Construction of World Tables of Purchasing Power Parities and Real Incomes Based on Multiple Benchmarks and Auxiliary Information: A State-Space Approach

D.S. Prasada Rao, Alicia Rambaldi (*) and Howard Doran
School of Economics
University of Queensland
Brisbane, Australia

Paper prepared for the 2008 World Congress on National Accounts and Economic Performance Measures for Nations

May 12–17, 2008

May, 2008

The authors gratefully acknowledge the support of the Australian Research Council (DP0557606) to pursue research on this topic. The authors are thankful to comments from the members of the Technical Advisory Group on ICP at the World Bank and also the participants of the EUROSTAT Workshop on Benchmarking held in Luxembourg in 2005. Expert research assistance from Renuka Ganegodage and Derek Headey is gratefully acknowledged.

(*) School of Economics, University of Queensland, Brisbane, Australia 4072. Email: a.rambaldi@economics.uq.edu.au

Construction of World Tables of Purchasing Power Parities and Real Incomes Based on Multiple Benchmarks and Auxiliary Information: A State-Space Approach

Abstract

The paper presents a new method for the construction of a consistent panel of Purchasing Power Parities (PPPs), and real incomes, using an econometric framework that combines available predictions of PPPs into a single model. The method improves upon the current practice used in the construction of the Penn World Tables, PWT, and similar tables produced by the World Bank as it integrates the various steps involved in the compilation of the Penn World Tables, makes use of all the PPP benchmark data from the various phases of the International Comparison Program, and ensures the model's prediction of the PPP for the reference country is identically one for all time periods. This general econometric model combines two sources of observations of PPP. The first and primary source of data is the ICP and the PPPs for benchmark years constructed by the ICP. The second are the PPP predictions from a model of the national price level (or exchange rate deviation index) which is able to produce predictions for all countries and years. The errors in the regression model are assumed to be spatially correlated across countries. These observations are combined using a state-space formulation so that the problem of obtaining optimum predictions of PPPs, from the available time series and cross sectional information, is treated as one of signal extraction. The smoothed PPP predictions (and standard errors) obtained through the state-space are produced for both ICP- participating and non-participating countries and non-benchmark years. Both sets of PPP observations (ie those based on model predictions as well as the ICP benchmarks) are assumed to suffer from measurement error. More developed countries are assumed to compile both national accounts data and ICP benchmark surveys with less measurement error. Therefore, the covariance structure of the state-space model is heteroskedastic and spatially correlated.

The objective of this paper is to present a number of analytical results to highlight some of the properties and flexibility of the approach. For example, we show that the resulting PPP predictions when the regression predictions are used only for non-benchmark countries in the benchmark years are weighted averages of past observations, and that the PPP estimates from earlier ICP benchmarks are weighted less than the more recent ones. We show how a series of constraints to the general model produce variants that: a) result in PPP predictions that deviate minimally from the available ICP's PPPs (benchmarks); or b) preserve the growth rates in price levels implicit in individual countries' national accounts data. A data set for 141 countries for the period 1970 to 2005 is used to illustrate the flexibility of the method proposed here and to produce a tableau of *PPP* estimates with standard errors for all countries over the period.

JEL Classification: C53, C33
Key words: Purchasing Power parities, Penn World Tables; State-space models; Spatial autocorrelation; Kalman Filter

1. Introduction

In a globalised world there is an ever increasing demand for internationally comparable data on major economic aggregates such as gross domestic product (GDP), private and government consumption, and gross fixed capital formation. Over the last four decades, there has been a consensus that market exchange rates are not suitable for converting economic aggregate data from different countries expressed in respective national currency units[1]. Instead, purchasing power parities (PPPs) of currencies which measure price level differences across countries are widely used for purposes of converting nominal aggregates into real terms.[2] PPP-converted real per capita incomes are used in influential publications like the World Development Indicators of the World Bank (World Bank, 2006 and other years) and the Human Development Report (UNDP 2006) which publishes values of the Human Development Index (HDI) for all countries in the world. The PPPs are also used in a variety of areas including: the study of global and regional inequality (Milanovic 2002); measurement of regional and global poverty using international poverty lines like $1/day and $2/day (regularly published in the Word Development Indicators, World Bank); the study of convergence and issues surrounding carbon emissions and climate change (Castles & Hendersen 2003; McKibbin & Stegman 2005); and in the study of catch-up and convergence in real incomes (Barro & Sala-i-Martin 2004; Durlauf, Johnson & Temple 2005; Sala-i-Martin 2002).

What are the main sources of PPP data? The only source for PPPs for the economy as a whole is the International comparison Program (ICP). The PPP data are compiled under the International Comparison Program which began as a major research project by Kravis and his associates at the University of Pennsylvania in 1968 and in more recent years being conducted under the auspices of the UN Statistical Commission. Due to the complex nature of the project and the underlying resource requirements, the project has been conducted roughly every five years beginning in 1970. The latest round of the ICP for the 2005 benchmark year has just been completed. The final results are available on the World Bank website:URL http://siteresources.worldbank.org/ICPINT/Resources/ICP_final-results.pdf. In the more recent years, beginning from early 1990's, the OECD and EUROSTAT have been compiling PPPs roughly every three years. The country coverage of the ICP in the past benchmarks has been limited with 64 countries participating in the 1993 benchmark comparisons. However this coverage has increased dramatically to 146 for the 2005 benchmark year. Details of the history of the ICP and its coverage are well documented in the recent report of the Asian Development Bank (http://adb.org/Documents/Reports/ICP-Purchasing-Power-Expenditures/default.asp).

Generally the coverage of countries in various ICP benchmarks has been limited[3]. However, international organizations such as the World Bank and the United Nations, as well as economists and researchers, seek PPP data for countries not covered by the ICP and also for the non-benchmark years. For most analytical and policy purposes there is a need for PPPs covering all the countries and a three

[1] For a detailed discussion of the issues relating to the use of exchange rates, the reader is referred to Kravis, I, Summers and Heston (1982) as well as the ICP Handbook available on the World Bank website. In addition the most recent publication from the Asian Development Bank on the 2005 comparisons in the Asia Pacific (http://adb.org/Documents/Reports/ICP-Purchasing-Power-Expenditures/default.asp) also provides an in-depth discussion on the use of exchange rates and purchasing power parities.

[2] Nominal values refer to aggregates expressed in national currency units, and, in contrast, real aggregates are obtained by converting nominal values using PPPs. These are termed "real' since the use of PPPs eliminates price level differences.

[3] A notable exception is the current 2005 Round of the ICP which has an impressive coverage of 146 countries. It covers the People's Republic of China for the first time and India participated in 2005 after its last participation in 1980.

to four-decade period[4]. The Penn World Tables has been the main source of such data. Summers and Heston are pioneers in this field. Summers and Heston (1991) provides a clear description of the construction of the earlier versions of the Penn World Tables. The most recent version, PWT 6.2, is available on URL: http://pwt.econ.upenn.edu, covers 170 countries and a period in excess of five decades starting from 1950. In addition to the PWT, there is the real gross domestic product (GDP) series constructed by Angus Maddison (Maddison 1995, 2007). The Maddison series is available on Groningen Growth and Development Centre website: www.ggdc.net/dseries/totecon.html and the series constructed by the World Bank. The Maddison series make use of a single benchmark and national growth rates to construct panel data of real GDP and no estimates are available for non-benchmark countries. The World Bank series are based on the methodology described in Ahmad (1996) and the series make use of a single benchmark year for which extrapolations to non-benchmark countries are derived using a regression-based approach. The benchmark and non-benchmark PPPs are extrapolated using national growth rates[5] in national prices.

The construction of PWT essentially uses a two-step method: (i) extrapolation of PPPs to non-benchmark countries in an ICP benchmark year using ICP benchmark data (normally from the most recent available exercise) and national level data through the use of cross-sectional regressions; and (ii) extrapolation to non-benchmark years. The second step combines the information from step (i) with GDP deflators from national accounts data, to produce the tables. Details of the PWT methodology can be found in Summers and Heston (1991) and Heston, Summers and Aten (2002).[6]

There are several important issues associated with the PWT methodology. First and foremost is the problem of time-space consistency of the data produced from different benchmarks. It is quite clear that a set of time-space comparisons can be derived using PPPs from just one benchmark and that such comparisons are not invariant to the choice of the benchmark data used. For example, use of 1990 benchmark data may result in one set of tables and the use of 1996 or 1999 may result in a very different set of tables of PPPs, real incomes and other aggregates. In solving this problem, the PPP data from the most recent benchmark comparison from the ICP is taken as the preferred starting point and the extrapolations across space and over time are derived using country-specific growth rates. This choice of a single benchmark to construct PWT means that a large body of data from other benchmarks are not utilised[7]. Even when attempts are made to make use of information from several benchmarks, no clear methodology for combining information from different benchmarks is currently available.

A related problem associated with the use of PWT and other available series is the absence of any measures of reliability such as standard errors. Most researchers using PWT data consider them to be similar to data from national accounts or other national or international sources. There is no general recognition that the data presented in the Penn World Tables are indeed based on predictions from regression models and that they are also projections over time. Thus the PWT data should be treated and used as predictions with appropriate standard errors. Though the PWT data provide an indication of

[4] For example, the Human Development Index is computed and published on an annual basis. Similarly, the World Development Indicators publication provides PPP converted real per capita incomes for all the countries in the world for every year.
[5] We define "national growth rates" in the next section.
[6] A description of the earlier attempts to construct panels of PPPs can be found in Summers and Heston (1988).
[7] Use is made of data from the earlier benchmark years for countries which are not in the latest benchmark but have participated in earlier benchmark comparisons.

the quality of data for different countries, there are no quantitative indicators of reliability in terms of confidence intervals for predictions.

The main objective of the paper is to propose a new method that adequately addresses problems associated with the PWT and other sources of extrapolated PPPs. In particular, the method proposed here will allow the use of data on PPPs from all the past benchmarks along with data available from national sources on price movements in the form national price deflators. The new method is designed to make efficient use of all the available information in obtaining optimal predictors of PPPs for all the countries and time periods. In addition, standard errors associated with the extrapolated PPPs can be derived using the approach suggested here.

The econometric model and the state-space formulation used are designed to generate predictions of PPPs over time and across countries that are broadly consistent with the benchmark data on PPPs and the observed country-specific temporal movements in prices. The model proposed is flexible enough for the user to place emphasis either on tracking benchmarks or tracking the observed national price movements accurately.

The paper also implements the econometric model and methodology on a data set that covers 141 countries over the period 1970 to 2005 including data from the most recently completed ICP benchmark comparison for the year 2005. A complete panel of PPPs for all the 141 countries and all the years is constructed using the predictions generated from the state-space model. Standard errors of the predicted PPPs are also presented.

The structure of the paper is as follows. Section 2 describes in detail all the sources of information used in the construction of the panel of PPPs. Section 3 presents an econometric formulation of the problem. Section 4 is devoted to a discussion of some of the special features of the proposed methodology. This section demonstrates the flexibility and generality of the model proposed in the paper. Section 5 describes the basic data used in the paper including a discussion of the benchmark data on PPPs and data on various socio-economic variables used in the regressions underlying the extrapolation. Section 6 presents results from the empirical implementation of the methodology proposed. Section 7 outlines the estimation procedure and the Kalman filter/smoother used in producing the predictions of PPPs. The paper is concluded with some remarks in Section 8. A set of appendices showing mathematical proofs of some of the analytical properties discussed in Section 4 are also included.

2. Principal sources of information for the Tableau

The econometric methodology proposed in the paper is designed to make optimal use of all the information available for the purpose of constructing a panel of PPPs. There are four principal sources of data available from national and international sources. First and foremost source is the data from the benchmark comparisons in the form of PPPs for the currencies of all the countries participating in various benchmarks of the ICP since its inception, i.e., from the first benchmark comparison in 1970 till to date. Due to differing degrees of participation of countries in different benchmarks and due to the fact that the benchmark comparisons are conducted roughly once in five years, we have an incomplete panel of PPPs.

The second source of information is a set of restrictions emanating from the concept of PPPs. In concept, the PPP of the currency of country, say India, with respect to the currency of a reference country, say the United States, is defined as the number of currency units of Indian rupees required to purchase the amount of goods and services purchased with one US dollar. Therefore, PPPs are always defined relative to the currency of a reference country. Hence, PPPs are determined only when the currency of a country is chosen as the *base* or *reference* currency. Therefore, by definition the PPP of the reference currency is always equal to *unity* in all periods. So, if country *k* is chosen as the base currency, then PPP_k is equal to 1 in all periods.

The third source of data is in the form of implicit GDP deflators which provide a measure of movements in prices in different countries over time. These deflators provide critical information on country-specific temporal movements in prices. The main source of data on deflators are the national accounts published by countries, generally on an annual basis. The fourth type of data is in the form of information on various socio-economic variables that are used in modeling national price levels or deviations of PPPs from the market exchange rates. These sources are more formally described below.

The variable of interest will be denoted by $p_{it} = \ln(PPP_{it})$ for country *i*= 1,. . . N and time *t* = 1, . . ., T where PPP_{it} represents the purchasing power parity of the currency of country *i* with respect to a reference country currency. Although it is directly unobservable, we can identify four noisy sources of information that can be combined to obtain an optimal prediction[8], p^*_{it}. They are: theory of national price levels used in predicting PPPs, derived growth rates in national prices that can be used in updating PPP information, PPPs from ICP benchmark exercises, and a constraint used for the reference country identification. We discuss each source in turn and formally develop an econometric model in the next section.

2.1 A model derived using the theory of national price levels

There is considerable literature focusing on the problem of explaining the *national price levels*. If ER_{it} denotes the exchange rate of currency of country *i* at time *t*, then the national price level for country *i* (also referred to as the *exchange rate deviation index*) is defined by the ratio:

$$R_{it} = \frac{PPP_{it}}{ER_{it}} \qquad (1)$$

For example, if the PPP and ER for Japan, with respect to one US dollar, are 155 and 80 yen respectively, then the price level in Japan is 1.94 indicating that prices in Japan are roughly double to that in the United States.

Most of the explanations of price levels are based on productivity differences in traded and non-traded goods across developed and developing countries. A value of this ratio greater than one implies national price levels in excess of international levels and *vice versa*. Much of the early literature explaining national price levels (Kravis & Lipsey 1983, 1986) has relied on the structural characteristics of countries such as the level of economic development, resource endowments, foreign trade ratios, education levels. More recent literature has focused on measures like openness of the economy, size of the service sector reflecting the size of the non-tradable sector and on the nature and extent of any barriers to free trade (Ahmad 1996; Bergstrand 1991, 1996; Clague 1988).

It has been found that for most developed countries the price levels are around unity and for most developing countries these ratios are usually well below unity. In general it is possible to identify a vector of regressor variables and postulate a regression relationship:

---

[8] We return to the optimality of the prediction in Section 3.3

$$r_{it} = \beta_{0it} + \mathbf{x}'_{it}\boldsymbol{\beta}_{sit} + u_{it} \tag{2}$$

where,

$r_{it} = \ln(PPP_{it} / ER_{it})$

$\mathbf{x}'_{it}$ a set of conditioning variables

β_{0it} intercept parameter

$\boldsymbol{\beta}_{sit}$ a vector of slope parameters

u_{it} a random disturbance with specific distributional characteristics.

Equation (2) is clearly not identified as it stands and identifying assumptions about the parameters will be made subsequently.

Provided estimates of β_{0it} and $\boldsymbol{\beta}_{sit}$ are available, model (2) can provide a prediction of the variable of interest consistent with price level theory.

$$\hat{p}_{it} = \hat{\beta}_{0it} + \mathbf{x}'_{it}\hat{\boldsymbol{\beta}}_{sit} + \ln(ER_{it}) \tag{3}$$

Thus, (3) states that price level theory provides a prediction, $\hat{p}_{it}$ of the variable of interest. We return to the estimation of β_{0it} and $\boldsymbol{\beta}_{sit}$ in Section 6.

2.2 The derived growth rates of PPPs

The movements in national price level, PPP_{it}/ER_{it}, can be measured through the gross domestic product deflator (or the GDP deflator) for period *t* relative to period *t-1* and through exchange rate movements. This is due to the fact that PPPs from the ICP refer to the whole GDP. GDP deflators are used to measure changes in PPP and the national price level. If the US dollar is used as the reference currency to measure PPPs and exchange rates, PPP of country *i* in period *t* can be expressed as:

$$PPP_{i,t} = PPP_{i,t-1} \times \frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}} \tag{4}$$

Equation (4) defines the growth rate of PPP_{it}.[9] GDP Deflators are computed from national accounts. The availability of resources to national statistical offices is likely to be positively related to the level of resources (technical and human) available in individual countries. Thus, we assume growth rates are measured with error. Taking logarithm of (4) and accounting for the measurement error:

$$p_{it} = p_{i,t\text{-}1} + c_{it} + \eta_{it} \tag{5}$$

where,

$$\mathrm{c}_{it} = \ln\left(\frac{GDPDef_{i,[t,t-1]}}{GDPDef_{US,[t,t-1]}}\right)$$

[9] Equation (4) simply updates PPPs using movements in the GDP deflator of the country concerned. Equation (4) would be a simple identity if PPPs were based on price of a single commodity. However in the case of PPPs at the GDP level, the same argument holds if GDP is treated as a composite commodity.

