



UNSW
THE UNIVERSITY OF NEW SOUTH WALES

Faculty of
Business

**School of Economics
UNSW, Sydney 2052
Australia**

<http://www.economics.unsw.edu.au>

**Convergence or Divergence: How to Get the Answer
You Want**

Robert J. Hill

School of Economics Discussion Paper: 2007/06

The views expressed in this paper are those of the authors and do not necessarily reflect those of the School of Economic at UNSW.

ISSN 1323-8949

ISBN 978 0 7334 2438 0

Convergence or Divergence: How to Get the Answer You Want

Robert J. Hill

School of Economics

University of New South Wales

Sydney 2052, Australia

E-Mail: r.hill@unsw.edu.au

February 18, 2007

The construction of spatial benchmark estimates of differences in per capita income across countries is more complicated and more important than is generally recognized in the convergence literature. Using data from the International Comparisons Program (ICP) for the years 1980, 1985 and 1996, and national growth rate data from the IMF it is shown that the choice of multilateral price index method can significantly affect the results, even to the extent of sometimes switching the outcome from convergence to divergence or vice versa. I show how the substitution bias inherent in some methods, such as the Geary-Khamis method that underlies the Penn World Table (PWT), can be exploited to push the results either towards convergence or divergence. Even methods, such as EKS, that are free of substitution bias are often undermined by the poor quality of the underlying benchmark data. I show how the reliability of the results can be increased by computing multiple estimates of each benchmark by extrapolating from earlier or later benchmarks using national growth rate data and then averaging the results. In particular, over the period 1980 to 1985, the standard EKS method shows convergence while, after benchmark averaging, it shows divergence. Benchmark averaging on EKS also significantly reduces the amount of divergence between 1980 and 1996 as compared with the PWT method (Geary-Khamis). (*JEL* C43, E31, O47)

KEY WORDS: Sigma Convergence; Penn World Table; Multilateral Price Index; International Comparisons Program; Benchmark Averaging; Substitution Bias

1. Introduction

The question of whether per capita income levels are converging or diverging over time is an important one that has attracted considerable attention in recent years. Empirical studies on convergence are useful for testing and refining growth models, and for predicting the allocation of global income in the future. These findings have implications for poverty reduction programs, as well as for international relations and the environment.

It is generally accepted that comparisons of per capita income across countries should be made using purchasing power conversion rates and not market exchange rates.¹ The main source of such purchasing power parity (PPP) data at a global level is the Penn World Table (PWT) (see Summers and Heston 1991). Researchers in the field, however, rarely pause to consider the reliability of the PWT. One weakness of the PWT that has received some attention is its reliance on the Geary-Khamis method for computing PPPs (see Nuxoll 1994, Dowrick and Quiggin 1997, Hill 2000, Neary 2004, and Dowrick and Akmal 2005). These authors show how Geary-Khamis is subject to substitution bias. This causes it to systematically underestimate differences in per capita income across countries. Also, if there is any convergence (divergence) of price relatives across countries over time, Geary-Khamis will tend to underestimate (overestimate) the rate of convergence as well.

The PWT is constructed by combining detailed spatial benchmarks from the International Comparisons Program (ICP) and other sources such as the OECD with data on real national growth rates. Real national growth rates are used to fill in gaps in the data (i.e., to generate results for non-benchmark years). The three most recent ICP benchmark years are 1980, 1985 and 1996. In this paper I explore the sensitivity of the convergence results over the 1980-96 period to the way the 1980, 1985 and 1996 spatial

¹The International Panel on Climate Change (IPCC), for example, has been heavily criticized for using market exchange rates in the construction of its projections for carbon dioxide emissions (see Castles and Henderson 2004). Assumptions about future rates of convergence also have a critical impact on these projections.

benchmarks are constructed. Results are computed using market exchange rates and three of the best known multilateral PPP methods: Geary-Khamis, EKS (see Eltetö and Köves 1964, and Szulc 1964) and ECLAC (see Economic Commission for Latin America and the Caribbean 1978). I also consider the extent to which the results can be manipulated by exploiting the substitution bias inherent in these methods.

A further concern is the reliability of the raw data underlying the ICP benchmarks. This is because of the severe measurement problems encountered when comparing prices across countries from different continents with hugely varying average incomes and consumption baskets. I show how national growth rate data can be used to improve the quality of the ICP spatial benchmarks themselves. I achieve this by generating multiple estimates of each benchmark by extrapolating from one benchmark to the next and then averaging the results. Overall, my preferred method is the benchmark averaged version of EKS. It generates results that differ quite significantly both from standard EKS and the Geary-Khamis method used by the PWT.

2. σ -Convergence

The concept of convergence addressed here is σ -convergence. σ -convergence occurs when the dispersion of per capita income across countries or regions falls over time. Dispersion is usually measured by the standard deviation of the logarithms of per capita income (see Barro and Sala-i-Martin 1992). The reason for taking logarithms before computing the standard deviation is to ensure that the results are invariant to the units of measurement. For example, the same σ coefficient should be obtained irrespective of whether income is measured in dollars or thousands of dollars.²

Almost all comparisons across countries use either the PWT (see Summers and

²An alternative dispersion measure that is also invariant to rescaling of per capita income is the coefficient of variation (see de la Fuentes 1997). The coefficient of variation is defined as the ratio of the standard deviation to the mean. Interestingly, Dalgaard and Vastrup (2001) find that the standard deviation of the logs and the coefficient of variation generate quite different results for a sample of 121 countries over the period 1960 to 1988, with the former showing convergence and the latter showing divergence.

Heston 1991) or Maddison's (1987) data. The former is used far more than the latter because of its (until recently) greater coverage of countries (168 at the last count) and greater transparency.³ The PWT, currently covers the years 1950-2000.

Since the vast majority of convergence studies have focused on the PWT, I will do likewise here. I use data from the International Comparisons Program (ICP) – that underlies the PWT – to explore the sensitivity and biases of the σ estimates and the resulting convergence trend to the way the spatial benchmarks are constructed.

3. Constructing Spatial Benchmarks

3.1 *Bilateral Price Indexes*

Market exchange rates are unsuitable for converting currencies into the same units to allow comparisons of per capita income across countries for two reasons. First, exchange rates are typically volatile with short-term movements driven largely by speculative trading. Second, exchange rate comparisons systematically overestimate differences in per capita income across countries. This systematic bias can arise either because nontraded services are more labor intensive in poorer labor abundant countries (see Bhagwati 1984), or because productivity increases that have been focused predominantly on the tradable goods sector have driven up wages in both sectors and hence prices in the nontraded service sector in richer countries (see Balassa 1964 and Samuelson 1964). Either way the implication is that nontraded services are cheaper in poorer countries.

The alternative to market exchange rates is to compare the purchasing power of currencies directly. To explain how these price indexes are constructed, it is first necessary to introduce some notation. The set of time periods is indexed by $t = 1, \dots, T$, the set of countries by $k = 1, \dots, K$ and the set of commodity headings by $n = 1, \dots, N$.

³Maddison however has gradually expanded his data set. It now covers more countries than the PWT, and also goes back to 1950 (and even to 1820 for some countries) – see Maddison (2001). The data set can be downloaded from <http://www.ggdcc.net/maddison/>. Updates and further extensions can be found at the Groningen Growth and Development Centre website: <http://www.ggdcc.net/Dseries/totecon.html>.

The price and quantity data of commodity heading n for country k in period t are denoted, respectively, by p_{kt}^n and q_{kt}^n .

A distinction can be drawn between bilateral and multilateral price indexes. Let $P_{j_s,kt}^i$ and $Q_{j_s,kt}^i$ denote, respectively, bilateral price and quantity index comparisons between country j in time period s and country k in time period t . Three important bilateral formulas are Paasche, Laspeyres and Fisher. These indexes are defined below:

$$\text{Paasche : } P_{j_s,kt}^P = \frac{\sum_{n=1}^N p_{kt}^n q_{kt}^n}{\sum_{n=1}^N p_{j_s}^n q_{kt}^n} \quad Q_{j_s,kt}^P = \frac{\sum_{n=1}^N p_{kt}^n q_{kt}^n}{\sum_{n=1}^N p_{kt}^n q_{j_s}^n}, \quad (1)$$

$$\text{Laspeyres : } P_{j_s,kt}^L = \frac{\sum_{n=1}^N p_{kt}^n q_{j_s}^n}{\sum_{n=1}^N p_{j_s}^n q_{j_s}^n} \quad Q_{j_s,kt}^L = \frac{\sum_{n=1}^N p_{j_s}^n q_{kt}^n}{\sum_{n=1}^N p_{j_s}^n q_{j_s}^n}, \quad (2)$$

$$\text{Fisher : } P_{j_s,kt}^F = \sqrt{P_{j_s,kt}^P P_{j_s,kt}^L} \quad Q_{j_s,kt}^F = \sqrt{Q_{j_s,kt}^P Q_{j_s,kt}^L}. \quad (3)$$

Paasche and Laspeyres price indexes are subject to substitution bias since they compare the cost of buying the same basket of goods and services in two different country-periods. A Paasche price index underestimates changes in the price level, while Laspeyres overestimates changes.^{4,5} A Fisher price index, by contrast, is superlative (i.e., approximates the underlying cost-of-living index to the second order) and hence free of substitution bias (see Diewert 1976).⁶

3.2 Multilateral Price Indexes

Bilateral price indexes (including superlative indexes), however, are inconsistent in multilateral comparisons (i.e., $P_{j_s,kt} \times P_{kt,lu} \neq P_{j_s,lu}$). A price index formula is defined as multilateral if it is transitive. Let P_{j_s} and P_{kt} denote multilateral price indexes for country j in period s and country k in period t , respectively. Multilateral indexes can be expressed as follows:

$$P_{j_s,kt} = \frac{P_{kt}}{P_{j_s}}.$$

⁴Strictly speaking, Paasche and Laspeyres bound the same cost of living index only when preferences are homothetic.

