An Econometric Approach to the Extrapolation of PPPs to Non-Benchmark Years

D.S. Prasada Rao* and Alicia N. Rambaldi†

School of Economics, The University of Queensland. St Lucia, QLD 4072. Australia.
version: October 2015 - 3rd Regional Coordinators Meeting

Abstract

The paper presents the development and implementation of a coherent econometric framework to extrapolate PPPs using information on PPPs from the benchmark data collected by the International Comparison Program as well as the information on price deflators from national sources. The paper describes the complete version of a method originally presented in Rao et al. [2010], referred to as RRD from here on. The basic framework of the method including a discussion of the stochastic assumptions are presented as a preamble to the analytical properties of the methodology with proofs of the main propositions provided in the Appendix. In particular, we are able to show that the methodology is invariant to the choice of the reference country and that the model is flexible enough to either track the PPPs in different benchmarks or to track the national price movements. A snapshot of the constructed panels resulting from the methodology and now available to users from a dedicated website is presented. The website provides PPPs at the GDP level and for the three main components, viz., Consumption, Investment and Government Expenditure for comparisons over time available in the form of aggregates at constant prices. Measures of reliability of PPPs are available in the form of standard errors. These series are available from 181 countries over the period 1970-2012.

JEL: C53, C33.

Keywords: real income, current and constant prices, Purchasing Power Parities, UQICD

*Corresponding author. School of Economics, The Universtiy of Queensland, St Lucia, QLD 4072, Australia. Email: p.rao@economics.uq.edu.au

†This paper is a part of a larger project on the construction of panels of PPPs undertaken by Rao and Rambaldi at the University of Queensland. The authors acknowledge the contribution of H.E Doran, co-author of the original methodology, Dr K.R Ganegodage, database manager, L.T Huynh, PhD student working on a number of extensions of the project.

Cross-country comparisons of growth and economic performance require economic aggregates such as the gross domestic product, consumption and investment in real terms and expressed in a common currency unit. Purchasing power parities (PPPs)[1] reflecting the relative price levels in different countries are considered superior to the market exchange rates for this purpose[2]. Catch-up and convergence of real incomes has been a subject of intense research among development economists (Pritchett, 1997; Quah, 1997; Durlauf et al., 2005; Barro and Sala-i Martin, 2004) – these analyses typically rely on real per capita income data derived using PPPs. The recent debates on globalization and inequality are anchored on PPP-converted income data (Theil, 1989; Milanovic, 2002; Sala-i Martin, 2006; Chotikapanich et al., 2007. The Human Development Index (UNDP, various) and estimates of regional and global incidence showing the number of people under \$1/day and \$2/day regularly published in WDI (various years) rely on PPP data.

The diverse range of applications listed above as well as researchers, analysts and policy makers at the national and international levels critically depend upon the availability of reliable PPP data for a large number of countries covering a long period of time. The reality is that PPP data are sparse and are available for only selected benchmark years from the International Comparison Program (ICP)[3] (see Deaton and Heston [2010] for an overview of the theory and practice of constructing PPPs). The Penn World Tables pioneered by Summers and Heston [1988, 1991] providing PPPs and real incomes covering in excess of 180 countries and a 50-year period largely filled the gap[4]. Consequently, PWT has become one of the most widely used and cited source of data in economics. Details of the construction of PWT (up to and including 7.1) are available from Summers and Heston [1991] and Heston et al. [2006]. The New Generation PWT has commenced with PWT8.0, and a description of the methodology of construction can be found in Feenstra et al. [2015].

Despite the popularity enjoyed by the PWT, there are several directions in which the methodology that underpins the PWT can be improved. The original PWT were anchored

[1]PPP of a currency is defined as the number of currency units required to buy the same goods and services that can be bought with one unit of reference/numeraire currency. For example, a PPP of 130 Japanese yen per US dollar means that we need 130 yen to purchase the same goods and services that can be bought with one US dollar. The Big Mac index is a good example of a PPP which is based only on a single product.

[2]Issues relating to the use of PPPs versus market exchange rates are discussed in detail in Kravis and Lipsey, 1983, in the final report of the 2005 ICP available on the World Bank `http://siteresources.worldbank.org/ICPINT/Resources/ICP_final-results.pdf` and the final report of ICP Asia-Pacific (Asian Development Bank, 2007).

[3]The ICP began as a research project at the University of Pennsylvania conducted by Kravis, Summers and Heston. Currently ICP has grown into a major international statistical program conducted by the World Bank under the guidance of the UN Statistical Commission. The first benchmark comparison was in 1970 and the most recent being the 2005 round of the ICP.

[4]Maddison [1995, 2007] provides real income series for a large number countries covering a long period of time since 1820 expressed in 1990 international dollars.

on a single PWT (versions 6.1, 6.2 and 6.3 based on the 1996 PPP data from the ICP, and 7.1 on the 2005 PPP data from the ICP). The new generation PWT8.0 is based on ICP data from 1970, 1975, 1980, 1985, 1996 and 2005 and OECD/Eurostat benchmarks, and uses an interpolation between ICP benchmark years and an extrapolation using National Accounts, which is described as a special case of Rao et al. [2010]. Given the non-econometric nature of the PWT extrapolations, there are no measures of reliability[5], for example in the form of standard errors, associated with them.

