# An impossibility theorem concerning multilateral international comparison of volumes

van Veelen, M.

**DOI**
10.1111/1468-0262.00280

**Publication date**
2002

**Document Version**
Final published version

**Published in**
Econometrica

*Econometrica*, Vol. 70, No. 1 (January, 2002), 369–375

# AN IMPOSSIBILITY THEOREM CONCERNING MULTILATERAL INTERNATIONAL COMPARISON OF VOLUMES

By Matthijs van Veelen[1]

## 1. introduction

Comparing wealth of nations—or across years—is something an economist naturally wants to do. The classic problem we then face, and which the standard GDP ignores, is that prices are not necessarily the same in different countries. Trying to correct for this, quite a few methods of comparison have been designed, but index theorists have always been aware that all of them have their drawbacks, which for instance is reflected in overviews of Balk (1996) and Hill (1997). As long as we only want to compare two countries we do have fairly acceptable index numbers, but all known multilateral indices suffer from unwanted peculiarities. A natural question is then whether we can expect a perfect method ever to be discovered. The statement of this paper is that this infallible method does not exist and that the apples and oranges character of multilateral comparison is more fundamental.

In Section 2, four conditions will be formulated that one could regard as minimum prerequisites for such a method to be satisfactory and it will be shown that it is impossible that all four are satisfied by one method. To link up with the existing literature, two comments are to be made. First I would like to mention Diewert (1986, 1999), who also has designed two sets of properties one could expect methods of comparison to have, the second of which is an extension of the first. It should be noted that the set of four requirements for which impossibility is proved here does not coincide with his nine (1986) or twelve (1999) commandments. The third section of this paper is devoted to the discussion of the relation between the different requirements and there I will argue that mine are quite modest.

The other remark is that it is generally acknowledged that the step from bilateral to multilateral comparison creates a serious problem. In spite of this general awareness, I do think that this impossibility theorem is a contribution, since so far it has never been proved that there is no method of multilateral comparison behaving in a way we can reasonably expect it to do.

## 2. the apples and oranges theorem

### 2.1. *Notation*

Prices and quantities of $M$ goods that are being consumed in $K$ countries are known and ordered in matrices $P$ and $Q$.[2] These quantities can be either aggregated or per capita numbers, determining whether we are talking about an aggregate or a per capita

[1] I would like to thank Daniel van Vuuren, who pointed out a mistake in an earlier version of the proof, Gerard van der Laan, Bert Balk, Prasada Rao, the two referees, and the co-editor.

[2] Without exception we consider consumption and not production. The theorem is just as valid if we would have chosen otherwise.

measure.[3] The rows of these matrices indicate the countries; their columns indicate the different goods. It may also turn out to be useful to have superscripts to represent the row indices and subscripts to represent the column indices. So both $P$ and $Q$ are $K \times M$ matrices with

$$[P]_{kj} = p_j^k, \text{ representing the price of good } j \text{ in country } k \text{ and}$$

$$[Q]_{kj} = q_j^k, \text{ representing the quantity of good } j \text{ consumed in country } k.$$

Accordingly, the $k$th row of $P$ will be denoted by $p^k$ and the $k$th row of $Q$ by $q^k$.

What we are looking for is a method to be applied to a group of countries that gives each of those countries a single number, compressing the information contained in the matrices to a vector, entries of which can easily be compared. To put it differently: for a fixed number of goods $M$ and a fixed number of countries $K$ we want to have a function $F$ that is defined on the set of all possible combinations of nonnegative prices and quantities; $F : \mathbb{R}_+^{2\times K\times M} \to \mathbb{R}^K$. This function should capture the notion of wealth; a higher function value is to indicate higher aggregate consumption. We say that a function does so properly if it meets the following four conditions.

### 2.2. *Properties*

The first property we expect such a function $F$ to have, is that the relative ranking of countries is not sensitive to arbitrarily small changes.