η_{it} is a random error accounting for measurement error in the growth rates

2.3 PPPs computed by the ICP for each benchmark year.

Due to the complexity in the design and collection of the ICP benchmark data (see Chapters 4-6 of the ICP Handbook which can be found on the World Bank ICP website: www.worldbank.org/data/ICP), the observed PPPs are likely to be contaminated with some measurement error. As the surveys for these benchmark exercises are conducted by national statistical offices, the argument made above in relation to measurement errors applies here also. Thus, ICP benchmark observations are given by

$$\tilde{p}_{it} = p_{it} + \xi_{it} \qquad (6)$$

where,
$\tilde{p}_{it}$ is the ICP benchmark observation for participating country *i* at time *t*
ξ_{it} is a random error accounting for measurement error and $E(\eta_{it}\xi_{it}) = 0$

2.4 Reference Country Definition

The definition of PPP requires a choice of reference country. The reference country is defined to have a PPP of one for all time periods.[10] Thus, we know the value of the variable of interest for the reference country for all time periods. As the USA is taken as the reference country, it then follows that for all *t*

$$p_{US,t} = 0 \qquad (7)$$

In the next section we provide an econometric model that is designed to take into account all the information described in this section.

3. Econometric formulation of the problem

The objective is to produce a panel of predictions of p_{it} (denoted by p^*_{it}) which optimally uses all relevant available data accompanied by standard errors, and is internally consistent in a sense to be defined subsequently.

As a matter of notation, for any quantity a_{it} we define the N-vector $\mathbf{a}_t$ as

$$\mathbf{a}_t = (a_{1t}, a_{2t}, \ldots, a_{Nt})'.$$

This notation will be used throughout without further definition. Matrices will be defined in upper case and bold face.

3.1 Assumptions

a) The errors u_{it} in the regression relationship (2) are assumed to be spatially correlated. We assume an error structure of the form

[10] PPPs between currencies of two countries are invariant to the choice of the base country. In the current study, we use US dollar as the reference currency which, in turn, gives equation (7). The method proposed here is invariant to the choice of the reference currency. This invariance result is available from the authors upon request.

$$\mathbf{u}_t = \phi \mathbf{W} \mathbf{u}_t + \mathbf{e}_t \qquad (8)$$

where $\phi < 1$ and $\mathbf{W}$ $(N \times N)$ is a spatial weights matrix. That is, its rows add up to one and the diagonal elements are zero.

It follows that $\mathrm{E}(\mathbf{u}_t \mathbf{u}_t')$ is proportional to $\mathbf{\Omega}$, where $\mathbf{\Omega} = \left[(\mathbf{I} \text{-} \phi \mathbf{W})(\mathbf{I} \text{-} \phi \mathbf{W})' \right]^{-1}$

b) The measurement errors in the observation of ln(PPP_{it}) during benchmark years, equation (6), are assumed spatially uncorrelated, but might be heteroskedastic. Thus, if ξ_{it} is a measurement error associated with country *i* at time *t*, then

$$\begin{aligned} E(\xi_{it}) &= 0 \\ E(\xi_{it}^2) &= \sigma_\xi^2 V_{it} \\ E(\xi_{it}^2 \xi_{jt}^2) &= 0 \qquad j \neq i \end{aligned} \qquad (9)$$

where σ_ξ^2 is a constant of proportionality[11].

c) The measurement error in the growth rates are assumed spatially uncorrelated, but might be heteroskedastic. Thus, η_{it} in (5) is assumed

$$\begin{aligned} E(\eta_{it}) &= 0 \\ E\left(\eta_{it}^2\right) &= \sigma_\eta^2 V_{it} \\ E(\eta_{it} \eta_{jt}) &= 0 \qquad j \neq i \end{aligned} \qquad (10)$$

where σ_η^2 is a constant of proportionality[12].

3.2 An Econometric Model

The econometric problem is one of signal extraction. That is, we need to combine all sources of "noisy" information and extract the signal from the noise. A state-space (SS) is a highly suitable representation for this type of problem. We start by extending equation (5) to define the 'transition equation' of the SS:

$$\mathbf{p}_t = \mathbf{p}_{t\text{-}1} + \mathbf{c}_t + \boldsymbol{\eta}_t \qquad (11)$$

where,

$\mathbf{c}_t$ is the observed growth rate of $\mathbf{p}_t$ (see equation (4) in Section 2.2)

$\boldsymbol{\eta}_t$ is an error with $E(\boldsymbol{\eta}_t)$=0 and $E\left(\boldsymbol{\eta}_t \boldsymbol{\eta}_t'\right) \equiv \mathbf{Q}_t = \sigma_\eta^2 \mathbf{V}_t$

Equation (11) simply updates PPPs in period *t-1* using the observed price changes over the period represented by $\mathbf{c_t}$.

---

[11] In the empirical section we model V_{it} as inversely related to GDP_{it}. This means that reliability of an observed PPP is lower for low-income countries.

[12] See footnote 2.

Also, as previously discussed, noisy observations of $\mathbf{p}_t$ are given by (3), a prediction from the regression model, and (6) a measurement by the ICP. Equations (2) and (3) relate the conditioning variables, $\mathbf{X}_t$, to the price level ratio. As we wish to relate the conditioning variables to the variable of interest, $\mathbf{p}_t$, re-writing of (2) and (3) to eliminate $\ln(ER_{it})$ is necessary.

From equation (2)

$$r_{it} = p_{it} - \ln(ER_{it}) = \boldsymbol{\beta}_{0it} + \mathbf{x}_{it}'\boldsymbol{\beta}_{sit} + u_{it}$$

and if $\hat{p}_{it}$ denotes the prediction of p_{it}, then

$$\hat{p}_{it} = p_{it} + (\hat{\beta}_{0it} - \beta_{0it}) + \mathbf{x}_{it}'(\hat{\boldsymbol{\beta}}_{sit} - \boldsymbol{\beta}_{sit}) - u_{it} \tag{12}$$

Throughout the paper we will reserve the symbol θ to represent the error in a current estimate of a parameter β.

Thus,

$$\hat{\theta}_{0it} = \hat{\beta}_{0it} - \beta_{0it} \text{ and } \hat{\boldsymbol{\theta}}_{sit} = \hat{\boldsymbol{\beta}}_{sit} - \boldsymbol{\beta}_{sit} \tag{13}$$

it is always possible to write equation (14) in the form

$$\hat{\mathbf{p}}_t = \mathbf{p}_t + \mathbf{X}_t\boldsymbol{\theta} + \mathbf{v}_t \tag{14}$$

where,

$\boldsymbol{\theta} = [\boldsymbol{\theta}_1{}',\ldots,\boldsymbol{\theta}_T{}']'$

$v_{it} = -u_{it}$

Because the explicit form of $\mathbf{X}_t$ depends on the particular identifying restrictions imposed on β_{0it} and $\boldsymbol{\beta}_{it}$, we will define it later in the context of a particular case.

Finally, in order to express these different observations as a single equation, it is convenient to define three 'selection matrices',

$\mathbf{S}_1 = [1, \mathbf{0}'_{N-1}]$ (selects the reference country $i = 1$)[13]

$\overline{\mathbf{S}}_1 = [0, \mathbf{I}_{N-1}]$ (selects countries $i = 2, 3, \ldots, N$)

$\mathbf{S}_t$, is a known $[N_t \times (N-1)]$ matrix which selects N_t participating countries (excluding the reference country) in the benchmark year t.

We are now able to consolidate these sources of information into a single equation on an 'observation vector' $\mathbf{y}_t$, viz

$$\mathbf{y}_t = \mathbf{Z}_t\mathbf{p}_t + \mathbf{B}_t\mathbf{X}_t\boldsymbol{\theta} + \boldsymbol{\xi}_t \tag{15}$$

with variables defined as follows:

[13] The selection matrix can be appropriately amended if a country other than country 1 is selected as the numeraire country.

i) Non-benchmark years:

$$\mathbf{y}_t = \begin{bmatrix} 0 \\ \overline{\mathbf{S}}_1 \hat{\mathbf{p}}_t \end{bmatrix}, \mathbf{Z}_t = \begin{bmatrix} \mathbf{S}_1 \\ \overline{\mathbf{S}}_1 \end{bmatrix}, \mathbf{B}_t = \begin{bmatrix} 0 \\ \overline{\mathbf{S}}_1 \end{bmatrix}, \zeta_t = \begin{bmatrix} 0 \\ \overline{\mathbf{S}}_1 \mathbf{v}_t \end{bmatrix} \tag{16}$$

$$\mathrm{E}(\zeta_t \zeta_t') \equiv \mathbf{H}_t = \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \sigma_u^2 \overline{\mathbf{S}}_1 \mathbf{\Omega} \overline{\mathbf{S}}_1' \end{bmatrix} \tag{17}$$

with σ_u^2 a constant of proportionality, and in (18) the countries are ordered so that the reference country is the first row[14]

ii) Benchmark years

$$\mathbf{y}_t = \begin{bmatrix} 0 \\ \overline{\mathbf{S}}_1 \hat{\mathbf{p}}_t \\ \tilde{\mathbf{p}}_t \end{bmatrix}, \mathbf{Z}_t = \begin{bmatrix} \mathbf{S}_1 \\ \overline{\mathbf{S}}_1 \\ \mathbf{S}_t \end{bmatrix}, \mathbf{B}_t = \begin{bmatrix} 0 \\ \overline{\mathbf{S}}_1 \\ \mathbf{0} \end{bmatrix}, \zeta_t = \begin{bmatrix} 0 \\ \overline{\mathbf{S}}_1 \mathbf{v}_t \\ \boldsymbol{\xi}_t \end{bmatrix} \tag{18}$$

$$E\left(\zeta_t \zeta_t'\right) \equiv \mathbf{H}_t = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \overline{\mathbf{S}}_1 \mathbf{\Omega} \overline{\mathbf{S}}_1' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\zeta^2 \mathbf{S}_t \mathbf{V}_t \mathbf{S}_t' \end{bmatrix} \tag{19}$$

$\tilde{\mathbf{p}}_t$ is an $N_t \times 1$ vector of benchmark observations.

Again, σ_u^2 and σ_ζ^2 are constants of proportionality and the first row is the reference country.

3.3 A State-Space Representation

Equations (11) and (15), together with the matrix definitions (16) to (19), constitute the 'transition' and 'observation' equations, respectively of a state space model for the unobservable 'state vector' $\mathbf{p}_t$.

Given the unknown parameters, $\boldsymbol{\theta}$ and hyperparameters $\phi, \sigma_u^2, \sigma_\eta^2, \sigma_\zeta^2$ and the distribution of the initial vector, $\mathbf{p}_0$, under Gaussian assumptions[15], the Kalman filter computes the conditional (on the information available at time *t*) mean $\hat{\mathbf{p}}_t$, and covariance matrix, $\mathbf{\Psi}_t$, of the distribution of $\mathbf{p}_t$. Further, $\hat{\mathbf{p}}_t$ is a minimum mean square estimator (MMSE) of the state vector $\mathbf{p}_t$. When Gaussian assumptions are dropped, the Kalman filter is still the optimal estimator in the sense that it minimizes the mean square error within the class of all linear estimators (see Harvey (1990, pp. 100-12), Durbin and Koopman (2001) Sections 4.2 and 4.3).

[14] The inclusion of the reference country constraint is a necessary condition for invariance of the results to the chosen reference country.

[15] The disturbances and initial state vector are normally distributed.

4 Special Features of the generic model

The state-space model formulated in Section 3 is a flexible model that can easily accommodate a number of common approaches to the production of PPPs. We demonstrate how the model can be used in making sure that the extrapolated PPPs can be made to track the observed PPPs for the benchmark years or how the model can be made to track and preserve the movements in the implicit GDP deflator.

4.1 Constraining the model to track benchmark PPPs

In many practical situations, it may be considered desirable that, in benchmark years, and for participating countries, the estimates of the PPPs (produced automatically by the proposed algorithm) should coincide with the ICP benchmark figures. As PPPs for currencies of the participating countries are determined using price data collected from extensive price surveys, one may consider it necessary that the extrapolated PPPs from the state-space model described above track the benchmark PPPs accurately. This can be achieved simply by setting $\sigma_{\zeta}^{2}=0$ in (19). The last line in (18) then becomes a constraint, guaranteeing that predicted PPPs are identical with corresponding benchmark observations. This particular property of Kalman filter predictions follows from the results presented in Doran (1992).

4.2 Constraining the model to preserve the movements in the implicit GDP deflator

A standard requirement considered in international comparisons of prices using PPPs is that the PPPs in different years preserve the movements in national price levels as measured by the implicit GDP deflators. As the GDP deflator data are provided by the countries and such deflators are compiled using extensive country-specific data, it is considered important that the estimated *PPPs* preserve the observed growth rates implicit in the GDP Deflator[16]. This essential feature can be achieved by setting $\sigma_{\eta}^{2}=0$ in (11) (see also Section 2.2). This result is proved in Appendix 1.

4.3 Flexibility in use of regression predictions

An important feature of the model is that the information provided by relevant socio-economic variables can be utilized in all time periods, both benchmark and non-benchmark through the regressors $\mathbf{x}_{it}'$ in (2). If we wish to produce estimates that use only growth rates between benchmark years, the second line of the observation vectors (16) are removed. The algorithm will then automatically update predictions between benchmarks using only growth rates in deflators. We present an illustration of the results obtained under this simplified model in Section 6.

4.4 Kalman Filter predictions as a 'weighted averages' of benchmark year only predictions

As mentioned earlier, current methodology for the estimation of a panel of PPPs is a two step procedure. First, in a benchmark year, observations on participating countries are obtained and then used to extrapolate to non-participating countries through regression relationships. Thus, in benchmark years predictions for the whole cross-section are obtained.

[16] Preserving movements in the implicit deflator will ensure that the growth rates in GDP at constant prices (real) and growth in per capita income reported and used at the country level are preserved in the international comparisons.

The second step consists of completing the panels by using growth rates obtainable from national accounts.

If there are M+1 benchmark years (j = 0, …, M)[17], applying growth rates to benchmark PPPs will produce M+1 different panels of PPP estimates. Faced with the dilemma of which panel to use, two possible approaches (of many) would be to: (a) use the panel based on the most recent benchmark year; or (b) to take some sort of average of the M+1 different panels.

An important property of our method is that in the case that benchmark year estimates and growth rates are used, but no information is introduced for years in between benchmark years, the panel of PPP estimates produced is a 'weighted average' of the M+1 panels discussed above. More specifically, suppose $\breve{\mathbf{p}}_{t,j}$ is the vector of PPP estimates in year t obtained by applying growth rates to the j^{th} benchmark. Then, denoting the corresponding Kalman Filter estimates by $\breve{\mathbf{p}}_t$, we have

$$\breve{\mathbf{p}}_t = \sum_{i=0}^{M} \mathbf{W}_i^{(M)} \breve{\mathbf{p}}_{t,j} \tag{20}$$

where the $\mathbf{W}_i^{(M)}$ are the product of positive definite matrices, and

$$\sum_{i=0}^{M} \mathbf{W}_i^{(M)} = \mathbf{I}_N \tag{21}$$

It is in this sense the prediction in (20) is considered as a 'weighted average' although it is not generally true that the elements of $\breve{\mathbf{p}}_t$ are a weighted average of those of the $\breve{\mathbf{p}}_{t,j}$.

However, in a very important special case the elements of $\breve{\mathbf{p}}_t$ are a weighted average of the <u>corresponding</u> elements of the M+1 'benchmark only' panels. Suppose that measurement errors in growth rates and benchmark PPPs are uncorrelated across countries. Then, it can be shown that

$$\breve{p}_{it} = \sum_{j=0}^{M} w_j^{(M)} \breve{p}_{it,j} \tag{22}$$

where, $w_j^{(M)} > 0$ ($j = 0,1,...,M$) and $\sum_{j=0}^{M} w_j^{(M)} = 1$

The above result demonstrates that the Kalman filter estimates are a weighted average of all the corresponding elements of the M+1 panels. Furthermore, the weights are not chosen in some arbitrary way, but derived from the covariance properties of the model resulting in optimal predictions for PPPs. See Appendix 2 for the derivation of the above properties.

[17] It will be convenient for the algebraic derivations presented shortly to set the number of benchmarks to M+1.

5. Data compilation and data construction

This section describes the data set used in this study. The data set covers 141 countries over the years 1970 to 2005. Table DA.1 lists the 141 countries included in the study. This table also lists the currency of each country and the years each country has participated in the ICP Benchmark comparisons. The empirical analysis in this paper includes PPP data from the ICP 2005 round. Given that the 2005 PPPs were only recently released, we have been able to include only 110 countries of the 146 that participated in the round. We will be adding the remaining countries to produce a revision of there estimates later this year. Table DA.2 gives definitions and sources of the variables used in the study, while Table DA.3 provides some basic descriptive statistics of the variables.

5.1 Treatment of missing data

The dimensions of the data set were largely determined by data availability. That is, a number of countries were excluded because of missing data (see the notes for Table DA.1), and the time frame 1970-2005 was likewise chosen because of poor data availability prior to 1970. Many variables which were initially considered for the analysis were also excluded due to data unavailability. For the countries and variables chosen for inclusion, there were still a number of socio-economic variables which did not cover all countries for all years. Missing cell data were imputed in a number of ways:

Step 1. If a country had no observations for a given variable, we first tried to find alternative sources, such as the CIA Factbook (CIA 2007), or various UN sources. When this approach failed, we considered replacing the missing values for the country in question with values for a similar country, or in a few cases, a regional average. For example, Bermuda is missing a great deal of data (exceptionally so), so Bermuda's data were often replaced with data for the Bahamas. Generally such "nearest neighbour" replacements were rare, and this was only done for variables in which we were confident that a "nearest neighbour" provided a good proxy for missing data for a given country. For example, this type of replacement was carried out with "level of development" variables (literacy rates, infrastructure levels) whose levels and trends follow fairly standard patterns, but not with complex trade or financial data, the movements of which over time could not be justifiably imputed from the experience of similar "neighbours".