⁵A Paasche quantity index likewise underestimates the change in real income while Laspeyres overestimates changes in real income.

⁶Again strictly speaking, one can only say that superlative indexes are free of substitution bias when preferences are homothetic.

A large number of multilateral formulas have been proposed in the price index literature (see Balk 1996, Hill 1997, and Diewert 1999 for surveys of this literature). Here I focus attention on three classes of multilateral methods.

Average-Price Methods

Average-price methods compare each country with an artificially constructed average country. By implication, the underlying structure of such methods is a star graph with an artificial average country at the center of the star, as depicted in Figure 1.

Insert Figure 1 Here

Most average-price methods use the Paasche price index formula to make each bilateral comparison in the star, with the artificial country as the base. In the context of a spatial comparison (i.e., for a fixed value of t), the price index of country k in time period t , P_{kt} , is calculated as follows:

$$P_{kt} = P_{Xt,kt}^P = \frac{\sum_{n=1}^N p_{kt}^n q_{kt}^n}{\sum_{n=1}^N p_{Xt}^n q_{kt}^n} \quad \text{for } k = 1, \dots, K, \quad (4)$$

where p_{Xt}^n denotes the price of commodity heading n in the artificially constructed average country in period t . As a result of using the Paasche formula, there is no need to define an average basket q_{Xt} . The most widely used average-price method is Geary-Khamis (see Geary 1958 and Khamis 1972), which underlies the PWT and has also been used to make comparisons across the OECD countries. The Geary-Khamis average-price vector, p_{Xt} , and Paasche price indexes, $P_{Xt,kt}^P$, are obtained by solving the system of $N + K$ simultaneous equations in (4) and (5).

$$p_{Xt}^n = \sum_{k=1}^K \left(\frac{q_{kt}^n}{\sum_{j=1}^K q_{jt}^n} \frac{p_{kt}^n}{P_{Xt,kt}^P} \right) \quad \text{for } n = 1, \dots, N \quad (5)$$

Defining the average price vector in this way ensures that total expenditure on each product heading is the same when measured at international prices (i.e., p_X) as when expenditures are converted into units of the same currency using the Paasche price indexes. This can be seen by moving the term $\sum_{j=1}^K q_{jt}^n$ from the righthand side to the lefthand side of (5).

Average-price methods that use the Paasche price index formula suffer from substitution bias which may seriously distort estimates of both per capita income differentials

at a point in time and convergence rates over time (see Nuxoll 1994, Dowrick and Quiggin 1997, Hill 2000, Neary, 2004, and Dowrick and Akmal 2005). This is because the price vector of the artificial country at the center of the star will not be equally representative of the prices faced by all of the countries in the comparison. The more different p_{kt} is from p_{Xt} , the greater will tend to be the downward bias on $P_{Xt,kt}^P$. What matters ultimately is the bias in the price indexes of the K countries relative to each other, rather than relative to the artificial country X . When compared with each other they cannot all have a downward bias. Overall, a particular country's price index will tend to have a downward bias if its price vector is more different than average from the average country's price vector. Given that, by construction, the Geary-Khamis average price vector tends to more closely approximate the price vectors of richer countries, it follows that Geary-Khamis has a systematic tendency to underestimate differences in per capita incomes across countries.

Average-Basket Methods

Average-basket methods also compare each country with an artificially constructed average country. Most average-basket methods use the Laspeyres price index formula to make each bilateral comparison in the star, with the artificial country as the base. In the context of a spatial comparison, the price index of country k in time period t , P_{kt} , is calculated as follows:

$$P_{kt} = P_{Xt,kt}^L = \frac{\sum_{n=1}^N p_{kt}^n q_{Xt}^n}{\sum_{n=1}^N p_{Xt}^n q_{Xt}^n} \quad \text{for } k = 1, \dots, K, \quad (6)$$

where p_{Xt}^n and q_{Xt}^n denote the price and quantity of commodity heading n in the artificially constructed average country in period t . In practice it is not necessary to define the average price vector p_{Xt} since the denominator of (6) cancels when we take the ratio P_{kt}/P_{jt} . That is,

$$\frac{P_{kt}}{P_{jt}} = \frac{\sum_{n=1}^N p_{kt}^n q_{Xt}^n}{\sum_{n=1}^N p_{Xt}^n q_{Xt}^n} \frac{\sum_{n=1}^N p_{Xt}^n q_{Xt}^n}{\sum_{n=1}^N p_{jt}^n q_{Xt}^n} = \frac{\sum_{n=1}^N p_{kt}^n q_{Xt}^n}{\sum_{n=1}^N p_{jt}^n q_{Xt}^n}.$$

The most widely used average-basket method is the ECLAC method, which has been used to make comparisons in Latin America (see ECLAC 1978 and Hill 1997).

The ECLAC average quantity vector is defined as follows:

$$q_{Xt}^n = \sum_{k=1}^K q_{kt}^n \quad \text{for } n = 1, \dots, N. \quad (7)$$

Average-basket methods that use the Laspeyres price index formula also suffer from substitution bias, although it now acts in the opposite direction. Given that the ECLAC average basket more closely approximates the baskets of richer countries, it follows that ECLAC has a systematic tendency to overestimate differences in per capita incomes across countries.

EKS-Type Methods

The third class, which includes EKS (Eltetö and Köves 1964, and Szulc 1964) and CCD (Caves, Christensen and Diewert 1982a), makes bilateral comparisons between all possible pairs of countries. However, to obtain an internally consistent set of multilateral price indexes, the bilateral price indexes must be transitivized using a formula first proposed by Gini (1931). Alternatively, EKS-type methods can be thought of as the combination of K star graphs, each of which has a different country at the center. The EKS-type price indexes are obtained by taking the geometric mean of the price indexes generated by these K star graphs. The price index of country k in time period t , P_{kt} , is calculated as follows:

$$P_{kt} = \prod_{j=1}^K [(P_{jt,kt})^{1/K}],$$

where $P_{jt,kt}$ denotes the result of a bilateral comparison between countries j and k in period t . The EKS and CCD methods use the Fisher and Törnqvist formulas respectively to make each bilateral comparison. The EKS method is free of substitution bias since it is constructed from superlative indexes.⁷ The EKS method is the preferred method of both the OECD and Eurostat (see OECD 2002).

⁷Again, the discussion of substitution bias in relation to average-price, average-basket and EKS-type methods relies on homothetic preferences, since in the nonhomothetic case Paasche and Laspeyres bound the cost of living index at different utility levels (see Samuleson and Swamy 1974), and the second order approximation results for superlative indexes no longer apply except in some special cases (see Caves, Christensen and Diewert 1982b).

3.3 *Extrapolating from One Benchmark to the Next*

Choosing the appropriate multilateral method for constructing spatial benchmarks is not the only decision that must be made. One must also decide how to extrapolate from one benchmark to the next. The empirical application considered later in the paper focuses on the period from 1980 to 1996. Suppose spatial benchmarks are available in 1980, 1985 and 1996 (the actual situation considered later). Alternative benchmarks for 1996 can be constructed by combining the 1980 spatial benchmark with growth rates of real per capita income for each country between 1980 and 1996, or the 1985 spatial benchmark with national growth rates between 1985 and 1996. It is also possible to extrapolate backwards. For example, the 1996 spatial benchmark could be combined with growth rates of real per capita income between 1980 and 1996 to obtain an alternative spatial benchmark for 1980. The construction of alternative benchmarks is illustrated in Figure 2. The ovals refer to multilateral spatial benchmarks computed using the Geary-Khamis, EKS or ECLAC methods. Each vertex represents a particular country in a particular year. An edge connecting two vertices denotes a bilateral comparison between them.

