Chapter 9

CONVERSION FACTORS IN RELATIVE PRODUCTIVITY CALCULATIONS: THEORY AND PRACTICE

by

Mary O'Mahony
National Institute of Economic and Social Research, London

Introduction

International comparisons of standards of living and labour productivity are measured as bilateral ratios of income or output per capita or per unit of labour input. Since income or output are generally measured in each country as values in domestic currencies, it is necessary to employ conversion factors to convert values to a common currency, usually US dollars. Therefore reliable international comparisons require dependable conversion factors. The latter should measure relative prices in two countries but the price measures required will depend on the purpose of the comparison. Thus, the price ratios needed to obtain international measures of standards of living, which should approximate differences in the cost of living, will not in general be identical to the price ratios required for international comparisons of productivity, which should measure differences in output prices.

In practice there are two types of conversion factors which have been employed in the literature. The first are the purchasing power parities (PPPs) calculated under the International Comparisons Project begun by Kravis and his associates in 1950 and now estimated by international bodies such as Eurostat. These results underlie the Penn World Table which is discussed in Summers and Heston (1991). The second type of conversion factors, which have an equally long history, are unit value ratios (UVRs) based on information in each country on the sales values and quantities of goods and services produced. These have been employed by Paige and Bombach (1959), underlie the work of the ICOP centre at the University of Groningen (for example in Maddison and van Ark, 1988; van Ark, 1993; Pilat, 1994; and Mulder, 1994 – to name just a few), and also have an extensive history in research carried out at The National Institute of Economic and Social Research in London (for example in Smith, Hitchens and Davies, 1982; O'Mahony, 1992).[1]

The purpose of this chapter is to scrutinise the available conversion factors in the light of certain criteria which they should satisfy in order to provide reliable estimates of comparative productivity. The chapter will not be concerned with the properties of conversion factors required for standard of living comparisons, although some criticisms of the conversion factors used for productivity comparisons will apply equally to standard of living comparisons. The chapter begins with a discussion of the theory underlying the use of conversion factors in measuring relative productivity. The third section lists a number of criteria which the conversion factors should satisfy. The following

five sections explain the importance of these criteria and examine how well they are satisfied by both the PPPs and UVRs. The final section includes some concluding remarks.

Productivity and relative price measurement

This section begins with a discussion of the theory underlying measurement of comparative productivity. This chapter will only be concerned with labour productivity. Other measures such as multi-factor productivity also create serious estimation problems, in particular in the measurement of capital stocks, a consideration of which are beyond the scope of this paper.

Define the level of labour productivity in country j relative to country u in sector i by:

$$RLP_{iju} = \{(Q_{ij}/L_{ij})/(Q_{iu}/L_{iu})\} \quad [1]$$

where Q is the quantity of output and L is labour input (hours worked). Equation [1] is generally estimated using data on the money value of output data. Therefore relative labour productivity is frequently measured by:

$$RLPiju = \{(Y_{ij}/L_{ij})/(Y_{iu}/L_{iu})\}\{P_{iu}/P_{ij}\} \quad [2]$$

Where $Q_{ij} = Y_{ij} / P_{ij}$, Y_{ij} is nominal output and P_{ij} are measures of the prices of goods and services produced in sector i in country j. Conceptually the price ratios should correspond to those included in the calculation of nominal output. From an economic theory standpoint, output should be measured by gross output since this is the measure consistent with a production function framework (see discussions in Jorgenson, Gollop and Fraumeni, 1987; and Oulton and O'Mahony, 1992). In this case the price ratios should correspond to a weighted average of the prices of final sales, *i.e.* suppose output in sector i consists of k separate commodities or services, and let P_i be the ratio of prices in country j relative to country u in sector i. Then gross output prices are given by:

$$P^G_i = \Sigma w_{ik} P_{ik}$$

where the weights w depend of the share of each output in the total value of sales.

In many international comparisons of productivity the more readily available value added definition of output is used, which implies that prices should be double-deflated allowing for differences in the cost of intermediate input. Suppose there are l different types of materials or services used in the production process. Then value added prices are given by:

$$P^V_i = \Sigma w_{ik} P_{ik} - \Sigma s_{il} P_{il}$$

where s are the shares of each type of input in the total value of intermediate input used in the production process.

In measuring relative productivity in equation [2], a primary problem is to measure the price ratios.[2] It is now well-recognised that the exchange rate is not an appropriate measure since it is influenced by short-term capital movements. It is difficult to decide which year's exchange rate to use. Using the exchange rate in two alternative benchmark years will generally imply growth rates very different from that derived using growth in output for each country deflated by price indices. Also, price ratios will vary across industries even in traded sectors due to different degrees of

monopoly power, lags in response to exchange rate movements, etc. In non-traded sectors, there is no reason to suppose price ratios are ever close to the exchange rate.

As outlined in the introduction, there is in practice a choice between two sources of information on relative prices. Purchasing power parities are measures of final expenditure prices. Unit value ratios are generally measures of the prices faced by producers. Both are gross output prices. In most productivity comparisons these price ratios are applied to sectoral value added rather than applying theoretically correct value-added prices.

There have been some attempts to incorporate materials prices in the calculation of relative prices, *e.g.* in Gersbach and van Ark (1994). In some manufactured products – for example, computers – inputs were priced at the market exchange rate since each country faced competitive input markets. In the case of computers this adjustment raised the German/US price ratio by about 10 per cent but had a much smaller impact on the Japanese/US price ratio. Use of the market exchange rate would not be appropriate for service sectors where the inputs consist largely of other services. In principle it should be possible, using input-output tables, to adjust gross output prices for material inputs – this is the approach advocated by Dale Jorgenson and his collaborators in their studies of international productivity. However in many sectors, material input prices are either unavailable or are crudely measured. Hill (1971) suggests that the use of double deflation when material input prices are measured with error may be more misleading than using single deflation.