Step 2. Once every country had at least one observation, series were extrapolated within countries using a variety of methods. For series that were relatively complete (e.g. every 5 years), we used time trend interpolations and extrapolations. This was satisfactory for most "level of development" series (literacy rates, tractors per worker, life expectancy), which tend to change over time in relatively smooth and predictable ways. However, measures of trade (such as trade balance and trade as a percentage of GDP) vary substantially over time within countries, so extrapolation methods were often unsatisfactory, as were imputation methods. So in many instances we simply kept the nearest value fixed. This means that these series are measured with considerable error relative to other series.

The information is available from the authors with country- and variable-specific details of the imputation methods used.

5.2 PPP Data

The state variable in the state space model is $\ln(PPP_{it})$, and observed values (which define the dependent variable in the measurement equation) are obtained from all the benchmarks conducted so far. Thus PPP data are drawn from the early benchmarks of 1975, 1980 and 1985 as well as from more recent benchmark information for the years 1990, 1993, 1996, 1999, 2002 and 2005. Several features of the PPP data are noteworthy. The first benchmark covered 13 countries. The 1980, 1985 and the recent

2005, benchmarks represent truly global comparisons with PPPs computed using data for all the participating countries. For the years beginning from 1990 to 2002, data are essentially from the OECD and EU comparisons with the exception of 1996[18]. The 1996 benchmark year again is a global comparison with PPPs for countries from all the regions of the world. However, the 1996 benchmark may be considered weaker than the 1980, 1985 and 2005 benchmark comparisons as no systematic linking of regional PPPs was undertaken. In terms of reliability, one would consider the 1996 benchmark PPPs to be less reliable. Another related point of interest is the fact that PPPs for all the benchmarks prior to 1990 were based on the Geary-Khamis method and PPPs for the more recent years are all based on the EKS method of aggregation.[19] In the current empirical analysis, we have not made any adjustments to the PPP data but propose to make the series comparable through the use of the same aggregation methodology in the next revision of the paper.

5.3 Socio-Economic Variables included in the Regression

Table DA.2 includes a description of the socio-economic variables that are included in the regressions in the study. The variables used come under two categories. We use a set of variables that are essentially dummy variables designed to capture country-specific episodes that may influence the exchange rates or PPPs or both as well as time dummies. The second set of variables are more of a structural nature commonly discussed in the works of Kravis and Lipsey (1983; 1986), Clague (1988), Bergstrand (1991; 1996) and Ahmad (1996).[20]

5.4 Covariance Variables

Measuring spatial correlation: The spatial weights matrix, $\mathbf{W}_t$, used in modeling spatial autocorrelation is derived by identifying the five nearest "neighbours." We define neighbours by bilateral trade flows. For each country, i, a maximum of five columns have non-zero values that correspond to the major trading partners. A value of unity is assigned to each partner country and the matrix is row normalized (ie rows in the $\mathbf{W}_t$ add up to one and therefore assign equal weight to the five closest trading partners to any given country). Contiguity matrices were constructed for the years 1970, 1975, 1980, 1985, 1990, 1995, 2000 and 2005. This approach was taken due to time constraints, and therefore we assume that the trading partners of a country do not change within each of the five year periods, and therefore the weights matrix remains constant over that period[21]. These matrices are compiled using the data from Rose and IMF Trade Directions.

Accuracy of benchmarks and national accounts' growth rates: The model specification allows for the modeling of accuracy of benchmark PPPs and national growth rates (equations (5) and (6)). We assume that the measurement errors in both cases have variances that are inversely proportional to the per

---

[18] We are indebted to Ms Francette Koechlin (OECD) for providing ICP benchmark data for these years. PPPs for those countries which joined in the Euro zone, the pre-Euro domestic currencies were converted using the *1999 Irrevocable Conversion Rates*
(Source: http://www.ecb.int/press/date/1998/html/ pr981231_2.en.html). The irrevocable conversion rate of the drchma vis a vis the euro was set at GRD 340. 750 Source: http://www.bankofgreece.gr/en/euro.

[19] This was brought to our attention by Steve Dowrick who attended a seminar on the topic presented at the Australian National University in October 2007.

[20] We are conscious of the fact that serious multicollinearity issues may be present here as the variables are potentially correlated. As the main purpose of inclusion of these variables is to improve the quality of the predictions, we decided to leave the variables in the model with the view that the model results in better predictions.

[21] Yearly weight matrices will be constructed for the next revision.

capita GDP expressed in US dollars. This means that countries with higher per capita incomes are expected to have more reliable data, as reflected by lower variances associated with them.[22]

6. Estimation

In order for the Kalman filter to deliver a predictor of the state vector and its covariance matrix, we require estimates of the unknown parameters and a distribution of the initial state vector. The estimation of the parameters of a state-space system can be handled with likelihood based methods (Harvey 1990, pp. 125-46) or Bayesian methods (see for instance Durbin and Koopman(2002), Koop and van Dijk(2000), and Harvey, Trimbur and van Dijk (2005)). The results presented in this paper are obtained using likelihood based methods. The distribution of the initial state vector, $\mathbf{p}_o$, is assumed to be centered at zero and its covariance has been derived as follows.

Distribution of the Initial State Vector

For this specification we can derive a non-diffuse covariance for the initial state vector, $\mathbf{p}_o$ by making use of equation (3). Suppose at $t = 0$ we have socio-economic data, $\mathbf{X}_o$. Then we can define,

$$\mathbf{p}_o = \mathbf{X}_o\boldsymbol{\beta} + \ln(\mathbf{ER}_o) + \mathbf{u}_o \qquad (23)$$

where,

$$\boldsymbol{\beta} = [\boldsymbol{\beta}_{oo}\ \boldsymbol{\beta}_{so}]'$$

$$\mathbf{p}_o = \begin{bmatrix} \mathbf{p}_o^{(1)} \\ \mathbf{p}_o^{(2)} \end{bmatrix}$$

$$\mathbf{X}_o = \begin{bmatrix} \mathbf{X}_o^{(1)} \\ \mathbf{X}_o^{(2)} \end{bmatrix}$$

$\mathbf{X}_o^{(1)}$ and $\mathbf{p}_o^{(1)}$ represent the partition containing the observations from participating countries.

Then a prediction of $\mathbf{p}_o$ and its associated covariance are given by

$$\hat{\mathbf{p}}_o = \mathbf{X}_o\hat{\boldsymbol{\beta}} + \ln(\mathbf{ER}_o) \qquad (24)$$

$$\text{cov}(\hat{\mathbf{p}}_o) = \boldsymbol{\Psi}_o = \sigma^2\mathbf{X}_o(\mathbf{X}_o^{(1)\prime}\mathbf{X}_o^{(1)})^{-1}\mathbf{X}_o' \qquad (25)$$

We use the expression in (25) to obtain an estimate of the covariance of the initial state vector for the *constrained* and *unconstrained* models.

---

[22] We make use of exchange rate converted per capita incomes to overcome the problem of possible endogeniety arising out of the use of PPP converted exchange rates. These data are drawn from the UN sources. Given the systematic nature of the exchange rate deviation index (ratio of PPP to ER), use of exchange rate converted per capita GDP is likely to magnify differences in per capita incomes.

We note that under normality of the disturbances, the conditional distribution of the observation vector $\mathbf{y}_t$ is given directly by the Kalman filter[23] (we refer the reader to Harvey (1990) for details).

Identification of Regression Parameters

The unknown parameters of the state-space model given by (11) and (15) cannot be estimated as they stand. Specifically, the vector $\boldsymbol{\theta}$ is not identified without further structure. In this paper identification is achieved by using a time varying intercept and time invariant slopes. The time dummies shift the intercept at each benchmark year since the 1980 benchmark.

Algorithm

There are two types of parameters to be estimated in the SS, namely, hyperparameters, and coefficients associated with explanatory variables and the level shifts. Hyperparameters are those associated with the covariance structure. In our case these are: $\phi, \sigma_u^2, \sigma_\eta^2, \sigma_\xi^2$. These parameters must be estimated by numerical maximization of the likelihood function (in a likelihood based estimation). The other parameters, $\boldsymbol{\theta}$ in our case, can be estimated by a generalised least squares procedure (GLS) in conjunction with the numerical maximization of the likelihood function (see Harvey (1990, pp. 130-3)), which we denote by KF/GLS, or placed in the transition equation and estimated with the state vector[24].

The algorithm we use can be described in 5 steps.

Step 1: Obtain an initial estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^0$**,** by regressing $\mathbf{r}_t$ on $\mathbf{X}_t$ and construct an initial prediction, $\hat{p}_{it}^0$, using equation (3).

Step 2: Run SS through KF (or KF/GLS) to obtain estimates of the hyperparameters and $\boldsymbol{\theta}$.

Step 3: Use updated estimate of $\boldsymbol{\beta}_{it}$, $\hat{\beta}_{0t} = \hat{\beta}_t^0 - \hat{\theta}_{0t}$ $\hat{\boldsymbol{\beta}}_t = \hat{\boldsymbol{\beta}}_t^0 - \hat{\boldsymbol{\theta}}_t$, to obtain an updated $\hat{p}_{it}$

Step 4: Repeat 2 and 3 until $\hat{\boldsymbol{\theta}}$ are sufficiently close to zero.

Step 5: Run KF and Kalman smoother one more time to obtain p_{it}^* and standard errors.

A prediction of PPP_{it} is given by:

$$P\hat{P}P_{it} = e^{p_{it/T}^*} \tag{26}$$

where,

$p_{it/T}^*$ is the corresponding Kalman smoothed element

The standard errors for the predicted PPPs are computed as follows[25]:

$$se(P\hat{P}P_{it}) = \sqrt{e^{2 p_{it/T}^*} e^{\psi_{ii,t}^*} (e^{\psi_{ii,t}^*} - 1)} \tag{27}$$

where,

---

[23] The log likelihood is written in *prediction error decomposition form*

[24] The code for the empirical estimations was written by the authors in GAUSS and includes a procedure to evaluate the likelihood function when some of the parameters are obtained by a KF/GLS approach.

[25] The standard errors are computed under the assumption of the lognormality of the predictions.

$\psi^*_{ii,t}$ is the *ith* diagonal element of the estimated covariance of the state vector, $\mathbf{p}^*_{t/T}$.

7. Empirical Results

7.1 Introduction

The empirical results presented in this section are designed to illustrate the flexibility of our modeling approach. We will present the extrapolation of PPPs under different variations of the model, namely, *PPPs from estimation of an unconstrained model*, *PPPs from estimation of a benchmark constrained model* and *PPPs from estimation of a model without regression information for non-benchmark years*. In addition for each set of estimates we present the smoothed results that do not preserve implicit price movements as well as those that do. We first introduce each of these alternative specifications and then present the detailed results.

PPPs from estimation of an unconstrained model: This is a completely flexible model that is only constrained to obey the reference country constraint in equation (7). Otherwise, the errors in ICP benchmarks, σ^2_ς , and observed growth rates, σ^2_η , are not restricted and determined completely by the data. The estimation follows the steps presented in the Algorithm described in Section 6. For this case we present two sets of estimates. The first set has been smoothed without constraining growth rates to track the implicit GDP deflators, and the second set of estimates are obtained so that the estimates accurately track price movement in the implicit GDP deflators (see Section 4.2 and Appendix 1). The second set will be labeled with the prefix "*GRC*" in the graphs presented next. Parameter estimates from this model are presented in Table 1, Model (3).

PPPs from estimation of a benchmark constrained model: This is a model that obeys the reference country constraint in equation (7) and fixes the benchmark error to be very small through setting the value of σ^2_ς close to zero (See Section 4.1). In this case the algorithm is implemented as specified, however, in Step 2, the hyperparameter σ^2_ς is not estimated, but fixed at a given value (1e-4 in the results presented). Similar to the previous case, estimates are presented for the case when the smoother does not preserve growth rates as well as the case when the estimates track the price movement in the implicit GDP deflators ("*GRC*"). Parameter estimates from this model are presented in Table 1, Model (4).

PPPs from estimation of a model without regression information for non-benchmark years: As stated in Section 4.3, the use of the regression can be restricted only to the prediction of *PPPs* for non-participating countries in benchmark years. For this simplified case we have been able to show that the resulting *PPP* time series is a weighted average of the extrapolations from all the benchmark information and the weights depend on the estimated covariance structure (ie it is a function of the data). The results associated with this model will be labeled "*SIM*" in the graphs presented below. To obtain these results a regression with spatially correlated errors is estimated and used to produce predictions for non-participating countries in each benchmark year (Model (2) in Table 1). Thus, at each benchmark year a PPP value is available for each country in the sample. For those countries that participated are given the observed ICP *PPP*, $\tilde{\mathbf{p}}_{t(j)}$, *j*=1,…,M benchmarks; for those countries that did participate in the benchmark, predictions from the spatial regression model, $\hat{\mathbf{p}}_{t(j)}$, *j*=1,…,M are assigned. The later predictions are assumed to have a prediction error given by the estimated least squares variance, and the former are assumed to have a variance that suffers from measurement error, i.e. identical to that in the previous two models. Through the state space representation of this simplified model two hyperparameters, σ_{ζ}^2 and σ_{η}^2 are estimated. Therefore, regression predictions, ICP benchmarks and growth rates are assumed to be measured with error. The resulting estimates are constrained to obey the reference country constraint and can be smoothed to preserve growth rates. These estimates are labeled "SIM-GRC" in the graphs below. Hyperparameters estimates are presented in Table 1, Model (5).

7.2 Results and Discussion

Table 1 presents the estimates corresponding to the three alternative scenarios described above. The first column (model (1)) reports the least squares estimates that are used as starting values to the algorithm to obtain models (3) and (4) estimates. The second column reports the estimates from the regression with spatially correlated errors which is used to provide *PPP* predictions to non-participating countries in benchmark years in the model with no regression information in non-benchmark years. These predictions are then used in the state space model to estimate the two remaining covariance parameters and these estimates appeared on the last column of Table 1. The fit of the regressions is good, especially the spatial regression model (2) with a generalized R^2 of 0.85. The state space model estimates (3) and (4) can be compared by their likelihood value as these are the same model. It is clear that the benchmark constrained model (4) is rejected by the data as the likelihood value is considerable lower than that of the unconstrained model.

Table 1: Estimates of Parameters from Different Model Specifications

	REGRESSION ESTIMATES				STATE SPACE ESTIMATES					
	No Spatial Errors (1)		Spatial Errors (2)		Benchmarks Unconstrained SS Spatial Errors (3)		Benchmarks Constrained SS Spatial Errors (4)		No-Regression for non-benchmark years (5) (*)	
Variable	**Estimate**	S.e.	**Estimate**	S.e.	**Estimate**	S.e.	**Estimate**	S.e.	Estimate	S.e
Intercept	**-0.811**	0.457	**-0.886**	0.862	**0.668**	0.240	**0.527**	0.177		
dum80_84	**0.257**	0.297	**0.230**	0.559	**-0.043**	0.023	**-0.029**	0.021		
dum85_89	**-0.139**	0.298	**-0.165**	0.561	**-0.135**	0.026	**-0.236**	0.022		
dum90_92	**0.228**	0.301	**0.244**	0.571	**-0.046**	0.029	**-0.128**	0.026		
dum93_95	**0.060**	0.301	**0.062**	0.570	**-0.120**	0.029	**-0.215**	0.025		
dum96_98	**0.069**	0.297	**0.064**	0.560	**-0.185**	0.029	**-0.220**	0.024		
dum99_01	**-0.177**	0.301	**-0.180**	0.570	**-0.323**	0.031	**-0.392**	0.027		
dum02_04	**-0.292**	0.291	**-0.291**	0.546	**-0.318**	0.030	**-0.406**	0.026		
dum05	**-0.013**	0.295	**-0.011**	0.554	**-0.205**	0.036	**-0.237**	0.033		
D_anz	**-0.714**	0.222	**-0.724**	0.415	**-0.501**	0.141	**-0.824**	0.099		
D_asean	**-0.046**	0.077	**-0.071**	0.146	**0.339**	0.059	**0.186**	0.039		
D_cac	**0.000**	0.153	**0.005**	0.300	**-0.175**	0.110	**-0.285**	0.068		
D_euro	**0.127**	0.045	**0.122**	0.086	**0.050**	0.026	**0.152**	0.023		
D_mercsr	**-0.089**	0.077	**-0.091**	0.150	**0.013**	0.050	**0.322**	0.036		
D_nafta	**-0.209**	0.085	**-0.238**	0.161	**-0.265**	0.050	**-0.166**	0.045		
D_scucar	**0.192**	0.149	**0.206**	0.279	**0.204**	0.082	**0.223**	0.055		
D_spr	**0.594**	0.208	**0.581**	0.390	**0.406**	0.136	**0.843**	0.091		
D_usd	**0.038**	0.068	**0.033**	0.126	**0.327**	0.039	**0.385**	0.028		
AGEDEP	**0.694**	0.185	**0.715**	0.347	**-0.167**	0.126	**-0.128**	0.091		
AGVAGUN	**-0.009**	0.002	**-0.009**	0.004	**-0.006**	0.002	**-0.008**	0.001		
TRACTORPW	**0.083**	0.062	**0.098**	0.116	**0.185**	0.034	**0.263**	0.031		
LABPOP	**0.000**	0.003	**4.4E-04**	0.006	**-0.007**	0.002	**0.001**	0.002		
LIFE	**-0.005**	0.003	**-0.004**	0.006	**-0.013**	0.002	**-0.014**	0.002		
LITERATE	**1.2E-04**	1.E-04	**1.4E-04**	2.2E-04	**3.4E-04**	9.9E-05	**1.1E-04**	6.3E-05		
NTRVAG2	**-0.004**	0.002	**-0.004**	0.005	**-0.004**	0.002	**-0.002**	0.001		
BLACKIND	**0.043**	0.034	**0.056**	0.064	**0.039**	0.017	**0.109**	0.013		
EXPG	**-0.002**	0.003	**-0.002**	0.006	**-0.007**	0.001	**-0.011**	0.001		
PHONES	**0.001**	2.E-04	**0.001**	3.4E-04	**0.001**	9.7E-05	**0.002**	8.1E-05		
RADPCCN	**7.0E-06**	7.E-06	**6.0E-06**	1.3E-05	**1.6E-06**	4.3E-06	**-1.2E-05**	3.8E-06		
RURPOP	**-0.005**	0.001	**-0.004**	0.002	**-0.004**	0.001	**-0.004**	0.001		
TRADEGUN	**3.0E-04**	0.002	**1.2E-04**	0.003	**0.002**	0.001	**0.004**	0.001		
MANUFEXP	**-4.9E-04**	0.001	**-3.5E-04**	0.001	**-0.001**	3.9E-04	**1.1E-04**	3.2E-04		
MANUFIMP	**0.002**	0.001	**0.002**	0.002	**0.002**	0.001	**-0.001**	0.001		
R^2	0.726		0.851							
logL					-5365.00		-11508.35		-4042.09	
Sample	449		449		5076		5076		5076	
σ_η^2 (Growth Rates)					**0.067**	0.001	**0.027**	2E-04	**0.025**	2.4e-03
σ_u^2 (Regression)	**0.080**		**0.074**		**0.070**	0.002	**0.080**	4-E04		
σ_ζ^2 (Benchmarks)					**0.040**	0.004	**1E-04**	---	**0.042**	1.1e-02
ϕ (Spatial Auto)			**0.185**	0.117	**0.094**	0.003	**0.150**	0.156		

(*) Predictions of PPPs to non-participating countries in Benchmark years obtained from model (2) are included in the state-space with a variance equal to the least squares predictor variance.