The current paper describes the complete version of the method in contrast to that described in Rao et al. [2010], referred to as RRD from here on. The RRD version is designed to disseminate a preliminary version of the work and it mainly focuses on the details of the regression model used in explaining national price levels[6] and on a special case of the general methodology described here. The method has been implemented and PPPs for GDP, household consumption expenditures (C), general government expenditures (G), gross capital formation (I) are available from the dedicated website UQICD (). The current version UQICDv2.1.0 is based on ICP data 1970, 1975, 1980, 1985, 1996 and 2005, 2011 and OECD/Eurostat benchmarks.

The structure of the paper is as follows. Section 1 establishes the relevant notation and provides a statement of the problem. In section 2 we establish the basic framework of the method. Section 3 is devoted to the econometric formulation of the model including a discussion of the stochastic assumptions. The econometric model is given a state-space representation which makes it possible for us to use Kalman filtering and smoothing methods to generate optimal predictions. The analytical properties of the methodology proposed are discussed in section 4. Proofs of the main propositions are provided in the Appendix. In particular, we are able to show that the methodology proposed here is invariant to the choice of the reference country and that the model is flexible enough to either track the PPPs in different benchmarks or to track the national price movements[7]. Details of estimation of the state-space system are provided in section 5. Section 6 provides a snapshot of the newly constructed panels available from UQICD which provides PPPs at the GDP level as for the three main components, viz., Consumption, Investment and Government Expenditure. PPPs and real expenditures for comparisons over time are available in the form of aggregates at constant prices. Measures of reliability of PPPs are available in the form of standard errors.

[5]PWT, however, provide some indication of the reliability as perceived by the compilers and are usually expressed in the form a quality rating in the range A to D.

[6]We avoid duplicating this material by simply referring to the corresponding sections of RRD.

[7]Due to inherent inconsistencies between benchmarks and national price deflators (and hence the national growth rates), it is not possible for the methodology to generate a series that is capable of tracking both at the same time.

The paper concludes with a few remarks in section 7.

1 Notation and a Statement of the Problem

We consider the general case with N countries (i=1,2,. . . ,N) and T time periods (t=1,2,. . . T). Let PPP_{it} and ER_{it}, respectively denote the purchasing power parity and exchange rate of the currency of the ith country expressed in terms of the currency units of a reference country. PPPs reflect the general price levels whereas the exchange rates reflect the value of the respective currencies and generally reflect the demand for the currencies involved. Based on these, we define the *national price level* for country i in period t, as:

$$R_{it} = \frac{PPP_{it}}{ER_{it}} \tag{1}$$

where R_{it} is the national price level for country i in period t. A value of the price level above 1 indicates that prices are higher relative to the reference country and vice versa.

The Problem

Given the importance attached to PPPs in converting national economic aggregates into a common currency unit, the ideal situation is where PPP_{it}s are observed for every country in all time periods. However, the reality is quite different. The main source of these PPPs is the International Comparison Program which oversees the collection of data and compilation of PPPs for all the countries participating in the Program. Given the resource intensive nature of the underlying price collections, participation in the ICP has been limited and, consequently, the matrix of available PPPs is quite sparse as shown in Table 1.

Table 1: Number of Countries Participating in ICP (various Phases) and OECD/EuroStat Only Comparisons

ICP Phase	Benchmark Year	Total No. of Participating Countries
I	1970	10
II	1973	16
III	1975	34
IV	1980	60
V	1985	64
OECD/EUROSTAT	1990	24
VI	1993/1996	117
OECD/EUROSTAT	1999	28
OECD/EUROSTAT	2002	28
VII	2005	146
VIII	2011	199

Source: Asian Development Bank, 2007 and authors
Note: ICP coverage in different phases is global with participating countries from different regions.

In practice we wish to have PPPs for all the countries spanning a long period of time. For example, if we have 180 countries and a period covering 1970 to 2015, then it is easy to see that the information available from the ICP and the OECD-Eurostat sources would sparsely cover the space-time tableau of PPPs. So our main problem is to develop an econometric technique that allows us to generate optimal predictions for the missing PPP data in the tableau. A simple approach to this problem would be to take PPPs from a single benchmark, say the 2011 benchmark and extrapolate over time using relative price movements. However, the problem with this simple approach is that different benchmarks will produce different extrapolations and using a single benchmark amounts to discarding valuable information from the remaining benchmarks. The new generation PWT produces series that treat benchmark data as without error and adjusts the growth rates to distribute the discrepancy with equal weighting. The approach does not have any proved properties. The approach we propose here addresses this problem by treating both benchmarks and growth rates as measured with error within an econometric framework. The properties of the method can be derived and statistical measures of uncertainty attached to the extrapolated PPPs.

2 The Basic Framework

The basic building blocks that underpin the econometric methodology are described here. This description draws from a more elaborate exposition of the material in RRD. In this paper we consider the most general econometric model of which the case discussed in RRD would be a special case. We highlight the significant differences when we discuss the econometric model and its state-space representation.