Insert Figure 2 Here

Alternative benchmarks are useful as a way of reducing the impact of errors in the data. These errors arise since it is difficult to match products across countries, particularly when the sample of countries is very diverse as is the case in ICP comparisons. This problem is compounded by the fact that national statistical offices devote far more effort to the measurement of real GDP than to international comparisons. By using all three ICP benchmarks (i.e., 1980, 1985 and 1996) and extrapolating from national growth rate data, two alternative benchmarks can be constructed for each year. By averaging over each original benchmark and its two alternatives, the impact of errors in the ICP data can be reduced thus generating more reliable overall benchmarks. This point is demonstrated in the next section.

4. Measuring Convergence using Benchmark Averaging

Given that ICP spatial benchmarks are available in 1980, 1985 and 1996, and that

real national growth rates are also available for each country, we are faced by a problem of overdeterminacy in a convergence study, as illustrated in Figure 3.

Insert Figure 3 Here

A problem of overdeterminacy arises because of the presence of cycles in the graph. For example, a comparison between France and Germany in 1996 can be made directly from the 1996 spatial benchmark, or via extrapolation from another benchmark. Each path will generate a different answer. The three possible paths are as follows:⁸

- (i) FRA96-GER96
- (ii) FRA96-FRA85, FRA85-GER85, GER85-GER96
- (iii) FRA96-FRA80, FRA80-GER80, GER80-GER96

One solution to this problem is to ignore the real national income data and use only path (i). This scenario corresponds to Figure 4.⁹ Given, however, that far more resources are invested in the computation of real national income than in the spatial benchmarks, this is a waste of useful information.

Insert Figure 4 Here

Using either paths (ii) or (iii) by themselves is also unattractive since we can probably have greater confidence in path (i), unless there is reason to suspect that the spatial benchmark in 1996 is less reliable than the 1980 or 1985 benchmarks.¹⁰

It should be possible to improve on all three answers by taking their geometric mean. Suppose spatial benchmarks are available in years r , s and t . The price index between countries j and k in period t is computed as follows:

$$\frac{P_{kt}^*}{P_{jt}^*} = \left[P_{jt,jr} \left(\frac{P_{kr}}{P_{jr}} \right) P_{kr,kt} \right]^{\lambda^t} \left[P_{jt,js} \left(\frac{P_{ks}}{P_{js}} \right) P_{ks,kt} \right]^{\mu^t} \left(\frac{P_{kt}}{P_{jt}} \right)^{1-\lambda^t-\mu^t}, \quad (8)$$

where λ^t , μ^t and $1 - \lambda^t - \mu^t$ denote the weights on years r , s and t .

⁸I do not consider paths via third countries.

⁹The choice of link country between spatial benchmarks in Figure 4 will have no effect on the estimated rate of convergence, although it will affect the price indexes between countries in different years (see Hill 2004).

¹⁰In the empirical section it turns out that the 1985 benchmark is more reliable and hence should be given more weight.

The ratio P_{kt}/P_{jt} refers to the estimate obtained from the transitive spatial benchmark in year t . $P_{ks,kt}$ is a bilateral (and hence intransitive) temporal price index for country k between years s and t .¹¹ It is important to check that the adjusted spatial benchmarks P_{kt}^*/P_{jt}^* are transitive (i.e., internally consistent). That is, it must be the case that $P_{kt}^*/P_{jt}^* \times P_{lt}^*/P_{kt}^* = P_{lt}^*/P_{jt}^*$. It can be verified that this is the case as long as the temporal price indexes satisfy the time reversal test (i.e., $P_{ks,kt} = 1/P_{kt,ks}$), which they will since only one set of growth rates are available for each country which are used to extrapolate forwards and backwards.

The corresponding adjusted spatial benchmarks for years r and s are computed as follows:

$$\frac{P_{kr}^*}{P_{jr}^*} = \left(\frac{P_{kr}}{P_{jr}}\right)^{\lambda^r} \left[P_{jr,js} \left(\frac{P_{ks}}{P_{js}}\right) P_{ks,kr} \right]^{\mu^r} \left[P_{jr,jt} \left(\frac{P_{kt}}{P_{jt}}\right) P_{kt,kr} \right]^{1-\lambda^r-\mu^r}, \quad (9)$$

$$\frac{P_{ks}^*}{P_{js}^*} = \left[P_{js,jr} \left(\frac{P_{kr}}{P_{jr}}\right) P_{kr,ks} \right]^{\lambda^s} \left(\frac{P_{ks}}{P_{js}}\right)^{\mu^s} \left[P_{js,jt} \left(\frac{P_{kt}}{P_{jt}}\right) P_{kt,ks} \right]^{1-\lambda^s-\mu^s}. \quad (10)$$

That averaging over all three spatial benchmarks will generate more accurate results can be demonstrated with an example. I assume that the spatial price indexes have the following error structures:

$$\text{Spatial : } \ln\left(\frac{P_{ku}}{P_{ju}}\right) = \alpha_{ju,ku} + \varepsilon_{ju,ku}, \quad \text{for } u = r, s, t, \quad (11)$$

where $\alpha_{ju,ku}$ is the true spatial price index (in logs) for a comparison between countries j and k in year u . I assume the focus of attention here is restricted to measurable price differences. For example, differences in black market prices in countries j and k are not captured in $\alpha_{ju,ku}$. Hence $\varepsilon_{ju,ku}$ is a random variable with the following properties:

$$E(\varepsilon_{ju,ku}) = 0 \quad \text{and} \quad \text{var}(\varepsilon_{ju,ku}) = (\sigma_{jk}^u)^2.$$

It follows from the log specification that $\sigma_{jk}^u = \sigma_{kj}^u$. The variance of the errors $(\sigma_{jk}^u)^2$ depends on three factors. The first factor is the accuracy of measurement of the underlying price and expenditure data in each country. The higher the expertise of the

¹¹Since it is not transitive, we cannot rewrite $P_{ks,kt}$ as P_{kt}/P_{ks} .

staff at the national statistical offices of countries j and k and the greater the resources at their disposal, the lower will be the measurement error and hence the variance. The second factor is the extent of mismatches of products across countries. That is, the variance depends on the extent to which countries j and k are outliers in the sample of countries in terms of the basket of goods and services consumed. For example, suppose country j is an outlier. This means that it is hard to find products purchased in country j that are also purchased in the other countries. It must be remembered that we are dealing here with multilateral comparisons that require all countries to supply expenditure data on the same list of basic heading aggregates (such as cereals, milk products, etc). Each country also supplies price data at a more disaggregated level. The ICP aggregates the price data to the basic heading level using the Country-Product-Dummy (CPD) method (see Summers 1973). While the use of price data below the basic heading level in combination with the CPD method is a sensible response to the mismatch problem, it does not eradicate it completely. Hence it follows that outlier countries will tend to generate higher variances. The third factor is formula spread. For countries that face very different relative prices, the results of a comparison are more sensitive to the choice of price index formula. Hence the comparison has a higher variance. In a bilateral context, this idea is captured by the spread between a Paasche and Laspeyres index. For example, this spread will almost certainly be larger in a comparison between France and Nigeria, than it will be in a comparison between France and Belgium, even when there is no measurement error or mismatches of products across countries.

The variance $(\sigma_{jk}^u)^2$ will also differ from one benchmark u to the next depending on the resources invested in each benchmark comparison and the sample of countries. In particular, the variance of each bilateral comparison will tend to be an increasing function of the heterogeneity of the set of countries. This is again due to the problem of matching at the elementary and basic heading level (the second factor above).

In general, the assumption that $E(\varepsilon_{ju,ku}) = 0$ requires that the multilateral method used for constructing the spatial benchmark is free of substitution bias, as will be the case if EKS is used.

In a similar manner, the temporal price indexes are assumed to have the following error structure:

$$\text{Temporal : } \ln P_{ku,kv} = \beta_{ku,kv} + \epsilon_{ku,kv}, \quad \text{for } u, v = r, s, t, \quad u \neq v, \quad (12)$$

where $\beta_{ku,kv}$ is the true temporal price index (in logs) for a comparison between years u and v in country k and $\epsilon_{ku,kv}$ is a random variable with the following properties:

$$E(\epsilon_{ku,kv}) = z_{uv}^k \quad \text{and} \quad \text{var}(\epsilon_{ku,kv}) = (\phi_{uv}^k)^2.$$

Again, it follows from the log specification that $\phi_{uv}^k = \phi_{vu}^k$. The expected error z_{uv}^k and variance $(\phi_{uv}^k)^2$ depend on the expertise and resources of the staff in the national statistical office of country k . The error also depends on which pair of time periods are being compared. In particular, the further apart u and v are, the larger the likely error and associated variance (this is the temporal equivalent of factor 3 above). In general, $E(\epsilon_{ku,kv}) < 0$ when $u < v$, because most countries compute their GDP deflators using the Paasche formula which has a downward substitution bias. I will assume that $z_{vu}^k = -z_{uv}^k$. This is because if $P_{uk,vk}$ is a Paasche index, then $P_{vk,uk}$ is a Laspeyres index, and the two biases should approximately offset each other. If instead a superlative index is used, then $E(\epsilon_{ku,kv}) = 0$.