As the panel for all the countries and the years is quite large, a small set of countries is chosen for purposes of presenting estimates of *PPP* for each of the estimation and smoothing alternatives discussed in 7.1.

Figure 1 presents the *ICP Benchmark PPP* (**BPPP**), and two sets of PPP estimates generated from the unconstrained state space model (model (3) in Table 1). *Unconstrained Smoothed* (**PPP-UN**) are those produced by the Kalman Smoother without imposing the restriction that they follow Implicit Price Deflator movements. *Constrained Smoothed* are those obtained when the Kalman smoother constraints the results to follow the Implicit Price Deflator movements (**PPP-GRC**). The two-standard error bound for PPP-GRC is also included. As a reference we also plot the corresponding estimates from Penn World Tables 6.2 (**PWT6.2**).

Figure 1a presents our results for Australia. Australia will illustrate throughout the results the case of a developed country that has consistently participated in most of the global as well as OECD comparisons; and, it will illustrate the case when all sources of available information (national accounts and benchmark data) seem to provide a consistent picture. Note the consistency between the implied price deflator movement and the PPP benchmarks since 1990 shown by the PPP-GRC.

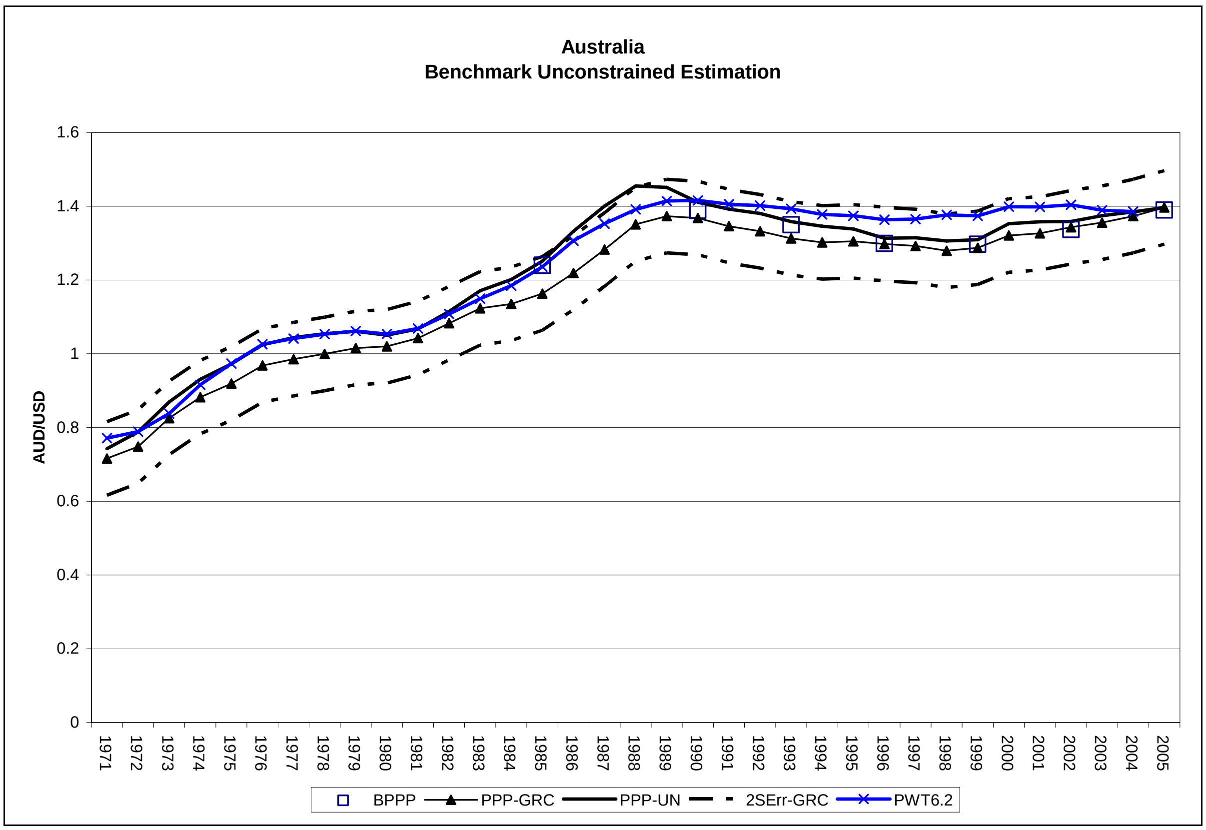

Figure 1a

Figure 1b is for China. China participated in the 2005 ICP comparison for the first time. A few important points can be made from this picture. First, the unconstrained estimates that have not been smoothed to follow the published GDP Deflator movements (PPP-UN) differ substantially from the series obtained when this is imposed (PPP-GRC), indicating that internationally available data on socio-economic variables for China, especially for the years before 1990, provide a different picture than that available through the movements in the latest available data on GDP Deflator. Further, and as expected, the standard error of the estimates is large, and the estimates from PWT 6.2 also differ substantially from ours.

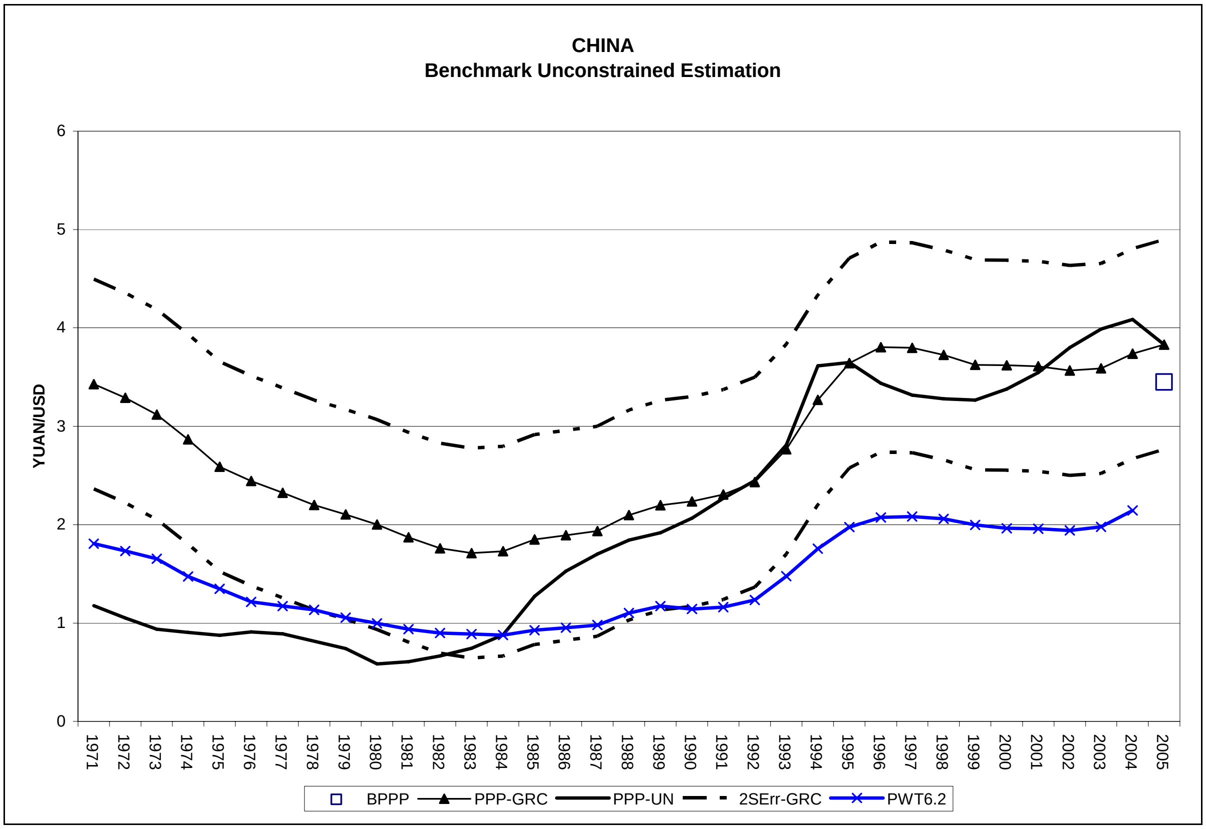

Figure 1b

Figure 1c presents our estimates for India. India had participated in the earlier ICP comparisons and has again participated in the 2005 round. The differences between PPP-UN and PPP-GRC are not as large as in the case of China which would indicate that available data on socio-economic variables and GDP Deflator movements provide a relatively consistent picture. Nevertheless, the standard errors are large and our estimate for 2005 is Rupee 16.98 instead of the benchmark value of Rupee 14.67.

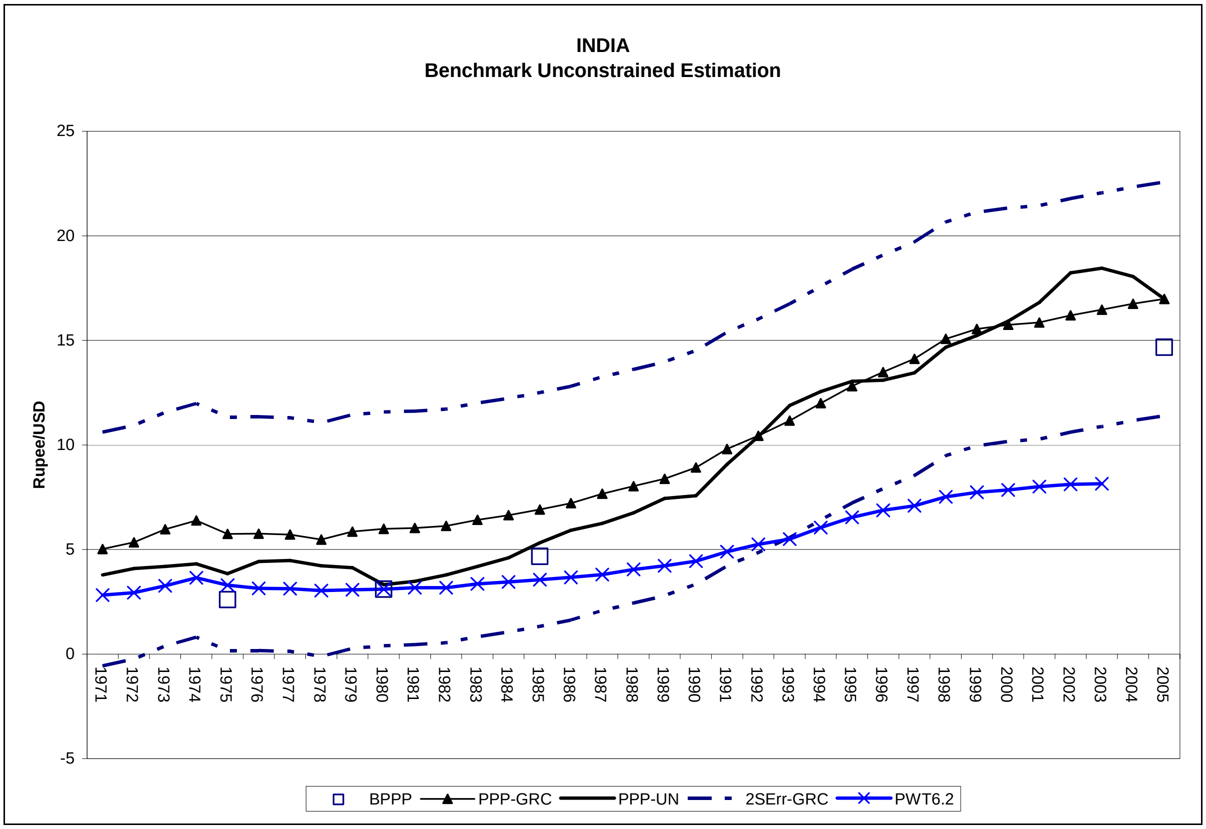

Figure 1c

The Central American countries did not participate in the 2005 ICP round. Figure 1d presents Honduras, which participated only in the 1980 round. The two series (PPP-UN and PPP-GRC) are fairly consistent which indicates consistency between socioeconomic data and the GDP Deflator. The standard errors are very large due to the lack of benchmark information for Honduras.

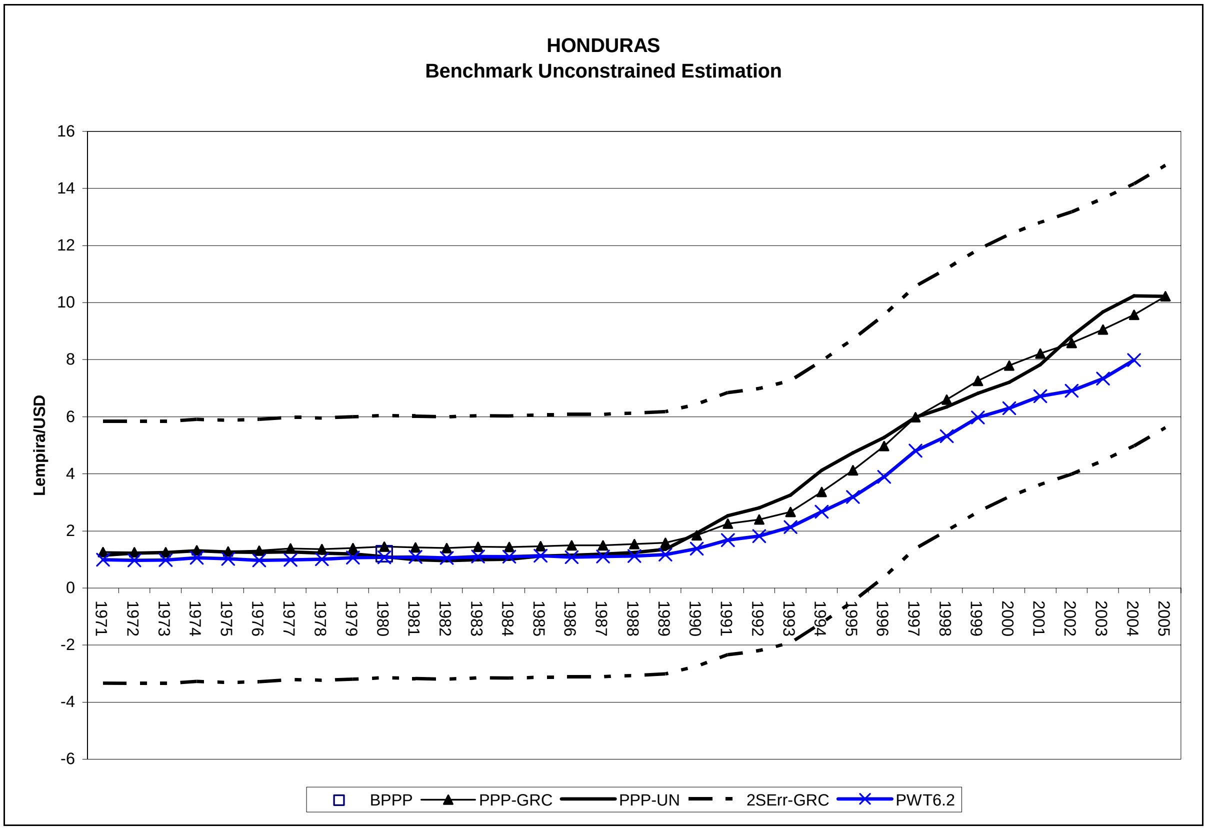

Figure 1d

Nigeria is presented in Figure 1e. Nigeria has participated in four benchmark exercises, including the 2005 round. There is some discrepancy between the PPP-UN series and the PPP-GRC series indicating some inconsistency between the most recent revision of the GDP Deflator and some of the socioeconomic variables available through international sources, which introduces a level of uncertainty shown in the standard errors. Our PPP-GRC estimates are very similar to those in PWT6.2.