1. (*Weak Continuity*): $F_k(P, Q) > F_l(P, Q) \Rightarrow \exists\, \varepsilon > 0$ *such that* $\|(R, S) - (P, Q)\| < \varepsilon \Rightarrow F_k(R, S) > F_l(R, S)$.[4]

Furthermore we demand that the prices play a role in the comparison of countries. We say that prices do not play a role in the comparison if another function, that disregards prices, leads to the same rankings for all possible combinations of prices and quantities.

2. (*Dependence on Prices*): *There is no* $G : \mathbb{R}_+^{K\times M} \to \mathbb{R}^K$ *such that* $\forall\, P, Q : F_k(P, Q) \geq F_l(P, Q) \Leftrightarrow G_k(Q) \geq G_l(Q)$.

The third condition we impose is that a country that consumes more of every single good than another country should also be ranked higher than this other country. This very modestly formalizes our notion of the information that should be contained in a volume indicator and restricts the relative ranking of countries on subsets of the domain that can be seen as the 'far out' parts, where relative ranking is obvious.

3. (*Weak Ranking Restriction*): $q^k > q^l \Rightarrow F_k(P, Q) > F_l(P, Q)$, *where* $q^k > q^l$ *means that* $q_j^k > q_j^l,\ j = 1, \ldots, M$.

Finally we consider it reasonable if the relative ranking of two countries is not dependent on prices or quantities in countries other than those two. Formally it says that if $P$ and $R$ as well as $Q$ and $S$ share their $k$th and $l$th row, then the rest of the matrices should

[3] Aggregate here means aggregated over people, not to be confused with the main problem of this article, which is aggregation over goods.

[4] Here we use the Euclidian norm on $\mathbb{R}^{2\times K\times M}$.

not make a difference for their relative ranking. A difference in the rows other than $k$ and $l$ can be seen as either a change in the rest of the world or a change in the composition of the group in which countries $k$ and $l$ are being compared to each other.

4. (*Independence of Irrelevant Countries*): *If for matrices* $P, Q, R,$ *and* $S$, $r^k = p^k$, $r^l = p^l$, $s^k = q^k$, *and* $s^l = q^l$, *then* $F_k(P,Q) > F_l(P,Q) \Leftrightarrow F_k(R,S) > F_l(R,S)$.

### 2.3. *Impossibility*

THEOREM: *There is no function* $F : \mathbb{R}_+^{2\times K\times M} \to \mathbb{R}^K$ *that satisfies Properties* 1, 2, 3, *and* 4, *with* $M \geq 2$ *and* $K \geq 3$.

PROOF: Let $F$ be a function that satisfies weak continuity (1), the weak ranking restriction (3), and independence of irrelevant countries (4). Then we will show that it must violate dependence on prices (2). To do so we make a detour by going from the function $F$ to binary relations on $\mathbb{R}_+^{2\times M}$ and derive some properties of those relations using 1, 3, and 4. Then we define a function $G$ again that represents $F$ equally well with identical components and finally we will infer that we have arrived at a violation of property 2.

For given $F$ we define $K(K-1)$ binary relations $R_{kl}$, $1 \leq k,\ l \leq K$, $k \neq l$ on $\mathbb{R}_+^{2M} = \{(r,s) | r \in \mathbb{R}_+^M, s \in \mathbb{R}_+^M\}$ as follows:

For $k \neq l$, $(r,s)R_{kl}(t,u)$ if and only if $\exists\, P$ and $Q$ with $p^k = r$, $p^l = t$, $q^k = s$, and $q^l = u$ such that $F_k(P,Q) \geq F_l(P,Q)$. First note that by independence of irrelevant countries (4) the following statements are equivalent:

(a) $\exists\, P, Q$ with $p^k = r$, $p^l = t$, $q^k = s$, and $q^l = u$ such that $F_k(P,Q) \geq F_l(P,Q)$;

(b) $\forall\, P, Q$ with $p^k = r$, $p^l = t$, $q^k = s$, and $q^l = u$, $F_k(P,Q) \geq F_l(P,Q)$ holds.