It now follows from equations (8), (11) and (12) that

$$\begin{aligned} \ln \left(\frac{P_{kt}^*}{P_{jt}^*} \right) &= \lambda^t \left\{ \ln \left[P_{jt,jr} \left(\frac{P_{kr}}{P_{jr}} \right) P_{kr,kt} \right] \right\} + \mu^t \left\{ \ln \left[P_{jt,js} \left(\frac{P_{ks}}{P_{js}} \right) P_{ks,kt} \right] \right\} \\ &\quad + (1 - \lambda^t - \mu^t) \left\{ \ln \left(\frac{P_{kt}}{P_{jt}} \right) \right\} \\ &= \lambda^t (\beta_{jt,jr} + \epsilon_{jt,jr} + \alpha_{jr,kr} + \varepsilon_{jr,kr} + \beta_{kr,kt} + \epsilon_{kr,kt}) + \mu^t (\beta_{jt,js} + \epsilon_{jt,js} + \alpha_{js,ks} + \varepsilon_{js,ks} + \beta_{ks,kt} \\ &\quad + \epsilon_{ks,kt}) + (1 - \lambda^t - \mu^t) (\alpha_{jt,kt} + \varepsilon_{jt,kt}). \end{aligned} \quad (13)$$

The variances of $\ln(P_{kt}/P_{jt})$ and $\ln(P_{kt}^*/P_{jt}^*)$ can be compared. It follows immediately from (11) that

$$\text{var}[\ln(P_{kt}/P_{jt})] = (\sigma_{jk}^t)^2. \quad (14)$$

To compute $\text{var}[\ln(P_{kt}^*/P_{jt}^*)]$ it is first necessary to define the covariance matrix for the error terms. Each spatial benchmark is computed independently, as are the growth

rate data for each country. In other words, the error on a comparison between countries j and k in year r should be independent of the error in year t (remembering again that we are only considering measurable price differences). Similarly, the error in a comparison between years u and v in country j should be independent from the error in country k . The growth rate data and spatial benchmarks should also be independent. That is, even though it is true that the variance on both spatial and temporal price indexes will be higher for a country with an under-resourced national statistical office, there is no reason to expect the spatial and temporal errors to be either positively or negatively correlated. It follows that $\text{cov}[\varepsilon_{jr,kr}, \varepsilon_{jt,kt}] \approx \text{cov}[\varepsilon_{jt,jr}, \varepsilon_{kr,kt}] \approx \text{cov}[\varepsilon_{jr,kr}, \varepsilon_{kr,kt}] \approx 0$. The only case where the independence assumption is violated is when we compare the errors in two different temporal comparisons for the same country, e.g., $\text{cov}(\varepsilon_{jt,jr}, \varepsilon_{jt,ks})$. This covariance will tend to be positive, as will the corresponding covariance for the first benchmark r , i.e., $\text{cov}(\varepsilon_{jr,js}, \varepsilon_{jr,jt})$.

For the t benchmark case, I assume therefore that the covariance matrix takes the following form:

$$\Omega_{jk}^t = \begin{matrix} & \varepsilon_{jt,jr} & \varepsilon_{jr,kr} & \varepsilon_{kr,kt} & \varepsilon_{jt,js} & \varepsilon_{js,ks} & \varepsilon_{ks,kt} & \varepsilon_{jt,kt} \\ \begin{matrix} \varepsilon_{jt,jr} \\ \varepsilon_{jr,kr} \\ \varepsilon_{kr,kt} \\ \varepsilon_{jt,js} \\ \varepsilon_{js,ks} \\ \varepsilon_{ks,kt} \\ \varepsilon_{jt,kt} \end{matrix} & \left(\begin{array}{cccccccc} (\phi_{rt}^j)^2 & 0 & 0 & \text{cov}_{tr,ts}^j & 0 & 0 & 0 & 0 \\ 0 & (\sigma_{jk}^r)^2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (\phi_{rt}^k)^2 & 0 & 0 & 0 & \text{cov}_{rt,st}^k & 0 \\ \text{cov}_{tr,ts}^j & 0 & 0 & (\phi_{st}^j)^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (\sigma_{jk}^s)^2 & 0 & 0 & 0 \\ 0 & 0 & \text{cov}_{rt,st}^k & 0 & 0 & (\phi_{st}^k)^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & (\sigma_{jk}^t)^2 \end{array} \right) \end{matrix}.$$

It follows that

$$\text{var}[\ln(P_{kt}^*/P_{jt}^*)] = (w^t)^T \Omega_{jk}^t w^t, \quad (15)$$

where

$$(w^t)^T = (\lambda^t \quad \lambda^t \quad \lambda^t \quad \mu^t \quad \mu^t \quad \mu^t \quad 1 - \lambda^t - \mu^t).$$

Substituting for Ω_{jk}^t and w^t in (15) we obtain that

$$\begin{aligned} \text{var}[\ln(P_{kt}^*/P_{jt}^*)] &= (\lambda^t)^2[(\phi_{rt}^j)^2 + (\sigma_{jk}^r)^2 + (\phi_{rt}^k)^2] + (\mu^t)^2[(\phi_{st}^j)^2 + (\sigma_{jk}^s)^2 + (\phi_{st}^k)^2] \\ &\quad + (1 - \lambda^t - \mu^t)^2(\sigma_{jk}^t)^2 + 2\lambda^t\mu^t(\text{cov}_{tr,ts}^j + \text{cov}_{rt,st}^k). \end{aligned} \quad (16)$$

Given estimates for the parameters in the covariance matrix it would be possible to compute optimal weights. I do not pursue this path for two reasons. First, it is not possible to derive plausible estimates of these parameters from the available data. This is because the covariance matrix Ω_{jk}^t is unique to each pair of countries for a particular benchmark. When there are 38 countries and three benchmarks – the case considered in a later section – this translates to 2109 different covariance matrices. Second, the optimal weights will differ for each bilateral comparison. However, to maintain transitivity it is necessary that the same weights are applied to all bilateral comparisons. Hence even if it were possible to compute optimal bilateral weights, these would then need to be averaged and hence would no longer be strictly optimal.

The approach outlined above is related to the consistentization (or benchmark reconciliation) approach used by Summers and Heston (1988) in the Penn World Table. Summers and Heston adjust their spatial and temporal indexes using an errors in variables model. In the three benchmark case, the Summers and Heston model only requires the estimation of a single five-by-five covariance matrix, irrespective of the number of countries in the comparison, as compared with the 2109 seven-by-seven covariance matrices considered here. The problem with the Summers-Heston approach is that it assumes that the errors on all the spatial indexes in a particular benchmark are drawn from the same distribution, as are the errors on the temporal indexes. This assumption is problematic since some countries are more outliers than others in terms of their relative prices and expenditure patterns, and because the level of resources and expertise also differ significantly across national statistical offices.

A second problem with consistentization is that it requires the alteration of national growth rate data which is often unpopular with national statistical offices and some users. For this reason it was abandoned in version 6.1 of the PWT. This was an

unfortunate development for the convergence literature since consistentization almost certainly improved the reliability of the spatial results in the PWT. There is an inevitable conflict between temporal and spatial price indexes (see Hill 2004). By trying to improve the quality of both simultaneously, consistentization is forced to compromise on both. Benchmark averaging, by contrast, focuses exclusively on improving the reliability of the spatial benchmarks and hence does not distort the national growth rate data.

Given that it is not possible to compute optimal weights, I consider the properties of the benchmark averaging method when all benchmarks are given equal weight (i.e., we set $\lambda^t = \mu^t = 1/3$). The covariance terms in Ω_{jk}^t have the following bounds:

$$\text{cov}_{tr,ts}^j \leq \phi_{rt}^j \phi_{st}^j, \quad (17)$$

$$\text{cov}_{rt,st}^k \leq \phi_{rt}^k \phi_{st}^k. \quad (18)$$

If in addition it is assumed that the benchmarks are equally reliable (i.e., $\sigma_{jk}^r = \sigma_{jk}^s = \sigma_{jk}^t = \sigma_{jk}$), and making use of the covariance bounds (17) and (18), it follows that

$$\text{var}[\ln(P_{kt}^*/P_{jt}^*)] \leq \frac{(\sigma_{jk}^t)^2}{3} + \frac{(\phi_{rt}^j + \phi_{st}^j)^2}{9} + \frac{(\phi_{rt}^k + \phi_{st}^k)^2}{9}.$$

Combining this with the inequality

$$2\phi_{rt}^j \phi_{st}^j \leq (\phi_{rt}^j)^2 + (\phi_{st}^j)^2, \quad (19)$$

it follows that

$$\text{var}[\ln(P_{kt}^*/P_{jt}^*)] \leq \frac{(\sigma_{jk}^t)^2}{3} + \frac{2[(\phi_{rt}^j)^2 + (\phi_{st}^j)^2 + (\phi_{rt}^k)^2 + (\phi_{st}^k)^2]}{9}. \quad (20)$$