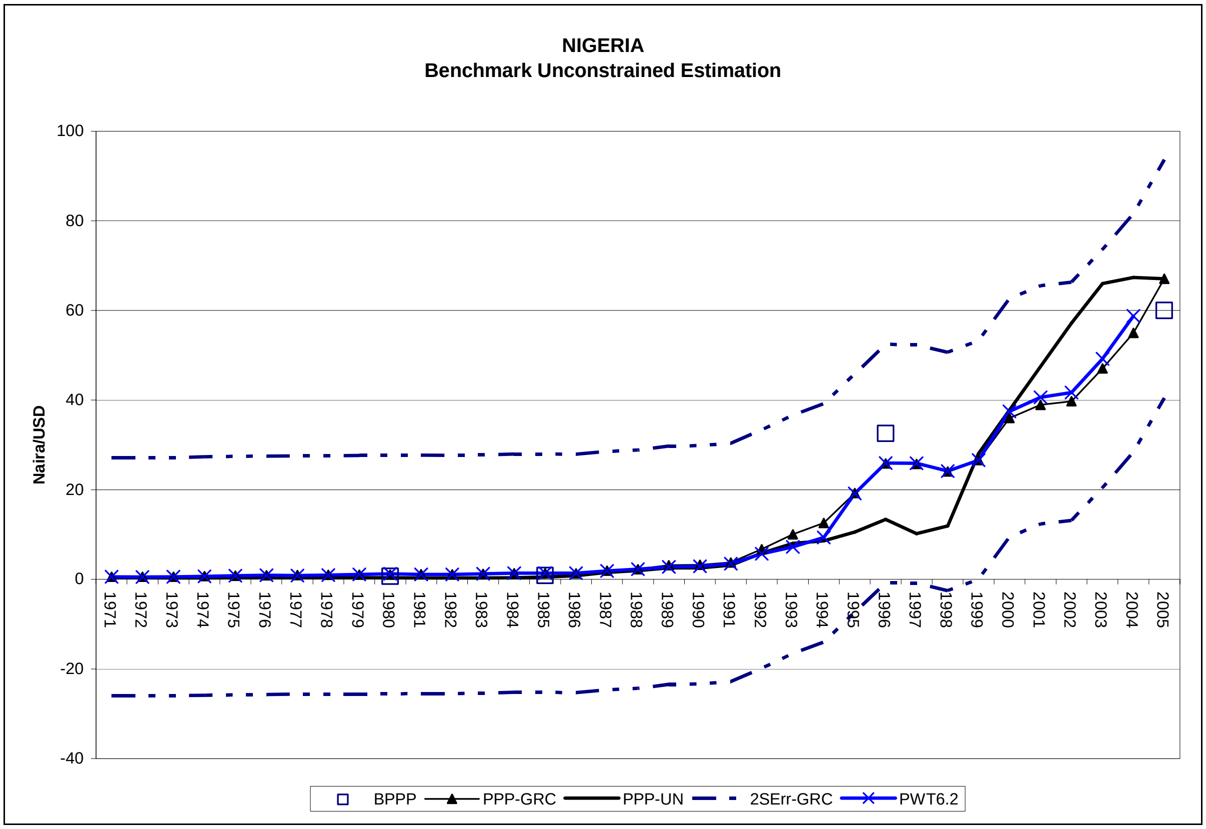

Figure 1e

Figure 2 presents estimates that are generated from the "benchmark constrained model" (Table 1 model (4)). Figure 2a shows the estimated series (PPP-CON) and corresponding two-standard error bound obtained by constraining the model to go through benchmarks (the smoother does not adjust for movements in the GDP deflator). Figure 2b shows how PPP-CON changes to obey the movements in the GDP deflator through the smoothing process (PPP-CON-GRC). We note that once the series has been obtained under the assumption that benchmarks had a very small error, and then smoothed to obey GDP Deflator movements, the standard errors will be reduced *substantially*. This is also visible for China and India (Figures 2c and 2d respectively). However, the constraint that benchmarks suffer from no error is rejected by the data.

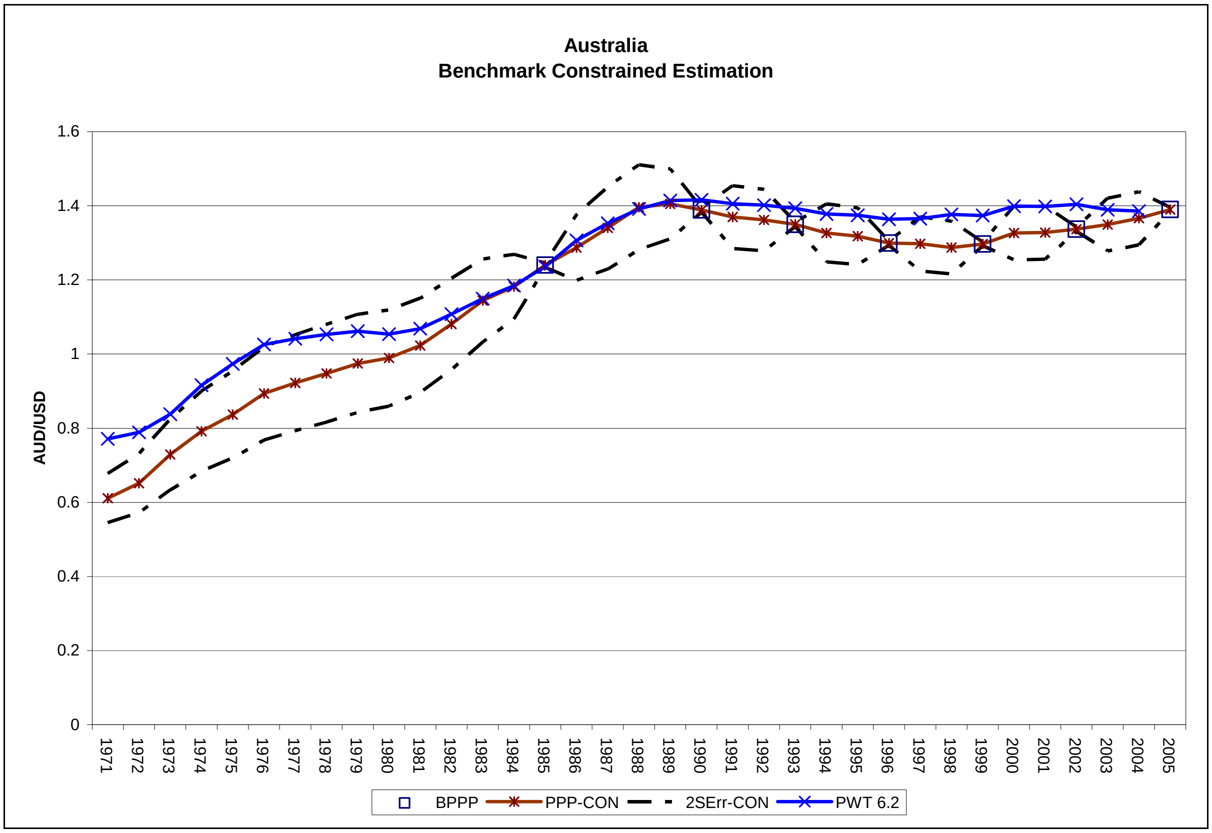

Figure 2a

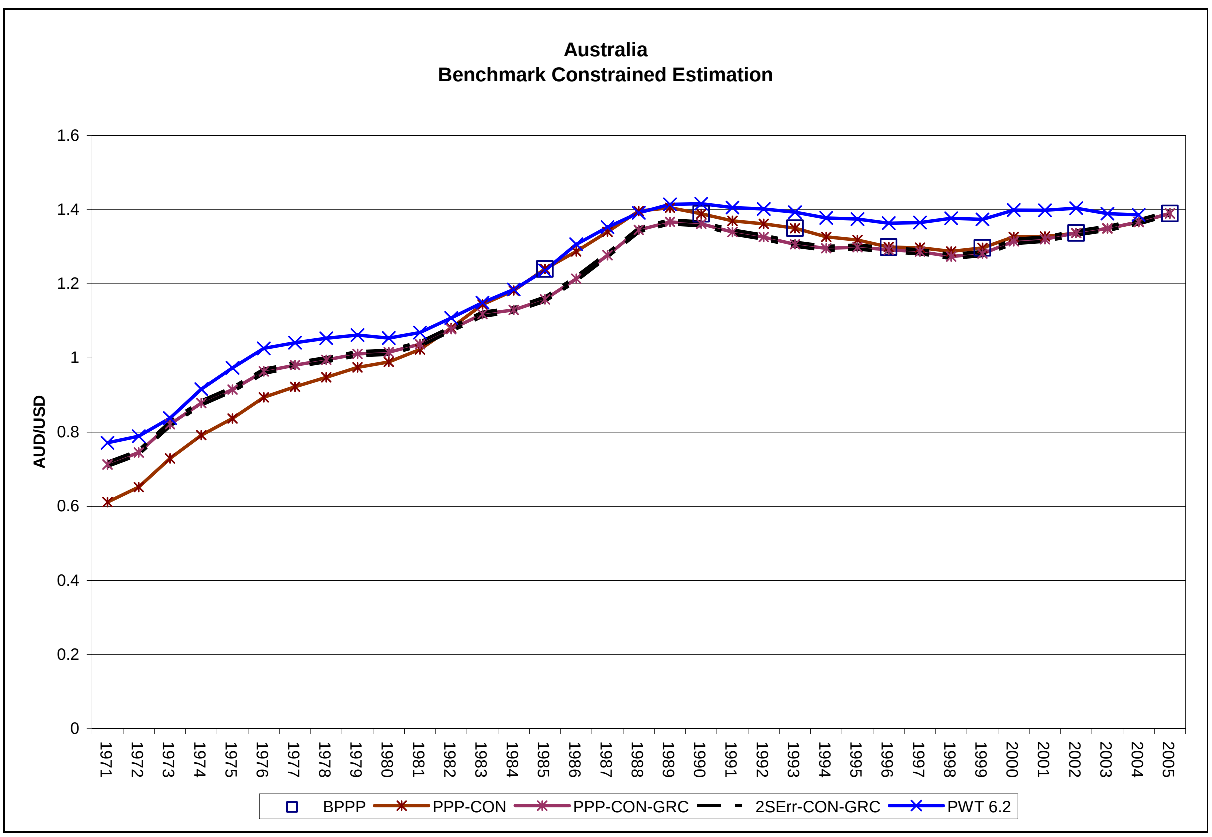

Figure 2b

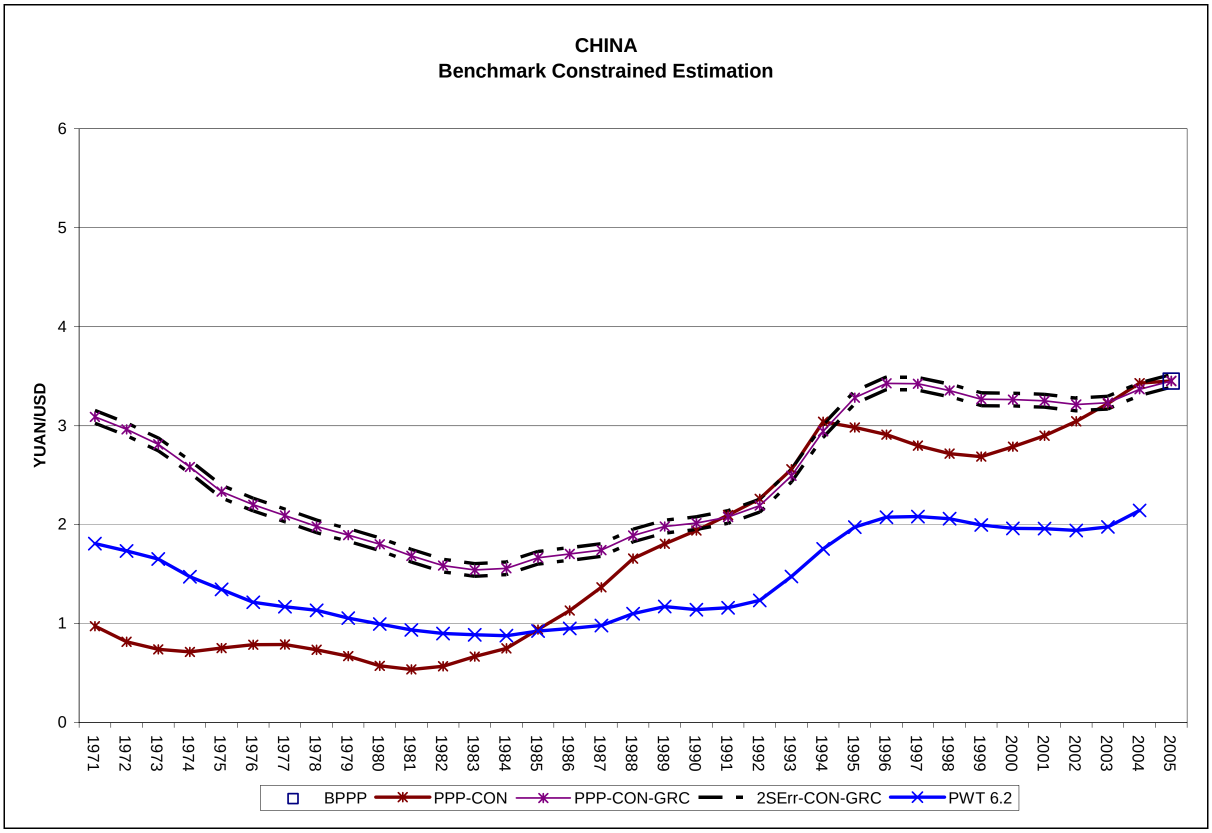

Figure 2c

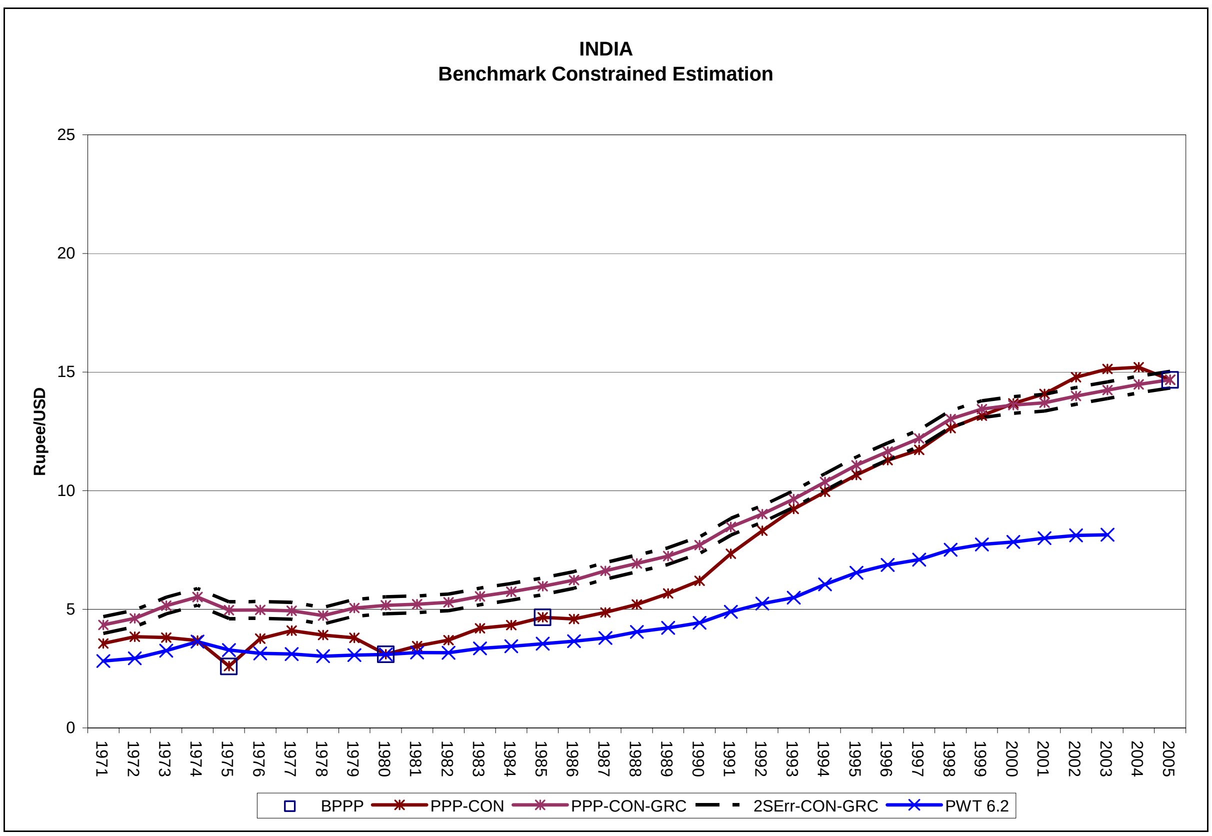

Figure 2d

The last set of results presented is from the model estimated without regressions in the non-benchmark years (Table 1, models (2) and (5)). Figure 3a shows the estimated series when the smoother does not adjust the movements to those of the GDP deflator (PPP-SIM), and the series that does obey the movement (PPP-SIM-GRC) with the two-standard error bound. For Australia, the estimates PPP-UN-GRC (Figure 1a) and PPP-SIM-GRC (Figure 3a) are virtually identical. This is very expected, as the number of benchmark points available for Australia as well as the consistency of information across data sources already mentioned.

