On the Reliability of Unit Value Ratios in International Comparisons

Timmer, Marcel P.

Publication date:
1996

On the Reliability of Unit Value Ratios in International Comparisons

Research Memorandum GD-31

Marcel Timmer

December 1996

On the Reliability of Unit Value Ratios in International Comparisons

Marcel Timmer*
Department of Technology and Development Studies
Eindhoven University of Technology

December 1996

Building DG 1.10
Department of Technology and Development Studies
Faculty of Technology Management
Eindhoven University of Technology
PO Box 513
5600 MB Eindhoven
The Netherlands

tel: +31-40-2473087
fax: +31-40-2449171
E-mail: M.P.Timmer@TM.TUE.NL

* I am indebted to Bart van Ark for extensive discussion on the subject and many stimulating comments. Also I am most grateful to Prasada Rao for his insightful remarks on an earlier draft. Thanks to Erik Monnikhof, Eddy Szirmai, Mark Huisman and Klaas Veenma for their help and advice. I also received useful suggestions of the participants of the CES Brown Bag Seminar of the US Bureau of the Census, March 14, 1996, Washington D.C. USA.

Abstract

In this paper reliability indicators are developed for the Laspeyres and Paasche index type currency conversion factors used in the industry of origin approach to international comparisons of output. Given the fact that these conversion factors are calculated on basis of a sampling technique, sampling theory is used to develop formulae for standard errors. A comparison with the stochastic approach to index numbers is made. Empirical evidence for five bilateral comparisons in manufacturing shows that reliability differs as both sample coverage and the degree of price variability play a role. Ninety percent confidence intervals for the total manufacturing Laspeyres and Paasche unit value ratios ranged from ± 4% to ± 11% for the comparisons involving high productivity countries only, and up to ± 27% in a comparison of a low and a high productivity country. At branch level the differences are even more pronounced but no particularly (un)reliable branches could be discerned.

Another issue addressed in this paper is whether products should be grouped together before estimation of the indices. Stratified sampling theory suggests that if the grouping is carried out properly, more precise estimates can be obtained. Guidelines for this intermediate grouping are given. Empirical evidence suggests that the grouping of products into four digit ISIC industries improves the estimation of the aggregate indices.

Section 1. Introduction

In international comparisons of output and productivity use is often made of the Laspeyres and Paasche type of indices to calculate currency conversion factors. As these indices are estimates based on prices or unit values of a sample of products, it would be desirable to have variance estimates attached to them in order to assess their reliability. The main purpose of this paper is to develop a comprehensive measure of reliability, based on standard errors. Two ways will be explored to do this. One approach is the *stochastic approach* to index number theory which has recently received renewed attention in a book by Selvanathan and Prasada Rao (1994). Product prices are modelled as signals from which to extract the price relative for a group. Another approach which will be developed in this paper is based on *stratified sampling theory*. It makes use of the possibility to group products into homogeneous categories. Indices based on samples taken from these categories can be estimated more precisely if the grouping is done properly. This approach is particularly suitable for index numbers which are built up in a stagewise weighting procedure.

The first part of the paper gives a new interpretation of the Laspeyres and Paasche type of indices used in the industry of origin method from a *stratified sampling perspective*. The industry of origin approach as practised and refined in the ICOP (International Comparisons of Output and Productivity) project is taken as a starting point (sections 2 and 3). The interpretation reinforces the standard ICOP method, especially the reweighting procedures involved. It also gives some suggestions for improvement. Section 4 gives an alternative description of relative price indices from a *stochastic perspective.*

The second part of the paper is devoted to the development of a reliability measure of the indices. Reliability measures can be used for several purposes: 1. identification of product groups for which indices are (relatively) unreliable, and where returns to a further search for currency conversion factors will be highest, 2. the detection of 'outliers', and 3. the establishment of confidence limits. The latter can be used in testing a host of hypotheses such as whether output in country X is higher than in country U, etc.

Unreliability originates from different causes, of which the quality problem and incomplete coverage (sampling) of the products are the most important. These factors create both biases and variance in the estimation of the indices. This is discussed in section 5. The measurement of the sample variance is the main topic of section 6. Using both the stochastic and the stratified sampling approach, formulae for the variances of indices are developed and differences and similarities are discussed. Finally, in section 7 the reliability measure, defined as the coefficient of variation and based on sampling variances, is applied to five bilateral manufacturing output comparisons.

2. The ICOP industry of origin Approach

Currency Conversion Factors for International Comparisons of Output

Studies aimed at comparing output and productivity levels across countries require a conversion factor to express output values in a common currency. The most obvious candidate for this is the exchange rate. However, there are a number of strong objections against the use of exchange rates. Firstly, an exchange rate reflects only the comparative price levels of *tradable* goods and services in an economy. Secondly, exchange rates are subject to other forces than price relatives of goods and services only. Especially in recent decades, these rates have been volatile because of capital movements, and speculation on currency markets. Also some governments try to maintain an under- or overvalued exchange rate because of political pressures. Thirdly, an exchange rate is an average for all tradables in an economy taken together. Studies aimed at comparing real output by industry however require industry-specific conversion factors.

Since the late 1960s alternative conversion factors became available on a large scale through the work of the International Comparisons Project (ICP)[1], which provided Purchasing Power Parities (PPPs) using the expenditure approach. ICP concentrates on comparisons of national accounts categories such as private consumption, government consumption and capital formation. PPPs are derived at a detailed item level by gathering a list of consumer prices of a sample of finely specified products for each country . Multilateral PPPs are derived from these item prices, which are subsequently aggregated into higher level PPPs. Expenditure PPPs are now made available on a regular basis by the UN, EUROSTAT and the OECD.

However, expenditure PPPs are less useful for international comparisons by industry of origin as they only apply to final output. The output of intermediate products, which in manufacturing accounts for at least one third in value, is not covered at all. Further drawbacks are that expenditure PPPs include marketing margins, and indirect taxes and subsidies. Also they include import prices, while excluding export prices. Attempts have been made to apply the expenditure PPPs in the industry of origin approach (the socalled proxy PPPs) by adjusting these PPPs to a domestic output factor price basis, and allocating expenditure PPPs to specific industries. However, only rough adjustments could be made[2]. The coverage of intermediate goods sectors remains problematic.

Alternatively for comparisons by sector and industry, the industry of origin approach can be applied. This approach computes industry specific conversion factors by using output data at producer level instead of final consumption data[3]. Ideally, these industry data should be based on specific product prices. However, output prices are not available on a large international comparable scale. As an

[1] See for example Kravis, Summers and Heston (1982).
[2] See for example Hooper and Vrankovich (1995).
[3] For a more detailed overview of these issues, see van Ark (1996).

alternative unit value ratios (UVRs) are used. Unit values are computed by dividing the value of output by produced quantity which are derived from producer census data. Subsequently the ratios of the products' unit values are taken cross-country. These UVRs are used to convert output by industry into a common currency. In this way a comparison can be made of real output produced in both countries. Subsequently labour or total factor productivity levels can be compared.

Outline of the ICOP industry of origin approach.

The industry of origin approach has been applied and refined in the ICOP research work (International Comparisons of Output and Productivity) at the University of Groningen since 1983.[4] In this method unit value ratios (UVRs) are computed on the basis of Laspeyres and Paasche index formulae. First, product UVRs are computed based on (bilateral) matching of broadly defined products with similar characteristics, for example shoes, cigarettes, cheese and rubber inner tubes. On the basis of output value and output quantity as given in the census, product unit values are calculated. Unit values (uv) are computed by dividing produced quantity (q) into produced output value (o), according to the following formula for product i:

$$uv_i = \frac{o_i}{q_i} \tag{2.1}$$

The unit value can be considered as an average price, averaged throughout the year for all producers and across a group of nearly similar products. The unit values for these matched products are used to derive the unit value ratios (UVRs):

$$UVR_i^{XU} = \frac{uv_i^X}{uv_i^U} \tag{2.2}$$

with X and U the countries being compared, U being the base country (in most cases the USA). UVRs indicate the relative producer price of the matched goods in the two countries.

Product UVRs are aggregated in a stagewise procedure to higher levels: industry, branch and finally to total manufacturing level. An industry is defined here as the lowest level at which economic activities can be compared between countries, that is where output, value added and labour input data are available for both countries. Examples are the dairy industry, men's wearing apparel, agricultural fertilizers or rubber tyres and tubes. These are mostly 4-digit industry groups in the International Standard Industrial Classification (ISIC). Branches correspond to 2-digit divisions or a group of 3-digit major industry groups. Examples of branches are food manufacturing, textile manufacturing and

---

[4] Until now the manufacturing sector, on which this paper also focusses, has been covered for some 30 countries. On a smaller scale, the agricultural and service sectors are covered as well. See Maddison and van Ark (1994) for an overview of the ICOP-research work.

basic and fabricated metals manufacturing. The reweighting procedure is performed for two reasons: 1. to derive industry and branch output conversion factors which are interesting in themselves, and 2. to ensure that original product UVRs are reweighted according to their relative importance in the aggregate.

Aggregation Step 1: Industry UVRs

The computation of industry UVRs is based upon two alternative price indexes: the Laspeyres, using the quantity weights of the the base-country ($UVR^{XU(U)}$) and the Paasche, using the quantity weights of the other country ($UVR^{XU(X)}$). They are expressed below, respectively, for an industry j. As not all products in an industry can be matched it is assumed that the UVR based on the matched products ($1,..., I_j(M)$) is representative for the UVR based on all products ($1,..,I_j$):

$$UVR_j^{XU(U)} \equiv \frac{\sum_{i=1}^{I_j} uv_{ij}^X q_{ij}^U}{\sum_{i=1}^{I_j} uv_{ij}^U q_{ij}^U} = \frac{\sum_{i=1}^{I_j(M)} uv_{ij}^X q_{ij}^U}{\sum_{i=1}^{I_j(M)} uv_{ij}^U q_{ij}^U} \qquad (2.3)$$

at quantity weights of base country U, and:

$$UVR_j^{XU(X)} \equiv \frac{\sum_{i=1}^{I_j} uv_{ij}^X q_{ij}^X}{\sum_{i=1}^{I_j} uv_{ij}^U q_{ij}^X} = \frac{\sum_{i=1}^{I_j(M)} uv_{ij}^X q_{ij}^X}{\sum_{i=1}^{I_j(M)} uv_{ij}^U q_{ij}^X} \qquad (2.4)$$

at quantity weights of country X.

However, in ICOP the assumption of representativeness is not always held true. In case the coverage percentage of the matched products in terms of total output value within the industry is lower than 25%, the assumption is not deemed justified (the socalled 25%-rule of thumb[5]). To use the UVRs of these products nevertheless they are grouped in an industry 'Others' within each branch. The UVR of this industry is based on *all product matches made within branch k*. The 'Others' industry is subsequently viewed as any other industry. This adhoc treatment will be discussed later on.