This chapter highlights important measurement problems in constructing gross output prices which need to be addressed in international comparisons of productivity. Many of the problems addressed here apply equally to the measurement of the intermediate input component of value-added prices. Thus estimating the latter involves measurement problems squared. The remainder of this chapter will not specifically address the problem of single versus double deflation but this is not meant in any way to undermine its importance.

Before analysing properties of the conversion factors, it is useful to give some indication of how productivity results vary according to which conversion factor is employed. As an illustration consider the following table reproduced from Oulton (1994) for manufacturing, which uses unit value ratios constructed by van Ark (1993). In Table 1 the variation in productivity levels arises purely from differences in conversion factors. It is clear from Table 1 that the use of different conversion factors does alter the comparative productivity picture, in particular when comparing Japan with other industrial nations.

Table 1. Alternative comparative productivity estimates, benchmark years

Output per employee, United Kingdom = 100

Country	Benchmark year	Using 1985-based proxy PPP	Using 1990 based proxy PPP	Using unit value ratio
France	1984	107.9	101.1	117.0
Germany	1987	109.5	109.6	113.5
Japan	1987	108.3	119.4	159.6
United States	1987	174.8	174.3	186.7
United Kingdom	1984 or 1987	100.0	100.0	100.0

Note: The PPPs have been extrapolated to benchmark years using national GDP deflators.
Source: Oulton (1994), Table 7, p. 52.

Desirable properties of conversion factors

In attempting to evaluate the pros and cons of each type of conversion factor, it is useful to set out a list of criteria which they should satisfy. This chapter is not concerned with the extension to international price measurement of index number theory, which evaluates price indexes in the light of their consistency with economic theories of utility maximisation or cost minimisation (see for example the papers in Diewert, 1990). Rather it is concerned with the more mundane, but also more practical, problem of how the prices are measured. Index number theory generally incorporates the assumption that prices are measured without error. The discussion in this paper will generally not be concerned with theoretical properties of index numbers, except to the extent that they influence the choice of measurement method as discussed in the context of property five below.

This chapter identifies five measurement properties which I think the conversion factors should satisfy. These are:

- *Conceptually correct*. The price ratios should be measured so that they are compatible with the output data used in the calculation of relative productivity.

- *Complete coverage*. The price ratios should be based on a representative sample of goods or services for each sector for which productivity comparisons are estimated.

- *Intertemporally consistent*. The price ratios for different benchmark years should be consistent with the growth in relative prices in the intervening period implied by each country's domestic price indices.

- *Quality adjusted*. The conversion factors should correct for differences in the quality of goods and services across countries.

These four properties refer to the estimation of price ratios at a detailed level. In many productivity comparisons, the price ratios require aggregation to broad sectoral levels. In this context a fifth set of properties can be defined as:

- *Index number properties*. The conversion factors should satisfy certain well-regarded properties of index numbers, in particular transitivity. Other desirable properties include absence of substitution bias and additivity.

The following sections discuss each of these criteria and attempt to evaluate how well they are satisfied by both PPPs and UVRs.

Conceptually correct measures of relative prices

In equation [2] the price ratios should be compatible with the quantity of output being compared across countries. If the purpose of the comparison is to compare standards of living across countries, then final expenditure prices are appropriate. In sectoral productivity comparisons we are often concerned with prices faced by the producer in that sector rather than final consumption prices. The latter differ from producer prices in that they are affected by distribution margins, include imported goods and exclude exports and are measured at market prices, *i.e.* they include taxes and exclude subsidies.

The final expenditure PPPs are conceptually the correct prices only in sectors where the output being compared consists of final retail sales. Thus, they are perfectly reasonable as measures of price ratios in sectors such as the distributive trades, and have been used in this sector in a number of comparative studies such as Smith and Hitchens (1985), Mulder (1994) and O'Mahony (1995).[3] However, the PPPs are not generally conceptually correct in productivity comparisons if the outputs consist of products which are not final sales to consumers, *e.g.* in manufacturing.

It is possible in theory to correct the PPPs to take account of distribution margins, taxes and the degree of international involvement in the market. This is the approach pursued by Hooper (1996). The most serious estimation problem involves the adjustment for import and export prices since the required price and expenditure share data are not generally available. Jorgenson and Kuroda (1992) in their study of US and Japanese productivity correct for differences in distribution margins and net taxes but ignore the international trade element. Also the adjustments are dependent on the existence of reliable information on distribution margins and net taxes by sector.

In contrast the unit value ratios do correctly measure producer prices since they are calculated as the sales value of the products divided by the quantity sold. In this respect the UVRs would appear to be better measures of relative prices in those sectors where producer prices are required.

How important are these conceptual differences in conversion factors in practice? The chapter by Peter Hooper in this volume presents PPPs adjusted for the spread between expenditure and output prices and compares these to unit value ratios constructed by van Ark and Pilat. Their results show that the adjustments can be large, in particular, for Japan. However, the results also show that, far from narrowing these gaps, the adjustment increases the discrepancy between the PPP-based estimates and UVRs. Therefore, in practice, differences between manufacturing PPPs and UVRs are not primarily due to these conceptual differences.

Coverage of price ratios

When considering the coverage of the price ratios, there are essentially two separate issues: the extent to which the estimates cover all sectors of the aggregate economy; and the sampling of products within each sector.

Expenditure PPPs, based on final consumption prices, exclude all intermediate products. Unit value ratios, on the other hand, are capable of constructing price ratios for intermediate manufactured goods. In this respect the UVRs are a better measure for manufacturing. There are sectors of the economy which are covered only by expenditure prices, such as distribution, hotels and catering and personal and professional services. In these cases the output produced is a service to customers which is intangible and hence not amenable to the construction of unit values. There are also sectors which are covered by both price measures, such as transport and communications.