(i) We begin with PPP data from the ICP and postulate that the observed PPPs are true values contaminated with noise and measurement error. Let $p_{it} = \ln(PPP_{it})$ be the logarithm of the true PPPs. The observed PPPs from the ICP are related to the true PPPs through the following equation:

$$\tilde{p}_{it} = p_{it} + \xi_{it} \tag{2}$$

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 with $E(\xi_{it}) = 0$[8]

(ii) We take into account the fact that the numerical value of the PPP for the reference/numeraire country is set at 1, we have the condition in logarithms as:

$$p_{1,t} = 0, \qquad t = 1, 2, \ldots T \tag{3}$$

where country 1 is taken to be the reference country without loss of generality.

(iii) The most important element in the extrapolation strategy is the theoretical model that provides a link between national price levels and a range of observable socio-economic variables within the countries. Drawing on the vast literature in this area (Kravis and Lipsey, 1983; Clague, 1988 and Bergstrand, 1991, 1996), we postulate a log-linear relationship between the national price levels and a set of control variables[9]. We postulate the following model:

$$r_{it} = \beta_{0t} + \boldsymbol{x}'_{it}\boldsymbol{\beta}_s + u_{it} \quad \text{for all}\, i = 1, 2, ..., N \,\text{and}\, t = 1, 2, ..., T \tag{4}$$

where $r_{it} = ln(PPP_{it}/ER_{it})$; $\boldsymbol{x}'_{it}$ is a set of conditioning variables; β_{0t} intercept parameters; $\boldsymbol{\beta}_s$ a vector of slope parameters; u_{it} a random disturbance with specific distributional characteristics.

[8] We discuss the distributional assumption of disturbances in Section 4.

[9] We discuss in detail the regression model specification and selection of variables in RRD

If estimates of β_{0t} and $\boldsymbol{\beta}_s$ are available, model (4) can provide a prediction of the variable of interest consistent with price level theory.

$$\hat{p}_{it} = \hat{\beta}_{0t} + \boldsymbol{x}'_{it}\hat{\boldsymbol{\beta}}_s + ln(ER_{it}) \tag{5}$$

where $\hat{p}_{it}$ is a prediction; $\hat{\beta}_{0t}$ and $\hat{\boldsymbol{\beta}}_s$ are estimates, and $\hat{p}_{it}$ is a prediction of $ln(PPP_{it})$.
In this paper we make use of (5) to generate predictions for all the cells (all countries in all years) in the panel of PPPs. This is a more general approach than that used in PWT or in RRD. In RRD, regression predictions are used for extrapolating PPPs only in benchmark years and that too only for the non-benchmark countries. This approach has implications to the specifications of the "observation equations" used in the state-space formulation in Section 3.2. The estimation of $\hat{\beta}_{0t}$ and $\hat{\boldsymbol{\beta}}_s$ is discussed in Section 5.

(iv) The main source of information used in the extrapolation of PPPs over time relates to the implicit price deflators data from the national accounts. The PPP of the currency of the ith country in period t, relative to the reference country, the US for example, can be updated to period $t+1$ by adjusting PPP_{it} for movements in the GDP deflators in country i and in the United States. Thus we have,

$$PPP_{i,t} = PPP_{i,t-1} \times \frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}} \tag{6}$$

Equation (6) defines the growth rate of PPP_{it}. We assume that the relationship (6) holds with a random disturbance term. The model in terms of $p_{it} = \ln(PPP_{it})$ is given by

$$p_{it} = p_{i,t-1} + c_{it} + \eta_{it} \tag{7}$$

where,
$c_{it} = ln\left(\frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}}\right)$; η_{it} is a random error accounting for measurement error in the growth rates
Equation (6) can be used in updating PPPs for any given year (especially a benchmark year) to all the years covered in the panel. Thus equation (6) can be independently used as a tool to fill the missing PPPs in the time-space panel of PPPs and it is used in the PWT and Maddison series. equation (7) plays a key role in the state-space representation of our model as it provides the "transition" equation for the model (see Section 3.2 for details).

3 Econometric Formulation and State-Space Representation of the Model

The objective is to produce a panel of predictions of p_{it} which optimally combines the information arising out of the four sources (i) to (iv) outlined in Section 2 above. The

regression model explaining national price levels in equation (4) is incorporated into the following econometric model which is subsequently expressed in state-space form.

3.1 The Stochastic Assumptions

1. We start with the model used to explain the national price levels presented in equation (4). The errors u_{it} in (4) are assumed to be spatially correlated. The error follows a Spatial Error Model (see for example Chapter 2 of LeSage and R.K.Pace, 2009)

$$\mathbf{u}_t = \phi \mathbf{W}_t \mathbf{u}_t + \mathbf{e}_t \tag{8}$$

where $|\phi| < 1$ and $\mathbf{W}_t$ $(N \times N)$ is a spatial weights matrix. That is, its rows add up to one and the diagonal elements are zero. The term spatial in the present context refers to socio-economic distance rather than the traditional geographical distance[10]. It follows that $E(\mathbf{u}_t \mathbf{u}_t')$ is proportional to $\mathbf{\Omega}_t$, where $\mathbf{\Omega}_t = (\mathbf{I} - \phi \mathbf{W}_t)^{-1} (\mathbf{I} - \phi \mathbf{W}_t)^{-1'}$.