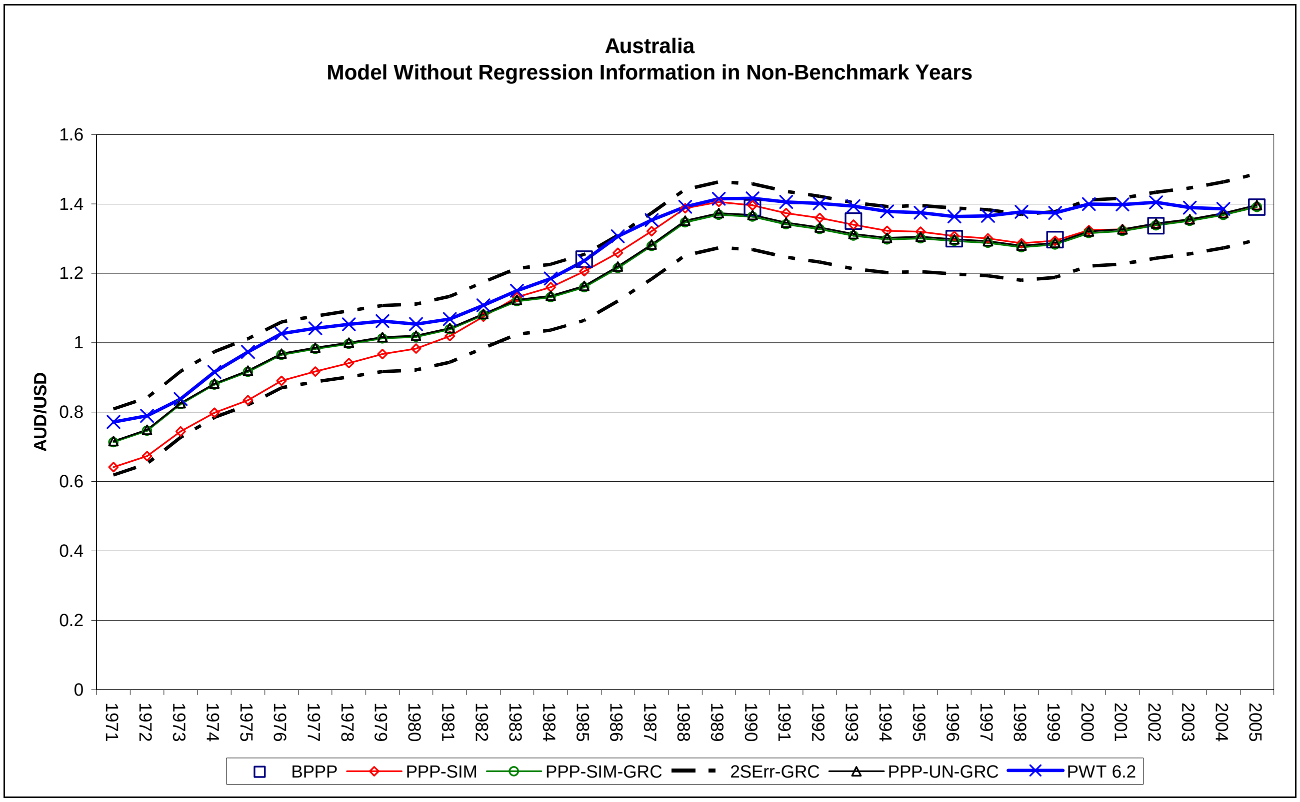

Figure 3a

Figures 3b and 3c present respective estimates for India and China. For India the model with regressions in all years produces estimates that are higher than those of the model without regressions. The main reason for this is that the latter is by design much more anchored on the benchmark values. The computed standard errors for the cases when the movements in GDP Deflator are obeyed (see Appendix 1) are larger for the simpler model, Rupee 3.12 and Rupee 2.79 for the full model. Similarly for China, the full model estimates that the PPP is higher than that obtained by the simpler model. The standard error is also slightly higher for the simpler model (Yuan 5.33 vs Yuan 5.73)

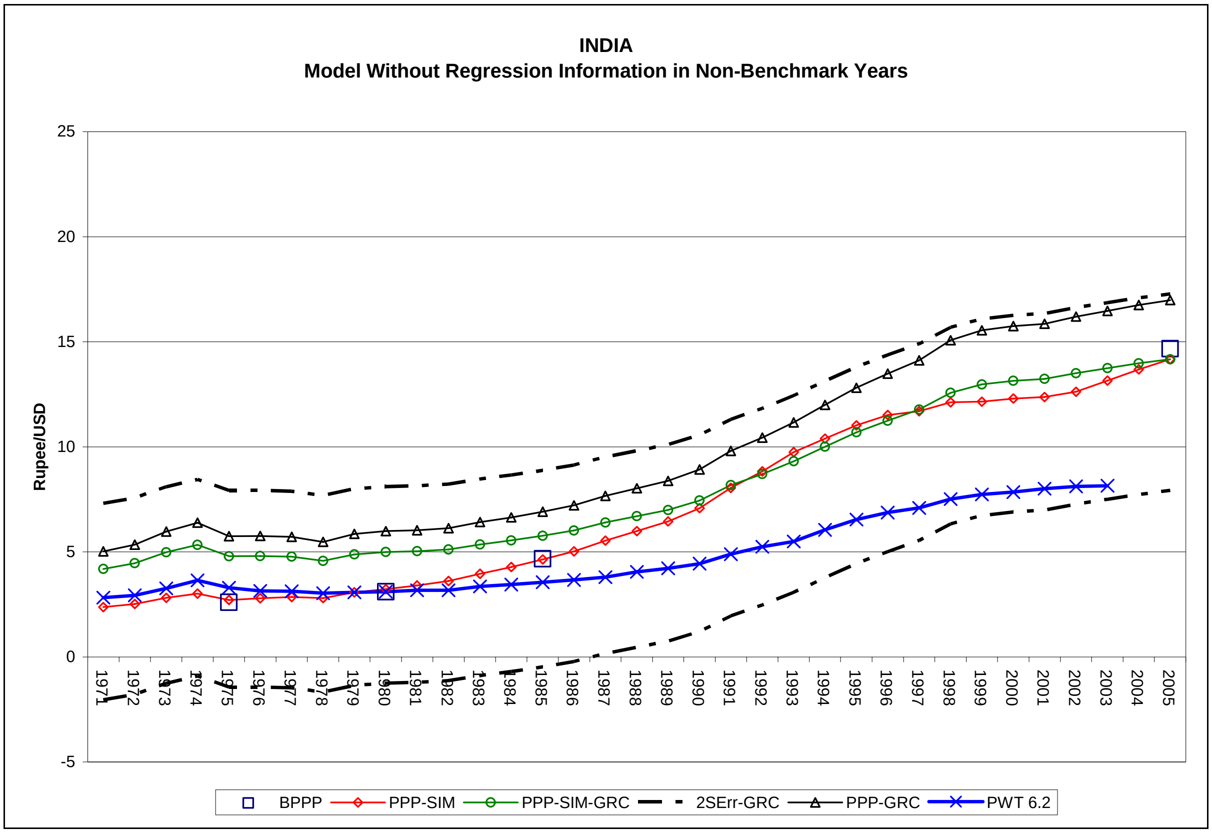

Figure 3b

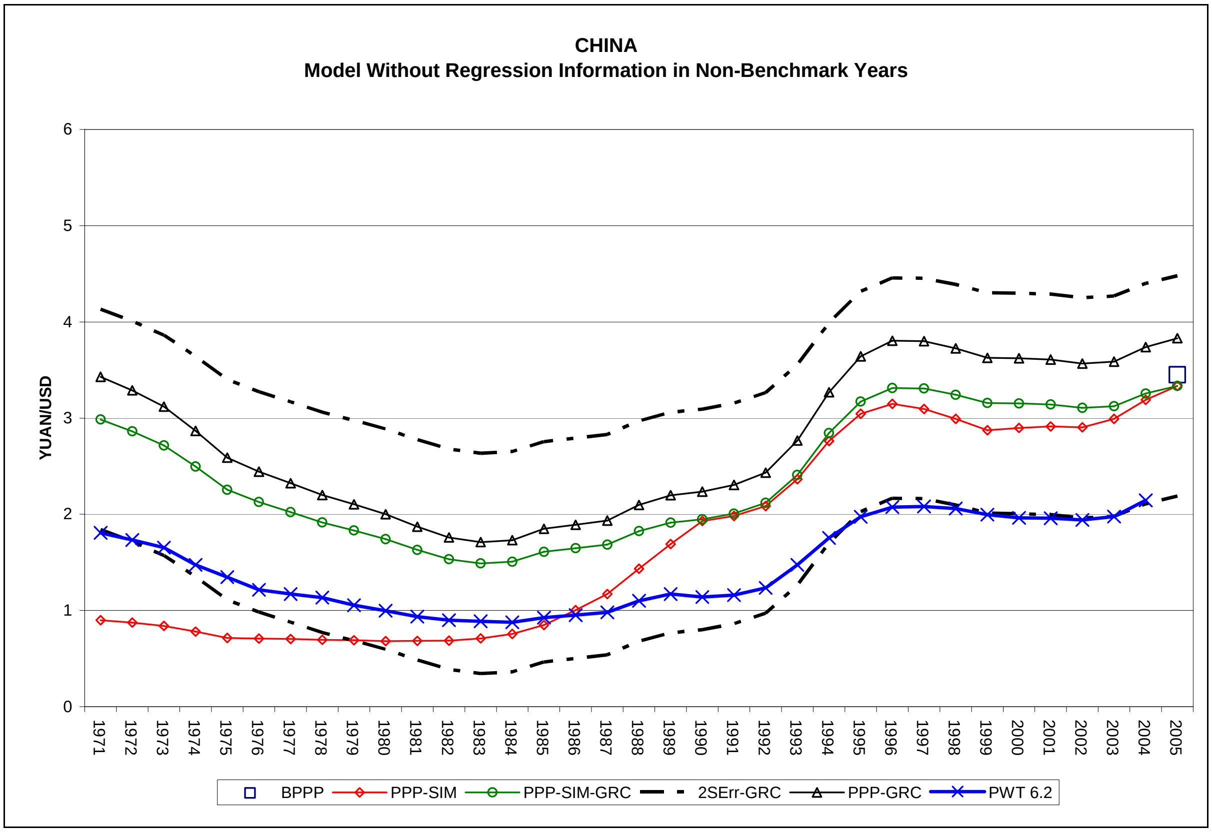

Figure3c

8. Conclusions

The econometric methodology suggested in the paper for the construction of a consistent panel of purchasing power parities represents the only attempt to provide a clear and coherent approach since the first attempt of Summers and Heston in 1988. The approach used in the paper is designed to make use of all the principal and auxiliary information available for the purpose of extrapolation of benchmarks. The first source used in the study is the data on PPPs from all the benchmark comparisons undertaken within the auspices of the International Comparison Program since 1970 including the latest round for the year 2005. The second source of data used for the purpose of constructing the panel of PPPs are the data on implicit price deflators at the GDP level published in all the countries included in the study. In addition to these two sources of data, an analytical constraint that requires the PPP of the reference country to be unity is also used as an additional piece of information. The forth source of information is for the purpose of extrapolating PPPs to countries not participating in the benchmark comparisons and to all countries in non-benchmark years. Data on a host of socio-economic variables are utilized.

The econometric model is expressed as a state-space model for the purpose of estimation and for generating optimal predictions. The parameters of the model are estimated using a maximum likelihood approach and the predictions are generated using the Kalman filter and smoother. The component of the model used to extrapolate to non-participating countries/years is a national price levels' model and is assumed to spatially auto-correlated disturbances. PPP benchmarks and growth rates are assumed to suffer from measurement error which is inversely proportional to the development level of each country. The paper demonstrates that the approach proposed here is flexible in that it can be used to consider a number of scenarios including the use of constraints on some variance parameters to generate extrapolations that track the observed PPPs in benchmark years; the observed price movements over time for different countries; and those that track both. An explicit form of the estimator is derived to show the estimates are weighted sums of past information. The estimator is a weighted average of past benchmark PPPs under simplified assumptions.

The methodology proposed is applied to a large data set covering 141 countries and a thirty-five year period 1970 to 2005 for generating predictions. The results from the empirical estimation are presented and analysed using PPP series generated for selected countries, including China, India, Australia and Nigeria, to examine the plausibility of the extrapolations. The results from the new methodology are contrasted with the published PPPs from the Penn World Table Version 6.2. The results are satisfactory and very encouraging. Further analysis and study of the results for all the 141 countries is currently underway and it is expected that the full panel of PPPs can be released for public use in the not too distant a future.

9. Reference

Ahmad, S 1996, 'Regression Estimates of per Capita GDP Based on Purchasing Power Parities ', in J Salazar-Carrillo & PDS Rao (eds), *International Comparisons of Prices, Output and Productivity:Contributions to Economic Analysis Series*, Elsevier, North Holland.

Asian Development Bank 2005, 'Purchasing Power Parities and Real Expenditures', <www.adb.org/Documents/Reports/ICP-Purchasing-Power-Expenditures/default.asp>.

Banks, AS 2006, *Cross-National Time-Series Data Archive*, Databanks International, Bronx, NY.

Barro, RJ & Sala-i-Martin, XX 2004, *Economic Growth*, Second edn, MA: MIT Press, Cambridge.

Bergstrand, JH 1991, 'Structural Determinants of Real Exchange Rates and National Price Levels: Some Empirical Evidence', *American Economic Review*, vol. 81, no. 1, pp. 325-34.

---- 1996, 'Productivity, Factor Endowments, Military Expenditures, and National Price Levels ', in PDS Rao & J Salazar-Carrillo (eds), *International Comparisons of Prices, Ouput and Productivity*, Elsevier Science Publishers B.V., North Holland.

Castles, I & Hendersen, D 2003, 'IPPC Issues: A Swag of Documents', viewed 14 December, 2007, < http://www.lavoisier.com.au/papers/articles/IPPCissues.html>.

CIA 2008, *The CIA World Factbook* The Central Intelligence Agency, Langley, Virginia, U.S.A., <https://www.cia.gov/library/publications/the-world-factbook/>.

Clague, C 1988, 'Explanations of National Price Levels', in PDS Rao & J Salazar-Carrillo (eds), *World Comparisons of Incomes, Prices and Product: Contributions to Economic Analysis series: no 173* Elsevier, North Holland.

Doran, HE 1992, 'Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions', *Review of Economics and Statistics*, vol. 74, no. 3, pp. 568-72

Doran, HE, Rambaldi, AN & Rao, PDS 2006, *Construction of Consistent Panels of Purchasing Power Parities (PPPs) for Comparisons of Real Incomes across Countries: A State-Space approach*, CEPA Working Paper Series 06/2006, University of Queensland, Brisbane.

Durbin, J & Koopman, SJ 2001, *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

---- 2002, 'A Simple and Efficient Simulation Smoother for State Space Time Series Analysis ', *Biometrika*, vol. 89, pp. 603-16.

Durlauf, SN, Johnson, P & Temple, J 2005, 'Growth Econometrics', in P Aghion & SN Durlauf (eds), *Handbook of Economic Growth*, Elselvier, Amsterdam vol. 1, pp. 555-677.

Easterly, WR 2006, *Global Development Network Growth Database*, NYU DRI, <http://www.nyu.edu/fas/institute/dri/global development network growth database.htm>.

European Central Bank 1998, *31 December 1998 - Determination of the euro conversion rates*, Press Release, <http://www.ecb.eu/press/pr/date/1998/html/pr981231_2.en.html>.

FAO 2006, *AGROSTAT*, Food and Agriculture Organization, <http://faostat.fao.org/default.aspx>.

Harvey, AC 1990, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press Cambridge.

Harvey, AC, Trimbur, TM & van Dijk, HK 2005, *Trends and Cycles in Economic Time Series: A Bayesian Approach*, manuscript.

Heston, A, Summers, R & Aten, B 2006, *Penn World Table Version 6.2*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, <http://pwt.econ.upenn.edu/php_site/pwt_index.php>.

IMF 2007, 'International Financial Statistics (IFS)', no. The International Monetary Fund.

Koop, G & van Dijk, HK 2000, 'Testing For Integration Using Evolving Trend and Seasonal Models: A Bayesian Approach', *Journal of Econometrics*, vol. 97, pp. 261-91.

Kravis, IB & Lipsey, RE 1983, *Toward an Explanation of National Price Levels, Princeton Studies in International Finance, No. 52* Princeton University Press, Princeton.

---- 1986, 'The Assessment of National Price Levels', paper presented to Eastern Economic Association Meetings, Philadelphia, April.

Kravis, IB, Summers, R & Heston, A 1982, *World Product and Income*, Johns Hopkins University Press Baltimore.

Maddison, A 1995, *Monitoring the World Economy, 1820-1992,* , OECD, Paris.

---- 2007, *Contours of the World Economy 1—2030 AD: Essays in Macro-Economic History,* , Oxford University Press, Oxford.