Aggregation Step 2: Branch Level UVRs

The following step is to derive branch level UVRs. These are obtained through a weighted averaging of the UVRs of industries belonging to a particular branch, using the industries' shares in the branch gross value added (GVA) as weights. With this reweighting procedure one assures that industries which are important in value will get a greater weight in the branch UVR, irrespective of their

[5] See van Ark (1993, p.28).

percentage of matched output (the coverage ratio). Let J_k be the number of industries in branch k ($j=1,..,J_k$). Then the UVR for branch k is given by:

$$UVR_k^{XU(U)} = \frac{\sum_{j=1}^{J_k} GVA_j^{U(U)} \times UVR_j^{XU(U)}}{\sum_{j=1}^{J_k} GVA_j^{U(U)}} \tag{2.5}$$

at value added weights of base country U, and:

$$UVR_k^{XU(X)} = \frac{\sum_{j=1}^{J_k} GVA_j^{X(X)}}{\sum_{j=1}^{J_k} \frac{GVA_j^{X(X)}}{UVR_j^{XU(X)}}} \tag{2.6}$$

at valued added weights of country X. If no matches are made in a branch, the total manufacturing UVR is thought to be representative.

Aggregation Step 3: Manufacturing UVR

The manufacturing sector UVR (UVR_{manu}) is derived by aggregating branch UVRs in the same way as the aggregation from industry to branch level. Let K be the number of branches in the manufacturing sector ($k= 1,..,K$), then

$$UVR_{manu}^{XU(X)} = \frac{\sum_{k=1}^{K} GVA_k^{X(X)}}{\sum_{k=1}^{K} \frac{GVA_k^{X(X)}}{UVR_k^{XU(X)}}} \tag{2.7}$$

at value added weights of country X, and:

$$UVR_{manu}^{XU(U)} = \frac{\sum_{k=1}^{K} GVA_k^{U(U)} \times UVR_k^{XU(U)}}{\sum_{k=1}^{K} GVA_k^{U(U)}} \tag{2.8}$$

at valued added weights of base country U.

The Laspeyres and Paasche indices are combined into a Fisher index when a single currency conversion factor is required. It is defined as the geometric average of the Laspeyres and the Paasche.

3. A Stratified Sampling Interpretation of the ICOP Industry of Origin Approach

The stagewise weighting procedure method used by ICOP as described above can be interpretated as a *(multi staged) stratified sampling method.* In the stratified sampling approach the heterogeneous population (all products produced in the manufacturing sector) is divided into more homogeneous subpopulations, called *strata.* Strata are defined nonoverlapping and together comprise the whole of the population. In this case the heterogeneous total manufacturing sector is subdivided into branches, which are subsequently subdivided into homogeneous industries. This is illustrated by Figure 1.

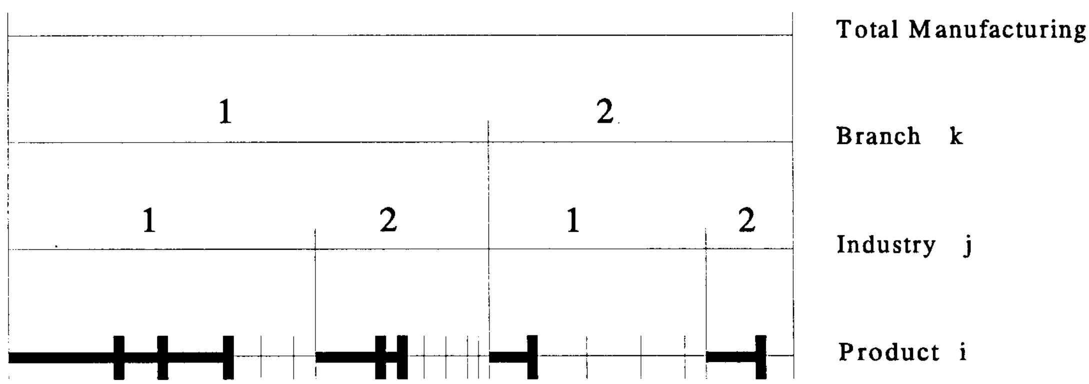

Figure 1 Simplified representation of the four levels of aggregation for which UVRs are being computed within ICOP.

Figure 1 shows the four levels which are being distinguished within ICOP. The four level lines in the figure can be thought of as representing manufacturing output value. The total manufacturing output is the sum of branch output, which is the sum of industries' output value. The output value of an industry is the sum of the value of output of its products. In the binary comparsion some of these products can be matched, but not all. This is because of lack of value or volume data (mainly for reasons of confidentiality), difficulties in finding good corresponding products, the existence of country-unique products etc. Bold lines at the product level in the figure indicate the total output value of the matched products in an industry. Thus matched products in an industry can be seen as a sampled subset of all products in the industry.

Three issues must be discussed in this stratified sampling context:
1. The definition of the population.
2. The homogeneity of the strata (industries and branches). Only in the case that strata are homogeneous, there is a possible gain in the precision of the estimation of the population parameters (manufacturing UVR) in comparison with estimation without stratification.
3. The randomness of matched products.

Ad 1. There are two alternatives for defining the population. The first most obvious alternative is to define all products in the manufacturing sector as the population. Out of this population some products are sampled, that is being matched, and each product has an equal weight. However, in that case the idea is lost that for the determination of the total manufacturing UVR, product UVRs should have different weights. Their output values, and therefore their importance in the total manufacturing output, differ and this should be reflected in their weights. The manufacturing UVR is not derived as a price 'signal' on its own, but with the explicit aim to serve as a currency conversion factor for total manufacturing *output*. Therefore I propose to define the population as the total output value of the manufacturing sector, or more precisely, as the set of output value units, say $. Of this output value a certain part is covered by sampled products: the sampling fraction.

Ad 2. The issue whether product UVRs are more similar *within* industries than *across* industries depends ultimately on the classification of products within an industry. There are two alternative general approaches for economic classifications: the supply-side approach and the demand-side approach[6]. In the supply-side approach products are classified according to the similarities in the production processes that are used to make them. A demand-side classification concept on the other hand, yields a classification system based on the use of the products. International differences in prices are of course due to both supply and demand forces. But given the fact that the hypothesis of identical consumer preferences across the world cannot be rejected (see e.g. Kravis, Summers and Heston (1982)), one might conclude that differences in product UVRs are mainly caused by differences in the (efficiency of) production processes and differences in input prices, which are probable more similar in industries when grouped according to the supply-side approach.

As resources do not allow individual researchers to reclassify national product and industry data, one is bound by the industrial classifications used by the national agencies. These national classifications are all variations from the International Standard Industrial Classification (ISIC). The observation of Triplett (1990) that a substantial portion of the present US System of Industrial Classification (SIC) is already grouped more or less by the supply-side approach is therefore strengthening. Indeed, the empirical evidence presented in section 7 of this paper will show that the UVRs within industries show less variance than UVRs across industries, and that therefore stratification is useful.

Ad 3. To apply sampling theory the assumption of randomness is of importance. However, products which are matched are typically *non-random* for the following reasons:

1. Matched products are products that are produced in both countries. Unique goods will never be matched.
2. There might be a selection bias in the census data as a country will put less effort in collecting or publishing data on unimportant products.
3. Especially in the case of small countries, product data are sometimes suppressed because of confidentiality reasons, so that data of products produced by a small number of firms is excluded.

[6] See Triplett (1990) for an overview of this issue.

In comparisons involving a low and a high productivity country like for example China and the USA, the sample of matched products tends to be biased towards low quality, homogeneous products because of these reasons. China will not produce most of the specialized, high quality goods produced in the USA, or only in quantities not noted by the census. Assuming that the USA has a bigger advantage in producing high quality goods, UVRs (in Yuan per US$) for these not matched goods would be high. Consequently, the industry UVR based on the matched goods is downwardly biased. Ideally, this bias should be taken into account, but as to now little information is available to do this. The assumption of randomness finds its counterpart in the ICOP approach assumption of representativity of the UVR of the sampled part of an industry for the non-sampled part, as will be shown below. Each aggregation step in the ICOP method as described in section 2 will now be reexamined from a stratified sampling perspective.

Aggregation Step 1 Industry UVRs

According to stratified sampling theory[7] the best estimate for the industry UVR (UVR_j) is given by the mean of the sampled products (assuming randomness). Because of the chosen definition of population as value, product UVRs are weigthed by their value:

$$UVR_j = \frac{\sum_{i=1}^{I_j(M)} O_{i,j} \times UVR_{i,j}}{\sum_{i=1}^{I_j(M)} O_{i,j}} \tag{3.1}$$

With i=1,.., $I_j(M)$ denoting the matched products in industry j and $O_{i,j}$ the value of output of product i. In bilateral comparisons the weights of either country can be used. It can be easily shown that the use of base country value weights leads to the Laspeyres index as used in ICOP. Substituting base country U weights in (3.1) gives:

$$UVR_j^{XU(U)} = \frac{\sum_{i=1}^{I_j(M)} O_{i,j}^{U} \times UVR_{i,j}^{XU}}{\sum_{i=1}^{I_j(M)} O_{i,j}^{U}} \tag{3.2}$$

Substituting the identity: $O^U = uv^U \times q^U$,gives

$$UVR_j^{XU(U)} = \frac{\sum_{i=1}^{I_j(M)} UVR_{ij}^{XU} \times uv_{ij}^{U} \times q_{ij}^{U}}{\sum_{i=1}^{I_j(M)} uv_{ij}^{U} \times q_{ij}^{U}} \tag{3.3}$$

[7] See e.g. Cochran (1977), Chapter 5.

Finally substituting uv_i^X / uv_i^U for UVR_{ij}^{UX} gives the Laspeyres index used in ICOP to calculate the industry UVR (see formula 2.3).

$$UVR_j^{XU(U)} = \frac{\sum_{i=1}^{I_j(M)} uv_{ij}^X q_{ij}^U}{\sum_{i=1}^{I_j(M)} uv_{ij}^U q_{ij}^U} \qquad (3.4)$$

The derivation of the Paasche index is less straightforward. Using the other country value weights would give

$$UVR_j^{XU(X)} = \frac{\sum_{i=1}^{I_j(M)} O_{i,j}^X \times UVR_{i,j}^{XU}}{\sum_{i=1}^{I_j(M)} O_{i,j}^X} \qquad (3.5)$$

which is not a Paasche index. Instead we need to use in the stratified sampling formula the other country quantities valued at base country prices ($O^{XU} = q^X \times uv^U$) as weights. This ensures also that the population sampled is the same for both the Laspeyres and the Paasche (both populations are then defined in \$).

$$UVR_j^{XU(X)} = \frac{\sum_{i=1}^{I_j(M)} UVR_{ij}^{XU} \times uv_{ij}^U \times q_{ij}^X}{\sum_{i=1}^{I_j(M)} uv_{ij}^U \times q_{ij}^X} \qquad (3.6)$$

which gives the Paasche (formula 2.4)

$$UVR_j^{XU(X)} = \frac{\sum_{i=1}^{I_j(M)} uv_{ij}^X q_{ij}^X}{\sum_{i=1}^{I_j(M)} uv_{ij}^U q_{ij}^X} \qquad (3.7)$$

This shows that the Laspeyres and Paasche indices used in ICOP to calculate industry UVRs can be interpretated as stratified sampling formulas using resp. base country output values and other countries quantity valued at base country prices as weights under the assumption of random sampling.