2. The measurement errors in the observation of $\ln(PPP_{it})$ during benchmark years, equation (2), are assumed spatially uncorrelated, but might be heteroskedastic. Thus, if ξ_{it} is a measurement error associated with the PPP of country i at time t, then

$$E(\xi_{it}) = 0; \quad E(\xi_{it}^2) = \sigma_\xi^2 V_{it} \tag{9}$$

where σ_ξ^2 is a constant of proportionality and $\boldsymbol{V}_t$ is defined below.

3. The measurement error in the growth rates, c_{it}, are assumed spatially uncorrelated, but might be heteroskedastic. Thus, η_{it} in (7) is assumed to have:

$$E(\eta_{it}) = 0; \quad E\left(\eta_{it}^2\right) = \sigma_\eta^2 V_{it} \tag{10}$$

where σ_η^2 is a constant of proportionality and $\boldsymbol{V}_t$ is defined below.

4. The measurement error variance-covariance is of the form

$$\boldsymbol{V}_t = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_{1t}^2 \boldsymbol{j}\boldsymbol{j}' + diag(\sigma_{2t}^2, ..., \sigma_{Nt}^2) \end{bmatrix} \tag{11}$$

[10] In the empirical section we test for cross-sectional dependence and specify a model of socio-economic distance to obtain the weights.

where, $\mathbf{j}$ is a vector of 1's, σ_{it}^2 is the variance of country i at time t, which we measure as the inverse of the a country's degree of development, and σ_{1t}^2 is the variance of the reference country.

This form of the covariance was derived from the definition of PPP (see RRD for detailed derivation) and it is sufficient for the invariance of the method to the choice of reference country (see Section4.5 and Appendix A for details on the invariance of the method). In the empirical implementation we model σ_{it}^2 as inversely related to GDP_{it}, the nominal per capita measured in \$US (exchange rates adjusted)[11]. This means that reliability of an observed PPP or growth rate is lower for low-income countries.

3.2 The State Space Representation

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 suitable representation for this type of problems. We start by writing equation (7) for the N countries to define the 'transition equation' of the SS:

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \boldsymbol{\eta}_t \tag{12}$$

where $\boldsymbol{\alpha}_t$ is the $N \times 1$ vector of unknown $\ln(PPP_{it})$ and it is the state vector in this representation; $\boldsymbol{c}_t$ is the observed growth rate of $\boldsymbol{\alpha}_t$ (this follows from equation (7) in Section 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 (12) simply updates period $t-1$ $PPPs$ using the observed relative price changes, represented by $\boldsymbol{c}_t$,over the period .

As previously discussed, noisy observations of $\boldsymbol{\alpha}_t$ are given by: , a prediction $\hat{\mathbf{p}}_t$from the regression model (5),; a measurement $\tilde{\mathbf{p}}_t$by the ICP (2)„ and by the condition in 3. Equation (5) relates the conditioning variables, $\boldsymbol{X}_t$, to the prediction $\hat{\mathbf{p}}_t$. Since the form of the observation equation of a SS model relates the observations (p_{1t}, $\hat{\mathbf{p}}_t$, $\tilde{\mathbf{p}}_t$) to the state vector $\boldsymbol{\alpha}_t$, it is convenient to re-write equation (4). Subtracting equation (5) from (1) we obtain:

$$\hat{p}_{it} = \alpha_{it} + (\hat{\beta}_{0t} - \beta_{0t}) + \boldsymbol{x}_{it}'(\hat{\beta}_s - \beta_s) - u_{it} \tag{13}$$

where α_{it} is an element of $\boldsymbol{\alpha}_t$.

[11] we use exchange-rate converted per capital income instead of PPP-adjusted per capita income mainly to avoid possible problems of endogeneity

Throughout the paper we will reserve the symbol $\boldsymbol{\theta}$ to represent the error in a current estimate of a parameter $\boldsymbol{\beta}$. Thus,

$$\hat{\theta}_{0t} = \hat{\beta}_{0t} - \beta_{0t} \text{ and } \hat{\theta}_s = \hat{\beta}_s - \beta_s \tag{14}$$

It is then possible to write equation (13) in the form

$$\hat{\mathbf{p}}_t = \boldsymbol{\alpha}_t + \mathbf{X}_t\boldsymbol{\theta} + \mathbf{v}_t \tag{15}$$

where $\boldsymbol{\theta} = [(\theta_{01}, ..., \theta_{0T})', \theta_s']'$; $v_{it} = -u_{it}$; $\mathbf{X}_t$ is an $N \times K$ matrix with columns including time dummy variables and socio-economic variables.

