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

By Matthijs van Veelen and Roy van der Weide

July 19, 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, $\mathbf{F} : \mathbb{R}_{+}^{2\times m\times n} \rightarrow \mathbb{R}^{m}$. 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 $\mathbf{F} : \mathbb{R}_+^{2\times m\times n} \to \mathbb{R}^m$ is matrix consistent if $F_j(\mathbf{P},\mathbf{Q}) = \langle \mathbf{p}(\mathbf{P},\mathbf{Q}), \mathbf{q}^j \rangle$ for every j.[1]

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.

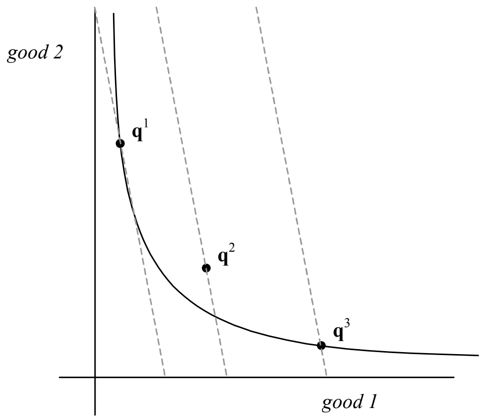

FIGURE 3. NO CHOICE OF A PRICE VECTOR RANKS $\mathbf{q}^2$ OVER THE OTHER TWO CONSUMPTION BUNDLES

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

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 $F_j(\mathbf{P}, \mathbf{Q}) = \langle \mathbf{p}(\mathbf{P}, \mathbf{Q}), \mathbf{q}^{j*}(\mathbf{P}, \mathbf{Q}) \rangle$ 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

[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 $\mathcal{U}$.

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 practicality 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

[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'.

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, 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.