Aggregation Step 2 Branch Level UVRs

In the stratified sampling interpretation the heterogeneous branches consist of a set of more homogeneous industries each with their own characteristics, i.e. own mean and own variance. The theory of stratified sampling suggests that if in every stratum (industry j in branch k) the sample estimate of the mean is unbiased, then the stratum weighted mean of all industries' UVRs ($UVR_{j,k}$) is an unbiased estimate of the branch mean (UVR_k).

$$UVR_k^{XU} = \frac{\sum_{j=1}^{J_k} O_{j,k} \times UVR_{j,k}^{XU}}{\sum_{j=1}^{J_k} O_{j,k}} \qquad (3.8)$$

with $O_{j,k}$ the output in industry j in branch k, as strata are defined over output (j= 1,..,J_k). Using output weights from the base country and the industry UVRs at base country weights, gives as a result the Laspeyres index used in ICOP to calculate branch UVRs (compare formula 2.5).

$$UVR_k^{XU(U)} = \frac{\sum_{j=1}^{J_k} O_{j,k}^{U(U)} \times UVR_{j,k}^{XU(U)}}{\sum_{j=1}^{J_k} O_{j,k}^{U(U)}} \qquad (3.9)$$

To arrive at the Paasche index the output of country X valued at base prices should be substituted. This gives:

$$UVR_k^{XU(X)} = \frac{\sum_{j=1}^{J_k} O_{j,k}^{X(U)} \times UVR_{j,k}^{XU(U)}}{\sum_{j=1}^{J_k} O_{j,k}^{X(U)}} \qquad (3.10)$$

which can be rewritten as:

$$UVR_k^{XU(X)} = \frac{\sum_{j=1}^{J_k} O_{j,k}^{X(X)}}{\sum_{j=1}^{J_k} \frac{O_{j,k}^{X(X)}}{UVR_{j,k}^{XU(X)}}} \qquad (3.11)$$

which is the same as the ICOP formula for the Paasche branch UVR (2.6). Notice that in ICOP use is made of value added instead of output weights. This cannot be justified from a stratified sampling perspective.

The Problem of Defining Industries

It has to be noted that from a stratified sampling perspective, industries are defined as an intermediate level between product and branch to improve the estimation of branch UVRs. Therefore, the only criterium for defining an industry is whether the products in an industry are homogeneous or not. In the traditional ICOP approach as described in section 2, industries for which output is covered by less than 25% are excluded from the reweigthing procedure as the UVRs of these industries are considered as not representative. However, this is an adhoc procedure without theoretical underpinning. Considerations about representativity should not be based on coverage ratios alone, but also on the degree of group homogeneity (see section 6). Moreover, if one wishes to limit the possibility of an 'outlier' product UVR getting a high industry weight, one should formulate the exclusion rule in terms of number of product matches made, and not in output covered[8].

Stratified sampling theory suggests that industries should be formed and reweighted *irrespective of their coverage ratio.* Even in the case when only one small product match can be made, it should be included in an industry, that is, grouped with products which have more or less the same UVR, as this will always improve the estimate of the branch mean. The part of a branch not covered by the defined industries should be treated as a non-sampled stratum. This stratum cannot be included in the computation of mean and variances. J_k (in formulae 3.8-11) is then the number of industries for which at least one product match has been made.

Aggregation Step 3 Total Manufacturing UVRs

The total manufacturing sector is divided into branches. Similar reasoning as used for branches applies to the aggregation from the branch to the total manufacturing level UVR. Base country output weights are used to arrive at the Laspeyres index (3.9), and the other country quantity weighted at base prices to arrive at the Paasche (3.11).

Conclusions

The ICOP industry of origin approach can be fruitfully described as a stratified sampling approach. On the basis of this theoretical interpretation the following suggestions can be made:

- Industries should be ideally defined on a priori expectations about the degree of homogeneity of the products included in the industry. In practice, the industries as classified by the SIC will serve this purpose to a considerable extent.
- Each product match should be included in only one industry. Thus one avoids the ad hoc treatment of product matches made outside industries (the 25%-rule). This makes the method both simpler and theoretically sounder.
- Industry UVRs should be weighted with their value of output instead of their value added as *value of output* is stratified. This is also true for the reweighting of branches into total manufacturing. An additional practical advantage of weighting with output is that in defining industries one is not bound by the unavailability of value added figures at a detailed level.
- The assumption of random sampling of products in an industry is problematic. In the case of comparisons involving low and high productivity countries, we suspect that there is a downward bias in the UVR estimates, which magnitude is unknown. Probably, it will be small in industries producing

[8] Better still, in case of serious doubt about a product UVR, is to include it in a group consisting of this product alone. Thus it will have only a small weight.

basic homogeneous goods like the cement, paper, or basic metals industries. These industries cover a major part of manufacturing output. But especially in sophisticated industries like machinery and electronic equipment, the bias caused by non-random sampling is likely to be severe. For these industries the standard ICOP approach should be supplemented or replaced by an approach more capable of indicating the UVRs for sophisticated products.

4. A Stochastic Interpretation of the ICOP Industry of Origin Approach[9]

The stratified sampling interpretation makes it possible to compute variances of the Laspeyres and Paasche index formulae. This will be shown in the next section. In this section, attention will be turned to the *stochastic approach* to index numbers, which enables the computation of variances as well.

In the stochastic approach each product UVR is looked at as a "signal" for the true underlying group (industry) UVR, explicitly accepting that each product UVR is contaminated by some random error. Therefore the industry UVR can be computed by means of simple OLS regression analysis, specifying the model as:

$$UVR_{i,j} = UVR_j + e_{i,j} \tag{4.1}$$

with $UVR_{i,j}$ the UVR of product i in industry j, UVR_j the UVR in industry j and $e_{i,j}$ a random error term symmetrically distributed around zero, whose variability depends on the product in question. In the simplest case one assumes uncorrelated error terms with a common variance and mean 0 (homoscedasticity). Then the best linear unbiased estimator (BLUE) of the industry UVR is the unweighted average of the product UVRs:

$$U\hat{V}R_j = \frac{1}{I_j(m)} \sum_{i=1}^{I_j(m)} UVR_{i,j} \tag{4.2}$$

with I_j (m) the number of product matches made in industry j.

However the assumption of homoscedastic error terms is doubtful in the case at hand. The stochastic approach allows for taking into account more reasonable error structures. One commonly made assumption is that errors will probably be greater for unimportant products (in terms of output value) than for more important products.[10]

---

[9] This section is inspired by Selvanathan and Prasada Rao (1994).

[10] See Selvanathan and Prasada Rao (1994), or Koszerek (1985a,b) in the context of trade indices.

Including this additional information, the regression equation (4.1) is transformed back into a homoscedastic form by multiplying both sides with the square root of the share in output of product i in total output ($w_{i,j} = O_{i,j} / O_j$). First we take the weights from the base country U:

$$\sqrt{w_{i,j}^{U}} \times UVR_{i,j}^{XU} = \sqrt{w_{i,j}^{U}} \times UVR_{j}^{XU(U)} + e_{i,j}^{*} \qquad (4.3)$$

Now the BLUE of UVR_j is given by Generalized Least Square (GLS):

$$UVR_{j}^{XU(U)} = \sum_{i=1}^{I_j(m)} w_{i,j}^{U} \times UVR_{i,j}^{XU} \qquad (4.4)$$

This is equal to the Laspeyres index used in ICOP as the industry UVR is defined as a base country output weighted mean of the product UVRs.

To arrive at the Paasche index we have to chose as weights the product shares of output in country X, valued at *base* country prices: $w_{i,j}^{XU} = O_{i,j}^{XU} / O_j^{XU}$

$$\sqrt{w_{i,j}^{X(U)}} \times UVR_{i,j}^{XU} = \sqrt{w_{i,j}^{X(U)}} \times UVR_{j}^{XU(X)} + e_{i,j}^{*} \qquad (4.5)$$

Now the BLUE of the industry UVR at country x weights is given by:

$$UVR_{j}^{XU(X)} = \sum_{i=1}^{I_j(m)} w_{i,j}^{X(U)} \times UVR_{i,j}^{XU} \qquad (4.6)$$

This is equal to the Paasche index used in ICOP.

Thus it is shown that under the assumption of diminishing variability of the product UVRs with increasing product importance in output, the ICOP approach and the stochastic approach deliver the same index formulae for the estimation of the industry UVRs.

However, it remains to be seen whether in practice the assumption of this specific type of heteroscedasticity is warranted. Until now the assumption of variance being inversely proportional to output has not found empirical validation. A more plausible error structure assumption may be based on the Gerschenkron effect. Van Ark, Monnikhof and Timmer (1996) have shown that the Gerschenkron effect is at work even at the detailed product level, by regressing relative prices and quantities produced for 26 bilateral comparisons. Products that are produced in relatively large (small) quantities in country X have relatively low (high) UVRs, i.e. relative to other products. This means that the error is correlated with relative quantities produced (q^X / q^U) and that therefore both sides of the regression model (4.1) should by multiplied by $\sqrt{(q^X / q^U)}$ to develop the GLS-estimator of the industry UVR. This gives rise to a new kind of index number, hitherto unknown. Its development is outside the scope of this paper and will not be pursued further.

5. Sources of Unreliability of Unit Value Ratios

Before developing reliability indicators for the Laspeyres and Paasche indices, first the determinants of unreliability will be discussed. Unreliability of unit value ratios (UVRs) originate from three main sources:

1. Product value and quantity basic data errors,
2. Quality problems, and
3. Incomplete coverage of products in the comparison.

The first two sources influence the reliability of the product level UVRs, the last one is responsible for the unreliability introduced when product UVRs are aggregated into higher level UVRs.

Ad 1. Product UVRs are based on information for output value and quantity produced in a benchmark year as gathered from the national censuses or surveys of manufacturing. Census data are collected directly from the enterprises through a huge data collection effort of the national statistical agencies. Errors can arise at every stage in collecting and reporting the data, but little can be said about their magniture or direction. They depend on the methods and efforts of the statistical agencies. A special type of error can be caused by the incomplete coverage of the enterprises, as studied in a temporal context by Allen (1975). For example, estimates of product prices used in price indices are often based on price data taken from a *sample* of enterprises or outlets. Allen develops formulae for sample variances of these price indices assuming a stratified sampling of enterprises. However, estimates of UVRs are based on census data which typically cover (nearly) all output of a product. Therefore this kind of variance of product UVRs will be small and not studied further here.