Finally, in order to express these different types of observations (viz, those given by (2) and (15)) as a single equation, it is convenient to define the following 'selection matrices',

$\mathbf{S}_1 = [1, \mathbf{0}'_{N-1}]$ (selects the reference country $i = 1$)[12]; $\mathbf{S}_p$ is a known matrix which selects the N_t participating countries (excluding the reference country) in the benchmark year t; $\mathbf{S}_{np}$ is a known matrix which selects $(N - 1 - N_t)$ non-participating countries in the benchmark year t[13].

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

$$\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{B}_t\mathbf{X}_t\boldsymbol{\theta} + \boldsymbol{\zeta}_t \tag{16}$$

with variables defined as follows:

i) Non-benchmark years:

$$\mathbf{y}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\hat{\mathbf{p}}_t \end{bmatrix}, \ \mathbf{Z}_t = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_{np} \end{bmatrix}, \ \mathbf{B}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np} \end{bmatrix}, \ \boldsymbol{\zeta}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\mathbf{v}_t \end{bmatrix} \tag{17}$$

$$\mathrm{E}(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t') \equiv \mathbf{H}_t = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \boldsymbol{S}_{np}\boldsymbol{\Omega}_t\boldsymbol{S}'_{np} \end{bmatrix} \tag{18}$$

with σ_u^2 a constant of proportionality, and in (16) the countries are ordered so that the reference country is the first row[14]. In non-benchmark years $\mathbf{S}_{np}$ is an $N - 1 \times N$ matrix by definition and $\hat{\mathbf{p}}_t$ is an $N \times 1$ vector of regression predictions for all countries[15].

[12] Without loss of generality country 1 is the reference country.

[13] The subscript t is omitted to keep the notation simple. In practice the list of non-participating countries varies from one benchmark to another. For non-benchmark years S_{np}remains the same for all t

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

[15] For invariance to hold it is necessary that the observation for participating countries in benchmark years be the ICP benchmark observations. The estimation of $\boldsymbol{\theta}$, to produce $\hat{\boldsymbol{p}}_t$, is based on all N countries in the sample. See Appendix A and Section 5.2 for details.

ii) Benchmark years

$$\mathbf{y}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\hat{\mathbf{p}}_t \\ \tilde{\mathbf{p}}_t \end{bmatrix}, \ \mathbf{Z}_t = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_{np} \\ \mathbf{S}_p \end{bmatrix}, \ \mathbf{B}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np} \\ \mathbf{0} \end{bmatrix}, \ \boldsymbol{\zeta}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\mathbf{v}_t \\ \mathbf{S}_p\xi_t \end{bmatrix} \tag{19}$$

$$E\left(\boldsymbol{\zeta}_t\boldsymbol{\zeta}_t'\right) \equiv \mathbf{H}_t = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2\boldsymbol{S}_{np}\boldsymbol{\Omega}_t\boldsymbol{S}_{np}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\xi^2\boldsymbol{S}_p\boldsymbol{V}_t\boldsymbol{S}_p' \end{bmatrix} \tag{20}$$

$\tilde{\mathbf{p}}_t$ is an $N_t \times 1$ vector of benchmark observations; $\hat{\mathbf{p}}_t$ is an $N \times 1$ vector of regression predictions for all countries. Again, σ_{u}^2 and σ_ξ^2 are constants of proportionality and the first row is the reference country.

Equations (12) and (16), together with the matrix definitions (17) to (20), constitute the 'transition' and 'observation' equations, respectively of a state space model for the unobservable 'state vector', $\boldsymbol{\alpha}_t$.

Given the unknown parameters, $\boldsymbol{\theta}$, ϕ, σ_u^2, σ_η^2, σ_ξ^2 and the distribution of the initial vector, $\boldsymbol{\alpha}_0$, under Gaussian assumptions, the Kalman filter computes the conditional (on the information available at time t) mean $\hat{\boldsymbol{\alpha}}_t$, and covariance matrix, $\mathbf{P}_t$, of the distribution of $\boldsymbol{\alpha}_t$. Further, $\hat{\boldsymbol{\alpha}}_t$ is a minimum mean square error predictor (MMSE) of the state vector, $\boldsymbol{\alpha}_t$. Even when Gaussian assumptions do not hold, the Kalman filter is still the optimal predictor in the sense that it minimizes the mean square error within the class of all linear predictors (see Harvey [1989], pp. 100-12, Durbin and Koopman [2001] Sections 4.2 and 4.3).

4 Special Features and Properties of the Method

We provide several analytical results including the important result that the constructed series are invariant to the reference country and in certain special cases of the model they are weighted averages of previous observations (benchmarks and regression predictions).

The econometric approach, and its state-space representation, encompass a number of models sought by practitioners and international organizations. In this section we demonstrate the versatility of the model by showing how we can adapt the model to suit different scenarios.