A comparison of (14) and (20) reveals that a sufficient condition for $\text{var}[\ln(P_{kt}^*/P_{jt}^*)] < \text{var}[\ln(P_{kt}/P_{jt})]$ is that

$$\frac{(\phi_{rt}^j)^2 + (\phi_{st}^j)^2 + (\phi_{rt}^k)^2 + (\phi_{st}^k)^2}{4} < \frac{3\sigma_{jk}^2}{4}.$$

That is, even under the worst case scenario where (17) and (18) hold with equality (i.e., there is perfect correlation between $\epsilon_{jt,jr}$ and ϵ_{jt,j_s} and between $\epsilon_{kr,kt}$ and $\epsilon_{ks,kt}$), and

(19) holding with equality, passive benchmark averaging (i.e., setting $\lambda^t = \mu^t = 1/3$) will increase the accuracy of the benchmarks as long as the average temporal variance $\bar{\phi}^2$ is less than three quarters the size of the spatial variance σ_{jk}^2 . This will almost certainly be the case. $\ln(P_{kt}^*/P_{jt}^*)$ will also be an unbiased estimator of $\ln(P_{kt}/P_{jt})$ since $E(\epsilon_{jt,jr}) + E(\epsilon_{kr,kt}) = 0$ and $E(\epsilon_{jt,js}) + E(\epsilon_{ks,kt}) = 0$.

The fact that the covariances $\text{cov}_{tr,ts}^j$ and $\text{cov}_{rt,st}^k$ are nonzero is a direct result of the GDP deflator being computed using the Paasche formula. If instead a superlative formula such as Fisher were used, these covariances should be approximately zero. Over time it is likely that other countries will follow the lead of the US and make the switch to a superlative index. Under such a scenario where $\text{cov}_{tr,ts}^j \approx \text{cov}_{rt,st}^k \approx 0$, $\text{var}[\ln(P_{kt}^*/P_{jt}^*)]$ reduces to the following:

$$\text{var}[\ln(P_{kt}^*/P_{jt}^*)] = \frac{\sigma_{jk}^2}{3} + \frac{[(\phi_{rt}^j)^2 + (\phi_{st}^j)^2 + (\phi_{rt}^k)^2 + (\phi_{st}^k)^2]}{9}. \quad (21)$$

A comparison of (14) and (21) now reveals that $\text{var}[\ln(P_{kt}^*/P_{jt}^*)] < \text{var}[\ln(P_{kt}/P_{jt})]$ as long as

$$\frac{(\phi_{rt}^j)^2 + (\phi_{st}^j)^2 + (\phi_{rt}^k)^2 + (\phi_{st}^k)^2}{4} < \frac{3\sigma_{jk}^2}{2}.$$

In this case, passive benchmark averaging will be preferable to not using any benchmark averaging even when the average variance on the temporal price indexes is up to 50 percent larger than the average spatial variance.

Suppose now that every bilateral spatial comparison in 1980, 1985 and 1996 is made using benchmark averaging. This method can be represented graphically. Figure 2 shows how each of the three spatial benchmarks can be extrapolated to cover the whole 1980-1996 period. The overall results are obtained by taking the geometric mean of the results generated by these three graphs. The attraction of this type of method is that it makes full use of all the available data. Assuming that EKS is used to compute the spatial benchmarks, it can also be viewed as a natural extension of EKS, which as was noted earlier computes a spatial benchmark by putting each country in turn at the center of a star and then takes a geometric average of these results. Benchmark averaging extrapolates using each spatial benchmark in turn as the reference, and then

takes a geometric average.

The benchmark-averaging method can also be derived as the solution to a least-squares minimization problem. The minimization problem is described below for the three benchmark case, with the benchmark years denoted by r , s and t .¹² Consider first the optimization problem for the adjusted benchmark for year t .

$$\begin{aligned} \min_{\ln(P_{kt}^*/P_{jt}^*)} & \left\{ \lambda^t \left[\ln \left(\frac{P_{kt}^*}{P_{jt}^*} \right) - \ln P_{jt,kt}^r \right]^2 + \mu^t \left[\ln \left(\frac{P_{kt}^*}{P_{jt}^*} \right) - \ln P_{jt,kt}^s \right]^2 \right. \\ & \left. + (1 - \lambda^t - \mu^t) \left[\ln \left(\frac{P_{kt}^*}{P_{jt}^*} \right) - \ln \left(\frac{P_{kt}}{P_{jt}} \right) \right]^2 \right\}, \end{aligned} \quad (22)$$

where

$$P_{jt,kt}^r = P_{jt,jr} \left(\frac{P_{kr}}{P_{jr}} \right) P_{kr,kt}, \quad P_{jt,kt}^s = P_{jt,js} \left(\frac{P_{ks}}{P_{js}} \right) P_{ks,kt}.$$

Solving this problem treating P_{kr}/P_{jr} , P_{ks}/P_{js} and all the temporal indexes as given, P_{kt}^*/P_{jt}^* as defined in equation (8) emerges as the solution. The corresponding optimization problem for year r is as follows:

$$\begin{aligned} \min_{\ln(P_{kr}^*/P_{jr}^*)} & \left\{ \lambda^r \left[\ln \left(\frac{P_{kr}^*}{P_{jr}^*} \right) - \ln \left(\frac{P_{kr}}{P_{jr}} \right) \right]^2 + \mu^r \left[\ln \left(\frac{P_{kr}^*}{P_{jr}^*} \right) - \ln P_{jr,kr}^s \right]^2 \right. \\ & \left. + (1 - \lambda^r - \mu^r) \left[\ln \left(\frac{P_{kr}^*}{P_{jr}^*} \right) - \ln P_{jr,kr}^t \right]^2 \right\}, \end{aligned} \quad (23)$$

where

$$P_{jr,kr}^s = P_{jr,js} \left(\frac{P_{ks}}{P_{js}} \right) P_{ks,kr}, \quad P_{jr,kr}^t = P_{jr,jt} \left(\frac{P_{kt}}{P_{jt}} \right) P_{kt,kr}.$$

Solving this problem now treating P_{ks}/P_{js} , P_{kt}/P_{jt} and all the temporal indexes as given, the solution is P_{kr}^*/P_{jr}^* as defined in equation (9). Finally, for year s , the optimization problem is

$$\begin{aligned} \min_{\ln(P_{ks}^*/P_{js}^*)} & \left\{ \lambda^s \left[\ln \left(\frac{P_{ks}^*}{P_{js}^*} \right) - \ln P_{js,ks}^r \right]^2 + \mu^s \left[\ln \left(\frac{P_{ks}^*}{P_{js}^*} \right) - \ln \left(\frac{P_{ks}}{P_{js}} \right) \right]^2 \right. \\ & \left. + (1 - \lambda^s - \mu^s) \left[\ln \left(\frac{P_{ks}^*}{P_{js}^*} \right) - \ln P_{js,ks}^t \right]^2 \right\}, \end{aligned} \quad (24)$$

¹²The problem generalizes in a straightforward manner to the case of K benchmarks.

where

$$P_{js,ks}^r = P_{js,jr} \left(\frac{P_{kr}}{P_{jr}} \right) P_{kr,ks}, \quad P_{js,ks}^t = P_{js,jt} \left(\frac{P_{kt}}{P_{jt}} \right) P_{kt,ks}.$$

Treating P_{kr}/P_{jr} , P_{kt}/P_{jt} and all the temporal indexes as given, the solution is P_{ks}^*/P_{js}^* as defined in equation (10).

The benchmark-averaging method, therefore, minimizes the least squares difference in logarithms of each bilateral spatial comparison from its original estimate, P_{ku}/P_{ju} , and the estimates extrapolated via the other two benchmarks.¹³

One problem with benchmark averaging is that it violates temporal fixity (see Hill 2004). That is, the results for all available benchmarks will change when a new benchmark becomes available. This means that benchmark averaging as described thus far is probably not suitable for constructing data sets such as the PWT, where users do not appreciate retrospective revisions of the data. This problem, however, does not arise for a researcher attempting to determine whether per capita incomes converged or diverged over a particular time interval, since she does not have to worry about how the appearance of a new benchmark, after the project has been completed, will change the results.

In cases where a violation of temporal fixity is deemed a problem, a slightly different approach is required. Temporal fixity can be imposed by averaging only over spatial benchmarks chronologically preceding the spatial benchmark of interest. In the three spatial benchmark case discussed above, this means that no adjustment would be made to the earliest benchmark (1980). An adjusted benchmark for 1985 is constructed by averaging over the 1985 benchmark and an alternative benchmark extrapolated from 1980, while the 1996 adjusted benchmark is constructed by averaging over the 1996 benchmark and alternative benchmarks extrapolated from 1980 and 1985. It follows that when a new benchmark appears for 2006 (as seems likely), the adjusted benchmarks for 1980, 1985 and 1996 are unaffected, while the adjusted benchmark for 2006 makes use of all four benchmarks. This method could be referred to as the temporally-fixed benchmark-averaging method.