Ad 2. Between two countries products are being matched on the basis of the product descriptions given in the censuses. These are broad specifications of the products, not including information on brand labels, package size, quality etc. Therefore it is typically a group of products that is being matched. For example "Cigarettes" encompass different brands and sizes, and "Men's leather shoes" include boots, dress and casual shoes, workshoes etc., all of different sizes and qualities.
Two sorts of quality problems arise[11]. First, the matching is based on a (often short) description of the products given in the census. Census descriptions are sufficient to identify comparable *common* goods, i.e. "goods which are used widely in both countries and serve the same purpose" (Gilbert and Kravis, 1954, p.75). But one cannot always be sure whether they are truely *identical*, i.e. "have the same specifications and characteristics" (ibid). One might call this the *'product content'* problem, or the *real* quality problem. An example is computers, which characteristics can differ greatly between countries. This problem may be generally severe for high technology products like aircrafts, machinery, precision instruments etc., but less so for basic and intermediate products like food, paper, cement, steel etc. which constitute the major bulk of products in the manufacturing sector.
Besides the product content problem there is also a *product mix problem* which is caused by the grouping of (census level) products before matching. This grouping is necessary because individual

[11] See also Gilbert and Kravis (1954) and van Ark (1993, p.34-36).

products could not be matched, e.g. because one of the countries gives only data for the group as a whole. This type of quality problem does not arise because the individual products are different in terms of content, but because of different weights of individual products in the group which is 'matched'. For example the unit value ratio of "Men's leather shoes" will be biased unless the weights of more detailed items like boots, dress and casual shoes, workshoes etc. are the same in both countries, or when the unit values of these more detailed items are equal in each of the two countries. An indicator of the severity of this latter type of quality problem can be derived from the ratio of the total output matched per matched product. Suppose that in a comparison 20% of the output of a country is matched with 100 matches, and in another comparison with 400 matches, one has an indication that the product mix problem is less severe in the second case, as more detailed matches were made. Thus, an indicator for the seriousness of this problem is the percentage branch output covered per product match. The higher this percentage, the more problematic the quality problem will be as it indicates that broader product groups have been matched. Table 1 gives the indicator for the five comparisons studied in this paper.

Table 1 Coverage per Product Match for Five Manufacturing Comparisons (in % covered).

	China/USA 1985		France/USA 1987		Japan/USA 1987		UK/USA 1987		Germany /USA 1992	
	China (% per match)	USA (% per match)	France (% per match)	USA (% per match)	Japan (% per match)	USA (% per match)	UK (% per match)	USA (% per match)	Germ (% per match)	USA (% per match)
Food Products	3.81	2.93	(a)	(a)	0.86	0.70	0.67	0.63	0.63	0.69
Beverages	13.04	27.49	(a)	(a)	10.90	10.07	22.15	14.30	5.00	5.92
Tobacco	3.54	5.07	(a)	(a)	86.00	80.70	31.67	28.86	15.50	11.00
Textiles	11.70	7.68	2.37	4.29	1.85	2.78	1.73	3.05	1.45	2.68
Wearing Apparel	(a)	(a)	2.74	2.04	2.36	3.38	1.51	1.34	1.38	1.85
Leather Products and Footwear	45.18	41.20	51.46	33.97	8.53	7.33	4.86	7.17	4.78	6.89
Wood Products and Furniture	(a)	(a)	3.52	0.96	9.75	3.95	1.97	1.11	2.23	1.54
Paper and Printing	17.41	2.50	(a)	(a)	1.46	1.67	1.35	1.56	1.38	1.69
Chemicals	3.52	2.66	0.75	0.73	0.50	0.46	0.48	0.58	0.34	0.34
Oil Products	(a)	(a)	(a)	(a)	10.82	12.77	8.80	14.46	3.17	14.00
Rubber and Plastic Products	8.85	0.66	8.97	7.10	1.23	1.90	0.78	0.87	0.83	1.33
Stone, Clay and Glass Products	20.07	2.34	3.37	3.19	3.67	3.09	2.14	1.16	1.33	0.94
Basic and Fabricated Metal Products	26.95	6.87	1.89	1.09	0.73	0.67	2.55	1.55	0.72	0.62
Machinery and Transport Equipment	2.91	4.88	1.35	0.87	0.78	0.68	1.22	1.49	0.13	0.13
Electrical Equipment	5.28	1.14	(a)	(a)	0.58	0.56	0.56	0.52	0.38	0.79
Other Manufacturing	(a)	(a)	(a)	(a)	(a)	(a)	(a)	(a)	0.41	0.12
TOTAL MANUFACTURING	0.75	0.32	0.29	0.19	0.10	0.11	0.10	0.11	0.05	0.07

Notes: (a) No product match made.

Sources: Germany/USA 1992 from *ICOP/LCRA estimates (1996)*. Underlying data can be obtained from Groningen Growth and Development Centre (E-mail: ggdc@eco.rug.nl.). China/USA from *Szirmai and Ruoen (1994)*, France/USA from *van Ark and Kouwenhoven (1994)*, UK/USA from *van Ark (1993)* and Japan/USA from *Pilat (1994)*.

Looking at total manufacturing in the last row, it appears that there is a much bigger quality problem in the China/USA comparison than in the other comparisons, whereas the problem is relatively smallest for the West-Germany/USA comparison. This is also true at branch level.

It should be noticed that the difference between the two types of quality problems is a matter of degree, depending on the level of classification. Product mix problems turn up at the most detailed level of census data if these are constructed from still finer classified data[12]. Also, if one could break down a census product into components each having different characteristics (by additional non-census information), a product mix problem is turned into a product content problem if these more detailed products could be matched. It is useful to make the distinction here as each problem requires a different solution. Product content problems require the use of methods like hedonic pricing, whereas product mix problems call for finer detailed product data.

For comparisons involving a low and a high productivity country quality problems may be severe. This is because low productivity countries usually have less detailed census reports which leads to product mix problems as only groups of products can be matched. But there is also a product quality problem as products which are matched differ in their quality, e.g. televisions, cars, tractors, telephones, lamps etc. Assuming lower quality products being produced in the low productivity countries, UVRs (stated in local currency per US$) will be underestimated, and subsequently output in these countries will be overestimated.

In comparisons involving two high (or two low) productivity countries the errors caused by quality problems appear to be non-systematic. In Gersbach and van Ark (1994) an assessment is made of the magnitude of errors caused by quality problems in the ICOP Germany/USA and Japan/USA comparisons for the year 1987. UVRs based on census information were adjusted where necessary, using additional information from other sources like trade sources, companies and industry experts. The extent and direction of the adjustments made differed between industries, but appeared to be non-systematic. The adjustments to the original UVRs varied from -34% to 41%, and were highest for industries producing investment goods like machinery and transport equipment. But it was also found that at higher levels the original census-based UVRs were fairly robust, as errors cancel out. The adjusted UVRs for total manufacturing were about 4% higher than the unadjusted UVRs for both comparisons.

Ad 3. Not all products in two countries can be matched. This is mainly caused by the lack of product value or volume data (for reasons of company confidentiality, or because of the unimportance of the

[12] Data on highly specified products is often collected for the construction of a Producer Price Index (PPI) in a country. PPIs, however, are only based on a small sample of products in an economy. This sample of products obviously differs considerably between countries. Therefore the data is only of limited use for making international comparisons for which product *matches* need to be made. It has to be noted however, that in an intertemporal context PPI data on product prices is superior to census unit value ratios, as measurement errors are much higher for the latter. For the case of the US, F.R. Lichtenberg and Z. Griliches (1989) found the estimated signal to noise ratios for the PPI and the UVR to be 2.72 and .53 respectively.

product), incomparable volume units (e.g. weight vs. metric units), difficulties in finding good corresponding products and the existence of country unique products. Because of this, output coverage will be incomplete. This has two effects.

Firstly, it creates a possible bias in the products matched, especially in comparisons involving a low and a high productivity country, as discussed in section 3. This is illustrated by Table 1. The coverage of Chinese output in the China/USA 1985 comparison is more as two times as high as the coverage of the USA output. This means that the products matched are biased towards basic, low tech goods which are produced in larger quantities in the low productivity country China. This is especially true for branches like Stone, glass and clay products and Basic and fabricated metal products. UVRs (in Yuan per US$) will be underestimated. But the other columns of Table 1 show that this is not true of comparisons involving two high productivity countries. At both the branch and total manufacturing level coverage is rather similar for these countries.

A second effect of incomplete coverage, and independent from the first, is uncertainty as the estimation of industry UVRs is based on a *sample* of products and not on *all* products. This uncertainty is measured by the *sampling variance,* which measurement will be the focus of the remaining part of this paper.

6. Reliability Measures of Unit Value Ratios

In this section reliability measures will be developed for industry, branch and total manufacturing level unit value ratios (UVRs). I define the reliability of a UVR as its coefficient of variation, that is its standard deviation divided by its mean. The basic idea is that because the estimate of an industry UVR is based on a sample of products, high variance in the product UVRs lowers the reliability of the industry UVR, as the sample is apparently taken from a high variance (heterogeneous) population. This unreliability is subsequently carried over to branch and total manufacturing UVRs.

Both the stratified sampling and the stochastic interpretation of the ICOP method enable us to compute variances. The difference between the two approaches concerns the question whether or not variances should *always* increase with the degree of relative price variability.

Consider the extreme case in which all products in an industry can be matched (100% coverage). If the UVRs of the products differ significantly from the industry UVR, should the industry UVR be considered as just as reliable as a same situation in which the product level UVRs are all equal to the industry UVR? The stochastic approach answer is no. "This agrees with the *intuitive* notion that when the individual prices move very disproportionately, the overall price index cannot be estimated more precisely.", according to Selvanathan and Prasada Rao (1994, p.48-49, my italics). Although I agree that in the case of high variance "the ability of the index to reflect the price change in all the commodities is dubious" (op. cit. p.5), I do not agree that therefore the *reliability* of the industry UVR product), incomparable volume units (e.g. weight vs. metric units), difficulties in finding good corresponding products and the existence of country unique products. Because of this, output coverage will be incomplete. This has two effects.

Firstly, it creates a possible bias in the products matched, especially in comparisons involving a low and a high productivity country, as discussed in section 3. This is illustrated by Table 1. The coverage of Chinese output in the China/USA 1985 comparison is more as two times as high as the coverage of the USA output. This means that the products matched are biased towards basic, low tech goods which are produced in larger quantities in the low productivity country China. This is especially true for branches like Stone, glass and clay products and Basic and fabricated metal products. UVRs (in Yuan per US$) will be underestimated. But the other columns of Table 1 show that this is not true of comparisons involving two high productivity countries. At both the branch and total manufacturing level coverage is rather similar for these countries.

A second effect of incomplete coverage, and independent from the first, is uncertainty as the estimation of industry UVRs is based on a *sample* of products and not on *all* products. This uncertainty is measured by the *sampling variance,* which measurement will be the focus of the remaining part of this paper.

6. Reliability Measures of Unit Value Ratios

In this section reliability measures will be developed for industry, branch and total manufacturing level unit value ratios (UVRs). I define the reliability of a UVR as its coefficient of variation, that is its standard deviation divided by its mean. The basic idea is that because the estimate of an industry UVR is based on a sample of products, high variance in the product UVRs lowers the reliability of the industry UVR, as the sample is apparently taken from a high variance (heterogeneous) population. This unreliability is subsequently carried over to branch and total manufacturing UVRs.

Both the stratified sampling and the stochastic interpretation of the ICOP method enable us to compute variances. The difference between the two approaches concerns the question whether or not variances should *always* increase with the degree of relative price variability.

