A note on different approaches to index number theory

By Matthijs van Veelen and Roy van der Weide*

August 31, 2007

The central question in a recent paper by J. Peter Neary (2004) concerning real incomes and the Penn World Tables is: what exactly do the numbers mean? This question is important, because these data are widely used in empirical research. More generally, index number data, including indicators of inflation and economic growth, are used so extensively in research throughout the profession, that it indeed seems helpful to understand what they actually indicate.

In the literature, two approaches to index numbers are distinguished: the axiomatic approach and the economic approach. In this note we would like to discuss the way in which these two approaches differ and how that affects what the numbers mean. In Neary's paper the difference is described as one between an approach that does and an approach that does not assume that quantities arise from optimizing behaviour. This is in line with a tradition in index number theory that can be traced back to Ragnar Frisch (1936). We will argue that a more accurate description is that the difference lies in whether or not optimizing agents, or representative consumers, are assumed to optimize the *same* utility function. It is exactly this distinction that sets the (different) limitations of both approaches for constructing a meaningful indicator of real income. Those limitations will be described in some detail and being aware of them is helpful for assessing the value of results from both the axiomatic and the economic approach. They also suggest ways in which the two approaches may be complementary for constructing meaningful index numbers.

Introducing the Geary-Allen International Accounts (GAIA) system, Neary combines taking the economic approach with judging the resulting indices in view of some axioms. We think that this makes his paper a suitable point of departure for the discussion of how the two approaches handle differences in relative prices on the one hand and differences in preferences on the other.

1 Preliminaries

The index number problem is described in Neary (2004) as follows. Suppose that there are m countries, labelled $j = 1, ..., m$, for which we have observations of prices and quantities consumed of n commodities, labelled $i = 1, ..., n$. Price and quantity vectors in country j are denoted by $\mathbf{p}^j$ and $\mathbf{q}^j$, with typical elements p_{ij} and q_{ij}, respectively. The goal is to use these data in order to arrive at a set of index numbers $Q_{jk}, j, k = 1, ..., m$, which gives the real income of each country j relative to every other country k.

One can assume or require these index numbers to be transitive - that is, $Q_{jk} = Q_{jl}Q_{lk}\ \forall\ j, k, l$. If so, then this restriction on the set of numbers Q_{jk} makes that together they provide a unique cardinal ranking of real incomes. If this restriction is taken to be of primary importance, then the problem can be redefined so that intransitive index numbers are excluded by construction. Rather than finding a set of m^2 index numbers and deriving a ranking of real incomes if these index numbers are transitive, one can also start at the other end, by looking for a vector of m real incomes, which by definition gives one cardinal ranking, and define index numbers Q_{jk} as quotients of the j'th and k'th elements of this vector. This is the way the problem is described in Matthijs van Veelen (2002), where price and quantity vectors are organised in matrices $\mathbf{P}$ and $\mathbf{Q}$ and the aim is to find a function $\mathbf{F} : \mathbb{R}_+^{2\times m\times n} \rightarrow \mathbb{R}^m$ of those prices and quantities.[1] Here, F_j is the real income of country j. This provides a useful way of formally describing some of the index numbers that follow.

The choice of a formulation is not a matter of taste if we want to keep the option of violating transitivity open, as for instance Laszlo Drechsler (1973) does. The real incomes that follow from the GAIA system that Neary (2004) introduces however are in fact such a function $\mathbf{F}$ and it therefore seems that transitivity is taken to be relatively important there. In Section 4 we will return to the question when transitivity is a reasonable requirement.

2 The 'economic' approach to index numbers

The defining feature of the economic approach is that it treats prices and quantities as observations that follow from optimization of one utility function under different budget sets. There is a subdivision possible, for one can try to recover this utility function with or without making further assumptions regarding the form it has. The non-parametric approach was developed and followed by, amongst others, Sydney Afriat (1967, 1981), W. Erwin Diewert (1973), Hal R. Varian (1983) and Steve Dowrick and John Quiggin (1994, 1997) and it works out what the restrictions are that the data impose on the set of utility functions that rationalize them. The parametric approach on the other hand looks at parametrized subsets

[1] In Neary (2004) Q_{jk} is already in use for denoting index numbers. We hope that we can use $\mathbf{Q}$ without subscript for the quantity martix without creating confusion. This should be possible, for the ij'th element of this matrix $\mathbf{Q}$ is already defined above as q_{ij}, which is the i'th element of vector $\mathbf{q}^j$.

of utility functions and uses the data to pinpoint parameter values. Here of course some parametrized subsets are better than others, so part of the task is to find a reasonable one. Douglas W. Caves, Laurits R. Christensen and W. Erwin Diewert (1982) is an example of this approach, and so is Neary (2004).

Within the context of the economic approach, there are different types of indices that are worth defining. The first one is an *exact* index, defined by Diewert (1976a) in a bilateral setting. Below, we will reproduce this definition in a multilateral setting with some additional and hopefully insightful formalities.

Definition 1 *A function* $\mathbf{F}$ *is exact for a set of utility functions* $\mathcal{S}$ *if there is a non-empty set* $\mathcal{D} \subset \mathbb{R}_+^{2\times m\times n}$ *such that for any* $(\mathbf{P}, \mathbf{Q}) \in \mathcal{D}$

(1) *there is a utility function* $u \in \mathcal{S}$ *for which* $\mathbf{q}^j = \arg\max u(\mathbf{q})$ *under*[2] $\langle \mathbf{p}^j, \mathbf{q}\rangle \leq \langle \mathbf{p}^j, \mathbf{q}^j\rangle$*, and a* $\lambda > 0$ *such that* $F_j(\mathbf{P}, \mathbf{Q}) = \lambda u(\mathbf{q}^j)$ *, for* $j = 1, ..., m$

(2) *for all* $u \in \mathcal{S}$ *for which* $\mathbf{q}^j = \arg\max u(\mathbf{q})$ *under* $\langle \mathbf{p}^j, \mathbf{q}\rangle \leq \langle \mathbf{p}^j, \mathbf{q}^j\rangle$*, there is a* $\lambda > 0$ *such that* $F_j(\mathbf{P}, \mathbf{Q}) = \lambda u(\mathbf{q}^j)$ *, for* $j = 1, ..., m.$

This definition may at first sight seem a bit different from Diewert's definition, but an example that features both in Diewert (1976a,b) and in Neary (2004) hopefully clarifies the sameness.