McKibbin, WJ & Stegman, A 2005, *Convergence and Per Capita Carbon Emissions*, Working Paper 04.05, Lowey Institute for International Policy, Sydney.

Milanovic, B 2002, 'True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone', *The Economic Journal*, vol. 112, pp. 51-92.

Rao, PDS, Rambaldi, AN & Doran, HE 2007, 'Constructing a Panel of Purchasing Power Parities (PPPs): A State-Space Approach', paper presented to Technical Advisory Group Meeting on International Comparison Program (ICP) Washington DC, 6th November.

Rose, A 2004, 'Bilateral trade data set', viewed January 2007, <http://faculty.haas.berkeley.edu/arose/RecRes.htm>.

Sala-i-Martin, XX 2002, *15 Years of New Growth Economics: What have we learnt?*, Unpublished manuscript.

Summers, R & Heston, A 1988, 'Comparing International Comparisons', in J Salazar-Carrillo & PDS Rao (eds), *World Comparisons of Incomes, Prices and Product, Contributions to Economic Analysis Series, no 173*, North-Holland, Amsterdam.

---- 1991, 'The Penn World Tables (Mark 5): An expanded set of international comparisons, 1950-88', *Quarterly Journal of Economics*, vol. 2, no. 1, pp. 1-45.

UN 2007, *National Accounts Main Aggregates Database*, United Nations, viewed June 2007, <http://unstats.un.org/unsd/snaama/>.

UNDP 2006 *Human Development Report*, United Nations Development Program, New York.

World Bank 2006 and various years, *World Development Indicators*, viewed May 2007 <http://devdata.worldbank.org/wdi2006/contents/cover.htm>.

Appendix 1: Preserving Movements in Implicit GDP Deflators through the Smoothing Filter

In this appendix we show that using a fixed interval smoother with **Q=0** (see Harvey 1990, p 154 for the relevant recursive formulae)[26], the smoothed estimates of the state vector, $\mathbf{p}^*_{t|T}$, preserve the movement in the implicit price deflator and its covariance matrix converges to the Kalman filter estimate of the covariance at time *T*.

The smoothed covariance matrix are given by

$$\boldsymbol{\Psi}^*_t = \boldsymbol{\Psi}_t \boldsymbol{\Psi}^{-1}_{t+1|t} \qquad \text{(A1.1)}$$

Where,

$\boldsymbol{\Psi}_t$ is the Kalman filter unconditional covariance of the state vector
$\boldsymbol{\Psi}^{-1}_{t+1|t}$ is the Kalman filter conditional covariance of the state vector

Now, if $\mathbf{Q}=\mathbf{0}$, $\boldsymbol{\Psi}_{t+1|t} = \boldsymbol{\Psi}_t$, which from (A1.1) implies $\boldsymbol{\Psi}^*_t = \mathbf{I}_N$. Therefore, $\mathbf{p}^*_{t|T} = \mathbf{p}^*_{t+1|T} - \mathbf{c}_{t+1}$, or

$$\mathbf{p}^*_{t+1|T} = \mathbf{p}^*_{t|T} + \mathbf{c}_{t+1} \qquad \text{(A1.2)}$$

That is, smoothed estimates, $\mathbf{p}^*_{t|T}$ preserve the movement in the implicit price deflator.

Now considering the covariance matrices, we have

$$\boldsymbol{\Psi}^*_{t|T} = \boldsymbol{\Psi}_t + \boldsymbol{\Psi}_t(\boldsymbol{\Psi}^*_{t+1|T} - \boldsymbol{\Psi}_{t+1|t})\boldsymbol{\Psi}'_t$$

Because,

$\boldsymbol{\Psi}_{t+1|t} = \boldsymbol{\Psi}_t$ and $\boldsymbol{\Psi}^*_t = \mathbf{I}_N$ it follows that $\boldsymbol{\Psi}^*_{t|T} = \boldsymbol{\Psi}^*_{t+1|T}$. Thus, $\boldsymbol{\Psi}^*_{t|T}$ is constant with respect to *t* and,

$$\boldsymbol{\Psi}^*_{t|T} = \boldsymbol{\Psi}^*_{T|T} = \boldsymbol{\Psi}_{T|T} \qquad \text{for all } t \qquad \text{(A1.3)}$$

[26] In our model Harvey's $\mathbf{T}_t = \mathbf{I}_N$. The recursive formulae of the smoother will be included for completeness in the next version of the paper.

Appendix 2

Suppose there are $M+1$ benchmark years at times $t(0), t(1), ..., t(M)$, where $t(0)=0,$ and no information is added between benchmark years.

Let $\breve{\mathbf{p}}_T$ be the Kalman filter estimate of $\mathbf{p}_T$ and $\breve{\mathbf{p}}_{T,j}$, $j=0,1,...,M$ be the $M+1$ different estimates of $\mathbf{p}_T$ obtained by applying growth rates to the benchmark observations until time $t=T$. Further, we define $\mathbf{G}(i)$, the *Kalman gain*[27] at $t=t(i)$ which in our case takes the form:

$$\mathbf{G}(i)=\begin{cases}\mathbf{\Psi}_{t|t-1}\mathbf{F}_t^{-1} & \text{for } i>0\\ \mathbf{I} & \text{for } i=0\end{cases} \qquad \text{(A2.1)}$$

Proposition

The Kalman filter estimate, $\breve{\mathbf{p}}_T$, is a weighted sum of the $\breve{\mathbf{p}}_{T,j}$, $j=0,1,...,M$.

That is,

$$\breve{\mathbf{p}}_T=\sum_{i=0}^{M}\mathbf{W}_i^{(M)}\breve{\mathbf{p}}_{T,i} \qquad \text{(A2.2)}$$

Where the weights $\mathbf{W}_i^{(M)}$ are defined as

$$\mathbf{W}_i^{(M)}=\begin{cases}\left[\prod_{j=1}^{M-i}(\mathbf{I}-\mathbf{G}(M-j+1))\right]\mathbf{G}(i) & \text{for } i=0,1,...,M-1\\ \mathbf{G}(i) & i=M\end{cases} \qquad \text{(A2.3)}$$

Lemma

The $\mathbf{W}_i^{(M)}$ defined in (A2.3) are the product of positive definite (pd) matrices and

$$\sum_{i=0}^{M}\mathbf{W}_i^{(M)}=\mathbf{I}_N \qquad \text{(A2.4)}$$

Proof of Lemma

From Harvey (Harvey 1990, p. 106)

$$\mathbf{\Psi}_{t|t-1}=\mathbf{\Psi}_{t-1}+\mathbf{Q}_t \qquad \text{(as here } \mathbf{T}_t,\mathbf{R}_t=\mathbf{I}_N\text{)}$$

with $\mathbf{\Psi}_{t-1}$ being positive semidefinite (psd) or pd and $\mathbf{Q}_t$ being pd. Therefore, $\mathbf{\Psi}_{t|t-1}$ is pd for all *t*. Also, by definition (see Harvey (1990) pp. 106) $\mathbf{F}_t=\mathbf{\Psi}_{t|t-1}+\mathbf{H}_t$ and must be pd as $\mathbf{H}_t$ is pd. Thus, $\mathbf{G}(i)$ is the product of pd matrices for all *i*.

---

[27] See Harvey (1990), p 110

Also, post-multiplying the above equation for $\mathbf{F}_t$ by $\mathbf{F}_t^{-1}$, we have

$$\begin{aligned}\mathbf{I}_N &= \mathbf{\Psi}_{t|t-1}\mathbf{F}_t^{-1} + \mathbf{H}_t\mathbf{F}_t^{-1} \\ &= \mathbf{G}(i) + \mathbf{H}_t\mathbf{F}_t^{-1}\end{aligned}$$

Therefore, $\mathbf{I}_N - \mathbf{G}(i) = \mathbf{H}_t\mathbf{F}_t^{-1}$, and is also the product of pd matrices for all i. Thus, it follows that by (A2.3) $\mathbf{W}_i^{(M)}$ is the product of pd matrices.

We will now establish that for $\mathbf{W}_i^{(M)}$ defined by (A2.3), (A2.4) holds. The proof will proceed by induction and we note that the form of $\mathbf{W}_i^{(M)}$ in (A2.3) implies that

$$\mathbf{W}_i^{(M)} = \left[\mathbf{I} - \mathbf{G}(M)\right]\mathbf{W}_i^{(M-1)} \tag{A2.5}$$

We will now assume that (A2.4) is true for $M-1$. That is,

$$\sum_{i=0}^{M-1} \mathbf{W}_i^{(M-1)} = \mathbf{I}_N \tag{A2.6}$$

Then from (A2.5) and (A2.4)

$$\begin{aligned}\sum_{i=0}^{M} \mathbf{W}_i^{(M)} &= \sum_{i=0}^{M-1} \mathbf{W}_i^{(M)} + \mathbf{W}_M^{(M)} \\ &= [\mathbf{I} - \mathbf{G}(M)]\sum_{i=0}^{M-1} \mathbf{W}_i^{(M-1)} + \mathbf{G}(M),\end{aligned}$$

And so by the assumption (A2.6)

$$\sum_{i=0}^{M} \mathbf{W}_i^{M} = \mathbf{I}_N$$

Therefore if (A2.4) is true for $M-1$, it is also true for M.

Now, set $M = 1$

$$\sum_{i=0}^{M} \mathbf{W}_i^{(M)} = \mathbf{W}_0^{(1)} + \mathbf{W}_1^{(1)}$$

From (A2.3) and (A2.1)

$$\mathbf{W}_o^{(1)} = (\mathbf{I} - \mathbf{G}(1)), \qquad \mathbf{W}_1^{(1)} = \mathbf{G}(1) \tag{A2.7}$$

Therefore, (A2.4) is true for $M = 1$ and so, by induction,

$$\sum_{i=0}^{M} \mathbf{W}_i^{(M)} = \mathbf{I}_N \qquad \text{for all } M \text{ as required.}$$

Proof of Proposition

In order to ease the notational burden, we will prove (A2.2) first for the case T= *t(M)* and then extend to the case T > *t(M)*

Assume (A2.2) and (A2.3) are true for T = *t(M-1)*.

That is,

$$\breve{\mathbf{p}}_{t(M-1)} = \sum_{i=0}^{M-1} \mathbf{W}_i^{(M-1)} \breve{\mathbf{p}}_{(M-1),i} \tag{A2.8}$$

Now, at $t = t(M)$ a benchmark observation, $\mathbf{y}(M)$, becomes available. By definition

$$\breve{\mathbf{p}}_{t(M),M} = \mathbf{y}(M).$$

The Kalman filter updating formula (see Harvey (1990), p 106) gives:

$$\breve{\mathbf{p}}_{t(M)} = (\breve{\mathbf{p}}_{t(M-1)} + \overline{\mathbf{c}}) + \mathbf{G}(M)[\mathbf{y}(M) - (\breve{\mathbf{p}}_{t(M-1)} - \overline{\mathbf{c}})] \tag{A2.9}$$

where $\overline{\mathbf{c}}$ is the cumulated growth rates from $t(M-1)$ to $t(M)$.

Thus,

$$\breve{\mathbf{p}}_{t(M)} = [\mathbf{I} - \mathbf{G}(M)]\left[\breve{\mathbf{p}}_{t(M-1)} + \overline{\mathbf{c}}\right] + \mathbf{G}(M)\breve{\mathbf{p}}_{t(M)|M}$$

Now, by assumption (A2.8)

$$\begin{aligned}\breve{\mathbf{p}}_{t(M-1)} + \overline{\mathbf{c}} &= \sum_{i=0}^{M-1} \mathbf{W}_i^{(M-1)} \breve{\mathbf{p}}_{(M-1),i} + \overline{\mathbf{c}} \\ &= \sum_{i=0}^{M-1} \mathbf{W}_i^{(M-1)} \left(\breve{\mathbf{p}}_{(M-1),i} + \overline{\mathbf{c}}\right) \qquad \text{(by A2.6)} \\ &= \sum_{i=0}^{M-1} \mathbf{W}_i^{(M-1)} \breve{\mathbf{p}}_{t(M),i}\end{aligned}$$

Thus,

$$\breve{\mathbf{p}}_{t(M)} = \sum_{i=0}^{M-1} [\mathbf{I} - \mathbf{G}(M)]\mathbf{W}_i^{(M-1)} \breve{\mathbf{p}}_{M,i} + \mathbf{G}(M)\breve{\mathbf{p}}_{t(M)|M}$$

$$= \sum_{i=0}^{M} \mathbf{W}_i^{(M)} \breve{\mathbf{p}}_{t(M),i}$$

And so if (A2.2) and (A2.3) are true for t($M-1$), then they are also true for t(M).

Now set $M = 1$. This implies two benchmark years, at t(0)=0 and t(1). By definition,

$\breve{\mathbf{p}}_{t(1),0} = \breve{\mathbf{p}}_0 = \mathbf{y}(0)$, and $\breve{\mathbf{p}}_{t(1),1} = \mathbf{y}(1)$.

Then, using the Kalman updating formula,

$$\breve{\mathbf{p}}_{t(1)} = \left[\mathbf{I} - \mathbf{G}(1)\right](\breve{\mathbf{p}}_0 + \overline{\mathbf{c}}) + \mathbf{G}(1)\mathbf{y}(1)$$

$$= \left[\mathbf{I} - \mathbf{G}(1)\right]\breve{\mathbf{p}}_{t(1),0} + \mathbf{G}(1)\breve{\mathbf{p}}_{t(1),1}$$

$$= \mathbf{W}_0^{(1)}\breve{\mathbf{p}}_{t(1),0} + \mathbf{W}_1^{(1)}\breve{\mathbf{p}}_{t(1),1} \qquad \text{(by (A2.3))}$$

Thus (A2.2) and (A2.3) hold for M=1, and hence, by induction, for all M.

We can now easily extend the result for T > t(M). If we denote the cumulated growth rates from t(M) to T by $\overline{\mathbf{c}}$, then

$$\breve{\mathbf{p}}_T = \breve{\mathbf{p}}_{t(M)} + \overline{\mathbf{c}}$$

$$= \sum_{i=0}^{M} \mathbf{W}_i^{(M)} \breve{\mathbf{p}}_{t(M),i} + \overline{\mathbf{c}}$$

$$= \sum_{i=0}^{M} \mathbf{W}_i^{(M)} (\breve{\mathbf{p}}_{t(M),i} + \overline{\mathbf{c}})$$

$$\breve{\mathbf{p}}_T = \sum_{i=0}^{M} \mathbf{W}_i^{(M)} \breve{\mathbf{p}}_{T,i}$$

Special case

If the elements of η_t and ξ_t are contemporaneously uncorrelated (that is, $\mathbf{Q}_t$ and $\mathbf{H}_t$ are diagonal) it is easily shown that the $\mathbf{W}_i^{(M)}$ are diagonal and positive definite for all $i = 1,...,M$, provided $\mathbf{\Psi}_0 = 0$.

Suppose that $\breve{p}_{jT}$ and $\breve{p}_{jT,i}$ are the Kalman filter and benchmark estimates (from the *ith* benchmark) of the PPP of country *j* at time *t* = T > *t*(*M*). Denote by $W_{jj,i}^{(M)}$ the *jth* diagonal element of $\mathbf{W}_i^{(M)}$. It then follows that

$$\breve{p}_{jT} = \sum_{i=0}^{M} W_{jj,i}^{(M)} \breve{p}_{jT,i}$$

Furthermore, because $\mathbf{W}_i^{(M)}$ is pd, and from (A2.4), it follows that $W_{jj,i}^{(M)} > 0$ and $\sum_{i=0}^{M} W_{jj,i}^{(M)} = 1$.

Thus, in this special case the Kalman filter estimate for country *j* is *weighted average* of the M+1 "benchmark only" estimates for that country. The weights are not arbitrary, but determined by the fundamental covariance matrices $\mathbf{Q}_t$ and $\mathbf{H}_t$.

Table DA.1 List of Countries Used in the Study

Country	IFS code	Currency	Exchange rate (2005)	Years Participated in the ICP Benchmark
African Region				
Algeria	612	Algerian Dinars	73.28	
Benin	638	Communaute Financiere Africaine Franc (XOF)[28]	527.47	1996,2005
Botswana	616	Botswana Pula	5.11	1980,1985,1996,2005
Burkina Faso	748	Communaute Financiere Africaine Franc (XOF)	527.47	2005
Burundi	618	Burundi Franc	1081.58	2005
Cameroon	622	Communaute Financiere Africaine Franc (XAF)[29]	527.47	1980,1985,1996,2005
Central African Republic	626	Communaute Financiere Africaine Franc (XAF)	527.47	2005
Chad	628	Communaute Financiere Africaine Franc (XAF)	527.47	2005
Congo, Dem. Rep.	636	Congolese Franc	473.91	2005
Congo, Rep.	634	Communaute Financiere Africaine Franc (XAF)	527.47	1985,2005
Cote d'Ivoire	662	Communaute Financiere Africaine Franc (XOF);	527.47	1980,1985,1996,2005
Ethiopia	644	Ethiopian Birr	8.65	1980,1985,2005
Gabon	646	Communaute Financiere Africaine Franc (XAF)	527.47	1996,2005
Gambia	648	Dalasi (GMD)	28.58	2005
Ghana	652	Ghana Cedi (GHC)	9072.54	2005
Guinea	656	New Franc Guineen	3640.04	1996,2005
Guinea-Bissau	654	Communaute Financiere Africaine Franc (XOF)	527.47	2005
Kenya	664	Kenyan Shilling (KES)	75.55	1975,1980,1985,1996,2005
Lesotho	666	Loti (LSL)	6.36	2005
Libya	672	Libyan Dinar (LYD)	1.31	
Madagascar	674	Madagascar Ariary	2003.03	1980,1985,1996,2005
Malawi	676	Malawi Kwacha	118.40	1975,1980,1985,1996,2005
Mali	678	Communaute Financiere Africaine Franc (XOF)	527.47	1980,1985,1996,2005
Mauritania	682	Ouguiya (MRO)	268.60	2005
Mauritius	684	Mauritius Rupee	28.94	1985,1996,2005
Morocco	686	Moroccan Dirham	8.86	1980,1985,1996,2005
Mozambique	688	Metical (MZM)	23060.97	2005
Namibia	728	Namibian Dollar	6.36	2005
Niger	692	Communaute Financiere Africaine Franc (XOF)	527.47	2005
Nigeria	694	Nigerian Naira	131.27	1980,1985,1996,2005
Rwanda	714	Rwanda Franc	557.82	1985,2005
Sao Tome and Principe	716	Dobra (STD)	10558.00	2005

[28] The responsible Authority of Communaute Financiere Africaine Franc (XOF) is the Central Bank of the West African States.
[29] The responsible Authority of Communaute Financiere Africaine Franc (XAF) is the Bank of the Central African States.