Consider the extreme case in which all products in an industry can be matched (100% coverage). If the UVRs of the products differ significantly from the industry UVR, should the industry UVR be considered as just as reliable as a same situation in which the product level UVRs are all equal to the industry UVR? The stochastic approach answer is no. "This agrees with the *intuitive* notion that when the individual prices move very disproportionately, the overall price index cannot be estimated more precisely.", according to Selvanathan and Prasada Rao (1994, p.48-49, my italics). Although I agree that in the case of high variance "the ability of the index to reflect the price change in all the commodities is dubious" (op. cit. p.5), I do not agree that therefore the *reliability* of the industry UVR as a *mean* of all product UVRs is less.[13] The difference comes down to whether or not to apply a finite population correction (fpc) to the variances computed. From the stratified sampling point of view it should be applied, whereas the stochastic approach argues for not. The argument in favour of applying the fpc is that with increasing coverage of the industry's output the variance diminishes, down to zero when the coverage is complete. This is an attractive property as the industry UVR is not derived as a price 'signal' on its own, but with the explicit aim to serve as a currency conversion factor for the industry output which is the sum of output of all products. Therefore, when the output of all products is covered the variance should be zero.

For this reason I prefer the stratified sampling approach to develop reliability indicators in this paper. If one opts for the stochastic interpretation, the same formulas as developed below are valid, ignoring the fpc.

Reliability of Industry UVRs

The estimated variance of a product UVR, denoted by $v(UVR_{i,j})$, is given by the weighted product variance around the estimated industry UVR[14]:

$$v(UVR_{i,j}) = \sum_{i=1}^{I_j(m)} \frac{O_{i,j}}{\sum_{i=1}^{I_j(m)} O_{i,j}} \times (UVR_{i,j} - \hat{UVR}_j)^2 \qquad (6.1)$$

Or in a computational easier form:

$$v(UVR_{i,j}) = \left(\sum_{i=1}^{I_j(m)} \frac{O_{i,j}}{\sum_{i=1}^{I_j(m)} O_{i,j}} \times UVR_{i,j}^2 \right) - \hat{UVR}_j^2 \qquad (6.2)$$

As weights O should be taken the base country values when the industry UVR is Laspeyres, or the other country values at base prices when Paasche.

As samples in industries are independently drawn, the variance of the industry UVR (recall that this UVR is a weighted mean of the product UVRs) is given by

[13] Unless one has the opinion that deviations of product UVRs from the industry mean UVR are mainly due to measurement errors and not to 'true' product specific deviations. The problem is that the relative importance of these two causes cannot be seperately identified. Their importance is therefore a matter of a priori belief. In this paper I assume that measurement errors are relatively unimportant.

[14] N.B. In order to compute this estimate at least two products have to be matched in each industry.

$$v(U\hat{V}R_j) = \left(1 - \frac{\sum_{i=1}^{I_j(m)} O_{i,j}}{O_j}\right) \times v(UVR_{i,j}) \times \sum_{i=1}^{I_j(m)} \left(\frac{O_{i,j}}{\sum_{i=1}^{I_j(m)} O_{i,j}}\right)^2 \tag{6.3}$$

with the first part between brackets being the finite population correction: one minus the output part of the industry which is matched (the industry coverage). The reliability of the industry UVR (r_j, the coefficient of variation) is given by:

$$r_j = \frac{\sqrt{v(U\hat{V}R_j)}}{U\hat{V}R_j} \tag{6.4}$$

The higher r_j the lower the reliability. Thus the higher the variance of the product UVRs in the industry, the lower the reliability of the industry UVR; and the higher the coverage, the higher the reliability.

Reliability of Branch UVRs

The reliability of an industry UVR is dependent on the industry boundaries. Narrowing down the definition of an industry to include the matched products only, would imply a 100% coverage and thus zero variance. In a normal stratification procedure such a 'gain' would automatically be counterweighted by a rise in the variance of some other widened industry, as all industries should comprise the total branch output. However, in the case at hand this correcting mechanism is not at work because there is typically always a part of the branch which is not covered by a product match. And narrowing down an industry would widen this residual branch part for which no variance can be computed. Take for example a branch in which two matches have been made. Define two industries, one including the matched products, the other including the remaining products. The variance of the first group will be zero because of 100% coverage, while the variance of the second is unknown. Branch variance, computed as a weighted average of industry variances, would be zero, which is obviously not desirable. Therefore the variance of branch k UVR, denoted by $v(UVR_k)$, is defined as a weighted average of the variances of the industry UVRs[15], which are calculated without the finite population correction (fpc).[16] Instead I apply the fpc at the branch level for which the boundaries are fixed.

---

[15] Industry variances are assumed to be uncorrelated. Industry *means* will probably by correlated as product prices are influenced by common factors like oil prices and governmental policies (e.g. national policy may lead to a rise in all prices in an economy by 2%). However, there is no a priori reason to suspect that sampling *variances* are correlated as well.

[16] As finite population corrections should be applied only once.

$$v(U\hat{V}R_k) = \left(1 - \frac{O_k(m)}{O_k}\right) \times \sum_{j=1}^{J_k} \left(\frac{O_{j,k}}{\sum_{j=1}^{J_k} O_{j,k}}\right)^2 \tilde{v}(U\hat{V}R_{j,k}) \tag{6.5}$$

with $O_{j,k}$ the total value of output in industry j, O_k the total output in branch k and $O_k(m)$ the total output in branch k which is covered by matched products. $O_k(m)$ is given by

$$O_k(m) = \sum_{j=1}^{J_k} \sum_{i=1}^{I_j(m)} O_{i,j,k}$$

The first part in (6.5) is the finite population correction. Industry variances are computed according:

$$\tilde{v}(U\hat{V}R_j) = v(UVR_{i,j}) \times \sum_{i=1}^{I_j(m)} \left(\frac{O_{i,j}}{\sum_{i=1}^{I_j(m)} O_{i,j}}\right)^2 \tag{6.6}$$

The reliability of the branch k UVR (r_k) is given by its coefficient of variation, analogue to formula (6.4). It depends on the variance of the industry UVRs (without fpc), but also on the coverage of the branch output. If the coverage ratio is higher, the reliability will be higher, and if the variance of the industry UVRs is higher, then the reliability will be lower.

Reliability of Total Manufacturing UVR

Because the sampling variances of the branches are uncorrelated, the sample variance of the total manufacturing sector, denoted by v (UVR_{manu}), is given by:

$$v(U\hat{V}R_{manu}) = \sum_{k=1}^{K} \left(\frac{O_k}{\sum_{k=1}^{K} O_k}\right)^2 v(U\hat{V}R_k) \tag{6.7}$$

The reliability of the total manufacturing UVR (r_{manu}) is given by its coefficient of variation, analogue to formula (6.4). If the branch UVR variances are high (low) the reliability of the manufacturing UVR will be low (high).

The formulae given are valid both for the variances of the Laspeyres and the Paasche UVRs. The formula for the variance of the Fisher, which is often used as a summary of the Laspeyres and the Paasche and defined as their geometric average, is not easily given. It cannot be analytically derived and depends on the covariance of the Laspeyres and Paasche. Simulation exercises[17] show that the geometric average of the standarddeviations of the Laspeyres and Paasche is an upperlimit for the

[17] Simulation results from Mark Huisman. Simulations performed for n=2500 and varying values for the correlation between the Laspeyres and the Paasche (from 0.21 to 0.92).

standarddeviation of the Fisher, being close when the correlation between the Laspeyres and Paasche is high, and slowly overestimating when correlation drops. Thus the geometric average of the standarddeviations of the Laspeyres and Paasche is a conservative estimate for the standarddeviation of the Fisher.

Using Variances for Constructing Confidence Limits

The sampling variances can be used to construct confidence limits for branch UVRs. However, this is only possible when the estimated branch UVR is normally distributed about the corresponding population value. The confidence limits for the branch UVR are given by:

$$[UVR_k - t \times s(\hat{UVR_k}) \; , \; UVR_k + t \times s(\hat{UVR_k})] \tag{6.8}$$

with s the square root of the estimated branch variance and t the critical value of the Student's t-distribution. As the number of matches in each branch are not particularly high from a sampling perspective (it varies from 1 to 30 at the most) the use of the Student's t-distribution instead of the normal distribution is of great importance. The degrees of freedom of the standard deviation (which are needed for the value of t) is not easily determined as its distribution is rather complex. It can be approximated by n_e, the effective number of degrees of freedom, which is given by [18]

$$n_e = \frac{\left(\sum_{j=1}^{J_k} g_j s_j^2\right)^2}{\sum_{j=1}^{J_k} \frac{g_j^2 s_j^4}{n_j - 1}} \tag{6.9}$$

with $g_j = O_j^2 / \sum O_j$. This number is inbetween the smallest number of product matches minus one in the industries of branch k, and the sum of all product matches made in branch k.

7. Empirical Implementation of the Reliability Measures

In this section the reliability measures developed will be empirically implemented for five manufacturing comparisons made according to the ICOP industry of origin approach. First, the 1992 West-Germany/USA comparison will be discussed extensively, as it is the most elaborate comparison made by now within ICOP (420 product matches). Then follow the aggregate results for four other bilateral comparisons.

[18] Based on Satterthwaite (1946) as described in Cochran (1977, p.96).

Reliability of Industry UVRs for West-Germany / USA Manufacturing Comparison 1992

The manufacturing sector was divided into the standard 16 ICOP branches. Table 2 provides the basic results.

Table 2 Unit Value Ratios and Details of the Matching Process for Manufacturing Branches West-Germany / USA Comparison 1992. [19]

Branch	Number of Unit Value Ratios	Matched value of output as % of branch gross value of output		UVR At German Quantity weights (Paasche)	UVR At US Quantity weights (Laspeyres)	Branch value of output as % of total manufacturing output	
		Germany (%)	US (%)	(DM/US$)	(DM/US$)	Germany (%)	US (%)
Food Products	71	43	47	2.06	2.15	7.88	11.51
Beverages	13	65	77	1.89	1.90	1.92	2.02
Tobacco	4	62	44	0.92	0.89	1.25	1.25
Textiles	20	32	59	2.51	2.74	1.53	2.19
Wearing Apparel	35	47	63	2.85	3.26	0.92	2.23
Leather Products and Footwear	9	43	62	2.07	2.12	0.38	0.32
Wood Products and Furniture	13	29	20	2.41	2.45	2.20	3.38
Paper and Printing (a)	19	22	27	1.86	2.02	4.34	7.05
Chemicals	34	12	12	2.50	2.57	7.82	10.43
Oil Products	6	19	84	1.77	1.93	4.20	5.21
Rubber and Plastic Products	6	5	8	1.90	2.13	4.34	3.83
Stone, Clay and Glass Products	19	24	17	1.74	1.85	3.34	2.66
Basic and Fabricated Metal	40	28	24	1.97	2.11	13.77	12.72
Machinery and Transport Equip.	56	7	7	1.71	2.07	31.92	21.86
Electrical Equipment	58	22	46	1.81	1.67	12.40	7.49
Other Manufacturing	17	7	2	1.90	2.13	1.79	5.84
Total Manufacturing	420	20	27	1.91	2.14	100.00	100.00

Note: (a) Excluding Printing

Source: Calculations based on ICOP/LCRA estimates, see Table 1.