Alexandr A. Konüs and S. S. Byushgens (1926) showed that if $m = 2$, that is, in bilateral comparisons, if prices and quantities follow from maximization of a homogeneous quadratic utility function $\left(\mathbf{q}^T \mathbf{A}\mathbf{q}\right)^{1/2}$, then the Fisher index is a quotient of utilities (see also Diewert (1976b) for the connection between Fisher's ideal index and revealed preference axioms). If we restate this result in terms of our new definition, $\mathcal{S}$ would be the set of homogeneous quadratic utility functions[3] and the set $\mathcal{D}$ consists of all the duo's of combinations of a price and a quantity vector that can be generated by a homogeneous quadratic utility function at a point where it is nondecreasing and quasi-concave. It is important to realize that $\mathcal{D}$ does not equal the whole of $\mathbb{R}_+^{2\times m\times n}$, for there are data-duo's that cannot be rationalized as maximizations of the same linear quadratic utility function.[4] The Fisher index can nonetheless be computed for all combinations of positive prices and quantities.

Note that, given the restriction that we only search within the set $\mathcal{S}$, the data (further) restrict the set of utility functions that rationalize them. A function $\mathbf{F}$ being exact for a given set of utility functions $\mathcal{S}$, however, means that it has the

[2] $\langle \cdot, \cdot \rangle$ denotes the inner product, so $\langle \mathbf{p}^j, \mathbf{q}\rangle = \sum_{i=1}^n p_i^j q_i$

[3] This set can for practical purposes be restricted by some normalization on the matrices $\mathbf{A}$, since scaling of the matrix $\mathbf{A}$ changes utility functions, but not the underlying preference relations.

[4] A further restriction is that data points must be rationalizable by a utility function that is locally nondecreasing and quasi-concave. The set $\mathcal{S}$ here contains functions that are not globally nondecreasing and quasi-concave, but $\mathcal{D}$ can allow for datapoints from those parts of the function that are.

extremely convenient feature that the quotient of utilities in the given quantities is constant across the remaining possibilities for rationalizing the data.

The second is a *true* index. Although this term is often used in the literature, we could not find a formal definition (the closest we got was Afriat's (1981, p138) description of a true price index. See also Dowrick and Quiggin, 1997 and Neary, 2004). We do think however that the following definitions provide a formalization of the idea behind it. The definition of GARP, which is used in this definition, can be found in Varian (1982).

Definition 2 *A function* $\mathbf{F}$ *is a true index if for all datasets* $(\mathbf{P}, \mathbf{Q})$ *that satisfy GARP, there is a concave, monotonic, continuous and non-satiated utility function* u *that rationalizes the data and for which the following holds:*
$F_j(\mathbf{P}, \mathbf{Q}) > F_k(\mathbf{P}, \mathbf{Q}) \Leftrightarrow u(\mathbf{q}^j) > u(\mathbf{q}^k)$

It is not unimportant to note that this definition does not impose restrictions on $\mathbf{F}$ if the data do not satisfy GARP. Whether the domain of $\mathbf{F}$ should include datasets that violate GARP is an issue that will be discussed in Section 4 and web appendix C.

The second definition accommodates the possibility that we restrict ourselves to a specific set of utility functions.

Definition 3 *A function* $\mathbf{F}$ *is a true index for the set of utility functions* $\mathcal{U}$ *if for all datasets* $(\mathbf{P}, \mathbf{Q})$ *for which there is a* $v \in \mathcal{U}$ *that rationalizes the data, there is a* $u \in \mathcal{U}$ *that rationalizes the data and for which the following holds:*
$F_j(\mathbf{P}, \mathbf{Q}) > F_k(\mathbf{P}, \mathbf{Q}) \Leftrightarrow u(\mathbf{q}^j) > u(\mathbf{q}^k)$

Having two utility functions feature in the definition may at first sight seem superfluous, but they are there in order to allow for the possibility that a dataset $(\mathbf{P}, \mathbf{Q})$ can be rationalized by different utility functions in $\mathcal{U}$. If these utility functions disagree on the ranking of two consumption bundles, $\mathbf{F}$ cannot be in line with both - or all - but it should be enough if $\mathbf{F}$ is consistent with at least one of them.

A third possibility is that one single utility function is simply given or assumed.

Definition 4 *A function* $\mathbf{F}$ *is a true index for utility function* u *if for all datasets* $(\mathbf{P}, \mathbf{Q})$ *that* u *rationalizes, the following holds:*
$F_j(\mathbf{P}, \mathbf{Q}) > F_k(\mathbf{P}, \mathbf{Q}) \Leftrightarrow u(\mathbf{q}^j) > u(\mathbf{q}^k)$

Here the choice of u does not depend on the data ($\mathbf{P}$ and $\mathbf{Q}$) and in this case the making of index numbers can be considered to be a (hypothetical) exercise where we not only assume that the data follow from optimization of one utility function, but where we also know which utility function it is that is maximized. This severely constrains the data, as quantity bundles $\mathbf{q}^j$ must optimize utility u with given relative prizes $\mathbf{p}^j$, which is unlikely if the data themselves are not constructed and therefore hypothetical too. Note that especially in the older literature this is the sense in which the term true index is used; see for instance Henry Schultz (1939) and Alexandr A. Konüs (1939).

Definitions 2 and 4 can be seen as special cases of Definition 3. If we choose $\mathcal{U}$ to be the set of all concave, monotonic, continuous, non-satiated utility functions, then that gives us Definition 2. If $\mathcal{U}$ is a singleton set, then we are in the situation of Definition 4. Being true in our definition therefore is a relative concept. One index can be true for a large set of utility functions, another index can be true for a smaller set. If an index number is true for a large $\mathcal{U}$, then it can handle data generated by many different utility functions, while if $\mathcal{U}$ is smaller then the definition allows for $\mathbf{F}$ to behave erratically for more datasets $(\mathbf{P}, \mathbf{Q})$.

It is worthwhile noting that being true (Definition 2) does not imply being true for any $\mathcal{U}$ that is a subset of the set of all concave, monotonic, continuous, non-satiated utility functions (Definition 3). If the data can be rationalized by a utility function $u_1 \in \mathcal{U}$, but also by $u_2 \notin \mathcal{U}$, then a true index $\mathbf{F}$ can be in line with u_2, but not with any $u \in \mathcal{U}$. Conversely, it is also not the case that if an index is true for a set $\mathcal{U}$ that is a strict subset of all concave, monotonic, continuous, non-satiated utility functions (Definition 3), it should also be true in the sense of Definition 2. An index that is true for $\mathcal{U}$ may return numbers that are not consistent with any utility function for data that *do* satisfy GARP, but that cannot be rationalized by any utility function in $\mathcal{U}$.