Furthermore it is evident that $(r,s)R_{kl}(t,u)$ and $(t,u)R_{lm}(v,w)$ implies that $(r,s)R_{km}(v,w)$ and that not $(r,s)R_{kl}(t,u)$ implies that $(t,u)R_{lk}(r,s)$.

We now want to show that these $K(K-1)$ relations must all be the same. First suppose that $\exists\, k \neq l \neq m \neq k$ and $(r,s)$, $(t,u)$ for which $(r,s)R_{kl}(t,u)$ but not $(r,s)R_{km}(t,u)$. Now take $P$ and $Q$ such that $p^k = r$, $q^k = s$, $p^l = p^m = t$, and $q^l = q^m = u$. Then it follows that $F_k(P,Q) \geq F_l(P,Q)$ and that $F_k(P,Q) < F_m(P,Q)$ and thus $F_m(P,Q) > F_l(P,Q)$. It is impossible to reconcile the last statement with weak continuity (1) and the weak ranking restriction (3), because weak continuity would imply that a strict increase in all entries of $q^l$ is possible without affecting the ordinality of $F_m$ and $F_l$, but this leads to a violation of the weak ranking restriction. Thus we arrive at the conclusion that $R_{kl} = R_{km}$. A similar story can be told to show that $R_{kl} = R_{ml}$, resulting in the final statement that we can be sure that $R_{kl} = R_{mn}$, $k \neq l, m \neq n$.[5] Furthermore we can see that this relation $R$ on $\mathbb{R}^{2\times M}$, hereafter stripped of its needless subscripts, is complete, reflexive, and transitive. This binary relation reflects the ordinality of any comparison of two countries, that is, for all $k \neq l$,

$$F_k(P,Q) \geq F_l(P,Q) \Leftrightarrow (p^k, q^k)R(p^l, q^l).$$

Now we go back from relations to functions again and see if we can find a function that reflects our relation $R$. Weak continuity (1) of $F$ implies the continuity of our relations

[5] Note that we need $K \geq 3$ here to establish $R_{kl} = R_{lk}$; $R_{kl} = R_{km} = R_{lm} = R_{lk}$.

$R$, and thus we know that there exists a continuous function $g$ such that

$$g(p^k, q^k) \geq g(p^l, q^l) \iff (p^k, q^k)R(p^l, q^l).^6$$

If we then define $G(P, Q)$ by $G_k(P, Q) = g(p^k, q^k)$, we find that

$$(*) \qquad F_k(P, Q) \geq F_l(P, Q) \Leftrightarrow G_k(P, Q) \geq G_l(P, Q).$$

Finally we will establish that $g$ does not depend on the price vector. To do so, assume that there are $(p, q), (p', q) \in \mathbb{R}^{2\times M}$ for which $g(p', q) \neq (p, q)$. Without loss of generality we can say that $g(p', q) < g(p, q)$. But then, by continuity of $g$, we can again find a $q'$ such that $q' > q$ and $g(p', q') < g(p, q)$,[7] where $q' > q$ means that $q'_j > q_j$, $j = 1, \ldots, M$. If we then take matrices $P$ and $Q$ such that for some $k$ and $l$, $p^k = p'$, $q^k = q'$, $p^l = p$, and $q^l = q$, this, by $(*)$, contradicts the weak ranking restriction (3), so that we can conclude that $g(p', q) = g(p, q) \,\forall\, p',\ p, q \in \mathbb{R}^M$. This means that we could just as well disregard actual prices and fix a price vector $\bar{p} \in \mathbb{R}^M$ to evaluate $g$; $G_k(P, Q) = g(p^k, q^k) = g(\bar{p}, q^k)$, $k = 1, \ldots, K$. But then we have actually found a function that depends on quantities only; if we redefine $G$ as a function from $\mathbb{R}_+^{K\times M}$ to $\mathbb{R}^K$, with $G_k(Q) = g(\bar{p}, q^k)$, $k = 1, \ldots, K$, we find that

$$F_k(P, Q) \geq F_l(P, Q) \Leftrightarrow G_k(Q) \geq G_l(Q),$$

and thereby that $F$ violates dependence on prices (2). *Q.E.D.*

## 3. DISCUSSION OF THE SET OF PROPERTIES

In this section I will discuss and defend the choice of the set of properties. Where possible, I will compare them to alternative conditions, most of which come from the set of tests proposed by Diewert (1986, 1999). I hope to show that the set of four we use here is quite slim; in order to arrive at impossibility we can do without some other, or stronger but still reasonable, assumptions.