4.1 Constraining the model to track benchmark PPPs

As the ICP is the main source of $PPPs$ for different benchmarks and the respective $PPPs$ are determined using price data collected from extensive price surveys, one may consider

it necessary that the econometric method proposed should generate predicted $PPPs$ that are identical to benchmark $PPPs$ in the benchmark years. This can be achieved simply by setting $\sigma_\xi^2 = 0$ in (20). The last line in (19) then becomes a constraint, guaranteeing that predicted $PPPs$ are identical to the 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 movements in the implicit GDP deflator

In the currently available series, including the Maddison and PWT, growth rates in real GDP as well as the implicit price deflators preserve the national level movements in prices and real income. As the GDP deflator data are provided by the countries and such deflators are compiled using extensive country-specific data, it is often considered more important that the predicted $PPPs$ preserve the observed growth rates implicit in the GDP deflator[16]. This essential feature can be guaranteed in our work by setting $\sigma_\eta^2 = 0$ in (12). In our work we choose to impose this restriction through the backward filter (smoother, see Section 5.2). It is trivial to show that national level movements in prices are preserved using the formulae for the fixed interval Kalman Smoother (see the Appendix B). Growth rate constrained predictions are demonstrated in the empirical section.

4.3 Flexibility in the use of regression predictions

An important feature of our model is that the predictions generated by our national price levels regression model (and information provided by relevant socio-economic variables through model (4)) can be utilized in both benchmark and non-benchmark periods. This is a more general framework than the approach presented in RRD where the regression predictions are used only for non-participating countries in benchmark years. *This is where this paper extends and presents a general form to the approach suggested in* RRD. In the case of RRD the *algorithm updates predictions between benchmarks using only movements in deflators.* Results obtained under this simplified model are also presented in Section 6.

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

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

In this sub-section we address the question of whether there are conditions under which our predictions can be interpreted as weighted averages of extrapolations from different benchmarks. We identify a set of sufficient conditions for this result to hold.

If there are $M+1$ benchmark years ($j = 0, \ldots, M$)[17], applying growth rates to benchmark $PPPs$ (along with extrapolation to non-participating countries) 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[18]; or (b) to take a weighted average of the $M+1$ different panels. PWT6.3 as well as the Maddison series use the approach in (a). Under (b), there is also a need to specify the weights given to different benchmarks..

Proposition 1: If $PPPs$ for benchmarks from the ICP and regression extrapolations are used only in the benchmark years and if national deflators are used for updating $PPPs$, then the predicted panel of PPP estimates produced by our approach is a 'weighted average' of the $M+1$ panels discussed above. More specifically, suppose $\overleftrightarrow{\boldsymbol{p}}_{t,j}$ is the vector of predicted PPP in year t obtained by applying growth rates to the jth benchmark. Then, denoting the Kalman Filter predictions under this scenario by $\hat{\boldsymbol{\alpha}}_t$, we have

$$\hat{\boldsymbol{\alpha}}_t = \sum_{j=0}^{M} \Upsilon_j^{(M)} \overleftrightarrow{\boldsymbol{p}}_{t,j} \tag{21}$$

where the $\Upsilon_j^{(M)}$ are positive definite matrices such that $\sum_{j=0}^{M} \Upsilon_j^{(M)} = \mathbf{I}_N$

It is in this sense the prediction in (21) is considered a 'weighted average' although it is not generally true that the elements of $\hat{\boldsymbol{\alpha}}_t$ are a weighted average of those of the $\overleftrightarrow{\boldsymbol{p}}_{t,j}$. In the special case when the measurement errors in growth rates and benchmark $PPPs$ are uncorrelated across countries, we can show that the elements of $\hat{\boldsymbol{\alpha}}_t$ are a weighted average of the corresponding elements of the $M+1$ 'benchmark only' panels. Then, we have been able to show that

$$\hat{\boldsymbol{\alpha}}_t = \sum_{j=0}^{M} \upsilon_{ii,j}^{(M)} \overleftrightarrow{\boldsymbol{p}}_{t,j} \tag{22}$$

[17]It is convenient for the algebraic derivations to set the number of benchmarks to $M+1$.

[18]This is the approach used in the WDI publication for 2008 and 2009. All the published figures are anchored on the results from the 2005 benchmark comparison.

where, $\upsilon_{ii,j}^{(M)}$ is the *ith* diagonal element of the matrix $\Upsilon_j^{(M)}$, $\upsilon_{ii,j}^{(M)} > 0$ and $\sum_{j=0}^{M} \upsilon_{ii,j}^{(M)} = 1$.

Here we note that the Kalman filter predictions automatically produce a set of weights, $\upsilon_{ii,j}^{(M)}$, for the averaging process. Elements $\upsilon_{ii,j}^{(M)}$ can be interpreted as reflecting the reliability of the *jth* benchmark. *A proof of the proposition has been provided in the Appendix of RRD.*

The above result has been explored by Harvey, 2000 and Koopman and Harvey, 2003 in the context of univariate time series models. RRD showed that for the state-space in (16) and (7), the Kalman filter predictors are weighted sums of all the corresponding elements of the $M + 1$ panels (result in (21)) in general, and weighted averages of the corresponding elements of the $M + 1$ 'benchmark only' panels in the special case (result in (22)).

4.5 Invariance of the Predictions PPPs to the Choice of Reference country

In the exposition of our model, we used country 1 as the reference or base country with $p_{1,t} = 0$ for all t. This condition is then incorporated into our model and its state-space representation. For this method to be meaningful, it is necessary that it the results are invariant to the choice of the reference country. A significant contribution of this paper is to provide a proof that our econometric approach is invariant to the choice of the reference country.