The PPPs are based on a large number of individual items; in 1990 about 2 500 goods and services were compared and were combined into over 200 basic headings which are the lowest disaggregation for which price ratios are made available to outside researchers. However, Eurostat or OECD do not attempt to quantify the proportion of expenditure covered by their sample of goods and services – a task which would be very difficult to implement. These sampled products are not equally spread across sectors, in fact only about 15 per cent of the basic headings refer to services. In final manufactured products it is likely that the coverage is reasonable and this is likely also to be the case in utilities and transport and communications. In other service sectors it is not clear if the coverage is representative. For example, in the case of the basic heading financial and insurance services, no

information is collected for the United Kingdom but a PPP is reported. This, in fact, has been based on some ICP rounds on wages in that sector and in most recent years on general consumer expenditure. It would be helpful if the number of items within each basic heading were published.

The nature of the UVR calculations, based on census information, allow the estimation of coverage percentages and hence are more transparent in this respect. The number of separate items for which UVRs are constructed is relatively small, generally amounting to a few hundred in manufacturing, and covers only about 20 per cent of manufactured output (see van Ark, 1993). In productivity level calculations, it is assumed that this 20 per cent is representative of total manufacturing.[4] Within manufacturing there are some sectors, such as machinery, where the coverage percentages are very low or where no products are matched.

This low coverage of the unit value ratios in a few industries has prompted some authors to use a mixture of UVRs and PPPs. For example O'Mahony (1992) and Smith, Hitchens and Davies (1982) both used PPPs to measure price ratios in most of the engineering sectors. More recently OECD (1996*b*) used a mixture in their productivity level comparison which employed the OECD's STAN database (OECD, 1995), but in this case the PPPs were adjusted for net indirect taxes and distribution margins. This latter "pragmatic" approach uses the best available information and so, in terms of coverage, would seem to have an advantage over estimation using purely PPPs or purely UVRs.

Intertemporal consistency

The conversion factors should be reasonably consistent over time with trends implied by each countries' retail or producer price indices. Here there are two issues:v intertemporal consistency at the basic heading level; and at the broad sectoral or whole economy level.

It is useful to begin with definitions of intertemporal consistency. Suppose there are two sets of conversion factors associated with time periods t and s. Let p_t be the conversion factor for time t and $p_{s(t)}$ the conversion factor using period s price ratios multiplied by the changes in each countries retail price between period t and period s. Let $RP(p_t)$ be a bilateral productivity ratio using price ratios for time t. A productivity ratio is *strongly* intertemporally consistent if it satisfies:

$$RP(p_t) = RP(p_{s(t)}) + \varepsilon$$

where ε is some small number which can be positive or negative.

Often in productivity comparisons, the concern is with identifying the productivity leader since this is an important consideration in issues of convergence and catch-up of productivity levels. Therefore, we can suggest a second definition of intertemporal consistency. Let $RP(p_t)$ be a vector of productivity ratios comparing countries with some base country. Let $SP(p_t)$ be a vector of rank orderings of the productivity ratios so that the elements of $SP(p_t)$ consist of numbers in ascending order of relative productivities. Then relative productivities are weakly intertemporally consistent if:

$$SP(p_t) = SP(p_{s(t)})$$

i.e. the rank ordering of relative productivity is not altered by changes in the conversion factors.

It is well-known that the successive rounds of ICP PPPs are not intertemporally consistent at the aggregate economy level. Thus, Blades and Roberts (1987) highlighted the inconsistencies between the 1980 and 1985 GDP PPPs, where the latter showed the United States in a more favourable light

relative to other OECD countries. Maddison (1995) examines the implications for GDP levels of successive ICP rounds and shows that large discrepancies occur when comparing ICP 5, for 1985, with other rounds. In general both the 1980 and 1990 PPPs reduce the US GDP level advantage relative to that implied by the 1985 PPPs. Heston and Summers (1993) examine in detail the productivity and price levels for a number of countries for each five-year benchmark from 1970 to 1990 compared to the growth rates implied by national accounts and domestic price indices. They point to many anomalies; for example, they show that the 1990 benchmark implies GDP in Japan was 112 per cent of the United Kingdom whereas a national accounts extrapolation of the 1970 benchmark would imply that Japan's GDP was 124 per cent of the United Kingdom in 1990.

This section concentrates on analysing discrepancies between the two most recent benchmark years, 1990 and 1993. It will first consider differences in the PPPs by broad category and then consider in more detail differences at the seven-digit basic heading level and their implications for productivity comparisons in some sectors. These two years were chosen both because the results are published on a consistent basis using the EKS aggregation method and because the time span between the two benchmarks is shorter than previous benchmarks which serves to highlight the differences.

Table 2 shows the log growth rate in the dollar 1993 PPPs relative to the 1990 PPPs for six countries. The growth in GDP PPPs in general are close to the growth (relative to the United States) in domestic retail prices for most countries. In terms of sectoral productivity comparisons, the most relevant comparison is between the PPPs for household consumption and retail prices indices. Here there is considerably more variation.