In total 420 product matches were made in 65 industries, of which in 59 industries more than one match was made. Table 3 shows the reliability of the Laspeyres and Paasche index for the 59 industries. Columns 4 and 5 give the coefficient of variation as calculated without the finite population correction according to formula (6.6). These columns indicate the degree of homogeneity of the industry, as they reflect the variance of the product UVRs only, irrespective the output percentage covered. This is interesting to assess whether the products in the industry are justifiable grouped together or not. Columns 6 and 7 are calculated according to (6.4) and include the finite population correction. These are useful when one is interested in the reliability of the industry UVRs per se. A number of conclusions can be drawn from this table.

[19] The UVRs for this comparison are based on the standard ICOP-method and therefore do not completely accord to the stratified sampling interpretation given in this paper.

Table 3 Reliability of Industry UVRs, West-Germany / USA Manufacturing Comparison 1992

	Number of Product Matches	Coverage ratio		Coefficient of Variation		Coefficient of Variation with finite population correction		Contribution of industry to total manufacturing variance (in % of total)	
		US	Germany	Laspeyres UVR	Paasche UVR	Laspeyres UVR	Paasche UVR	Laspeyres	Paasche
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
OPTHALMIC GOODS	3	3	2	0.62	0.70	0.61	0.69	8.56	0.24
OTHER METAL PRODUCTS	21	5	22	0.53	0.22	0.51	0.19	61.05	5.45
COMPUTERS AND OFFICE EQUIPMENT	7	50	7	0.51	0.48	0.36	0.47	0.53	0.75
COMMUNICATION EQUIPMENT	14	26	9	0.37	0.19	0.31	0.18	0.81	1.14
CIGARETTES	4	44	62	0.34	0.35	0.26	0.21	0.04	0.04
MEN'S AND BOY'S HEAD AND NECKWEAR	5	30	21	0.33	0.32	0.28	0.28	0.00	0.00
GLASS	7	31	47	0.32	0.17	0.27	0.13	0.49	0.18
ELECTRONIC CONPONENTS	13	29	23	0.32	0.40	0.27	0.35	0.13	2.46
PLASTICS	2	0	1	0.30	0.41	0.30	0.41	7.57	10.08
METEOROLOGICAL INSTRUMENTS	9	2	12	0.28	0.27	0.28	0.26	0.77	0.10
LEATHER FOOTWEAR	4	80	47	0.27	0.21	0.12	0.15	0.00	0.00
MOTOR VEHICLES AND THEIR ENGINES	4	7	2	0.26	0.35	0.25	0.34	5.61	53.88
AGRICULTURAL MACHINERY	16	34	24	0.26	0.36	0.21	0.32	0.02	0.06
KNITTING MILLS	6	40	11	0.25	0.34	0.19	0.32	0.04	0.08
RUBBER PRODUCTS	4	30	18	0.22	0.21	0.18	0.19	0.15	0.33
COFFEE AND OTHER FOOD PRODUCTS	7	23	23	0.22	0.12	0.19	0.10	0.06	0.04
LIME	3	84	0	0.22	0.24	0.09	0.24	0.00	0.00
GRAIN MILL PRODUCTS	4	68	71	0.21	0.22	0.12	0.12	0.01	0.00
SPRIRITS	8	72	49	0.21	0.15	0.11	0.11	0.02	0.01
WOVEN GOODS AND YARN, EXCL. WOOL	8	73	62	0.21	0.23	0.11	0.14	0.09	0.06
WET CORN MILLING	3	35	65	0.20	0.19	0.16	0.11	0.01	0.00
ELECTRIC HOUSEHOLD APPLIANCES	15	58	47	0.20	0.18	0.13	0.13	0.02	0.04
OTHER ELECTRONIC TOOLS	6	9	1	0.20	0.25	0.19	0.25	0.00	1.07
PACKAGING MATERIAL	4	14	40	0.19	0.19	0.18	0.15	0.81	0.16
UNDERWEAR	7	74	67	0.19	0.54	0.10	0.31	0.02	0.00
BASIC CHEMICALS, INCLUDING FERTILIZE	12	12	4	0.19	0.22	0.18	0.21	3.51	4.60
DAIRY PRODUCTS	8	59	60	0.17	0.15	0.11	0.09	0.09	0.11
PETROLEUM PRODUCTS	6	82	19	0.16	0.22	0.07	0.19	0.21	1.54
CONSTRUCTION MACHINERY	10	6	26	0.16	0.29	0.15	0.25	0.02	0.28
FINISHED LEATHER	2	46	69	0.16	0.17	0.12	0.09	0.00	0.00
BUILDERS' CARPENTRY	2	13	57	0.16	0.17	0.15	0.11	0.11	0.01
MEAT AND POULTRY PRODUCTS	6	63	53	0.16	0.16	0.09	0.11	0.34	0.11
CASTINGS	5	54	65	0.15	0.21	0.10	0.12	0.05	0.14
WOOLLEN YARN AND FABRICS	3	95	62	0.15	0.17	0.03	0.10	0.00	0.00
HOUSEHOLD, OFFICE FURNITURE	8	23	25	0.15	0.15	0.13	0.13	0.08	0.13
CLAY AND NON-CLAY REFRACTORIES	5	62	35	0.15	0.26	0.09	0.21	0.00	0.07
CONFECTIONERY PRODUCTS	6	15	8	0.14	0.16	0.13	0.15	0.00	0.03
LEATHER GOODS	3	36	19	0.14	0.12	0.11	0.11	0.00	0.00
MUSICAL INSTRUMENTS	3	22	26	0.14	0.18	0.12	0.15	0.00	0.00
OTHER MACHINES	21	3	5	0.14	0.15	0.14	0.15	8.15	16.32
CHEMICAL PRODUCTS	16	26	37	0.14	0.10	0.12	0.08	0.27	0.13
LAMPS	3	54	85	0.13	0.14	0.09	0.05	0.00	0.00
BEER	4	93	80	0.13	0.14	0.03	0.06	0.01	0.04
FISH	4	32	27	0.12	0.09	0.10	0.08	0.00	0.00
VEGETABLE PRODUCTS AND NUTS	19	34	59	0.11	0.22	0.09	0.14	0.05	0.02
PREPARED ANIMAL FEEDS	4	60	70	0.11	0.14	0.07	0.08	0.01	0.01
OTHER TEXTILE GOODS	3	9	5	0.10	0.09	0.10	0.08	0.00	0.00
BASIC PULP AND PAPER	9	72	52	0.09	0.11	0.05	0.08	0.12	0.08
SAWMILLING, PLANING, ETC. AND VENEE	3	27	39	0.09	0.11	0.07	0.08	0.03	0.01
MEN'S AND BOYS' APPAREL	22	88	57	0.08	0.11	0.03	0.07	0.03	0.02
NON-STEEL METAL PRODUCTS	7	46	58	0.06	0.08	0.05	0.05	0.03	0.06
SYNTHETIC YARN	3	52	89	0.06	0.06	0.04	0.02	0.00	0.00
SOAP	4	8	14	0.06	0.06	0.06	0.05	0.05	0.01
STEEL PRODUCTS	6	68	69	0.06	0.06	0.03	0.03	0.02	0.10
SANITARY AND OTHER PAPER PRODUCTS	3	36	20	0.04	0.04	0.03	0.04	0.00	0.00
BAKERY PRODUCTS	3	18	29	0.04	0.08	0.03	0.06	0.00	0.01
SUGAR	2	66	76	0.03	0.03	0.02	0.02	0.00	0.00
CARPETS AND RUGS	2	85	54	0.03	0.05	0.01	0.03	0.00	0.00
CEMENT	2	87	69	0.02	0.03	0.01	0.02	0.00	0.00

Source: Based on ICOP/LCRA estimates, see Table 1.

Firstly, the degree of homogeneity of industries differs considerably (see column 4 and 5). There are homogeneous industries like cement, sugar and soap as indicated by their low coefficient of variation. For these industries, it can be safely assumed that the UVRs of the covered part are representative of the non-covered part. Therefore they should be reweighted accordingly. But there also rather heterogeneous industries like opthalmic goods, other metal products and computers and office equipment. These heterogeneous industries are of two kinds. There are industries which have been defined too broadly as could be expected beforehand. An obvious example is the plastics industries, encompassing a wide range of 4-digit industries, just as other metal products. These industries require a further breakdown. But there are also industries which are already defined as single 4-digit industries like opthalmic goods, but which are simply not homogeneous, even not at this level. These products are unjustifiably taken together into an industry.

A second conclusion following from Table 3 is that reliability is correlated with the coverage ratio: compare columns 2 and 3 with columns 6 and 7 respectively. This is not obvious as the reliability measure of industries is not only influenced by the coverage ratio, but also by the (estimated) degree of homogeneity. The relationship is not strict however. Industries like leather footwear, glass and cigarettes appear to be unreliable although their coverage ratios are high (above 30%). And on the other hand lowly covered industries like bakery products, soap and sanitary and other paper products (20% or lower in at least one country) appear to be highly homogeneous and therefore are reliable nevertheless.

A third conclusion drawn from Table 3 is that reliability is uncorrelated with the number of matches made: compare column 1, giving the number of product matches, and columns 6 and 7. Men's and boys apparel has 22 matches and is highly reliable, just as cement having 2. On the other hand other metal products having 21 matches is unreliable, just as opthalmic goods having 3 matches.

Reliability of Branch UVRs for West-Germany / USA Manufacturing Comparison 1992

Table 4 gives the reliability of the Laspeyres and Paasche branch UVRs as computed according to formula (6.5). Columns 3 to 6 give the 90% confidence limits of the branch UVRs as follows from the reliability and the degrees of freedom computed according to formulae 6.8 and 6.9. Two observations can be made. Firstly, the reliability of branch UVRs differs greatly. As columns 1 and 2 show the "other manufacturing branch", basic and fabricated metal, tobacco and the rubber and plastic branch are the most unreliable branches, whereas food, beverages, wood and wearing apparel are relatively reliable. Secondly, the branch UVRs differ significantly. Columns 3 to 6 give the lower and upper limit of the UVRs of the 90% confidence interval. If the confidence intervals of two branch UVRs intersect, the UVRs are not significantly different (at 90% confidence). This is the case for, for example, the textiles, wood and chemical branch UVRs. On the other hand the UVRs of wearing apparel, tobacco and food products differ significantly from each other.

Table 4 Reliability of Manufacturing Branch UVRs for West-Germany/USA Comparison 1992.

