reflection` option (because the HM index is a bounded variable between 1 and 50). `kdens` is documented in Jann (2005) and is available on Statistical Software Components (note that `kdens` needs to be installed prior to using it).

Finally, we note that `hmindex` runs very quickly: It found a solution for every subject in at most 0.053 seconds. The mean running time over all subjects was 0.029 seconds with a standard deviation of 0.007 seconds.

5 Conclusions

In this article, we presented the command `hmindex`, which is an implementation of the Houtman–Maks index that gives a measure of how "close" consumer demand data are to satisfying utility-maximizing behavior. The command is formulated as a combinatorial algorithm and therefore converges in a finite number of steps. Consequently, `hmindex` can be implemented on rather large datasets. A natural extension for future work is to provide implementations of other disaggregated measures of goodness of fit for revealed preference tests. Some examples of such measures can be formulated as (mixed-integer) linear programming problems, whose computational complexity strongly depends on the algorithms used. Thus, in practice, solving such problems may not be a trivial task.

6 Acknowledgments

We thank the editor Stephen Jenkins and a reviewer for helpful comments and useful suggestions. Per Hjertstrand thanks Torsten Söderbergs Stiftelse for financial support.

7 Programs and supplemental materials

To install a snapshot of the corresponding software files as they existed at the time of publication of this article, type

```
. net sj 23-2
. net install st0720      (to install program files, if available)
. net get st0720          (to install ancillary files, if available)
```

`hmindex` is available on the Statistical Software Components Archive and can be installed by entering `ssc install hmindex` in the Stata command prompt.

8 References

Aguiar, V. H., P. Hjertstrand, and R. Serrano. 2020. A rationalization of the weak axiom of revealed preference. Bravo Working Paper 2020-016, Brown University and Research Institute of Industrial Economics (IFN) Working Paper No. 1321. https://doi.org/10.2139/ssrn.3543674.

Alston, J. M., and J. A. Chalfant. 1991. Can we take the con out of meat demand studies? *Western Journal of Agricultural Economics* 16: 36–48.

Banerjee, S., and J. H. Murphy. 2006. A simplified test for preference rationality of two-commodity choice. *Experimental Economics* 9: 67–75. https://doi.org/10.1007/s10683-006-4313-6.

Choi, S., R. Fisman, D. Gale, and S. Kariv. 2007. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review* 97: 1921–1938. https://doi.org/10.1257/aer.97.5.1921.

Demetry, M., P. Hjertstrand, and M. Polisson. 2022. Testing axioms of revealed preference in Stata. *Stata Journal* 22: 319–343. https://doi.org/10.1177/1536867X221106374.

Gross, J., and D. Kaiser. 1996. Two simple algorithms for generating a subset of data consistent with WARP and other binary relations. *Journal of Business and Economic Statistics* 14: 251–255. https://doi.org/10.1080/07350015.1996.10524652.

Heufer, J., and P. Hjertstrand. 2015. Consistent subsets: Computationally feasible methods to compute the Houtman–Maks-index. *Economic Letters* 128: 87–89. https://doi.org/10.1016/j.econlet.2015.01.024.

Houtman, M., and J. A. H. Maks. 1985. Determining all maximal data subsets consistent with revealed preference. *Kwantitatieve Methoden* 6: 89–104.

Jann, B. 2005. kdens: Stata module for univariate kernel density estimation. Statistical Software Components S456410, Department of Economics, Boston College. https://ideas.repec.org/c/boc/bocode/s456410.html.

Matzkin, R. L., and M. K. Richter. 1991. Testing strictly concave rationality. *Journal of Economic Theory* 53: 287–303. https://doi.org/10.1016/0022-0531(91)90157-Y.

Rose, H. 1958. Consistency of preference: The two-commodity case. *Review of Economic Studies* 25: 124–125. https://doi.org/10.2307/2296210.

Varian, H. R. 1982. The nonparametric approach to demand analysis. *Econometrica* 50: 945–973. https://doi.org/10.2307/1912771.

About the authors

Marcos Demetry is a PhD student at Linnaeus University in Växjö, Sweden, and an affiliated doctoral student at the Research Institute of Industrial Economics in Stockholm, Sweden.

Per Hjertstrand is an associate professor and research fellow at the Research Institute of Industrial Economics in Stockholm, Sweden.