Table 2. Growth in PPPs, 1990-93, selected countries

	ln (PPP_{1993}/PPP_{1990})					
	UK	GE	FR	IT	JA	CA
Household consumption	10.6	2.2	1.0	11.0	-5.2	-4.3
Food, drink & tobacco	5.8	0.6	-0.2	14.7	-3.8	-1.6
Clothing & footwear	-8.7	-7.2	-13.9	5.3	12.8	-4.3
Rent, fuel & power	14.5	0.0	15.3	24.1	-14.1	-6.7
Household equipment	23.6	7.4	1.6	12.1	14.1	8.1
Medical goods & services	10.2	-0.8	-11.3	-0.4	-4.0	0.9
Transport & communication	7.2	12.5	6.0	15.2	-1.3	1.7
Education and recreation	15.9	1.6	-2.0	10.4	5.8	-5.1
Miscellaneous goods & services	13.7	7.6	3.6	4.1	-18.3	-1.8
Government expenditure	18.0	2.5	-2.9	6.3	-8.8	2.3
Gross fixed capital formation	-19.5	-4.8	-5.4	-5.8	-3.6	-5.3
Construction	-28.6	0.0	-2.0	-2.5	-9.4	-6.9
Equipment	-8.4	-12.8	-9.4	-8.7	0.0	-4.7
Gross domestic product	6.2	1.1	0.0	8.2	-5.4	-2.6
Domestic retail prices	4.4	1.7	-1.4	5.8	-3.5	-0.9

Note: GE = Germany; FR = France; IT = Italy; JA= Japan; CA = Canada.

Source: PPPs: Purchasing Power Parities and Real Expenditures. EKS Volume 1, 1990, OECD, 1992 Paris, and Purchasing Power Parities and Real Expenditures. EKS Volume 1, 1993, OECD, 1995. Paris.

Retail prices: United States: "Consumer Price Index for all Urban Consumers", *Monthly Labour Review*, Bureau of Labour Statistics, Washington, DC; United Kingdom: "General Index of Retail Prices – Detailed Figures", *The Monthly Digest of Statistics*, CSO, London (*note*: mortgage interest payments are excluded); Germany: "Preisindex fuer die Lebenshaltung", *Statistiches Jahrbuch*, Statistisches Bundesamt, Wiesbaden; all other countries: "Consumer Price Index", *The National Institute Economic Review*, The National Institute of Economic and Social Research, London.

The results for the United Kingdom, and to a lesser extent for Italy, are not consistent with trends in domestic retail prices, whereas the results for the other four countries are reasonably close. Within the household expenditure category there are much larger changes, with rent, fuel and power being particularly volatile for most countries. The results in Table 2 do point to possible anomalies such as household equipment for the United Kingdom, and clothing and footwear for Japan.

Differences in PPPs across time can be affected by both changes in the weights and changes in the individual PPPs which underlie calculations at the most detailed basic heading level. Basic heading level PPPs are calculated as an average over a number of sampled products. At this level of disaggregation, price ratios are not weighted by expenditures, since the latter do not exist at such detail. However, the seven-digit categories are not calculated as a simple average over all sampled products within that category but involve a complex formula which gives higher weights to products which are deemed to form an important part of each country's consumption.

Details of the calculation of the PPPs at the basic heading level are outlined, together with examples, in OECD (1996*a*). Each country's statistical office chooses "starred" products which are deemed to be important in that country's consumption. Each country has at least one representative product which is priced in at least one other country. The first step is to calculate Laspeyres price ratios which only involve the "starred" products of the base country. Then Paasche price ratio are calculated using the "starred" products of the partner country and Fisher price ratios are calculated using the geometric mean of the Laspeyres and Paasche ratios. An adjustment is made to complete the Fisher matrix for cases where two countries do not have any common representative products. These Fisher ratios are not transitive, *i.e.* the price ratios for country A relative to B divided by the ratio of country B to C is not equal to the direct Fisher price ratio comparing country A with country C. The price ratios are made transitive by using the EKS (Eltetoe-Koeves-Szulc) method which involves the geometric means of both direct Fisher price ratios for two countries (say A and B) and indirect ratios involving all other countries (A compared to C divided by C compared to B, etc.). Thus the method involves all pairwise price comparisons but is designed to give a result close to the bilateral price ratios for representative products while at the same time maintaining transitivity.

It is interesting to examine the volatility of these basic heading PPPs. Table 3 shows summary statistics for the log growth of PPPs from 1990 to 1993 comparing both the United Kingdom and Germany with the United States. This was carried out for the 199 basic headings which were common to the two years' PPP calculations. Comparison of the mean PPPs suggests that prices in the United Kingdom relative to the Untied States in 1993 were on average about 3.6 per cent higher than were relative prices in the two countries in 1990. On average, the price rise in Germany was close to that in the United States. However, the standard deviation and the range show the extreme volatility of PPPs over that period. The coefficient of variation, measured as the standard deviation divided by the mean, is very large for both comparisons.

Table 3. Growth in PPPs, 1990-93

	UK£ per US$	German DM per US$
Mean	0.036	0.005
Standard deviation	0.425	0.377
Maximum	1.592	1.396
Minimum	-1.544	-1.570

The growth in PPPs in Table 3 need to be compared to movements in each country's domestic retail price indices. Table 4 shows the price rise for each country and summary statistics over individual items published by each country's statistical office. The coefficient of variation is

considerably lower in Table 4 than in Table 3. The range is high in Table 4 but the items with very large or small prices changes tend to be similar in the three countries. The number of items compared in the two tables differs significantly, with the PPPs based on about 200 items whereas the retail price indexes are based on about 30 items. Hence we should expect more variation in Table 3 than in Table 4, but the difference seems to be too large to be entirely explained by this.

Table 4. Growth in the retail price indices, 1990-93

	United States	United Kingdom	Germany
All retail items	0.0967	0.1404	0.1138
Mean	0.0876	0.1445	0.0924
Standard deviation	0.0613	0.0933	0.0401
Maximum	0.2294	0.3192	0.1736
Minimum	-0.528	-0.0911	0.0197

Source: See footnote to Table 2.