We will close this section with a few examples. A well known procedure for making index numbers is the following. First a reference basket is made by averaging consumption bundles over countries: $\overline{\mathbf{q}} = \sum_{j=1}^{m} w_j \mathbf{q}^j$ with $\sum_{j=1}^{m} w_j = 1$. Then these reference baskets are used to compute price levels and those in turn are employed to deflate incomes, that is, $F_j(\mathbf{P}, \mathbf{Q}) = \frac{\langle \mathbf{p}^j, \mathbf{q}^j \rangle}{\langle \mathbf{p}^j, \overline{\mathbf{q}} \rangle}$. This gives an index that is true for the set of all Leontief utility functions; for any $u(\mathbf{q}) = \min\left\{\frac{\mathbf{q}^1}{a_1}, ..., \frac{\mathbf{q}^n}{a_n}\right\}$ we get consumption bundles that are multiples of $(a_1, ..., a_n)$, which implies that $\mathbf{F}$ and u agree on the ranking of all data generated by u. Note that being true for this particular set $\mathcal{U}$ allows $\mathbf{F}$ to behave in any conceivable way outside a very limited set of possible datasets; Leontief production functions only produce datasets in which all consumption bundles are multiples of each other.

Other examples are the Geary, EKS and CCD index. These indices are true for, respectively, the set of Leontief, homogeneous quadratic and homogeneous translog utility functions (see Neary, 2004, Proposition 5).

Finally it may be worthwhile observing that we can see an index that is exact for $\mathcal{S}$ as an index that is true for $\mathcal{S}$ with part (2) of Definition 1 as an additional requirement. We would need to take for $\mathcal{D}$ all datasets that can be rationalized by utility functions in $\mathcal{S}$. In the example given below Definition 1 we would take $\mathcal{S}$ to be the set of all homogeneous quadratic functions that are *globally* nondecreasing and quasi-concave. From this example we can furthermore conclude that it is not always easy to see whether or not an index number is true for a nontrivial set of utility functions.

3 The GAIA system

The core of Neary (2004) is the introduction of the GAIA system, which is a recipe for making index numbers for a given utility function u. The system is defined as follows:

$$\begin{aligned} E_j &= \frac{e(\mathbf{\Pi}, u_j)}{e(\mathbf{p}^j, u_j)} = \frac{\sum_{i=1}^n \Pi_i q_{ij}^*}{\sum_{i=1}^n p_{ij} q_{ij}}, j = 1, ..., m \\ \Pi_i &= \frac{\sum_{j=1}^m E_j p_{ij} q_{ij}}{\sum_{j=1}^n q_{ij}^*}, i = 1, ..., n \end{aligned}$$

where the q_{ij}^*'s denote the "virtual" quantities that follow from minimizing costs, given the actual utility level u_j, under prices $\mathbf{\Pi}$. The E_j's can be seen as exchange rates, and the Π_i's as world prices. The corresponding real incomes are $z_j^* = \sum_{i=1}^n \Pi_i q_{ij}^*$. Note that the virtual quantities q_{ij}^*, and therefore also the real incomes, remain the same if another utility function that represents the same preference relation is chosen.

Part of the elegance of the GAIA system is that many well-known index numbers are nested as special cases (see properties 4 to 6 in his paper). It thereby allows for a variety of ways in which it can be true. If the GAIA system uses a given utility function u, then it is true for u by construction (Definition 4). In Neary's paper, however, the system is used in combination with different specifications of u, or of the expenditure function, where the parameters are estimated from the data. Therefore, rather than looking at the GAIA system for a given u, we can also consider the combination of GAIA and the way in which u depends on $(\mathbf{P}, \mathbf{Q})$, which adds up to a (compound) function $\mathbf{F} : \mathbb{R}_+^{2\times m\times n} \rightarrow \mathbb{R}^m$. Under which category the resulting index falls, depends on how u is estimated. If u is estimated parametrically, with $\mathcal{U}$ the set of utility functions that the parametrization allows for, then the resulting index is true for $\mathcal{U}$ (Definition 3, see also Proposition 5 in Neary, 2004). If a non-parametric approach is taken that leads to a u that lies within the Afriat bounds (see Dowrick and Quiggin, 1994, 1997), then this $\mathbf{F}$ is a true index in the sense of Definition 2.

Although the GAIA system follows the 'economic' approach, one can evaluate the behaviour of the real incomes it produces by looking at which axioms hold for them. That is also what Neary does in his paper. In the web appendix to this note we discuss the specific axioms he considers and include some more axioms from Van Veelen (2002) that we think are relevant too.

4 Limitations of either approach

In the article, Neary describes the economic and the axiomatic approach as an approach with and one without a foundation in economic theory. In line with a tradition that can be traced back to Paul A. Samuelson and Subramanian Swamy (1974) or even to Frisch (1936), the axiomatic approach is said to treat prices and quantities as independent variables, whereas the economic approach, by contrast, assumes that quantities arise from optimizing behaviour. We would however like to argue that a more accurate description of the distinction can be made. As we see it, all complications that make international real income comparisons problematic are caused by two things: differences in relative prices and differences between (representative) consumers. The way in which the economic approach tries to overcome the problems caused by differences in relative prices, is to treat the observations as if they follow from optimizing one and the same utility function. An axiomatic approach, on the other hand, can very well aim at making real income comparisons, even if (representative) consumers differ. It therefore does not necessarily see prices and quantities as independent variables, it only takes differences in taste as another part of the explanation of differences in consumption bundles, together with, obviously, differences in budget sets. (Note that in his article, Neary does not claim that (representative) consumers would in fact have identical preference. It is only the procedure of determining the preferences of the *reference* consumer that treats them as such.)

If the axiomatic approach seems uneconomic, that may be due to the choice of axioms rather than the approach in itself. As discussed in the web appendix to this note, the *Matrix Consistency* axiom for instance rules out indicators of real income on grounds that have little connection to economic theory. On the other hand, however, even in the presence of possibly differing consumers, for example the *Weak Ranking Restriction* - ‘more of everything is better’ - from Van Veelen (2002) is a reasonable property for a real income indicator to have, and does not exclude optimizing of well-behaved utility functions.

(a) *Weak Ranking Restriction:* if $q_{ij} > q_{ik}$ for all $i = 1, ..., n$, then $F_j(\mathbf{P}, \mathbf{Q}) > F_k(\mathbf{P}, \mathbf{Q})$.

One can also formulate other reasonable axioms that make economic sense and do not need assumptions about (shared) preferences. We could for instance require that if the budget set of one consumer includes the budget set of the other, it should be associated with a real income that is at least as high.