Proposition 2: If we denote by $\hat{\boldsymbol{\alpha}}_t^{(1)}$ the Kalman Filter predictions when the reference country is 1 (e.g. the USA), and by $\hat{\boldsymbol{\alpha}}_t^{(2)}$ the Kalman Filter predictions of the state vector when the reference country selected is 2 (e.g. the UK); then

$$\hat{\alpha}_{it}^{(2)} = \hat{\alpha}_{it}^{(1)} - \hat{\alpha}_{2t}^{(1)} \tag{23}$$

The proof is fairly complex and it is presented in the Appendix A. 23shows that the predicted PPP for country i with country 2 as the reference country is equal to the rebased PPP for country i computed using country 1 as the numeraire.

5 Estimation

In order for the Kalman filter and smoothing algorithms to deliver a predictor of the state vector and its covariance matrix, we require estimates of the unknown parameters and a distribution for the initial state vector. The parameters of the state-space system can be estimated using likelihood based methods (Harvey, 1989, pp. 125-46) or Bayesian methods (see for instance Durbin and Koopman, 2001, Koop and van Dijk, 2000, and Harvey et al.,

2007). The results presented in this paper are obtained using likelihood based methods (details are provided in Section 5.2). The distribution of the initial state vector, $\boldsymbol{\alpha}_o$, is derived as follows.

5.1 Distribution of the Initial State Vector

We specify a distribution with a non-diffuse covariance for the initial state vector, $\boldsymbol{\alpha}_o$, by making use of equation (5). Let $\mathbf{X}_o$ denote a matrix of data on socioeconomic variables in period $t = 0$, the selected benchmark year. Then we can define,

$$\hat{\boldsymbol{\alpha}}_o = \hat{\beta}_{00} + \mathbf{X}_o\hat{\boldsymbol{\beta}}_s + ln(\mathbf{ER}) \tag{24}$$

where,

$\hat{\beta}_{oo}$ is the intercept; $\hat{\boldsymbol{\beta}}_s$ is the estimated slope coefficient vector which is independent of t; $\hat{\boldsymbol{\alpha}}_o = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_0^{(1)} \\ \hat{\boldsymbol{\alpha}}_o^{(2)} \end{bmatrix}$; $\mathbf{X}_o = \begin{bmatrix} \mathbf{X}_o^{(1)} \\ \mathbf{X}_o^{(2)} \end{bmatrix}$; $\mathbf{X}_o^{(1)}$ and $\boldsymbol{\alpha}_o^{(1)}$ represent the partition containing the observations from participating countries.

Then covariance associated to the prediction of $\boldsymbol{\alpha}_o$ is given by

$$\text{cov}(\hat{\boldsymbol{\alpha}}_o) = \mathbf{P}_o = \hat{\sigma}^2\mathbf{X}_o(\mathbf{X}_o^{(1)\prime}\mathbf{X}_o^{(1)})^{-1}\mathbf{X}_o' \tag{25}$$

We use the expression in (25) to obtain an estimate of the covariance of the initial state vector in the empirical section using data for the year 1985 (ICP Phase V year). The intercept and slope estimates in (24) are obtained using a regression containing the participating countries in 1985.

5.2 Estimation of unknown parameters and completion of the panel

There are two types of parameters to be estimated in the state-space, namely, hyperparameters (associated with the covariance structure, which in our case are: ϕ, σ_u^2 , σ_η^2 , σ_ξ^2), and coefficients associated with explanatory variables. Hyperparameters are estimated by numerical maximization of the likelihood function (in a likelihood based estimation). The parameters $\boldsymbol{\theta}$ (those in the regression partition in (16)) can be included as part of the state vector or estimated by a generalized least squares procedure (GLS) in conjunction with the numerical maximization of the likelihood function. We implement the later, which we denote by KF/GLS as it involves running the Kalman filter through both $\mathbf{y}_t$ and the columns of $\mathbf{X}_t$

(see Harvey, 1989, pp. 130-3 or Section 6.2 of Durbin and Koopman, 2001 for more details of this procedure)[19].

One non-standard feature of the state space form in this paper is that some elements of the observation vector, $\mathbf{y}_t$, the $\hat{\mathbf{p}}_t$, are predictions from the price level regression which has errors that are spatially correlated. The observation vector contains these predictions for non-participating countries in benchmark years, and for countries $i = 2, 3, ..., N$ in non-benchmark years. An initial set of predictions to start our algorithm (described below) is obtained by estimating the price level regression for the unbalanced panel composed of participating countries in all benchmark years. These predictions are updated as part of our algorithm as estimates of the parameter ϕ and vector $\boldsymbol{\theta}$ are obtained using the available information on $\mathbf{X}_t$ and $\mathbf{W}_t$ for *all countries and all time periods*. This is achieved by using the Kalman filter algorithm to combine all the information for the purpose of estimation of these parameters.

We first use the state-space form and Kalman filter to obtain the estimates of the unknown parameters and then the filtering-smoothing step to obtain the complete panel, which is run once estimates of the parameters have been obtained.