Table 5 shows the number of basic heading PPPs which fall within percentage bounds around the relative change in domestic retail prices. Thus, comparing the United States with the United Kingdom, only 44 items (or just over 20 per cent) fall within 10 per cent of the difference in general retail prices in the two countries. A slightly greater percentage (about 30 per cent) fall within these bounds comparing Germany with the United States. A disturbing feature of Table 5 is the significant percentages which fall outside the extreme bounds of plus or minus 50 per cent. Thus over 15 per cent of items imply price rises in the United Kingdom relative to the United States – either 50 per cent more or 50 per cent less than the relative change in general retail prices in these two countries.

Table 5. Number of PPPs whose growth rate between 1990-93 fall within specified bounds

	United Kingdom/United States			Germany/United States		
Bounds	within	>	<	within	>	<
> or < 10 per cent	44	69	86	61	64	74
> or < 20 per cent	83	54	62	118	34	37
> or < 25 per cent	113	40	46	134	29	36
> or < 30 per cent	132	33	34	147	27	25
> or < 40 per cent	153	22	23	162	17	20
> or < 50 per cent	164	17	18	171	14	14

Source: Unpublished data derived from Eurostat.

How important are these changes for productivity calculations? The fact that as many items in Table 5 are greater than as are less than the bounds might imply that these extreme movements cancel each other in the aggregate. However, in sectoral productivity comparisons we often need to resort to the use of fairly detailed PPPs combining only a small number of basic headings. One extreme anomaly has been reported by O'Mahony, Oulton and Vass (1996*a*) in their study of comparative productivity in market services. In the hotels and catering sector, the 1993 PPPs implied that productivity in both the United Kingdom and Germany was about 45 per cent of the US level, whereas the productivity level in France was about 90 per cent of the US level. Thus in this sector the United States was the productivity leader with levels in France coming close but still below those in the United States. If the 1990 PPPs, updated to 1993 using domestic retail price indices for this sector, are used with the same output and labour input data, then the productivity levels relative to the United States turn out to be 65 per cent, 67 per cent and 131 per cent, respectively, for the United Kingdom, Germany and France. The 1990 PPPs imply a much smaller US lead over the United Kingdom and Germany, a change which does not satisfy the strong definition of intertemporal consistency mentioned above. But, more disturbing, the weak definition of intertemporal consistency

is also not satisfied comparing the United States and France – the 1990 PPPs imply a substantial French productivity lead in hotels and catering whereas the 1993 PPPs implied a slight US lead.

The expenditure shares used to weight the basic heading PPPs to broad sectors also show enormous volatility across this three-year period. Thus the coefficient of variation is close to ten for Germany, equals 14 for the United Kingdom and eight for the United States. If prices are truly as volatile as suggested by Table 5, then it might also be the case that the expenditure shares would also be highly variable. However, a large fall in the price of one good should induce increased demand, dampening the impact on expenditure shares. It is difficult to check consistency of the shares and PPPs since the former are calculated for each country whereas the latter are price relatives. In some cases, such as motor fuel, both PPPs and expenditure shares show large changes from 1990-93. Also, there are cases, such as hotels, where the PPP shows a dramatic fall in the US price relative to Germany, France and the United Kingdom, but US expenditure shares have shown only a very small rise. On the other hand, some products, such as butter, show small changes in relative prices but very large changes in shares. It is likely, therefore, that extreme volatility at the basic heading level may be compounded by large changes in shares when aggregating to broader groups.

So far the chapter has considered in detail intertemporal problems with the PPPs but it is also necessary to ascertain whether the UVRs behave any better in this respect. It is beyond the scope of this chapter to consider in as much detail the consistency of UVRs, primarily because benchmark estimates are carried out much less frequently and often by different authors. However there is some evidence that the UVRs are more stable over time. O'Mahony (1992) in a study of Anglo-German productivity compared her UVR for total manufacturing in 1987 with that derived by Smith, Hitchens and Davies (1982) for 1968 where the latter was brought forward to 1987 using changes in each country's producer price index. O'Mahony's unit value ratio was estimated at DM 3.5/£ whereas updating the Smith *et al.* UVR implied a price ratio of DM 3.6/£, a remarkable consistency over such a long time span.

Further evidence in favour of the stability of the UVRs is given by Broadberry (1993) who finds that benchmark estimates in the post-war period comparing US manufacturing with the United Kingdom and Germany can be reasonably tracked using time series on nominal output deflated by the producer price index and employment to bring forward the 1950 benchmark comparison reported in Paige and Bombach (1959). van Ark (1993) also examines time series extrapolations using real output indices from the national accounts for manufacturing and finds less agreement between benchmark and extrapolated values, although in general these discrepancies are not very pronounced. Both Broadberry (1993) and van Ark (1993) find relatively large differences in their extrapolated productivity levels and that derived by Smith, Hitchens and Davies (1982) comparing manufacturing in the United States and the United Kingdom in 1968.

Turning back to the differences in the 1990 and 1993 results, it proved possible to compare the consistency of PPPs and UVRs in communications. The PPP is based on an expenditure-weighted average of postal services and telecommunications, whereas the UVR is based on a quantity-weighted average of the unit values of items posted and number of telephone access lines. Table 6 shows the productivity ratios using the two measures of relative prices both for 1993 and 1990 prices updated to 1993. Note that the same output and labour input data, derived from the 1993 national accounts in each country, underlie all four sets of estimates. The table shows that, in the communications example, the UVR-based productivity levels are less dependent on which year's conversion factors are employed than are the PPP-based estimates. In the latter the results again do not pass the weak intertemporal consistency test since they show the United Kingdom leading Germany using the 1993 PPPs but Germany ahead of the United Kingdom using the 1990 PPPs. However, the UVR based

estimates comparing the United Kingdom and Germany, though less variable than those based on PPPs, also show some degree of intertemporal consistency.