¹³Eltető and Köves 1964, and Szulc 1964 derived an analogous result for the EKS method.

Although I have argued thus far that the temporal data (i.e., national real per capita growth rates) are more reliable than the spatial data, this does not mean that the temporal data are error free. In fact, the GDP deflator in most countries is computed using the Paasche price index formula. This implies that real GDP is measured using the Laspeyres quantity index. This situation is not ideal, since Paasche and Laspeyres indexes are subject to substitution bias. It would be preferable if all countries adopted the US approach and switched to using a superlative index such as Fisher or Törnqvist (and annual chaining) for computing the GDP deflator. The choice of base year is also important. The base year in most countries currently is 2000. This means that the formula for computing the change in real GDP between 1980 and 1996 is actually the ratio of two Paasche quantity indexes (since both years precede the base year), as shown below:

$$Q_{80,96} = \frac{\sum_{n=1}^N p_{00}q_{96}}{\sum_{n=1}^N p_{00}q_{80}} = \frac{\sum_{n=1}^N p_{00}q_{00} \sum_{n=1}^N p_{00}q_{96}}{\sum_{n=1}^N p_{00}q_{80} \sum_{n=1}^N p_{00}q_{00}} = \frac{Q_{80,00}^P}{Q_{96,00}^P}. \quad (25)$$

The quantity index $Q_{80,96}$ will tend to be too small since the downward bias on $Q_{80,00}^P$ will be bigger than on $Q_{96,00}^P$. This substitution bias could be a problem for extrapolation methods if either poorer or richer countries are more prone to it. This issue is explored further in the next section.

5. An Empirical Application using Data from the International Comparisons Program (ICP)

The PWT uses real GDP growth rates to extrapolate from one ICP benchmark to the next. Also, since the countries in each spatial benchmark often differ, some spatial extrapolation is also required. The ICP benchmark years are 1970, 1975, 1980, 1985 and 1996.¹⁴ Here I focus on the last three benchmarks. A total of 39 countries are present in all three benchmarks. Of these I delete Germany, since as a result of reunification it is not comparable before and after 1989. I compute spatial benchmarks for 1980, 1985 and 1996 using market exchange rates, and the Geary-Khamis, EKS and ECLAC

¹⁴The benchmark data can be downloaded from <http://pwt.econ.upenn.edu/Downloads/benchmark/benchmark.html>.

PPP methods. The total number of expenditure headings available in each year are 151 in 1980, 139 in 1985 and 31 in 1996. Examples of headings in 1980 and 1985 include “rice”, “flour and cereals”, “bread”, etc. In contrast, in 1996 the headings are more aggregated (e.g., “bread and cereals”). This difference in the level of aggregation may have an important impact on the results, since the substitution bias of Geary-Khamis and ECLAC tends to decrease the higher the level of aggregation of the data. In other words, these methods should be less prone to substitution bias in 1996 than in 1980 and 1985.

A further problem is that there are a few gaps in the data for certain countries. A missing expenditure heading can simply be set equal to zero. What are more problematic are missing prices. Simply setting a missing prices equal to zero will impart a downward bias to the price index for that country. I deal with this problem by using the Country-Product-Dummy (CPD) method to impute prices for missing headings (see Summers 1973 and Rao 2004) prior to the construction of any price indexes. The CPD method estimates a regression equation with price as the dependent variable and dummy variables for each heading and country. The imputed price for a particular heading in a particular country is equal to the product of the estimated parameters on the corresponding heading and country dummies.

The 38 countries considered here consist of 16 countries from Europe, 14 from Africa and 8 from the rest of the world.¹⁵ It must be emphasized that the choice of countries in the sample was driven purely by their presence in all three spatial benchmarks. The real income data in units of domestic currency for each country over the period 1980 to 1996 were obtained from the IMF Financial Statistics Yearbook available online at <http://ifs.apdi.net/imf>. Also, although the ICP provides population data, I chose to use population estimates from the IMF to convert the results in to per

¹⁵The full country list is as follows: United States, Canada, Japan, Korea, Hong Kong, the Philippines, Pakistan, Sri Lanka, Austria, Belgium, Denmark, Finland, France, Greece, Hungary, Ireland, Italy, Luxembourg, the Netherlands, Norway, Poland, Portugal, Spain, UK, Botswana, Cameroon, Kenya, Ivory Coast, Madagascar, Malawi, Mali, Morocco, Nigeria, Senegal, Tanzania, Tunisia, Zambia, Zimbabwe.

capita terms. This is because there are a few significant discrepancies between the two data sets. The biggest discrepancies arise for Botswana and Nigeria. The Botswana estimates differ by 16 percent in both 1980 and 1985. The Nigeria estimates differ by 24 percent in 1980 and by 27 percent in 1985.

I focus here on the case of σ convergence. I calculate the standard deviation of the logarithm of per capita income in 1980, 1985 and 1996 across the 38 countries in the sample for spatial benchmarks computed using 20 different methods. More precisely, I consider five different ways of extrapolating spatial benchmarks for four multilateral methods (market exchange rates, Geary-Khamis, EKS and ECLAC). The extrapolation methods considered are as follows:

- (i) No extrapolation (i.e., use only the three spatial benchmarks)
- (ii) Extrapolation from the 1980 spatial benchmark
- (iii) Extrapolation from the 1985 spatial benchmark
- (iv) Extrapolation from the 1996 spatial benchmark
- (v) Benchmark averaging ($\lambda^t = \mu^t = 1/3$).

The resulting σ coefficients for all 20 methods are shown in Table 1. The rankings of the σ coefficients across multilateral methods for each year for the “no extrapolation” case (i.e., type (i) methods) are exactly as expected. In all three years we observe the same ranking. The σ coefficient is always highest for market exchange rates, followed by ECLAC, EKS and Geary-Khamis, in that order. The result for exchange rates is explained by Bhagwati and Balassa-Samuelson. The order of the PPP results is attributable to the substitution bias inherent in the ECLAC and Geary-Khamis methods. That is, ECLAC systematically overestimates differences in per capita income levels across countries, while Geary-Khamis systematically underestimates differences.

Insert Table 1 Here

The substitution bias of average-price and average-basket methods can also be exploited to move the convergence results in a particular direction. The direction of bias in the results depends on whether price relatives have converged or diverged over the period of interest. Hill (2004) and Dowrick and Akmal (2005) both find evidence of

divergence in price relatives in recent years. If this is the case, it follows that the upward bias on the per capita incomes of poorer countries in a Geary-Khamis comparison may be rising over time, and hence that studies using the PWT may overestimate the rate of convergence. Conversely, the downward bias on the per capita incomes of poorer countries in an ECLAC comparison may be falling over time, and hence an ECLAC comparison may underestimate the rate of convergence.

Geary-Khamis and ECLAC, however, are not representative of all average-price and average-basket methods, respectively, with regard to the direction of the bias. An average-price method that gives greater weight to countries with lower incomes in the average price formula, will exhibit the same type of bias as the ECLAC method. Such a method is obtained for example by replacing the Geary-Khamis weighted arithmetic mean formula for the deflated prices $p_{kt}^n/P_{Xt,kt}^P$ in (5) with a weighted harmonic mean as shown below:

$$p_{Xt}^n = \left[\sum_{k=1}^K \left(\frac{q_{kt}^n}{\sum_{j=1}^K q_{jt}^n} \frac{p_{kt}^n}{P_{Xt,kt}^P} \right)^{-1} \right]^{-1} \quad \text{for } n = 1, \dots, N. \quad (26)$$

Similarly, an average-basket method that exhibits the same type of bias as the Geary-Khamis method is obtained by replacing the ECLAC arithmetic mean formula for the average basket in (7) with a harmonic mean as shown below:

$$q_{Xt}^n = \left[\sum_{k=1}^K (q_{kt}^n)^{-1} \right]^{-1} \quad \text{for } n = 1, \dots, N. \quad (27)$$

The results in Table 1 do not support the findings of Hill (2004) and Dowrick and Akmal (2005). This may be because the sample of countries and time horizon here differ significantly from those considered by these authors. If Dowrick and Akmal are correct, then the GK(i) σ coefficient should rise noticeably less than the EKS(i) σ coefficient between 1980 and 1996 (since EKS is not affected by substitution bias), and the ECLAC (i) σ coefficient should rise noticeably more than the EKS(i) σ coefficient. In fact, exactly the opposite is observed in Table 1. This suggests that price relatives may actually have converged somewhat over this period. I investigate this hypothesis below using a metric proposed in Hill (2004).