(b) *Budget Set Ranking:* If for all $\mathbf{q}$ for which $\mathbf{p}^k\mathbf{q} \leq \mathbf{p}^k\mathbf{q}^k$ also $\mathbf{p}^j\mathbf{q} \leq \mathbf{p}^j\mathbf{q}^j$ holds, then $F_j \geq F_k$

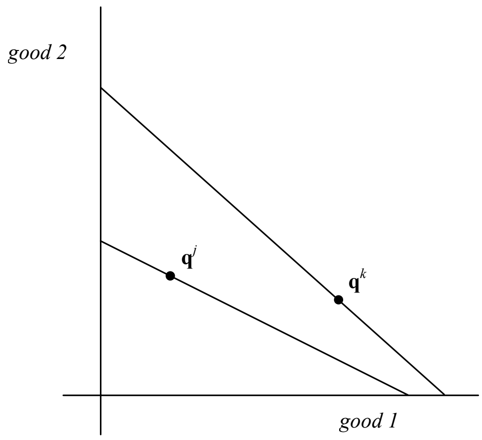

FIGURE 1. BUDGET SET RANKING WHERE PREFERENCES MIGHT DIFFER BETWEEN THE TWO OBSERVATIONS

Even if preferences differ, as they very well may in Figure 1, this axiom suggests that one budget set including the other is enough information to rank their real incomes. Note that this rules out rankings that the *Weak Ranking Restriction* allows for, and vice versa.

Obviously, if there were no differences in relative prices, then for any two budget sets, at least one of them would be a subset of the other. In that case one could easily define a function $\mathbf{F}$ that satisfies *Budget Set Ranking* and the *Weak Ranking Restriction*, as well as all other axioms mentioned in Van Veelen (2002). That is, on a restricted domain $\mathcal{D}_R = \{(\mathbf{P}, \mathbf{Q}) \subset \mathbb{R}_+^{2\times m\times n} \mid \forall\ j, k\ \exists\ \lambda_{jk} \text{ such that } \mathbf{p}^j = \lambda_{jk}\mathbf{p}^k\}$, an $\mathbf{F}$ defined by $F_j = \langle \mathbf{p}^1, \mathbf{q}^j \rangle$ satisfies all of the four axioms from Van Veelen, as well as the *Budget Set Ranking.* Unfortunately, however, relative prices do tend to differ, and this is where complications start. Once we lift the restrictions on the domain, we have the impossibility theorem of Van Veelen (2002), which states that there is no function $\mathbf{F}$ that satisfies four reasonable requirements - one of which is the *Weak Ranking Restriction.*

The economic approach on the other hand opens opportunities for handling combinations of budget sets for which one is not a subset of the other, if we can assume that consumption bundles follow from budget sets by optimization of a single utility function, as they very well may in Figure 2.

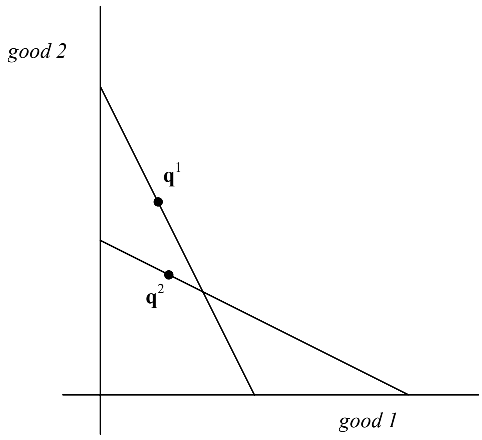

Figure 2. Possible observations from one preference relation under different price vectors

If these are observations from (representative) consumers optimizing one utility function over different budget sets, then the real income of country 1 can and should be ranked higher than country 2. This approach, however, is also not entirely unproblematic. First there may be data that violate GARP, implying that the assumption, at least in this simple version, does not hold. But even if we do not find violations of GARP, it can still be that preferences differ. These and related problems have gotten considerable attention in the literature, of which the web appendix contains a brief overview. The remaining question is therefore if we can satisfactorily handle the issue of possible heterogeneity in preferences.[5]

In the concluding section, Neary is careful and precise in discussing the meaning of the numbers. The approach is described as choosing a hypothetical reference consumer, and the data are said to provide an answer to the question: "How well off would the same reference consumer be in different countries?" The reasonable choice for a reference consumer is then one whose consumption patterns mimic world consumption behaviour as closely as possible (see also Proposition 6 of the paper). Neary therefore does not imply that the assumption does in fact hold, and the paper accurately describes the output of a procedure that only handles the data *as if* the assumption holds. Here we want to highlight that this procedure can

[5] We argue that the difference between the two approaches is that one does and the other does not assume homogeneity. The difference therefore is not just that the one uses axioms and the other does not. Moreover, what can be seen as characteristic properties of index numbers in the economic approach can generally also be stated as axioms. In Quiggin & Van Veelen (2007) consistency with GARP is stated as an axiom, and it is shown that this is inconsistent with two other axioms; the *Weak Ranking Restriction* and *Independence of Irrelevant Countries.* This indicates that the difference between assuming homogeneity and not assuming it can be reflected in axioms too; *Consistency with GARP* is after all an unreasonable axiom if there is heterogeneity, while *Independence of Irrelevant Countries* is an unreasonable restriction if there is homogeneity.

however give results that, when not interpreted with the care that Neary applies in his paper, point in the wrong direction.

Suppose that heterogeneity has a simple form, where there are two types of representative consumers. A reference consumer can be constructed by trying to best fit a utility function to the data. This estimated utility function will lie somewhere in between the two types. Consequently there will be a difference between type 1 and the reference consumer, which makes it possible that two countries, both of type 1, are ranked differently by the use of the reference consumer than they would be if we had used their true shared utility function. This can occur more generally if preferences of representative agents are heterogeneous: countries 1 and 2 may even have different preferences, yet they may both prefer the consumption bundle of country 1 to the consumption bundle of country 2, while the reference consumer ranks them the other way round. Basing policy implications for these two countries on the answer to the question how well off the reference consumer would be, may not be advisable in such a case. Note that if we want to exclude real incomes that rank one country's consumption bundle over the consumption bundle of the other, while both countries themselves would rank them the other way round, we will have to give up on *Transitivity* and make real income comparisons indexed by the country, or kind of country, that does the comparing.

Because heterogeneity may play a role for the interpretation and the use of the numbers, it would be useful to have some alarm bells going off if homogeneity is unlikely or impossible. This is discussed in the overview of test procedures in the web appendix. It would also be useful to think about a plan B if heterogeneity is more plausible than homogeneity.

Two tasks can be distinguished if there is likely to be heterogeneous preferences. One is to find out as much as we can about these preferences (as opposed to the assumed common preference relation.) Although heterogeneity and measurement error are in general hard to distinguish, it may not always be impossible. Allowing for dependence on observed characteristics would most probably broaden the possibilities. Combining data across countries with data across time might also help, if we think there is reason to assume that preferences are more or less constant across time (see also Robert J. Hill, 2004). Regrettably, relative prices are relatively constant across time too, which is likely to hamper the recovery of such utility functions.