Table 6. Productivity levels in communications: A comparison of alternative estimates

	1993 price ratios		1990 price ratios	
	United States/United Kingdom	Germany/United Kingdom	United States/United Kingdom	Germany/United Kingdom
PPPs	251.5	86.5	285.3	108.3
UVRs	328.8	100.8	312.9	110.3

Source: See O'Mahony, Oulton and Vass (1996*b*).

To summarise this section, it has been shown that the PPPs tend to vary considerably over time and that this is likely to cause problems in relative productivity calculations, in particular, if these are carried out at a detailed sectoral level. Unit value ratios may also be subject to some intertemporal inconsistencies, but these are probably not as pronounced as those using PPPs. This raises the important question of why these differences arise.

Heston and Summers (1993) suggest they may result from terms-of-trade effects, reflecting temporary changes in the market exchange rate. For example, they argue: "When the dollar appreciates relative to the SDR as in 1985, then the effective exchange rate declines because costs or prices in the United States have not changed relative to those in other countries by as much as exchange rates. In the short run of a benchmark comparison, it will make the United States look affluent relative to production. The more favourable dollar will allow a number of prices in the United States to decline relative to previous levels and permit an increase in income". Therefore Heston and Summers argue that when the dollar per SDR is falling significantly (and the effective exchange rate is rising), the benchmark estimates produce a PPP that is low relative to that obtained from extrapolations. However, their argument is uni-directional and, hence, should show more items with falling rather than rising relative prices. The results in Table 5 suggest that large changes in relative prices go in both directions.

It is possible that the large changes in relative prices may reflect differences in the samples of products used in the various ICP rounds. This may be due to refinements over time in attempting to adjust for quality or a more rigorous attempt to exclude outliers. It may also reflect the procedure underlying the calculation of PPPs at the basic heading level which gives high weight to "starred" products which are deemed to be important by each country's statistical office. The inclusion of more countries or more resources devoted in each country to the sampling of products can increase the number of starred products and, hence, can potentially lead to large differences over time. Also changes over time in attempts to complete the price comparison matrix where no actual price information exists can cause problems – the example of financial services in the United Kingdom, referred to above, is a case in point.

It is easy to visualise how differences in procedures might arise by considering the following simple example. Suppose that in comparing fares for a particular type of transport, say airlines, initially the PPPs in all countries excluded all discounts. At some stage, one country, say the United States, decides that discounted fares represent a large share of the market so they alter their estimates of the prices of air transport. In a further round statisticians in other countries, realising that the US procedure is a more realistic estimate of the cost of air transport, decide to change their estimates to include discounts. In this simple example the PPP in US dollars for country j would show a large rise in the second round, but would fall back again in the third round. These large changes could occur

even if the initial guidelines were fairly specific, *e.g.* the price of a particular class of travel over a specified distance by a major carrier.

In this respect it is interesting to note that the UVRs tend to be based on the same sample of products over time. Thus, in the manufacturing study comparing the United Kingdom and Germany, the products used to calculate unit value ratios were very similar in the 1968 and 1987 comparisons. In transport and communications, the unit values are collected for quantities – such as passenger and fright tonne/km for different modes of transport, number of pieces of mail or number of telephone lines – which do not vary over time. The one exception is that the product sample used by Smith *et al.* comparing the United States and the United Kingdom employed some PPPs for machinery and vehicles which were not used by van Ark in his later study, thus explaining some of the inconsistency referred to above.

The use of a constant sample of products is not always desirable, since it does not take account of the appearance of new, previously unavailable products. On the other hand, constantly changing samples of products leads to large jumps in the price ratios which undermine the credibility of international productivity comparisons. An intermediate position, where new products are introduced but are chain linked to previous samples would seem to be more appropriate – this is one method employed in constructing changes in prices over time. It would be useful if international bodies could engage in some analysis of intertemporal inconsistencies in PPPs and some reasons why they might occur. This would be very helpful to researchers in deciding which year's PPP to use in productivity comparisons.

Quality problems in measuring relative prices

Problems of intertemporal consistency relate to the growth in relative prices over some time period, quality problems relate to the levels of the price ratios at any point in time. Both PPPs and UVRs can suffer from problems in correctly measuring quality, although this is likely to be more serious in the latter. The PPPs match a very large number of carefully specified products but there is a danger that the only possible match in some cases is to match domestically produced goods in one country with imports in another – thus incorporating quality bias in productivity comparisons. The UVRs match broad ranges of goods (such as tonnes of plain biscuits) where it is difficult to specify any quality dimension. However, a judicious examination by researchers to exclude any obvious outliers circumvents this problem to some extent.

Comprehensive information on the quality of the products sampled is not available but there have been a number of studies which have attempted to quantify this for selected goods. The most thorough analysis was that carried out by McKinsey (1993) and reported in Gersbach and van Ark (1994) and Gersbach and Baily (1996). In this study quality was related to the valuation of the product by users but only insofar as it was associated with characteristics of the product or the production process. Thus, price differences which were due to differences in tastes or advertising were not taken into account. Also, quality premiums were only measured in markets where two products had equal access, otherwise it is not possible to distinguish quality premiums and price mark-ups.

Employing the above definition of quality, Gersbach and van Ark (1994) quote a price premium of about 8 per cent for Japanese relative to US passenger cars in 1990 and about a 5 per cent premium of German over US cars. In beer, they argue that the quality is the same in the United States and Germany. The argument in favour of this somewhat surprising result was as follows: "The treatment of both the United States and Germany as equal in quality rests on the ability that in most cases US

and German breweries could mass produce beer in Germany, using large-scale production methods, that tastes like existing German beer usually brewed in smaller establishments". Finally, a machinery survey by McKinsey showed that part of German machine tools commanded a price premium of about 8 per cent over comparable US products in Germany, but in total metal-working this quality premium only amounted to about 1 to 2 per cent.