After discussing what more we could have asked of such a function, it will also be interesting to check if we could have been less demanding and still arrive at impossibility. I will argue that the answer is no by giving some counterexamples.

In this discussion, one should bear in mind that the most striking difference between the tests proposed by Diewert and mine is that his tests focus on cardinal characteristics of this function $F$ while here we limit ourselves to the ordinality of the outcomes.

First note that instead of weak continuity (1) we could also have suggested continuity of $F$, as Diewert does.

5. (*Continuity*): *$F$ is continuous*.

[6] Note that the mere existence of $F$ does not imply the existence of such a function $g$; a lexicographic relation $R$ can very well represent a function $F$, but it does not come with a function $g$, both in the sense described above.

[7] Indeed we can find an $\varepsilon$ such that, if $\|(p', q') - (p', q)\| < \varepsilon$, then $|g(p', q') - g(p', q)| < |g(p', q) - g(p, q)|$.

It is straightforward that continuity of $F$ implies weak continuity, but the converse is not true; take for instance $F_k = g(\langle \bar{p}, q^k \rangle)$, where $g : \mathbb{R} \to \mathbb{R}$ is any discontinuous nondecreasing function.

As a second condition we have demanded prices not to sit mum. We could reformulate that so as to require that there exist at least one example of prices being decisive if everything else is equal. This is not a kind of test that is common in index theory.[8] Nonetheless one could argue that it is rather modest to expect actual prices not to be completely superfluous information; by not including this property, one would allow for anyone's guess on how to satisfactorily compare bundles of goods regardless of prices, even if that guess would imply, loosely speaking, that a car is to be weighted just as heavily as a tire. We naturally turn to prices to obtain weighing factors and this property demands that prices should render at least some useful information. (Another argument for considering this condition to be quite moderate is that it is not violated by any of the methods that Balk (1996) or Hill (1997) apparently considered of interest.)

The weak ranking restriction (3) is a clear example of the advantages of the ordinal approach we have taken. To illustrate this, we can compare it to a test by, again, Diewert:

6. (*Proportional Quantities Test*): *If* $q^k = \beta_k \bar{q}$ *for* $k = 1, \ldots, K$ *and* $\sum_k \beta_k = 1$, *then* $F_k(P, Q) = \beta_k$.

This demand is both stronger and weaker than our third restriction. It is weaker in the sense that it only puts a restraint on cases in which quantity vectors are exact multiples of each other and it is stronger in the sense that, for those cases, it specifies the outcome exactly. But I do think the weak ranking restriction preferable in both aspects; the relative ranking of two quantity vectors, one of which has every entry greater than the corresponding entry of the other, seems obvious, whether or not the difference is exactly proportional or not. On the other hand I do think it is a gain in modesty if we refrain from wanting specific things concerning the magnitude of the outcome and restrict our attention to the sign of the difference between them. Another alternative to the weak ranking restriction could be the following.

7. (*Strong Ranking Restriction*): $q^k \geq q^l \Rightarrow F_k(P, Q) \geq F_l(P, Q)$ *and* $q^k > q^l \Rightarrow F_k(P, Q) > F_l(P, Q)$, *where* $q^k \geq q^l$ *means that* $q^k_j \geq q^l_j \ \forall j$ *and* $q^k > q^l$ *means that* $q^k \geq q^l$ *and not* $q^k = q^l$.