The similarity, J_{jk}^t , of the price vectors of two countries j and k in period t can be measured by the variance of the logarithm of the price relatives, p_{kt}^n/p_{jt}^n , across the set of goods $n = 1, \dots, N$. The variance is weighted by the average expenditure share of each commodity heading.

$$J_{jk}^t = \sum_{n=1}^N \left(\frac{s_{jt}^n + s_{kt}^n}{2} \right) \left[\ln \left(\frac{p_{kt}^n}{p_{jt}^n} \right) - \overline{\ln \left(\frac{p_{kt}^n}{p_{jt}^n} \right)} \right]^2, \quad (28)$$

$$\text{where } \overline{\ln \left(\frac{p_{kt}^m}{p_{jt}^m} \right)} = \sum_{m=1}^N \left[\left(\frac{s_{jt}^m + s_{kt}^m}{2} \right) \ln \left(\frac{p_{kt}^m}{p_{jt}^m} \right) \right],$$

and s_{kt}^n denotes the expenditure share of good n in country k in time period t . That is, $s_{kt}^n = p_{kt}^n q_{kt}^n / (\sum_{m=1}^N p_{kt}^m q_{kt}^m)$. The logarithmic transformation ensures that J_{jk}^t is symmetric (i.e., $J_{jk}^t = J_{kj}^t$).

An overall measure of similarity of relative prices across all the countries in the sample in a given year is obtained by taking the arithmetic mean of J_{jk}^t , denoted by $Av(J^t)$, across all pairs of countries.

$$Av(J^t) = \frac{1}{K(K-1)} \sum_{j=1}^K \sum_{k \neq j}^K J_{jk}^t$$

Using this metric, I obtain the following results: $Av(J^{1980}) = 0.5058$, $Av(J^{1985}) = 0.5977$, $Av(J^{1996}) = 0.4093$. That is, I find divergence in price relatives from 1980 to 1985, and convergence thereafter. Furthermore, the convergence after 1985 swamps the divergence that precedes it.

These findings are consistent with the results in Table 1. GK(i) will tend to underestimate convergence when relative prices are converging, and to overestimate convergence when relative prices are diverging. That is, the fall in the GK(i) σ from 1980 to 1985 is too large since relative prices diverged over this period, and its rise from 1985 to 1996 is too large since relative prices converged over this period. Conversely, ECLAC(i) will tend to overestimate convergence when relative prices are converging, and to underestimate convergence when relative prices are diverging.

Hence, clearly, an index compiler has some scope to manipulate the results either in favor of convergence or divergence simply by the choice of multilateral price index

formula. In particular, ECLAC(i), EKS(i) and GK(i) all agree that per capita incomes converged between 1980 and 1985, and then diverged between 1985 and 1996, as they did also between 1980 and 1996.

Once we broaden the comparison to include extrapolation methods, the possibilities become more intriguing. Now we have a total of 20 methods. While there is unanimity regarding the finding of divergence from 1980 to 1996 and from 1985 to 1996, there is now disagreement over what happened from 1980 to 1985. According to seven of the 20 methods, per capita incomes converged between 1980 and 1985, while according to 13 methods, incomes diverged.

Might this lack of consensus in the extrapolation results over the 1980 to 1985 period reflect a problem with the extrapolation methods? It is possible that substitution bias in the national GDP deflators may be distorting the convergence results. This could happen, for example, if poor countries are more prone to higher levels of substitution bias in their GDP deflators. Under such a scenario, it follows from (25) that σ in 1996 may be too high in ECLAC(ii), EKS(ii) and GK(ii) relative to in 1980 (since the substitution bias here is of the downward variety). Conversely, σ in 1980 may be too low in ECLAC(iv), EKS(iv) and GK(iv) relative to in 1996. In both cases, this may cause divergence to be overestimated. Furthermore, benchmark averaging will not eliminate this problem, since in this case all the extrapolation methods will be subject to the same bias. If rich countries are more prone to a higher level of substitution bias in their GDP deflators, then the same arguments apply in reverse. That is, now divergence may be underestimated.

Neither of these problems seems to be relevant here. From Table 1 it can be seen that ECLAC(ii) shows more divergence than ECLAC(i) between 1980 and 1996 while ECLAC(iv) does not. EKS(ii) shows more divergence than EKS(i), while EKS(iv) shows less. Finally, GK(ii) shows the same divergence as GK(i) while GK(iv) shows less. Hence there is no clear evidence that differences in the magnitude of the substitution bias in the GDP deflators across countries is distorting the convergence results generated by extrapolation methods.

Overall, my preferred method is benchmark averaging on EKS (i.e., $EKS(v)$), since it combines multiple spatial benchmarks that are free of substitution bias with GDP deflator data to produce more robust results. Inspection of Table 1 reveals that the use of benchmark averaging is enough to switch the result for the 1980-85 comparison from convergence to divergence for all three PPP methods. That is, $ECLAC(i)$, $EKS(i)$ and $GK(i)$ all show convergence, while $ECLAC(v)$, $EKS(v)$ and $GK(v)$ all show divergence. This example clearly demonstrates how benchmark averaging can impact on the results. Also, the rate of divergence over the 1980-96 period is much smaller under $EKS(v)$ than it is under the method used by the PWT (i.e., $GK(i)$). This finding further supports the claim that unadjusted Geary-Khamis overestimates the rate of divergence for this set of countries over the 1980-96 period. The rate of divergence is also significantly lower for $GK(v)$, suggesting that it might be sufficient to adopt benchmark averaging to correct this particular bias.

One must be cautious drawing general conclusions about convergence over the 1980-1996 period from the results provided in Table 1. The sample of countries is unusual and hardly representative of the world as a whole. The key insights of this paper are methodological in nature. I have shown how sensitive the results of convergence studies can be to the way the spatial benchmarks are constructed, how substitution bias can affect the results, and proposed a new method for improving their reliability. The empirical results are provided to illustrate the significance of these findings.

6. Conclusion

Comparing per capita income levels across countries is an inexact science. Studies of convergence are inherently sensitive to the way the spatial price indexes are constructed. It is important that researchers in the field face up to this and accept that the results cannot be as precise as they might wish. That being said, it is also important that spatial benchmarks are made as reliable as possible. A switch from Geary-Khamis to EKS would be a good start since it would eliminate substitution bias. The spatial benchmarks can be further improved by computing multiple estimates of each benchmark by extrapolating from earlier or later benchmarks using national growth rate data

and then averaging the results. This process, which I refer to as benchmark averaging, reduces the impact of errors in the data. Benchmark averaging on EKS generates results that differ significantly from both standard EKS results and the Geary-Khamis results that underlie the Penn World Table.

References

- Balk, B. M. (1996), "A Comparison of Ten Methods for Multilateral International Price and Volume Comparison," *Journal of Official Statistics* 12(2), 199-222.
- Balassa, B. (1964), "The Purchasing Power Parity Doctrine: A Reappraisal," *Journal of Political Economy* 72, 584-596.
- Barro, R. and X. Sala-i-Martin (1992), "Convergence," *Journal of Political Economy* 100(2), 223-251.
- Bhagwati, J. N. (1984), "Why Are Services Cheaper in the Poor Countries?" *Economic Journal* 94, 279-286.
- Castles, I. and D. Henderson (2004), "Economics, Emissions Scenarios and the Work of the IPCC," *Energy and Environment* 14(4), 415-435.
- Caves, D. W., L. R. Christensen and W. E. Diewert (1982a), "Multilateral Comparisons of Output, Input and Productivity using Superlative Index Numbers," *The Economic Journal* 92, 73-86.
- Caves, D. W., L. R. Christensen and W. E. Diewert (1982b), "The Economic Theory of Index Numbers and the Measurement of Input, Outputs, and Productivity," *Econometrica* 50(6), 1393-1414.
- Dalgaard, C. J. and J. Vastrup (2001), "On the Measurement of σ -Convergence," *Economics Letters* 70, 283-287.
- Diewert, W. E. (1976), "Exact and Superlative Index Numbers," *Journal of Econometrics* 4, 115-145.

- Diewert, W. E. (1999), "Axiomatic and Economic Approaches to International Comparisons," in Heston, A., and R. E. Lipsey (ed.), *International and Interarea Comparisons of Income, Output and Prices*, NBER, University of Chicago Press: Chicago, 13-87.
- Dowrick, S. and M. Akmal (2005), "Contradictory Trends in Global Income Inequality: A Tale of Two Biases," *Review of Income and Wealth* 51(2), 201-229.
- Dowrick, S. and J. Quiggin (1997), "True Measures of GDP and Convergence," *American Economic Review* 87, 41-64.
- Economic Commission for Latin America and the Caribbean (ECLAC) (1978), "Series Históricas del Crecimiento de América Latina," *Cuadernos Estadísticos de la CEPAL* 3, Santiago de Chile.
- Eltető, O. and P. Köves (1964), "On a Problem of Index Number Computation Relating to International Comparison," *Statisztikai Szemle* 42, 507-518.
- Fuentes, A. de la (1997), "The Empirics of Growth and Convergence," A Selective Review," *Journal of Economic Dynamics and Control* 21, 23-73.
- Geary, R. G. (1958), "A Note on the Comparison of Exchange Rates and PPPs between Countries," *Journal of the Royal Statistical Society, Series A*, 121, 97-99.
- Gini, C. (1931), "On the Circular Test of Index Numbers," *International Review of Statistics*, Vol.9, No.2, 3-25.
- Hill, R. J. (1997), "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities," *Review of Income and Wealth* 43(1), 49-69.
- Hill, R. J. (1999), "Comparing Price Levels Across Countries using Minimum Spanning Trees," *Review of Economics and Statistics* 81(1), 135-142.
- Hill, R. J. (2000), "Measuring Substitution Bias in International Comparisons Based on Additive Purchasing Power Parity Methods," *European Economic Review* 44, 145-162.