One further study on quality is worth mentioning. Mason, van Ark and Wagner (1994) examine the quality of biscuits in four countries: the United Kingdom, Germany, the Netherlands and France. Based on the observation that German producers operate at the higher quality end of the market, they adjust crude tonnage per worker-hour productivity levels (which are close to those derived using UVRs) to quality adjusted measures. Their biggest adjustment is to turn a UK advantage of nearly 20 per cent over Germany to a German advantage of 40 per cent over the United Kingdom.

These results suffer from a methodological flaw in that they treat price differences as pure quality premiums. In their analysis, there is no role for the possibility that the difference between the price at the highest and lowest quality range may reflect differences in tastes, price premiums due to more aggressive advertising, or price mark-ups due to less competitive markets for higher quality products. Therefore the estimate of quality differences in Mason, van Ark and Wagner is likely to be exaggerated. Nevertheless, the results do point to some need to take account of quality differences in constructing unit value ratios.

Index number properties

It is well known that the calculation of bilateral price ratios are generally not transitive. Thus, for example, the price ratio derived by matching products in the United Kingdom relative to the United States divided by the price ratio in Germany relative to the United States will not generally equal the price ratio in the United Kingdom relative to Germany. This can occur both because the sample of products matched differs in each bilateral comparison, and may also be due to differences in the weights used to aggregate price matches to broad sectoral levels.

To overcome this problem international bodies such as Eurostat and OECD use multilateral schemes to render their results transitive. The two most commonly used aggregation methods are the EKS and GK (Geary Khamis) methods. Both lead to transitive results, but the EKS method gives results close to that derived by bilateral Fisher price ratios, whereas the GK method can give results which fall outside the range covered by the Laspeyres and Paasche price ratios since it is influenced by the relative size of countries (see the discussion in van Ark, 1993).

The EKS method is more consistent with economic theory, since it is less affected by substitution bias (Dowrick and Quiggan, 1994). Substitution bias occurs because price ratios are constructed using fixed quantities with variable prices. But optimal quantities (derived from utility or profit maximisation) vary with changes in prices. Thus price comparisons using the Laspeyres formula bias productivity results in favour of the base country, and Paasche ratios bias the results in favour of the partner country. The Fisher index is the theoretically correct or "ideal" index in a bilateral comparison. Dowrick and Quiggan's index which applies Afriat's ideal multilateral index to international comparisons is the "ideal" multilateral index. In comparing their index with both the EKS and GK methods, they find that the EKS has a low degree of substitution bias, since it is designed to yield results closest to bilateral comparisons while maintaining transitivity. In this respect the EKS is preferable to the GK method.

On the other hand the GK method yields results which satisfy additivity, a property not satisfied by the EKS method. Additivity allows a researcher to add the rows or columns of productivity comparisons to yield a measure of aggregate industry or country productivity. For the EKS the sum of productivity levels at the sectoral levels will not generally give the same answer as the use of the aggregate economy-wide PPP.

One problem with constructing multilateral price ratios is that they change with the addition of each new country in the comparisons. To overcome this, Eurostat has developed block fixity within the European countries ensuring that the addition of new countries outside the EU does not alter rankings within the EU. This is achieved by constructing EKS multilateral price ratios for the block of EU countries and then combining this with non-EU countries.

The productivity comparisons based on UVRs are generally not transitive. Hence van Ark's comparison for the United Kingdom/United States and Germany/United States imply a German productivity advantage over the United Kingdom nearly 20 per cent above that derived by O'Mahony (1992) comparing Germany directly with the United Kingdom. These industry of origin studies are carried out at different times and often by different authors but this should not preclude the construction of multilateral price ratios. In this respect van Ark (1993) has experimented with various multilateralisation methods for the countries which have been carried out by the ICOP group at the University of Groningen. In particular, it would be useful if some chain linking could be built in so that comparisons among industrial countries are not overly affected by recent ICOP studies on Eastern Europe or by studies covering developing countries.

There are obviously a range of theoretical properties which are satisfied by one or both of the two aggregation methods outlined above. Also these two methods do not exhaust the possible aggregation methods. Well-known results in index number theory show that there is no globally superior index number in most applications. The choice of aggregation method depends on the particular application being considered. Most researchers would opt for transitivity, otherwise it is possible to get almost any result, which undermines the credibility of the measurement exercise. A researcher interested in say a comparison of the United Kingdom and Germany would not wish the result to be significantly influenced by the inclusion of a third country, say the United States, in the comparisons. To achieve this they may be willing to forego other desirable properties such as additivity.

Conclusion

To summarise, the use of ICP purchasing power parities in international studies of comparative productivity has an advantage in that they are based on large samples of carefully matched products and, hence, are likely to be more representative and less affected by quality problems. PPPs are also, conceptually, the correct measure in some sectors and they address more rigorously theoretical and practical index number problems. On the other hand, unit value ratios have a conceptual advantage in some sectors such as manufacturing and cover many product areas not covered by the PPPs. This chapter has devoted considerable length to the problem of intertemporal consistency. In this important respect, UVRs appear to be more reliable than PPPs.

This leaves open the question of which type of conversion factor is most appropriate in measuring relative productivity. Since both types of conversion factor have advantages and disadvantages, it would seem that the most sensible procedure would be to attempt to use a mixture of the two, ideally combined with some independent price comparisons, along the lines carried out by McKinsey, in particular for goods and services where the PPPs and UVRs give conflicting results.