The other task is to find out if and how we can make real income indicators once we have information on these heterogeneous preferences. This is also an ambitious task, which might include meeting with a variety of impossibilities. The present economic approach however implicitly is one way of dealing with heterogeneity too and it is not on forehand clear that there are no other ways of dealing with it that have other, perhaps better properties. This task may well require an axiomatic approach that explores the limits of what we can and what we cannot do with this extra information.

5 Conclusion

The difference between the axiomatic and the economic approach in index number theory is that the economic approach treats prices and quantities as observations that result from optimizing one single utility function, while the other tries to make meaningful comparisons without the assumption of homogeneity. While the axiomatic approach may include axioms that obstruct finding a good index if this assumption actually is correct, the economic approach is vulnerable to constructing indices of limited use if the assumption turns out to be false. Results such as Neary's (2004) therefore should be handled with care. A challenge for the future is to explicitly allow for heterogeneity in index number theory.

References

[1] **Afriat, Sydney N.** 1967. "The construction of utility functions from expenditure data." *International Economic Review*, 8(1): 67-77.

[2] **Afriat, Sydney N.** 1981. "On the constructability of consistent price indices between periods simultaneously." In *Essays in the Theory and Measurement of Consumer Behaviour in Honour of Sir Richard Stone*, ed. Agnus S. Deaton, 133-161. Cambridge: Cambridge University Press.

[3] **Caves, Dougles W., Laurits R. Christensen, and W. Erwin Diewert.** 1982. "On the economic theory of index numbers and the measurement of input, output, and productivity." *Econometrica,* 50(6): 1393-1414

[4] **Diewert, W. Erwin.** 1973. "Afriat and revealed preference theory." *Review of Economic Studies,* 40(3): 419-425.

[5] **Diewert, W. Erwin.** 1976a. "Exact and superlative index numbers." *Journal of Econometrics,* 4(2): 115-145.

[6] **Diewert, W. Erwin.** 1976b. "Harberger's welfare indicator and revealed preference theory." *American Economic Review,* 66(1): 143-152

[7] **Diewert, W. Erwin.** 1978. "Superlative index numbers and consistency in aggregation." *Econometrica,* 46(4): 883–900.

[8] **Drechsler, Laszlo.** 1973. "Weighing of index numbers in multilateral international comparisons." *Review of Income and Wealth,* 19(1): 17-34.

[9] **Dowrick, Steve. and John Quiggin.** 1994. "International comparisons of living standards and tastes: A revealed-preference analysis." *American Economic Review,* 84(1), 41-64.

[10] **Dowrick, Steve. and John Quiggin.** 1997. "True measures of GDP and convergence." *American Economic Review,* 87(1): 41-64.

[11] **Frisch, Ragnar.** 1936. "Annual survey of general economic theory: the problem of index numbers." *Econometrica,* 4(2): 1-38.

[12] **Hill, Robert J.** 2004. "Constructing price indices across space and time." *American Economic Review,* 94(5): 1379-1410.

[13] **Konüs, Alexandr A.** 1939. "The problem of the true index of the cost of living," *Econometrica,* 7(1): 10-29. (Translation of an article in Russion in 1924).

[14] **Konüs, Alexandr A. and S. S. Byushgens.** 1926. "K probleme pokupatelnoi cili deneg," *Voprosi Konyunkturi* 2(1): 151–72. English title: "On the problem of the purchasing power of money".

[15] **Neary, J. Peter.** 2004. "Rationalizing the Penn World Table: True Multilateral Indices for International Comparison of Real Income." *American Economic Review,* 94(5), 1411-1428.

[16] **Quiggin, John, and Matthijs van Veelen.** 2007. "Multilateral indices: conflicting approaches?" *Review of Income and Wealth,* 53(2): 372-378.

[17] **Samuelson, Paul A., and Subramanian Swamy.** 1974. "Invariant economic index numbers and canonical duality: survey and synthesis." *American Economic Review,* 64(4), 566-593.

[18] **Schultz, Henry.** 1939: "A misunderstanding in index-number theory: the true Konus condition on cost-of-living index numbers and its limitations." *Econometrica,* 7(1): 1-9.

[19] **Varian, Hal R.** 1982. "The nonparametric approach to demand analysis."*Econometrica,* 5(4): 945-973.

[20] **Varian, Hal R.** 1983. "Non-parametric tests of consumer behaviour." *Review of Economic Studies,* 50(1): 99–110.

[21] **van Veelen, Matthijs.** 2002. "An impossibility theorem concerning multilateral international comparison of volumes." *Econometrica,* 70(1): 369-375.

Web appendix to: A note on different approaches to index number theory

By Matthijs van Veelen and Roy van der Weide

August 31, 2007

In this appendix we will do three things. First we compare the axioms that Peter J. Neary (2004) uses to those that feature in the impossibility theorem of Matthijs van Veelen (2002). In the discussion we will pay special attention to the economic sensibility of the different axioms. Second we evaluate which of these axioms hold - or in which way they do - for the GAIA system. Finally we give a brief review of ways in which one can test for heterogeneity (or homogeneity).

A Axioms as criteria for choosing between index numbers

We begin with describing a result from Van Veelen (2002).

The apples-and-oranges theorem involves a function $\mathbf{F}$ that is defined on the set of all possible combinations of non-negative prices and quantities, $\mathbf{F} : \mathbb{R}_{+}^{2\times m\times n} \to \mathbb{R}^{m}$. This function should capture the notion of wealth, and an index Q_{jk} that compares country j and k then would be F_j/F_k. The first axiom is *Weak Continuity*, which requires that if $\mathbf{F}$ ranks one country higher than another, then this ranking should not be sensitive to arbitrarily small changes in prices and/or quantities. The second axiom is *Dependence on Prices*, which rules out indices, or functions $\mathbf{F}$, in which prices do not play a role at all. The third axiom is the *Weak Ranking Restriction*, that states that consuming more of everything is better. We will return to this axiom, and therefore we state it formally:

(a) *Weak Ranking Restriction:* if $q_{ij} > q_{ik}$ for all $i = 1, ..., n$, then $F_j(\mathbf{P}, \mathbf{Q}) > F_k(\mathbf{P}, \mathbf{Q})$.

The fourth axiom is *Independence of Irrelevant Countries*, that requires the relative ranking of two countries not to be sensitive to changes in a third country.

The impossibility theorem states that these four relatively reasonable requirements are inconsistent:

Theorem *If $m \geq 3$ and $n \geq 2$ then there is no function* $\mathbf{F}$ *that satisfies Weak Continuity, Dependence on Prices, the Weak Ranking Restriction and Independence of Irrelevant Countries.*

In Neary (2004) the index number problem is described a little differently. There we are to find a set of index numbers $Q_{jk}, j, k = 1, ..., m$, which gives the real income of each country j relative to every other country k. Contrary to the description in Van Veelen (2002), this does allow for intransitive indices, so the first axiom of Neary is:

(i) *Transitivity* or *Circularity:* Country j's real income relative to country k's should be the same whether the two are compared directly or via an arbitrary intermediate country l: $Q_{jk} = Q_{jl}Q_{lk}$.