- Hill, R. J. (2004), "Constructing Price Indexes Across Space and Time: The Case of the European Union," *American Economic Review* 94(5), 1379-1410.
- Khamis, S. H. (1972), "A New System of Index Numbers for National and International Purposes," *Journal of the Royal Statistical Society, Series A*, 135, 96-121.
- Kravis, I. B., A. Heston, and R. Summers (1982), *World Product and Income: International Comparisons of Real Gross Product*. Published for the World Bank by Johns Hopkins University Press: Baltimore.
- Maddison, A. (1987), "Growth and Slowdown in Advanced Capitalist Economies: Techniques of Quantitative Assessment," *Journal of Economic Literature* 25, 649-698.
- Maddison, A. (2001), *The World Economy: Historical Statistics*. Paris: OECD.
- Neary, J. P. (2004), "Rationalizing the Penn World Table: True Multilateral Indices for International Comparisons of Real Income," *American Economic Review* 94(5), 1411-1428.
- Nuxoll, D. A. (1994), "Differences in Relative Prices and International Differences in Growth Rates," *American Economic Review* 84, 1423-1436.
- OECD (2002), *Purchasing Power Parities and Real Expenditures: 1999 Benchmark Year*, Paris: Organization for Economic Co-operation and Development.
- Rao, D. S. P. (2004), "The Country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power Parities in the ICP," Paper Presented at the SSHRC Conference on Index Numbers and Productivity, 30 June - 3 July 2004, Vancouver.
- Samuelson, P. A. (1964), "Theoretical Notes on Trade Problems," *Review of Economics and Statistics* 46, 145-154.
- Samuelson, P. A. and S. Swamy (1974), "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis," *American Economic Review* 64(4), 566-593.

- Summers, R. (1973), "International Price Comparisons Based upon Incomplete Data," *Review of Income and Wealth* 19(1), 1-16.
- Summers, R. and A. Heston (1988), "A New Set of International Comparisons of Real Product and Price Level Estimates for 130 Countries, 1950-1985," *Review of Income and Wealth* 34(1), 1-25.
- Summers, R. and A. Heston (1991), "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *Quarterly Journal of Economics* 106, 327-368.
- Szulc, B. (1964), "Indices for Multiregional Comparisons," *Przegląd Statystyczny* 3, *Statistical Review* 3, 239-254.

Table 1.– σ Coefficients for 38 Countries

| | 1980 | 1985 | 1996 |
|------------|-------------|-------------|-------------|
| ExR(i) | 1.4894 | 1.5177 | 1.8783 |
| ExR(ii) | 1.4894 | 1.5378 | 1.6366 |
| ExR(iii) | 1.4730 | 1.5177 | 1.6359 |
| ExR(iv) | 1.8589 | 1.7470 | 1.8783 |
| ExR(v) | 1.5974 | 1.5933 | 1.7103 |
| ECLAC(i) | 1.4534 | 1.4132 | 1.5620 |
| ECLAC(ii) | 1.4534 | 1.4884 | 1.6010 |
| ECLAC(iii) | 1.3841 | 1.4132 | 1.5279 |
| ECLAC(iv) | 1.4567 | 1.4744 | 1.5620 |
| ECLAC(v) | 1.4192 | 1.4458 | 1.5536 |
| EKS(i) | 1.3162 | 1.2482 | 1.4799 |
| EKS(ii) | 1.3162 | 1.3855 | 1.5037 |
| EKS(iii) | 1.2189 | 1.2482 | 1.3647 |
| EKS(iv) | 1.4347 | 1.4086 | 1.4799 |
| EKS(v) | 1.3087 | 1.3331 | 1.4376 |
| GK(i) | 1.1887 | 1.0946 | 1.4140 |
| GK(ii) | 1.1887 | 1.3034 | 1.4172 |
| GK(iii) | 1.0660 | 1.0946 | 1.2099 |
| GK(iv) | 1.3639 | 1.3590 | 1.4140 |
| GK(v) | 1.1869 | 1.2327 | 1.3312 |

FIGURE 1. — STAR GRAPH

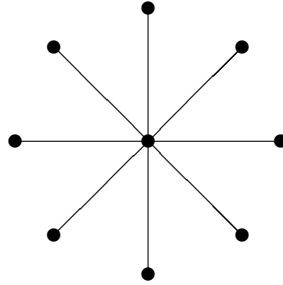


FIGURE 2. — EXTRAPOLATING FROM A SINGLE BENCHMARK

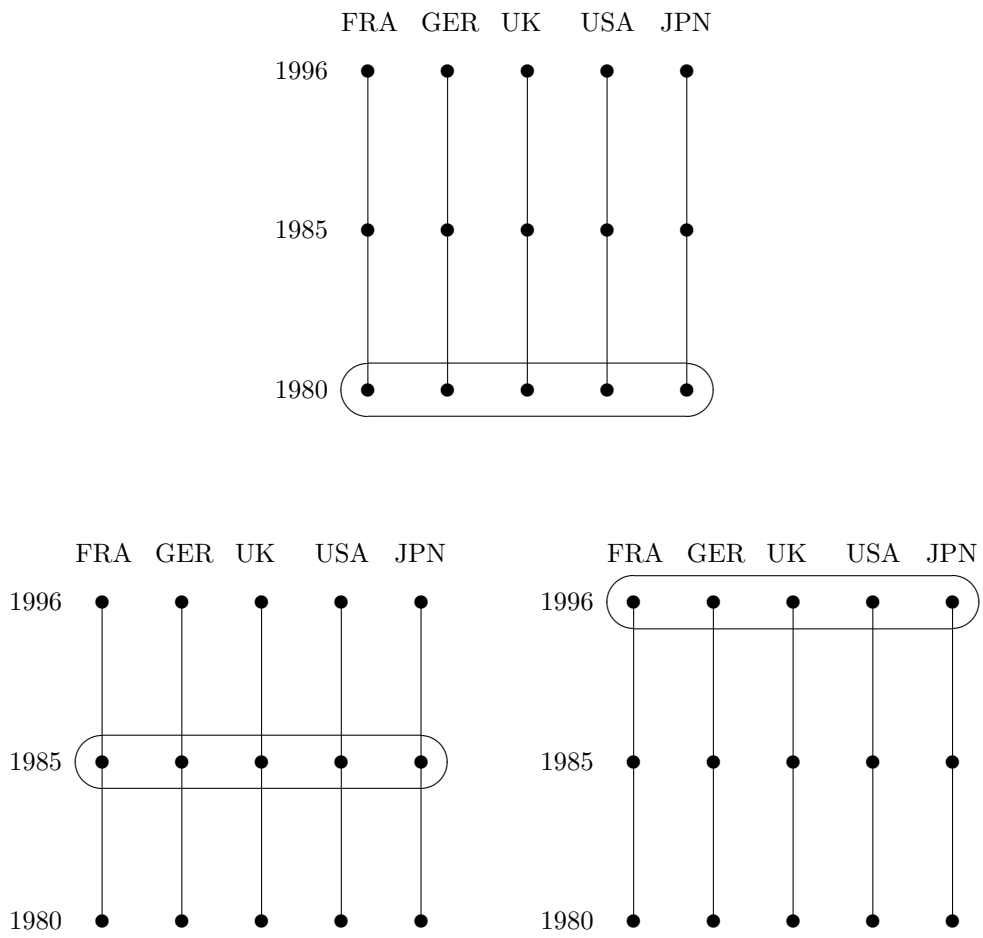


FIGURE 3. — THE INCONSISTENCY BETWEEN SPATIAL AND TEMPORAL PRICE INDEXES

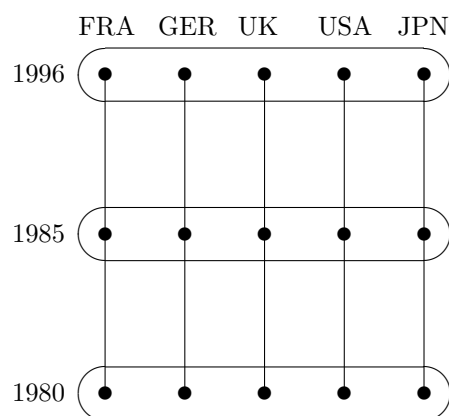


FIGURE 4. — USING EVERY SPATIAL BENCHMARK (LINKED THROUGH A SINGLE COUNTRY)