NOTES

1. There is also a third type of conversion factor, used successfully by McKinsey (1993), which involves researchers undertaking their own price comparisons. This paper, however, only considers those conversion factors which have enjoyed widespread use.
2. There are also problems in deciding on the appropriate data to use to measure output and labour input which are often as important as the choice of conversion factor. For example in manufacturing relative productivity is sensitive to the choice of censuses of production or national accounts data on output and employment. The emphasis in this chapter on conversion factors is not meant in any way to diminish the importance of these measurement issues.
3. In practice these studies include adjustments for differences across countries in sales taxes.
4. It is not clear that this coverage percentage is any less than that for PPPs. The UVRs cover a wider range of goods, *e.g.* litres of soft drinks, whereas the PPPs are much more detailed , *e.g.* cans of coke of a particular brand and size.

REFERENCES

ARK, B., van (1993), "International Comparisons of Output and Productivity: Manufacturing Productivity Performance in Ten Countries from 1950 to 1990", Groningen Growth and Development Centre, Monograph series No. 1.

ARK, B., van (1990), "Manufacturing Productivity Levels in France and the UK", *National Institute Economic Review*, No. 133, August.

ARK, B., van (1993), "Comparative Productivity in British and American Manufacturing", *National Institute Economic Review*, No. 146, November.

BLADES, D.W. and D. ROBERTS (1987), "A Note on the New OECD Benchmark Purchasing Power Parities for 1985", *OECD Economic Studies*, No. 9, Autumn, OECD, Paris.

BROADBERRY, S.N. (1993), "Manufacturing and the Convergence Hypothesis: What the Long-run Data Show", *The Journal of Economic History*, Vol. 53, pp. 772-795.

DIEWERT, W.E. (ed.)(1990), *Price Level Measurement*, *Contributions to Economic Analysis*, No. 196, North Holland, Amsterdam.

DOWRICK, S. and J. QUIGGAN (1995), "True Measures of GDP and Convergence: A Non-parametric Analysis of Multilateral Bounds", Australian National University.

GERSBACH, H. and B. van ARK (1994), "Micro Foundations for International Productivity Comparisons", Research memorandum, No. 572, Institute of Economic Research, University of Groningen.

GERSBACH, H. and M.N. BAILY (1996), "Explanations of International Productivity Differences: Lessons from Manufacturing", in K.Wagner and B. van Ark (eds), *International Productivity Differences*, *Contributions to Economic Analysis, No. 233*, North Holland, Elsevier Press, Amsterdam.

HILL, T.P. (1971), *The Measurement of Real Product: A Theoretical and Empirical Analysis of Growth Rates, for Different Industries and Countries*, OECD, Paris.

HESTON, A. and R. SUMMERS (1991), "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-88", *Quarterly Journal of Economics*, May.

HESTON, A. and R. SUMMERS (1993), "What Can Be Learned from Successive ICP Benchmark Estimates", in A. Szirmai, B. van Ark and D. Pilat (eds), *Explaining Economic Growth: Essays in Honour of Angus Maddison*, North Holland, Amsterdam.

HOOPER, P. (1996), "Comparing Manufacturing Output Levels Among the Major Industrial Countries", included in this volume.

JORGENSON, D.W. and M. KURODA (1992), "Productivity and International Competitiveness in Japan and the United States", *Economic Studies Quarterly*, Vol. 43, December.

JORGENSON, D.W., G. GOLLOP and B. FRAUMENI (1987), *Productivity and U.S. Economic Growth*, Harvard University Press.

McKINSEY (1992), *Service Sector Productivity*, McKinsey Global Institute, Washington, DC.

McKINSEY (1993), *Manufacturing Productivity*, McKinsey Global Institute, Washington, DC.

MADDISON, A. (1995), *Monitoring the World Economy: 1890-1992*, OECD Development Center, Paris.

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

MASON, G., B. van ARK and K. WAGNER (1994), "Productivity, Product Quality and Workforce Skills: Food Processing in Four European Countries", *National Institute Economic Review*, February.

MULDER, N. (1994), "New Perspectives on Service Output and Productivity: A Comparison of French and US Productivity in Transport, Communications, Wholesale and Retail Trade", Research Memorandum 575 (GD-14), University of Groningen, The Netherlands.

O'MAHONY, M. (1992), "Productivity Levels in British and German Manufacturing Industry", *National Institute Economic Review*, No. 139, February.

O'MAHONY, M. (1995), "International Productivity in Market Services: The Distributive Trades", mimeo, National Institute of Economic and Social Research, London.

O'MAHONY, M., N. OULTON and J. VASS (1996*a*), "Productivity in Market Services: International Comparisons", mimeo, National Institute of Economic and Social Research, London.

O'MAHONY, M., N. OULTON and J. VASS (1996*b*), "Productivity in Market Services: Transport and Communications", mimeo, National Institute of Economic and Social Research, London.

OECD (1995), *The OECD STAN Database for Industrial Analysis, 1974-1993*, Paris.

OECD (1996*a*), *Eurostat – OECD PPP Programme 1996 – The Calculation and Aggregation of Parities*, Paris.

OECD (1996*b*), "Competition, Wages and Productivity", internal memorandum, Paris.

OULTON, N. (1994), 'Labour Productivity and Unit Labour Costs in Manufacturing: the UK and its Competitors', *National Institute Economic Review*, May.

OULTON, N. and M. O'MAHONY (1994), *Productivity and Growth: A Study of British Industry, 1954-1986*, Cambridge University Press, Cambridge.

PAIGE, D. and G. BOMBACH (1959), *A Comparison of National Output and Productivity*, OEEC, Paris.

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

SMITH, A.D., D.M.W.N. HITCHENS and S.W. DAVIES (1982), *International Industrial Productivity: A Comparison of Britain, America and Germany*, Cambridge University Press, Cambridge.

SMITH, A.D. and D.M.W.N. HITCHENS (1985), *Productivity in the Distributive Trades*, Cambridge University Press.

SUMMERS, R. and A. HESTON (1991), The Penn World Table (Mark 5), *Quarterly Journal of Economics*, Vol. 106, No. 2, pp. 327-368.