To see that the strong ranking restriction (7) really is stronger than the weak ranking restriction (3), we can for instance take a function $F$ with $K = 2$ and

$$F_1(P,Q) = \begin{cases} \langle \bar{p}, q^1 \rangle & \|q^1\| < 1, \\ 2\langle \bar{p}, q^1 \rangle & \|q^1\| \geq 1, \end{cases} \quad \text{and} \quad F_2(P,Q) = \begin{cases} \langle \bar{p}, q^2 \rangle & \|q^2\| \leq 1, \\ 2\langle \bar{p}, q^2 \rangle & \|q^2\| > 1. \end{cases}$$

This function does satisfy 3, but fails to meet 7. There is also no equivalence between the strong ranking restriction (7) and the weak ranking restriction (3) combined with weak

[8] The test in the set by Diewert that comes closest is the *Proportional Prices Test*, which demands that in case all relative prices in all countries are equal, that is $p^k = \alpha_k \bar{p}$, then $F_k$ should be proportional to $\langle \bar{p}, q^k \rangle$.

continuity (1). Take for example $M = K = 2$ and

$$F_k(P, Q) = \begin{cases} 1 & \text{if } Q = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}, \\ \langle \bar{p}, q^k \rangle & \text{elsewhere.} \end{cases}$$

This function is weakly continuous (1), it satisfies the weak ranking restriction (3), but it does not satisfy the strong ranking restriction (7). In the opposite direction, an example is the function $F$ with $M = K = 2$ and $F_k(P, Q) = q_1^k + q_2^k + \mathbf{1}_{\{q_1^k > 1\}} + \mathbf{1}_{\{q_2^k \geq 1\}}$. This function satisfies the strong ranking restriction (7) and therefore the weak ranking restriction (3), but it is not weakly continuous (1); take for instance $q_1^1 = q_2^2 = 1$ and $q_2^1 = q_1^2 = a$, $a \neq 1$. This indicates that we cannot replace 1 and 3 by 7 in the theorem.

Independence of irrelevant countries (4) is the name Neary (1999) suggested for what Drechsler (1973) called characteristicity. We use an ordinal and formal variant of this test, the concept behind which actually dates back to Fischer (1922). This being the only test that needs more than two countries to be meaningful, footnote 4 shows at what point the proof would break down in the case of bilateral comparison. If indeed $K = 2$, this property could obviously never be violated and the other three can be met by functions derived from standard index numbers; we can, for instance, take $F_1(P, Q) = (\langle p^1, q^1 \rangle \langle p^2, q^1 \rangle)^{\frac{1}{2}}$ and $F_2(P, Q) = (\langle p^1, q^2 \rangle \langle p^2, q^2 \rangle)^{\frac{1}{2}}$, which is equivalent to using the quantity index of Fisher.[9] The content of the theorem is that such an index number cannot be generalized consistently to a multilateral setting.

For more conditions, some of which are actually quite appealing, I refer to Diewert (1986, 1999). The last one I will mention here is an ordinal variant of the condition that no country should get a special treatment.

8. (*Symmetric Treatment of Countries*): $F_k(OP, OQ) \geq F_l(OP, OQ) \Leftrightarrow [O \circ F]_k(P, Q) \geq [O \circ F]_l(P, Q)$ *for $O$ a permutation matrix.*

Please do note that the first half of the proof of the theorem can also be used to show that 1, 3, and 4 together imply 8.

Finally we can check whether any relaxation of the constraints 1, 2, 3, and 4 leads to failure of the impossibility theorem. If we forget about independence of irrelevant countries (4), we find that the function $F$ with components $F_k = \langle p^1, q^k \rangle$ does satisfy the remaining demands. A more elegant and symmetric counterexample is the well known Geary-Khamis method (see Geary (1958) and Khamis (1972)).[10] Suppose we leave out the weak ranking restriction (3); a function that does satisfy the other three is the standard GDP approach or exchange rate system, in which $F_k(P, Q) = \lambda^k \cdot \langle p^k, q^k \rangle$ with conversion factors $\lambda^k, k = 1, \ldots, K$. If we neglect dependence on prices (2), an example of a function that does not violate the rest is one with components $F_k(P, Q) = \langle \bar{p}, q^k \rangle$ in which $\bar{p}$ is a fixed weight vector. Not regarding weak continuity (1) needs a more sophisticated

[9] Equivalents to both the quantity index of Laspeyres and the one of Paasche would do as examples, but these are not symmetric.