	Coefficient of Variation of UVR		90%-Confidence Limits Laspeyres		90%-Confidence Limits Paasche		Contribution to total manufacturing variance	
	Laspeyres (1)	Paasche (2)	Lower (3)	Upper (4)	Lower (5)	Upper (6)	Laspeyres (7)	Paasche (8)
Food Products	0.04	0.04	2.00	2.22	1.92	2.20	0.57	0.32
Beverages	0.07	0.07	1.66	2.14	1.65	2.14	0.04	0.05
Tobacco	0.26	0.21	0.40	1.41	0.50	1.33	0.04	0.04
Textiles	0.09	0.11	2.33	2.92	2.01	3.01	0.13	0.14
Wearing Apparel	0.04	0.07	3.01	3.10	2.48	3.22	0.05	0.03
Leather Products and Footwear	0.09	0.10	1.79	2.41	1.68	2.46	0.00	0.00
Wood Products and Furniture	0.08	0.08	1.90	2.96	1.93	2.88	0.21	0.15
Paper and Printing	0.10	0.07	1.59	2.33	1.59	2.14	0.93	0.24
Chemicals	0.10	0.13	2.11	2.96	1.93	3.07	3.83	4.75
Oil Products	0.06	0.19	1.69	2.02	1.11	2.44	0.21	1.54
Rubber and Plastic Products	0.48	0.46	(a)	8.34	(a)	7.37	7.72	10.40
Stone, Clay and Glass Products	0.20	0.10	1.14	2.45	1.41	2.07	0.49	0.25
Basic and Fabricated Metal Products	0.41	0.10	0.62	3.46	1.62	2.32	61.15	5.75
Machinery and Transport Equipment	0.12	0.18	1.65	2.13	1.18	2.24	13.80	70.55
Electrical Equipment	0.14	0.12	1.29	2.20	1.44	2.18	1.50	5.46
Other Manufacturing	0.35	0.20	(a)	4.04	0.80	3.00	9.33	0.34
TOTAL MANUFACTURING	0.07	0.06	1.90	2.15	1.71	2.11	100.00	100.00

Note: (a) Negative value.

Source: See Table 2.

Reliability of Total Manufacturing UVR for West-Germany / USA Manufacturing Comparison 1992

The 90% confidence interval for the total manufacturing UVR is given in the last row of Table 4. For the Laspeyres and Paasche UVRs resp. [1.90 , 2.15] and [1.71 , 2.11], or ± 10.9% and. ± 10.2%. However, these intervals need a qualification as it it is shown that a major part of the unreliability originates from only a couple of branches, see columns 7 and 8 of Table 4. They give the contribution of individual branches to the total manufacturing variance. It shows that the variance of the manufacturing Laspeyres UVR originates for more than 60% from the basic and fabricated metal branch, whereas the variance in the Paasche originates for more than 70% from the machinery and transport branch. Taking the analysis one level lower, columns 8 and 9 in Table 3 give the contribution of industries to the total manufacturing variance. This contribution depends on the variance of UVRs in the industry, the industry output share in the manufacturing branch it belongs to, and the output share of this branch in total manufacturing. It shows that the Laspeyres variance originates for 61% from the other metal products industry, the opthalmic goods industry being second with 9%. Paasche variance originates from motor vehicles and their engines (54%), followed by other machines (16%) and plastics (10%).

This analysis suggests that the reliability of the manufacturing UVR might be improved by excluding only one or two industries. Indeed, when excluding "other metal products", the total manufacturing Laspeyres UVR has a 90% confidence interval of ± 6.9%, instead of ±10.9%. The Paasche UVR however can not be easily improved. Excluding the motor vehicle industry gives an interval of ± 9.4% which is only a minor improvement upon the original ±10.2%. This is because the other industries in the machinery and transport equipment branch are also rather unreliable, thus excluding the motor vehicles industry results only in a minor improvement. Industries in the metal branch on the other hand are highly homogeneous, except for other metal products. Therefore excluding this industry results in a large gain of precision in the Laspeyres UVR.

The Case of Japan/USA, UK/USA, France/USA and China/USA Manufacturing Comparisons.

The bilateral manufacturing comparisons of Japan, UK, France (for 1987) and China (for 1985) with the US reflect a wide range of ICOP comparisons. The number of product matches vary from 58 for China/USA and 63 for France/USA to 172 for UK/USA and 190 for Japan/USA.

Table 5 shows the reliability of the Laspeyres and Paasche branch UVRs and the coverage ratios involved. It shows that there is no clearcut distinction between reliable and unreliable branches. In Japan the most unreliable branch is food manufacturing, whereas for the UK it is beverages, chemicals in France, and rubber and plastics in China. Most reliable branches are wood manufacturing for Japan, oil products for the UK, textiles for France and tobacco for China. The table shows also that the coverage ratio is not the prime determinant of reliability. For example, one of the worst covered branches in the Japan/US comparison, wood products and furnitures, is at the same time the most reliable. Reliability depends both on the branch coverage ratio and the degree of homogeneity of its industries. Table 6 shows the resulting 90% confidence interval limits for the total manufacturing UVRs for each comparison when the degrees of freedom are calculated according to (6.9). It shows that the UK/USA comparison was the most reliable one with 90% probability that the Laspeyres and Paasche UVR deviate no more than resp. ±4.4% and ±4.5% from the estimated values On the other hand the 90% confidence interval is rather wide for the China/USA UVRs, resp ±27% and ±15%[20].

Gains of Stratification

An important issue which has to be addressed is whether the stratification of the total manufacturing sector into branches and industries has been useful. That is, did stratification lower the variance associated with the total manufacturing UVR? In theory stratification almost certainly leads to lower variances than simple random sampling under two conditions: 1. the characteristics (means, variances) of the strata differ significantly, and 2. the samples are taken optimal or at least

[20] This is caused not only because of a high coefficient of variation, but also because the number of matches in each branch is particularly low for this comparison. This forces down the effective degrees of freedom for the manufacturing UVR, which in turn has a magnifying influence on the critical value of the t-distribution as used in calculating the confidence limits (see (6.9)).

Table 5 Branch Reliability and Coverage Ratios for Four Manufacturing Comparisons.

	China/USA 1985				France/USA 1987				Japan/USA 1987				United Kingdom/USA 1987			
	Reliability Laspeyres UVR	Reliability Paasche UVR	Coverage Ratio USA (%)	Coverage Ratio China (%)	Reliability Laspeyres UVR	Reliability Paasche UVR	Coverage Ratio USA (%)	Coverage Ratio France (%)	Reliability Laspeyres UVR	Reliability Paasche UVR	Coverage Ratio USA (%)	Coverage Ratio Japan (%)	Reliability Laspeyres UVR	Reliability Paasche UVR	Coverage Ratio USA (%)	Coverage Ratio UK (%)
Food Products	0.344	0.114	44	57	(b)	(b)			0.178	0.170	11	14	0.055	0.035	16	18
Beverages	(a)	(a)	27	13	(b)	(b)			0.135	0.172	30	33	0.128	0.123	29	44
Tobacco	0.026	0.019	15	11	(b)	(b)			(a)	(a)	81	86	0.092	0.055	87	95
Textiles	0.126	0.096	54	82	0.001	0.001	9	5	0.070	0.085	39	26	0.039	0.060	64	36
Wearing Apparel	(b)	(b)			0.059	0.056	27	36	0.065	0.068	30	21	0.027	0.036	36	41
Leather Products and Footwear	(a)	(a)	41	45	(a)	(a)	34	51	0.158	0.150	29	34	0.066	0.077	50	34
Wood Products and Furniture	(b)	(b)			0.099	0.085	5	18	0.023	0.028	8	20	0.034	0.121	9	16
Paper and Printing	0.274	0.297	10	70	(b)	(b)			0.069	0.096	15	13	0.059	0.104	11	9
Chemicals	0.228	0.222	27	35	0.115	0.148	9	9	0.040	0.053	14	16	0.038	0.046	13	10
Oil Products	(b)	(b)			(b)	(b)			0.129	0.147	77	65	0.016	0.024	72	44
Rubber and Plastic Products	0.244	0.612	3	35	(a)	(a)	7	9	0.073	0.088	11	7	0.160	0.056	8	7
Stone, Clay and Glass Products	0.197	0.402	7	60	0.034	0.051	16	17	0.044	0.034	28	33	0.102	0.113	8	15
Basic and Fabricated Metal Products	0.570	0.165	14	54	0.038	0.040	7	11	0.051	0.035	23	25	0.041	0.046	12	20
Machinery and Transport Equipment	0.154	0.186	20	12	0.075	0.054	16	24	0.043	0.069	18	20	0.082	0.065	16	13
Electrical Equipment	0.334	0.450	5	21	(b)	(b)			0.070	0.104	11	12	0.061	0.060	5	5
Other Manufacturing	(b)	(b)			(b)	(b)			(b)	(b)			(b)	(b)		
TOTAL MANUFACTURING	0.093	0.078	19	43	0.034	0.037	12	18	0.037	0.034	20	19	0.021	0.021	19	18

Remark: Reliability is defined as the coefficient of variation.
Notes: (a) Only one product match made.
(b) No product match made.
Source: Based on product matches underlying Table 1.

Table 6 Reliability and Confidence Limits Total Manufacturing of Five Comparisons

	Reliability Laspeyres UVR	Reliability Paasche UVR	90% Confidence Limits for UVR (a) Laspeyres (+/- %)	Paasche (+/- %)
China/USA 1985	0.093	0.078	27.08	15.11
France/USA 1987	0.034	0.037	6.83	7.41
Japan/USA 1987	0.037	0.034	7.40	7.21
UK/USA 1987	0.021	0.021	4.41	4.53
West-Germany/USA 1992	0.066	0.061	10.90	10.20

Remark: Reliability is defined as the coefficient of variation.
Notes: (a) Left (-) and Right (+) limits are given as percentage deviation from estimated UVR.
Source: See Table 1 .

Table 7 Reliability of UVRs with and without Reweigthing Total Manufacturing of Five Comparisons

	With reweighting (a)				Without reweighting (b)			
	Laspeyres UVR	Paasche UVR	Reliability Laspeyres UVR	Reliability Paasche UVR	Laspeyres UVR	Paasche UVR	Reliability Laspeyres UVR	Reliability Paasche UVR
China/USA 1985	1.99	1.29	0.093	0.078	1.98	1.26	0.223	0.158
France/USA 1987	6.12	5.20	0.034	0.037	5.89	5.04	0.329	0.174
Japan/USA 1987	212	152	0.037	0.034	225	156	0.153	0.171
UK/USA 1987	0.74	0.65	0.021	0.021	0.70	0.65	0.070	0.061
West-Germany/USA 1992	2.14	1.91	0.066	0.061	2.12	1.96	0.062	0.043

Remark: Reliability is defined as the coefficient of variation. UVRs in local currency per US$.
Notes: (a) Based on product UVRs with industry and branch reweighting (the normal ICOP-procedure).
(b) Based on product UVRs without industry and branch reweighting.
Source: See Table 1 .

proportional. Condition 1 is fulfilled, as can be inferred from Tables 2 and 3. But sampling in ICOP is not proportional, let alone optimal. From the branch coverage ratios in Tables 2 and 5, it can be seen that samples have not been taken proportional as the ratios differ considerably across branches. Optimal stratification would imply that samples are taken not only proportional to size, but also to variance: branches with high variances should have a higher coverage ratio. But as pointed out earlier, reliability and coverage tend to correlate positively. This indicates that stratification is indeed far from optimal, and the gain of stratification is expected to be small, if positive at all.

Nevertheless, Table 7 provides evidence that stratification has been useful for four out of five comparisons. The last four columns give the results of the total manufacturing UVRs and their reliability when all product matches are taken together without industry and branch reweighting, thus without stratification. They can be compared with the results from the stratification procedure as given in the first four columns. It can be concluded that stratification has been successful in raising the reliability of the total manufacturing UVR estimates in all comparisons[21], except Germany/US 1992, for which the results are almost the same. It is also shown that the UVRs from the unweighted procedure are near the weighted ones. So the choice whether to reweight or not, is not really important for the estimation of the manufacturing UVR, but it is for the precision with which this estimate is made.