This restriction on the set of numbers $Q_{jk}, j, k = 1, ..., m$ makes that together they provide a unique cardinal ranking of real incomes. With this restriction, a set of index numbers $Q_{jk}, j, k = 1, ..., m$, all of which obviously are functions of prices and quantities, is equivalent to a function $\mathbf{F} : \mathbb{R}_+^{2\times m\times n} \rightarrow \mathbb{R}^m$.

Neary's second axioms is:

(ii) *Characteristicity* or *Independence of Irrelevant Countries:* Country j's real income relative to country k's should be unaffected by changes in a third country.

While this is the classic definition of *Independence of Irrelevant Countries*, Van Veelen (2002) uses an ordinal version of this axiom, where changes in a third countries are only demanded not to affect country j ranking higher or lower than country k.

The third axiom is *Matrix Consistency.* This axiom is stated informally in Neary's paper. For further discussion a more formal definition will be useful, so we will give one that seems to match the informal description (See also W. Erwin Diewert (1999) for a definition of additivity and Itsuo Sakuma, D.S. Prasada Rao and Yoshimasha Kurabayashi (2000) for the difference between additivity and matrix consistency).

(iii) *Matrix Consistency:* A function $\mathbf{F} : \mathbb{R}_+^{2\times m\times n} \rightarrow \mathbb{R}^m$ is matrix consistent if $F_j(\mathbf{P}, \mathbf{Q}) = \left\langle \mathbf{p}(\mathbf{P}, \mathbf{Q}), \mathbf{q}^j \right\rangle$ for every j.[1]

[1] $\langle \cdot, \cdot \rangle$ denotes the inner product, so $\left\langle \mathbf{p}(\mathbf{P}, \mathbf{Q}), \mathbf{q}^j \right\rangle = \sum_{i=1}^{n} p_i(\mathbf{P}, \mathbf{Q})\, q_{ij}$

Note that this definition makes that the consumption bundles of all countries j are evaluated by one and the same vector of weights $\mathbf{p}(\mathbf{P},\mathbf{Q})$, that itself is a function of $\mathbf{P}$ and $\mathbf{Q}$. It is natural, in the spirit of Geary, to call this a vector of world prices.

The reason to choose this axiom clearly is a practical one. It imposes a restriction on the form that $\mathbf{F}$ can have, namely that it should allow for consistent disaggregation by country as well as by commodity. When all bundles are weighted by the same set of prices, the 'values against world prizes' of the consumption of good i in country j can be summed up over the goods to get country incomes, or over countries to get world consumption levels. Yet however convenient it may seem to be if we can disaggregate by country as well as by commodity, it is good to be aware that in order to achieve this practical advantage, this axiom restricts us to a set of functions $\mathbf{F}$ - or a set of indices - that is conceptually rather peculiar if we want to make wealth comparisons within the economic approach. The following figure indicates why.

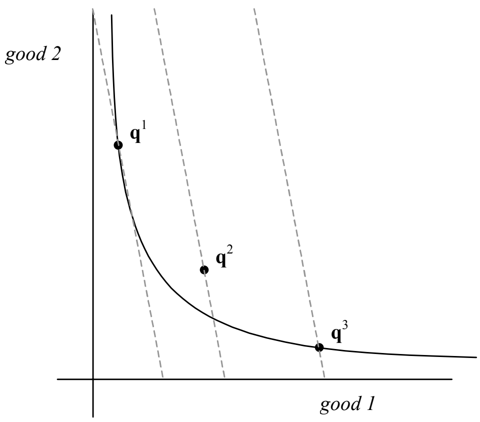

FIGURE 3. NO CHOICE OF A PRICE VECTOR RANKS $\mathbf{q}^2$ OVER THE OTHER TWO CONSUMPTION BUNDLES

In this picture there are two points on an indifference curve, and one above it. The lines through the three points represent curves of equal value of $\langle \mathbf{p}(\mathbf{P},\mathbf{Q}), \mathbf{q}\rangle$ for given $\mathbf{p}(\mathbf{P},\mathbf{Q})$. One can directly see that no vector $\mathbf{p}(\mathbf{P},\mathbf{Q})$ can make the consumption bundle above the indifference curve be ranked above both other consumption bundles. Indeed, for any choice of $\mathbf{p}(\mathbf{P},\mathbf{Q})$ - that is, for any slope of the lines through the consumption bundles - it is clear that the line through the point above the indifference curve will be under at least one of the other two lines. Therefore, if the observations result from optimization of one utility function, then *Matrix Consistency* goes against the assumption of convex preferences. In other words, even if we are in the comfortable situation where preferences are convex as well as identical for everyone, this axiom would prevent us from recovering them. This contradiction indicates that although 'consistent disaggregation' is a term that has a certain bookkeeping appeal, it is conceptually not really a desirable property. Therefore the practical advantage of being able to disaggregate incomes comes at the cost of leaving us with disaggregate numbers for which we have no useful interpretation when making wealth comparisons.

A weaker axiom that does seem reasonable in itself is the *Weak Ranking Restriction* mentioned above. *Matrix Consistency* implies the *Weak Ranking Restriction*, but there is a variety of functions $\mathbf{F}$ that does satisfy the latter, but not the former. Take for instance any function that does satisfy *Matrix Consistency*, that is, a function $\mathbf{F}$ for which there is a $\mathbf{p}(\mathbf{P},\mathbf{Q})$ such that $F_j(\mathbf{P},\mathbf{Q}) = \langle \mathbf{p}(\mathbf{P},\mathbf{Q}), \mathbf{q}^j \rangle$ for all j. Then $\mathbf{G}$ with $G_j = \left\langle \mathbf{p}(\mathbf{P},\mathbf{Q}), (\mathbf{q}^j)^2 \right\rangle$ for all j satisfies the *Weak Ranking Restriction* but not *Matrix Consistency.*

The remarks so far suggest that Van Veelen (2002) contains two less strict alternatives for two of the axioms. There is however also a way in which the axioms of Neary are not strict enough. Although none of the indices that feature in the paper satisfies the three axioms given, there is one that does, namely the index $Q_{jk} = \langle \overline{\mathbf{p}}, \mathbf{q}^j \rangle / \langle \overline{\mathbf{p}}, \mathbf{q}^k \rangle$, where $\overline{\mathbf{p}}$ is an arbitrary fixed price vector, or, equivalently, a function $\mathbf{F}$ with $F_j = \langle \overline{\mathbf{p}}, \mathbf{q}^j \rangle$. This being an index that many will disagree with as a choice for an indicator of real income, it is ruled out by *Dependence on Prices* from Van Veelen (2002).