Table DA.1 continued				
Country	**IFS code**	**Currency**	**Exchange rate (2005)**	**Years Participated in the ICP Benchmark**
African Region				
Senegal	722	Communaute Financiere Africaine Franc (XOF)	527.47	1980,1985,1996,2005
Seychelles	718	Seychelles Rupee (SCR)	5.50	
Sierra Leone	724	Sierra Leone Leone	2889.59	1985,1996,2005
South Africa	199	Rand (ZAR)	6.36	2005
Sudan	732	Sudanese Dinar (SDD)	243.61	2005
Swaziland	734	Swaziland Lilangeni	6.36	1985,1996,2005
Tanzania	738	Tanzania Shilling	1128.93	1980,1985,1996,2005
Togo	742	Communaute Financiere Africaine Franc (XOF);	527.47	2005
Tunisia	744	Tunisian Dinar	1.30	1980,1985,1996,2005
Uganda	746	Ugandan Shilling (UGX)	1737.23	2005
Zambia	754	Zambian Kwacha	4463.50	1975,1980,1985,1996,2005
Zimbabwe	698	Zimbabwe Dollar	22363.64	1980,1985,1996,2005
Asian/Pacific Regions				
Bangladesh	513	Bangladesh Taka	61.75	1985,1996,2005
China	924	Renminbi (RMB) (Yuan)	8.19	2005
Fiji	819	Fiji Dollar	1.69	1996,2005
Hong Kong	532	Hong Kong Dollar	7.78	1980,1985,1996,2005
India	534	Indian Rupee	44.27	1975,1980,1985,2005
Indonesia	536	Indonesian Rupiah	9705.00	1980,1996,2005
Iran	429	Iranian Rial	8963.96	1975,1985,1996,2005
Kiribati	826	Australian Dollar (AUD)	1.31	
Malaysia	548	Malaysian Ringgit	3.79	1975,2005
Maldives	556	Rufiyaa (MVR)	12.80	2005
Mongolia	948	Mongolia Tugrik	1205.30	1996,2005
Nepal	558	Nepalese Rupee	72.19	1996,2005
Pakistan	564	Pakistan Rupee	59.13	1975,1980,1985,1996,2005
Papua New Guinea	853	Kina (PGK)	3.10	
Philippines	566	Philippine Peso	55.09	1975,1980,1985,1996,2005
Samoa	862	Tala (SAT)	2.71	
Singapore	576	Singapore Dollar	1.66	1996,2005
Solomon Islands	813	Solomon Islands Dollar (SBD)	7.53	
Sri Lanka	524	Sri Lanka Rupee	100.50	1975,1980,1985,1996,2005
Thailand	578	Thailand Baht	40.22	1975,1985,1996,2005
Tonga	866	Pa'anga (TOP)	1.94	
Vanuatu	846	Vatu (VUV)	109.25	
Vietnam	582	Viet Nam Dong	15858.92	1996,2005
OECD and Eurostat				
Albania	914	Albanian Lek	99.86	1996,2005
Australia	193	Australian Dollar	1.31	1985,1990,1993,1996,1999,2002, 2005
Austria	122	Euro	0.80	1975,1980,1985,1990,1993,1996, 1999,2002,2005

Table DA.1 continued				
Country	**IFS code**	**Currency**	**Exchange rate (2005)**	**Years Participated in the ICP Benchmark**
OECD and Eurostat				
Belgium	124	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
Bulgaria	918	Bulgarian Lev	1.57	1996,2005
Canada	156	Canadian Dollar	1.21	1980,1985,1990,1993,1996,1999,2002,2005
Cyprus	423	Cypriot Pound (CYP)	0.46	2005
Denmark	128	Danish Krone	6.00	1975,1980,1985,1990,1993,1996,1999,2002,2005
Finland	172	Euro	0.80	1980,1985,1990,1993,1996,1999,2002,2005
France	132	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
Germany	134	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
Greece	174	Euro	0.80	1980,1985,1990,1993,1996,1999,2002,2005
Hungary	944	Hungary Forint	199.58	1975,1980,1985,1996,1999,2002,2005
Iceland	176	Iceland Krona	62.98	1990,1993,1996,1999,2002,2005
Ireland	178	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
Israel	436	Israel New Sheqel	4.49	1980,1996,2005
Italy	136	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
Japan	158	Japanese Yen	110.22	1975,1980,1985,1990,1993,1996,1999,2002,2005
Korea	542	South Korean Won	1024.12	1975,1980,1985,1999,2002,2005
Luxembourg	137	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
Malta	181	Maltese Lira (MTL)	0.35	2005
Mexico	273	Mexico Peso	10.90	1975,1996,1999,2002,2005
Netherlands	138	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
New Zealand	196	NZ Dollar	1.42	1985,1990,1993,1996,1999,2002, 2005
Norway	142	Norway Krone	6.44	1980,1985,1990,1993,1996,1999, 2002,2005
Poland	964	Polish Zloty	3.23	1975,1980,1985,1996,1999,2002,2005
Portugal	182	Euro	0.80	1980,1985,1990,1993,1996,1999,2002,2005
Romania	968	Romanian New Leu	2.91	1975,1996,2005
Spain	184	Euro	0.80	1975,1980,1985,1990,1993,1996,1999,2002,2005
Sweden	144	Swedish Krona	7.47	1985,1990,1993,1996,1999,2002,2005
Switzerland	146	Swiss Franc	1.25	1990,1993,1996,1999,2002,2005

Table DA.1 continued				
Country	**IFS code**	**Currency**	**Exchange rate (2005)**	**Years Participated in the ICP Benchmark**
OECD and Eurostat				
Turkey	186	Turkish Lira	1.34	1985,1990,1993,1996,1999,2002,2005
United Kingdom	112	British Pound Sterling	0.55	1975,1980,1985,1990,1993,1996,1999,2002,2005
United States	111	United States Dollar	1.00	1975,1980,1985,1990,1993,1996,1999,2002,2005
South, North, Central American regions and Caribbean Countries				
Antigua and Barbuda	311	East Caribbean Dollars	2.70	1996
Argentina	213	Argentina Peso Convertible	2.90	1980,1996,2005
Bahamas	313	Bahamian Dollars	1.00	
Barbados	316	Barbados Dollar	2.01	1996
Belize	339	Belize Dollar	2.00	1996
Bermuda	319	Bermudan Dollar	1.00	1996
Bolivia	218	Boliviano	8.07	1980,1996,2005
Brazil	223	Brazil Real	2.43	1975,1980,1996,2005
Chile	228	Chilean Peso	560.09	1980,1996,2005
Colombia	233	Colombia Peso	2320.75	1975,1980,2005
Costa Rica	238	Costa Rican Colon	477.79	1980
Dominica	321	East Caribbean Dollar	2.70	1996
Dominican Republic	243	Dominican Peso	30.41	1980
Ecuador	248	United States Dollar	1.00	1980,1996,2005
El Salvador	253	United States Dollar	1.00	1980
Guatemala	258	Guatemala Quetzal	7.62	1980
Guyana	336	Guyanese Dollar	199.88	
Haiti	263	Gourde (HTG)	40.45	
Honduras	268	Honduras Lempira	19.00	1980
Jamaica	343	Jamaica Dollar	62.28	1975,1996
Nicaragua	278	Gold Cordoba (NIO)	16.73	
Panama	283	United States Dollar	1.00	1980,1996
Paraguay	288	Paraguay Guarani	6177.96	1980,2005
Peru	293	Peru Sol Nuevo	3.30	1980,1996,2005
St. Kitts and Nevis	361	East Caribbean Dollar	2.70	1996
St. Lucia	362	East Caribbean Dollar	2.70	1996
St. Vincent and the Grenadines	364	East Caribbean Dollar	2.70	1996
Suriname	366	Surinam Dollar (SRD)	2.73	1985
Trinidad and Tobago	369	Trinidad and Tobago Dollar	6.30	1996
Uruguay	298	Uruguay Peso	24.48	1975,1980,1996,2005
Venezuela	299	Venezuela Bolivar	2089.75	1985,2005

Table DA.1 continued				
Country	**IFS code**	**Currency**	**Exchange rate (2005)**	**Years Participated in the ICP Benchmark**
West Asia				
Egypt	469	Egyptian Pound	6.00	1985,1996,2005
Jordan	439	Jordan Dinar	0.71	1996,2005
Kuwait	443	Kuwaiti Dinar (KD)	0.29	2005
Lebanon	446	Lebanese Pound	1507.50	1996,2005
Oman	449	Omani rial (OMR)	0.38	1996,2005
Qatar	453	Qatari Rial	3.64	1996,2005
Saudi Arabia	456	Saudi Riyal (SAR)	3.75	2005
Syria	463	Syrian Pound	52.14	1975,1996,2005
United Arab Emirates	466	Emirati Dirham (AED)	3.67	

Notes: Countries were excluded on the basis insufficient data (either in terms of 'x variables' or omission of sufficient trade data for the contiguity matrix). In addition, some countries are excluded due to their small size or political instability. The list of omissions includes mainly newly formed countries such as former Soviet Republics and those of former Yugoslavia and Czechoslovakia. Since North Korea is not included due to data unavailability, Korea refers to South Korea

Table DA.2. Definitions and sources of variables.

Code	Definitions & Notes	Source
Socio Economic Variables, ICP Benchmark and Exchange Rate Data		
Agedep	Age dependency ratio (dependents to working-age population): Age dependency ratio is the ratio of dependents - people younger than 15 and older than 64 - to the working-age population (those aged 15-64). For example, 0.7 means there are 7 dependents for every 10 working-age people.	WDI
Agvag	Agriculture, value added (as a percentage of GDP).	UN
Blackind	Index of distortions in exchange rates which takes the values 0, 1, 2. The index is based on the average black market premium in the exchange rate (BMP) for the last 5 years, starting in 1960. The index is equal to 0 if the BMP has been less than 20%; equal to 1 if greater than 20% and equal to 2 if greater than 100%.	Authors' calculations .
dum80_84	dummy for years 1980 to 1984	
dum85_89	dummy for years 1985 to 1989	
dum90_92	dummy for years 1990 to 1992	
dum93_95	dummy for years 1993 to 1995	
dum96_98	dummy for years 1996 to 1998	
dum99_01	dummy for years 1999 to 2001	
dum02_04	dummy for years 2002 to 2004	
dum05	dummy for year 2005	
D_anz	dummy for Australia-New Zealand ANZD agreement.	Rose
D_asean	dummy for ASEAN countries	Rose
D_cac	dummy for CACM countries (Central American)	Rose
D_euro	dummy for countries which have used the euro since 1999	
D_mercsr	dummy for MERCOSUR countries	Rose
D_nafta	dummy for NAFTA countries	Rose
D_scucar	Dummy for strict currency union in Caribbean	Rose
D_spr	dummy for South Pacific Trade and Economic Co-Operation Agreement	Rose
D_usd	dummy variable for countries with currencies either pegged to US for substantial amounts of time (including the post-Bretton Woods era (1973 onwards) or expected to move closely with the US (e.g. Central American countries).	Authors' calculations .

Table DA.2. Continued		
Code	**Definitions & Notes**	**Source**
Expg	Exports of goods and services (% of GDP): Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude labour and property income (formerly called factor services) as well as transfer payments.	UN
gdpdefl	The GDP implicit deflator is the ratio of GDP in current local currency to GDP in constant local currency. The WDI base year varied by country but was converted to a 2002 base year for all countries by the authors.	WDI and UN
Labpop	Labor force as percentage of total population. For developing countries the labour force is simply defined as the "economically active" population, which is itself based on age groups.	FAO/ILO
Life	Life expectancy at birth, total (years): Life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life.	WDI
Literate	population aged 15 and over which is literate (per 1,000 population)	CNTS
Manufexp	Manufactures exports (as a percentage of merchandise exports): Manufactures comprise commodities in SITC sections 5 (chemicals), 6 (basic manufactures), 7 (machinery and transport equipment), and 8 (miscellaneous manufactured goods), excluding division 68 (non-ferrous metals).	WDI
Manufimp	Manufactures imports (as a percentage of merchandise imports): Manufactures comprise the commodities in SITC sections 5 (chemicals), 6 (basic manufactures), 7 (machinery and transport equipment), and 8 (miscellaneous manufactured goods), excluding division 68 (nonferrous metals).	WDI
ntrvag2	Non-tradable sector value added (as a percentage of GDP) - definition 2: sum of Construction, Wholesale, retail trade, restaurants and hotels, Transport, storage and communication and "Other Activities"	UN
Phones	Telephone mainlines (per 1,000 people): Telephone mainlines are telephone lines connecting a customer's equipment to the public switched telephone network. Data are presented per 1,000 people for the entire country.	WDI
ICP PPP benchmarks	Contains data from original PPP surveys. All benchmarks were converted to the currency in which official exchange rates are currently expressed (i.e. the current currency of the countries in question). For instance, all Euro country PPPs were already expressed euros, except for 1975 benchmark, which was therefore converted to euro-expressed PPP for 1975 using the 1999 euro-domestic currency conversion rate.	STARS, World Bank
Radpccn	Radios per capita (rescaled).	CNTS
Rurpop	Rural population (as a percentage of total population): Rural population is calculated as the difference between the total population and the urban population.	WDI
Tractorpw	Agricultural machinery, tractors per agricultural worker: Agricultural machinery refers to the number of wheel and crawler tractors (excluding garden tractors) in use in agriculture at the end of the calendar year specified or during the first quarter of the following year.	WDI
Tradeg	Trade (as a percentage of GDP): Trade is the sum of exports and imports of goods and services measured as a share of gross domestic product.	UN

Table DA.2. Continued		
Code	**Definitions & Notes**	**Source**
ER	This is chiefly the Exchange rates series with some 'decimal point' corrections for Afghanistan, Zimbabwe and 1 or 2 other countries, and all countries euro countries converted to euro-based exchange rates. However, UN National Accounts sources (also based on IMF data) were used to fill in missing values for several countries.	WDI, IMF
Covariance Variables		
W Contiguity Matrix	A value of 1 is assigned to the five closest trading partners (ie five countries with the highest bilateral trade volume) of a particular country.	Rose, IMF
Ypccus	GDP per capita, constant USD (converted using exchange rates)	UN

Notes: Source definitions are as follows: WDI= World Development Indicators (WDI 2007); UN=United Nations National Accounts Main Aggregates Database (UN 2007); IMF=International Monetary Fund (IMF 2007) IMF trade directions (various years) ; PWT 6.1=Penn World Tables Version 6.1 (Summers, Robert & Heston 2002); Rose=Bilateral trade data from Andrew K. Rose (Rose, A 2004; Rose, AK 2004); CNTS=Cross-national Time Series data (Banks 2006); FAO=Food and Agricultural Organization (FAO 2006). Also note that *blackind* was calculated from black market premium data from Easterly (2006).

Table DA.3.Descriptive Statistics for Variables

Variable Code	Mean	Median	Max.	Min.	Std. Dev.
Agedep	0.75	0.78	1.35	0.30	0.19
Agvag	18.78	15.76	74.27	0.06	14.91
Blackind	0.33	0.00	2.00	0.00	0.63
Expg	35.95	29.00	244.30	1.00	25.3
Gdpdefl	0.50	0.46	34.73	0.00	0.70
Labpop	43.00	43.48	74.19	23.31	7.18
Life	62.86	66.26	82.08	34.22	11.60
Literate (0-1000)	712	810	999	40	271
Manufexp	36.45	26.95	107.27	0.00	30.39
Manufimp	65.90	66.46	152.93	0.19	14.11
Ntrvag2	29.80	29.26	56.12	3.06	8.55
phones	126.95	36.64	869.83	0.12	178.12
ICP PPP Benchmark	19.72	0.00	33068	0.00	512.41
Radpccn	3547.46	2523.5	25349.00	49.00	3353.05
rurpop	51.18	52.75	97.61	0.00	24.94
tractorpw	0.15	0.01	1.97	0.00	0.33
tradbg_un	-5.6	-3.00	79.00	-233	16.63
tradeg	77.54	66.00	463.00	5.00	48.74
ER	252.23	344	23782.27	0.00	1346.54
Ypccus	5472.53	1324.67	53315.79	37.61	8344

Notes: See Table DA.2 for definitions of variables. This table does not include descriptive statistics for dummies.