8. Conclusions and Suggestions

In this paper estimates of variances of Laspeyres and Paasche type of indices, as used in the industry of origin approach to international comparisons, were computed starting from the basic idea that these indices are based on a *sample* of products. Using stratified sampling theory it was possible to assess the reliability of branch and total manufacturing unit value ratios (UVRs). Reliability, defined as the coefficient of variation, depends on the degree of price variability and the percentage output covered by the sampled (matched) products. Application to five bilateral comparisons shows that there are no particularly (un)reliable branches, but that the reliability by branch differ greatly across comparisons. 90% Confidence intervals for the total manufacturing Laspeyres and Paasche UVRs ranged from ± 4% to ± 11% for the comparisons involving high productivity countries only; in case of the China/USA comparison confidence intervals were much larger: ± 15% and ± 27% for the Paasche and Laspeyres respectively. It was shown also that variance at the total manufacturing level originates from only a small number of unreliable industries.

A further issue that was addressed is the problem of reweighting. As branch UVRs are based on product UVRs, an aggregation step is required. In the simplest case one simply takes output weighted

[21] This is especially true when one dominant product match like cars has a major influence on reliability. Its influence is diminished by the reweighting procedure: by including it in the vehicles industry and subsequently machinery branch.

product UVRs for each branch. More sophisticated is to define an intermediate level, as done in the ICOP approach, which has been described in this paper from a stratified sampling perspective. Products are grouped into homogeneous industries and are subsequently reweighted by the industry output value. Stratified sampling theory suggests that estimates can be made more precise if the within group relative price variances are less than the between group variances. Four digit ISIC industries, although far from optimal, appear to satisfy this criterion. Industry reweighting greatly improved the reliability of the UVRs in four of the five cases studied.

Causes of unreliability of the estimated UVRs have been discussed in Section 5. These include the product mix problem, the real quality problem and incomplete coverage of products. All factors may create biases in the estimated UVRs. It was argued that in the case of comparisons involving a low and a high productivity country the product mix and real quality problems create a downward bias in the UVR estimations[22] as the quality of products in low productivity countries is lower than in the high productivity countries, and their product mixes consist of more low quality goods. Moreover, because of incomplete coverage the product sampling (matching) is biased towards basic, low tech products. This leads to a further underestimation of the UVR. These problems differ both across branches and across comparisons. Basic intermediate product branches will be less affected than branches like machinery and transport equipment. And the more the countries in a comparison differ in their level of development, the greater the problems will be.

The problems noted create both biases and variance in the estimation of UVRs. The part that will show up in the variance of the estimated UVRs is covered by the reliability indicator as defined in this paper as the coefficient of variation. Further research in the vein of Gersbach and van Ark (1994) to cover the part relating to biases is required.

[22] Stated in low productivity country local currency per high productivity country currency.

References

Allen, R.G.D. (1975), *Index Numbers in Theory and Practice,* Chicago.

van Ark, B. (1993), *International Comparisons of Output and Productivity*, Ph.D. thesis, Monograph Series No.1, Groningen Growth and Development Centre.

van Ark, B. (1996), "Issues in Measurement and International Comparison of Productivity- An Overview", in OECD, *Industry Productivity. International Comparisons and Measurement Issues*, Paris, pp.19-47.

van Ark, B. (1993) and R.D.J. Kouwenhoven (1994) , "Productivity in French Manufacturing: An International Comparative Perspective", *Research Memorandum* no. 571 (GD-10), Groningen Growth and Development Centre.

van Ark, B., E.J. Monnikhof and M.P. Timmer (1996), "Prices, Quantities and Productivity in Industry: A Study of Transition Economies in a Comparative Perspective", Paper presented at NBER/CRIW Conference on "International and Interarea Comparisons of Prices, Income and Output." 15/16 March 1996, Washington DC.

Cochran, W.G. (1977), *Sampling Techniques*, third edition, Wiley, New York.

Gersbach, H. and B. van Ark (1994), "Micro Foundations for International Productivity Comparisons", *Research Memorandum* GD-11, Groningen Growth and Development Centre.

Gilbert, M. and I.B. Kravis (1954), *An International Comparison of National Products and the Purchasing Power of Currencies*, OEEC, Paris.

Hooper, P. and E. Vrankovich (1995), "International Comparisons of the Levels of Unit Labor Costs in Manufacturing", *International Finance Discussion Papers,* Number 527, Board of the Governors of the Federal Reserve System, Washington D.C., Oktober 1995.

Koszerek, D. (1985a), "Eurostat's External Trade Indices", Chapter 9 in: *External Trade Indices.* Proceedings of a seminar held in Luxembourg, 6-8.11.1985, Eurostat News Special Edition.

Koszerek, D. (1985b), "The Relationship Between Price and Unit Value Movements and The Effect of Quality Changes" Chapter 7 in: *External Trade Indices.* Proceedings of a seminar held in Luxembourg, 6-8.11.1985, Eurostat News Special Edition.

Kravis, I.B., A. Heston and R.Summers (1982), *World Product and Income*, John Hopkins, Baltimore.

Lichtenberg, F.R. and Z. Griliches (1989), "Errors of Measurement in Output Deflators" in *Journal of Business & Economic Statistics*, January 1989, vol.7, No. 1.

Maddison, A. and B. van Ark (1988), "Comparisons of Real Output in Manufacturing", *Policy, Planning and Research Working Papers*, WPS5, World Bank, Washington DC.

Maddison, A. and B. van Ark (1994), "The International Comparison of Real Product and Productivity", *Research Memorandum GD-6*, Groningen Growth and Development Centre.

Pilat, D. (1994), *The Economics of Rapid Growth. The Experience of Japan and Korea*, Edward Elgar Publishers, Aldershot.

Selvanathan, E.A. and D.S. Prasada Rao (1994), *Index Numbers. A Stochastic Approach,* MacMillan, Basingstoke.

Szirmai, A. and R. Ruoen (1995), "China's Manufacturing Performance in Comparative Perspective", *Research Memorandum* no. 581 (GD-20), Groningen Growth and Development Centre.

Triplett, J. (1990) "The Theory of Industrial and Occupational Classifications and Related Phenomena" in: *Proceedings of 1990 Bureau of Census Annual Research Conference*, March 18-21, 1990, Arlington.

US Bureau of Census (1989), *1987 Census of Manufactures*, Industry Series, US Department of Commerce, Washington DC.

Papers issued in the series of the Groningen Growth and Development Centre:

536 (GD-1) Maddison, Angus and Harry van Ooststroom, The International Comparison of Value Added, Productivity and Purchasing Power Parities in Agriculture (1993)

537 (GD-2) Mulder, Nanno and Angus Maddison, The International Comparison of Performance in Distribution: Value Added, Labour Productivity and PPPs in Mexican and US Wholesale and Retail Trade 1975/7 (1993)

538 (GD-3) Szirmai, Adam, Comparative Performance in Indonesian Manufacturing, 1975-90 (1993)

549 (GD-4) de Jong, Herman J., Prices, Real Value Added and Productivity in Dutch Manufacturing, 1921-1960 (1993)

550 (GD-5) Beintema, Nienke and Bart van Ark, Comparative Productivity in East and West German Manufacturing before Reunification (1993)

567 (GD-6) Maddison, Angus and Bart van Ark, The International Comparison of Real Product and Productivity (1994)

568 (GD-7) de Jong, Gjalt, An International Comparison of Real Output and Labour Productivity in Manufacturing in Ecuador and the United States, 1980 (1994)

569 (GD-8) van Ark, Bart and Angus Maddison, An International Comparison of Real Output, Purchasing Power and Labour Productivity in Manufacturing Industries: Brazil, Mexico and the USA in 1975 (1994) (second edition)

570 (GD-9) Maddison, Angus, Standardised Estimates of Fixed Capital Stock: A Six Country Comparison (1994)

571 (GD-10) van Ark, Bart and Remco D.J. Kouwenhoven, Productivity in French Manufacturing: An International Comparative Perspective (1994)

572 (GD-11) Gersbach, Hans and Bart van Ark, Micro Foundations for International Productivity Comparisons (1994)

573 (GD-12) Albers, Ronald, Adrian Clemens and Peter Groote, Can Growth Theory Contribute to Our Understanding of Nineteenth Century Economic Dynamics (1994)

574 (GD-13) de Jong, Herman J. and Ronald Albers, Industrial Output and Labour Productivity in the Netherlands, 1913-1929: Some Neglected Issues (1994)

575 (GD-14) Mulder, Nanno, New Perspectives on Service Output and Productivity: A Comparison of French and US Productivity in Transport, Communications Wholesale and Retail Trade (1994)

576 (GD-15) Maddison, Angus, Economic Growth and Standards of Living in the Twentieth Century (1994)

577 (GD-16) Gales, Ben, In Foreign Parts: Free-Standing Companies in the Netherlands around the First World War (1994)

578 (GD-17) Mulder, Nanno, Output and Productivity in Brazilian Distribution: A Comparative View (1994)

579 (GD-18) Mulder, Nanno, Transport and Communication in Mexico and the United States: Value Added, Purchasing Power Parities and Productivity (1994)

580 (GD-19) Mulder, Nanno, Transport and Communications Output and Productivity in Brazil and the USA, 1950-1990 (1995)

581 (GD-20) Szirmai, Adam and Ren Ruoen, China's Manufacturing Performance in Comparative Perspective, 1980-1992 (1995)

GD-21 Fremdling, Rainer, Anglo-German Rivalry on Coal Markets in France, the Netherlands and Germany, 1850-1913 (December 1995)

GD-22 Tassenaar, Vincent, Regional Differences in Standard of Living in the Netherlands, 1800-1875. A Study Based on Anthropometric Data (December 1995)

GD-23 van Ark, Bart, Sectoral Growth Accounting and Structural Change in Postwar Europe (December 1995)

GD-24 Groote, Peter, Jan Jacobs and Jan Egbert Sturm, Output Responses to Infrastructure in the Netherlands, 1850-1913 (December 1995)

GD-25 Groote, Peter, Ronald Albers and Herman de Jong, A Standardised Time Series of the Stock of Fixed Capital in the Netherlands, 1900-1995 (May 1996)

GD-26 van Ark, Bart and Herman de Jong, Accounting for Economic Growth in the Netherlands since 1913 (May 1996)

GD-27 Maddison, Angus and D.S. Prasada Rao, A Generalized Approach to International Comparisons of Agricultural Output and Productivity (May 1996)

GD-28 van Ark, Bart, Issues in Measurement and International Comparison of Productivity - An Overview (May 1996)

GD-29 Kouwenhoven, Remco, A Comparison of Soviet and US Industrial Performance, 1928-90 (May 1996)

GD-30 Fremdling, Rainer, Industrial Revolution and Scientific and Technological Progress (December 1996)

GD-31 Timmer, Marcel, On the Reliability of Unit Value Ratios in International Comparisons (December 1996)

Groningen Growth and Development Centre Research Monographs:

No. 1 van Ark, Bart, International Comparisons of Output and Productivity: Manufacturing Productivity Performance of Ten Countries from 1950 to 1990 (1993)

No. 2 Pilat, Dirk, The Economics of Catch-Up: The Experience of Japan and Korea (1993)