[10] Although the equally well-known EKS multilateral index can be thought of as minimizing deviations from independence of irrelevant countries (4), it cannot serve as an example here, because it also violates the weak ranking restriction (3).

counterexample; we can take a fixed weight vector $\bar{p}$ again and define $K$ sets iteratively as follows:

$$S_1 = \{x \in \mathbb{R} | \exists\, k \in \{1, \ldots, K\} \text{ such that } \langle \bar{p}, q^k \rangle = x\},$$

$$S_{i+1} = S_i \backslash \{x | x = \min S_i\} \qquad (i = 1, \ldots, K-1).$$

These sets are not necessarily all nonempty. Then define $F$ iteratively by

$$\forall\, k \text{ such that } \langle \bar{p}, q^k \rangle = \min S_1, \qquad F_k(P, Q) = \langle p^k, q^k \rangle,$$

$$\forall\, k \text{ such that } \langle \bar{p}, q^k \rangle = \min S_{i+1}, \quad F_k(P, Q) = \langle p^k, q^k \rangle + \max_{\{l | \langle \bar{p}, q^l \rangle = \min S_i\}} F_l(P, Q).$$

Counterexamples given in the first and third case satisfy not only the weak (3) but also the strong ranking restriction (7) and all counterexamples but the very first do satisfy symmetric treatment of countries (8).

## 4. CONCLUSION AND REMARKS

What we have established here is that any method of comparing wealth of nations must violate at least one of four very reasonable requirements. This is a bit of a bleak prospect for index theory as far as multilateral comparison of volumes and price levels is concerned.

As a last remark, I would like to mention that this theorem can also be interpreted in an intertemporal way. All four conditions remain meaningful if we substitute years for countries.

*Department of Econometrics and Tinbergen Institute, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands; veelen@tinbergen.nl*

*Manuscript received January, 2000; final revision received January, 2001.*

## REFERENCES

BALK, B. M. (1996): "A Comparison of Ten Methods for Multilateral International Price and Volume Comparison," *Journal of Official Statistics*, 12, 199–222.

DIEWERT, W. E. (1986): "Microeconomic Approaches to the Theory of International Comparisons," Technical Working Paper No. 53, NBER, Cambridge, MA.

——— (1999): "Axiomatic and Economic Approaches to International Comparisons," in *International and Interarea Comparisons of Income, Output and Prices*, ed. by A. Heston and R. E. Lipsey. Chicago: University of Chicago Press, 13–87.

DRECHSLER, L. (1973): "Weighing of Index Numbers in Multilateral International Comparisons," *Review of Income and Wealth*, 19, 17–34.

FISHER, I. (1922): *The Making of Index Numbers*. Boston: Houghton Mifflin.

GEARY, R. C. (1958): "A Note on the Comparison of Exchange Rates and Purchasing Power Between Countries," *Journal of the Royal Statistical Society, Series A*, 121, 97–99.

HILL, R. J. (1997): "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities," *Review of Income and Wealth*, 43, 49–69.

KHAMIS, S. H. (1972): "A New System of Index Numbers for National and International Purposes," *Journal of the Royal Statistical Society, Series A*, 135, 96–121.

NEARY, J. P. (1999): "True Multilateral Indexes for International Comparisons of Real Income: Theory and Empirics," Mimeo., University College Dublin.

SEN, A. K. (1982): "Real National Income," in *Choice, Welfare and Measurement.* Oxford: Basil Blackwell, 388–415.

VAN IJZEREN, J. (1956): "Three Methods of Comparing the Purchasing Power of Currencies," *Statistical Studies*, No. 7. Zeist: The Netherlands Central Bureau of Statistics/De Haan.