B Which axioms does the GAIA system satisfy?

Although the GAIA system follows the 'economic' approach, one can evaluate the behaviour of the real incomes it produces by looking at which axioms hold for them. A question one can ask first is: when we check for these axioms to hold, do we look at GAIA for a given u, or do we look at the combination of GAIA and a way to arrive at the utility function u to be used, as a function of prices and quantities. We will consider them as applying to the latter, that is, to the whole procedure of making real incomes from price and quantity data. The former though can be seen as a special case of the latter, with a degenerate estimation procedure that returns the same utility function u for every possible $(\mathbf{P},\mathbf{Q})$. As we will see, which axioms are satisfied sometimes does and sometimes does not depend on how u behaves as a function of $\mathbf{P}$ and $\mathbf{Q}$.

By construction, the GAIA real incomes satisfy *Transitivity*, whatever way of arriving at the utility function is used. They do, on the other hand, not satisfy the cardinal version of *Independence of Irrelevant Countries.* The ordinal version is satisfied for a fixed u, but, again, not for sensible ways to estimate u from $\mathbf{P}$ and $\mathbf{Q}$. In Section IV as well as in Section VIII of Neary, it is stated that the system satisfies *Matrix Consistency*, albeit in terms of virtual rather than actual consumption levels. This is actually a bit too modest, because the fact that it does not satisfy *Matrix Consistency* in the sense defined above is only good, for reasons that we give there. On the other hand, there is no straightforward and meaningful definition of *Matrix Consistency* in terms of virtual quantities. It seems that it is meant to restrict the function $\mathbf{F}$ to have a form $F_j(\mathbf{P},\mathbf{Q}) = \left\langle \mathbf{p}(\mathbf{P},\mathbf{Q}), \mathbf{q}^{j*}(\mathbf{P},\mathbf{Q}) \right\rangle$ for every j. This, however, is no restriction at all; any function $\mathbf{F}$ can be rewritten in this way by appropriate choice of virtual quantities. To see how, take any function $\mathbf{F}$ and any function $\mathbf{p}(\mathbf{P},\mathbf{Q})$. Then we choose $\mathbf{q}^{j*}(\mathbf{P},\mathbf{Q}), j = 1, ..., m$ as follows: $q^*_{ij}(\mathbf{P},\mathbf{Q}) = q_{ij}\left(\frac{F_j(\mathbf{P},\mathbf{Q})}{\langle \mathbf{p}(\mathbf{P},\mathbf{Q}), \mathbf{q}^j \rangle}\right)\ \forall\ i, j$. Now we have constructed virtual quantities that, together with $\mathbf{p}(\mathbf{P},\mathbf{Q})$, make $\mathbf{F}$ matrix consistent; by construction the inner product of $\mathbf{p}(\mathbf{P},\mathbf{Q})$ and $\mathbf{q}^{j*}(\mathbf{P},\mathbf{Q})$ returns $F_j(\mathbf{P},\mathbf{Q})$. This can obviously be done for any $\mathbf{F}$, so matrix consistency in terms of virtual quantities does not impose restrictions on the indices, as long as we are free to choose the way prices and quantities translate to virtual quantities. If we however would want to turn this into an axiom that can also be violated, we would have to restrict the form that the $\mathbf{q}^{j*}$'s have as functions of $\mathbf{P}$ and $\mathbf{Q}$, possibly in relation to each other. This we think is an exercise of rather limited use, because what it would do is create an axiom that allows for the GAIA system and excludes all others, whereas the idea of an axiom is that it defines a property that is more or less reasonable in itself. This important aspect is lost if we try to fit an axiom to an index, rather than try to find indices that have desirable properties. Furthermore it is hard to think of any good conceptual reason why to demand *Matrix Consistency* and not just the *Weak Ranking Restriction*.

The GAIA real incomes do satisfy the *Weak Ranking Restriction* as long as $u(\mathbf{P},\mathbf{Q})$ returns quasi-concave, non-satiated utility functions, and it satisfies *Dependence on Prices* if $u(\mathbf{P},\mathbf{Q})$ truly depends on $\mathbf{P}$. Under reasonable restrictions on the way in which the utility function follows from $\mathbf{P}$ and $\mathbf{Q}$ - some equivalent of continuity - it also satisfies *Weak Continuity* (see van Veelen, 2002).

C Alarm bells: how to check for homogeneity

This appendix considers different indications one can look for when we would like to know whether or not homogeneity is a good assumption. Some of those indications are obviously more sophisticated than others. A standard revealed preference approach would be to determine the set of utility functions that rationalize the data. If we find out that it is empty, then one could conclude that the assumption of homogeneiy does not hold. Possibility or impossibility however is a relatively crude indication. On the one hand, even the slightest noise (measurement error, or imprecision in optimizing) can in principle produce datasets that do not allow for rationalization by one utility function. It may therefore be too demanding to simply want one utility function and nothing more to rationalize the data. On the other hand, there are possibly quite different ways to produce data that can still be rationalized by one utility function. The fact that we can rationalize the data therefore may not be very informative about how likely it is that we are correct in doing so.

Without further restrictions on the utility functions being used (apart from being non-satiated), the first alarm bells should nonetheless be revealed preference tests. It should be noted however that indices that are true for a set $\mathcal{U}$ in principle work in the same way, but there the search for utility functions that rationalize the data is restricted on forehand to a subset $\mathcal{U}$.[2] What is also good to remember is that the corresponding indices usually still can be calculated, even when there is no utility function in $\mathcal{U}$ that rationalizes the data. Neary (2004) however does not ignore this alarm bell; see page 1415.

Violations of GARP may be attributed to different reasons: measurement error, optimization error or heterogeneity. The literature suggests a number of different ways of accommodating a reasonable level of error. Two approaches may be distinguished; the statistical and the non-statistical. An early proposal of a statistical test can be found in Hal R. Varian (1985). A recent, but similar, alternative is a test by Philippe de Peretti (2005). The two are compared in Barry E. Jones and Philippe de Peretti (2005). Both test the same null hypothesis: the observed data is a noisy version of the true unobserved data set that satisfies GARP. There are however two differences. The first is that Varian (1985) and de Peretti (2005) use different adjustment procedures to turn the observed data into a data set that does satisfy GARP (which then comes with a utility function that rationalizes them). The second is that they choose different test procedures, which is a matter of prac-

[2] If the set of utility functions $\mathcal{U}$ is a parametrized set, which it typically is, then the drawbacks can be stated even a little more precisely. If the observations do indeed come from optimizing one utility function in $\mathcal{U}$, then the data obviously will be rationalizable by an element of $\mathcal{U}$. But if there is even the slightest of deviations, then the data will in general not be rationalizable as soon as there are more observations than parameters. On the other hand, the freedom to choose parameters (or, equivalently, utility functions within $\mathcal{U}$) may be large enough for heterogeneity to go unnoticed. What is different from revealed preference tests is that if $(\mathbf{P}, \mathbf{Q})$ cannot be rationalized by a $u \in \mathcal{U}$, then it is not on forehand clear whether this is due to optimization of one utility function being a wrong assumption, or to the true utility function not belonging to the set $\mathcal{U}$.

ticality and power. See also Larry G. Epstein and Adonis J. Yatchew (1985), Barry E. Jones, Donald H. Dutkowsky, and Thomas Elger (2005) and Adrian R. Fleissig and Gerald A. Whitney (2005).

Also statistical yet a very different test is that of John Gross (1995). Here heterogeneity is put at the center of attention. In contrast to the null of homogeneous optimizing behaviour, Gross (1995) formulates the null hypothesis as: "The data were generated by consumers with different preferences (or whose preferences have changed over time)". The design of the approach entails a partition of the data into two subsets, which are denoted CS and VS. Where CS is consistent with GARP, adding observations from VS will induce violations of GARP. By construction, observations in VS are inconsistent with the preferences that rationalize the observations in CS. Gross (1995) quantifies this divergence by estimating the expenditure wasted by those in VS when maximizing utility consistent with CS, which leads to the test statistic. Naturally as the heterogeneity in preferences becomes smaller the statistic will tend to zero. The null of heterogeneity is rejected when the statistic becomes too small.

Varian (1990) advocates a non-statistical approach. Statistical tests will reject the null if the observed value of the test statistic is improbable under the null. Varian (1990) notes that "given enough data we can always reject optimizing behaviour, even if it is 'nearly optimizing behaviour'. The value of the test statistic will typically give no clue as to whether the economic agent under examination is nearly optimizing or grossly nonoptimizing".[3] Accordingly, Varian (1990) proposes a function of the observed data that captures the seriousness of a violation of GARP without considering the likelihood of such a violation.

James Andreoni and William T. Harbaugh (2005) and references therein look at the power of the standard revealed preference test, which simply checks whether or not the data satisfy GARP. One general recipe is to look at ways to generate data other than optimization of a utility function. If such an alternative data-generating process implies that violations of GARP are relatively likely, then the power of the revealed preference test against this alternative is high. A classical example is Stephen G. Bronars (1987). It should be noted that the alternative hypotheses discussed in Andreoni and Harbaugh (2005) are different forms of non-rational behaviour. This naturally explores the power of GARP as a way to test whether or not one subject conforms to a model of optimizing behaviour. However,

[3] The exact phrase reads "given enough data we can always reject *non*optimizing behaviour, even if it is 'nearly optimizing behaviour'. The value of the test statistic will typically give no clue as to whether the economic agent under examination is nearly optimimizng or grossly nonoptimizing". Our guess is that '*non*optimizing' should have been 'optimizing'.

if it is assumed that *different* subjects optimize the *same* utility function, as we do here, then a natural alternative hypothesis will be heterogeneity. This can take different forms, against which the power of GARP will differ.

For part of the literature, heterogeneity is less of an issue. For instance in Richard W. Blundell, Martin Browning, and Ian A. Crawford (2003), GARP tests are performed on estimated demand functions to check whether aggregate demand still follows some regularities, thereby allowing heterogeneity to end up in the residuals. In Arthur Lewbel (2001), the aim is to make theoretical connections between rationality of individual demand functions and rationality of statistical demand functions, whereby the latter can be useful, even if it is estimated from data that are really observations from heterogeneous agents.

References

[1] **Andreoni, James, and William T. Harbaugh.** 2005. "Power indices for revealed preference tests," University of Wisconsin Madison Working Paper 2005-10

[2] **Blundell, Richard W., Martin Browning, and Ian A. Crawford.** 2003. "Nonparametric Engel curves and revealed preference." *Econometrica,* 71(1), 205-240.

[3] **Bronars, Stephen G.** 1987. "The power of non-parametric tests of preference maximization," *Econometrica,* 55 (3), 208-240.

[4] **Diewert, W. Erwin.** 1999. "Axiomatic and economic approaches to international comparisons." In *International and Interarea Comparisons of Income, Output and Prices,* ed. Alan Heston and Robert E. Lipsey, 13-87. Chicago: University of Chicago Press.

[5] **Epstein, Larry G., and Adonis J. Yatchew.** 1985. "Nonparametric hypothesis testing procedures and applications to demand analysis," *Journal of Econometrics* **30**, 149-169.

[6] **Fleissig, Adrian R., and Gerald A. Whitney.** 2005. "Testing for the significance of violations of Afriat's inequalities." *Journal of Business & Economic Statistics,* 23(3), 355-362.

[7] **Gross, John.** 1995. "Testing data for consistency with revealed preference." *The Review of Economics and Statistics,* 77(4), 701-710.

[8] **Jones, Barry E., Donald H. Dutkowsky, and Thomas Elger.** 2005. "Sweep programs and optimal monetary aggregation." *Journal of Banking and Finance,* 29(2), 483-508.

[9] **Jones, Barry E., and Philippe de Peretti.** 2005. "A comparison of two methods for testing the utility maximization hypothesis when quantity data are measured with error." *Macroeconomic Dynamics,* 9(5), 612-629.

[10] **Lewbel, Arthur.** 2001. "Demand systems with and without errors." *American Economic Review,* 91(3), 611-618.

[11] **Neary, J. Peter.** 2004. "Rationalizing the Penn World Table: True Multilateral Indices for International Comparison of Real Income." *American Economic Review,* 94(5), 1411-1428.

[12] **de Peretti, Philippe.** 2005. "Testing the significance of the departures from utility maximization." *Macroeconomic Dynamics,* 9(3), 372-397.

[13] **Sakuma, Itsuo, D.S. Prasada Rao, and Yoshimasha Kurabayashi.** 2000. "Additivity, matrix consistency and a new method for international comparisons of real income and purchasing power parities." Conference paper.

[14] **Varian, Hal R.** 1985. "Non-parametric analysis of optimizing behaviour with measurement error." *Journal of Econometrics,* 30(1-2), 445-458.

[15] **Varian, Hal R.** 1990. "Goodness-of-fit in optimizing models." *Journal of Econometrics,* 46(1-2), 125-140.

[16] **van Veelen, Matthijs.** 2002. "An impossibility theorem concerning multilateral international comparison of volumes." *Econometrica,* 70(1), 369-375.