Parameter Estimation Algorithm:

Step *1*: Obtain an initial estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^0$, by regressing $\boldsymbol{r}_t$ on $\boldsymbol{X}_t$ with the panel of benchmark observations without accounting for spatial errors, see equation (4), and construct an initial prediction, $\hat{\boldsymbol{p}}_{it}^0$, using equation (5). An initial estimate of ϕ can then be obtained by computing the correlation between the OLS residuals and lagged residuals (from the regression in (4)). A choice of starting values for σ_u^2, σ_η^2, σ_ξ^2 is also needed. We use a grid search over the range $1e^{-8}$ to $1e^4$ and check the value of the likelihood by running Step 2 to locate a neighborhood of the global maximum. Denote $\boldsymbol{\gamma}$ =(ϕ, σ_u^2, σ_η^2, σ_ξ^2) and $\hat{\boldsymbol{\gamma}}^0$ the vector of starting values.

Step *2*: Given a set of values for $\boldsymbol{\gamma}$, an estimate of $\boldsymbol{\theta}$ is obtained by a KF/GLS procedure (explained below) and conditioning on these estimates, denoted by $\boldsymbol{\theta}^j$, a Newton-Raphson iterative procedure is used to maximize the likelihood function and obtain a new estimate of $\boldsymbol{\gamma}$, $\hat{\boldsymbol{\gamma}}^j$. Thus, a set of MLE estimates of ϕ, σ_u^2 , σ_η^2, σ_ξ^2, $\boldsymbol{\theta}$ are obtained at each iteration j. We note that after the first iteration of this step, the estimates of all the parameters in (4) are based on data for the N countries in the sample and account for the spatial correlation structure of the error through the KF/GLS estimation of $\boldsymbol{\theta}$.

[19]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 the KF/GLS approach .

KF/GLS estimation: each of the columns of $\mathbf{X}_t$ and $\mathbf{y}_t$ ($K+1$ vectors of dimension N in each time period) are run through the Kalman filter equations to obtain a set of "innovations", denote by $\mathbf{y}_t^*$, $\mathbf{X}_t^*$. The vector $\mathbf{y}_t$ includes observations $p_{1t}=0$, $\tilde{\mathbf{p}}_t$ and $\hat{\mathbf{p}}_t$ if t is a benchmark year, or $p_{1t}=0$ and $\hat{\mathbf{p}}_t$ if t is a non-benchmark year. The matrix $\mathbf{X}_t$ is $N \times K$. The estimation uses data for all countries which is necessary for the invariance result to hold (see Appendix A). The GLS estimator of $\boldsymbol{\theta}$ is computed by regressing $\mathbf{y}_t^*$ on $\mathbf{X}_t^*$ (we refer to it as KF/GLS and the interested reader may consult Harvey, 1989, pp. 130-3 or Durbin and Koopman, 2001, pp. 122-123 for further details).

***Step** 3*: Using updated estimates, $\hat{\boldsymbol{\theta}}^j$, we obtain revised estimates of $\hat{\beta}_{0t}^j = \hat{\beta}_{0t}^{j-1} - \hat{\theta}_0^j$ and $\hat{\boldsymbol{\beta}}_s^j = \hat{\boldsymbol{\beta}}_s^{j-1} - \hat{\boldsymbol{\theta}}_s^j$, which are used to obtain an updated $\hat{p}_{it}^j = \hat{\beta}_{0t}^j + \mathbf{x}_{it}'\hat{\boldsymbol{\beta}}_s^j + \ln ER_{it} + \hat{u}_{it}^*$, where $\hat{\mathbf{u}}_t^* = \hat{\phi}\mathbf{W}_t\hat{\mathbf{u}}_t^j$. For invariance to hold the predictions require an adjustment by subtracting the base country's prediction, $\hat{p}_{it}^{adjusted} = \hat{p}_{it}^j - \hat{p}_{1t}^j$ (see Appendix A, Section A.2 for details). With the updated $\hat{p}_{it}^{adjusted}$, we update the relevant elements of $\mathbf{y}_t$.

***Step** 4*: Repeat 2 and 3 until the change in the estimates of $\hat{\theta}_0^j$ and $\hat{\boldsymbol{\theta}}_s^j$ from $j-1$ to j are sufficiently close to zero. In our empirical implementation this occurs after three to four iterations[20].

Upon convergence of the algorithm, the parameters in equations (16), (12) and associated covariances are replaced by their estimates. The observation equation is now given by:

$$\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\zeta}_t \tag{26}$$

with variables defined as follows:

i) Non-benchmark years:

$$\mathbf{y}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\hat{\mathbf{p}}_t^f \end{bmatrix},\ \mathbf{Z}_t = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_{np} \end{bmatrix},\ \boldsymbol{\zeta}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\mathbf{v}_t \end{bmatrix} \tag{27}$$

$$\hat{\mathbf{H}}_t = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}_u^2 \boldsymbol{S}_{np}\tilde{\boldsymbol{\Omega}}_t\boldsymbol{S}_{np}' \end{bmatrix} \tag{28}$$

where $\hat{\mathbf{p}}_t^f = (\tilde{\beta}_{0t} + \boldsymbol{X}\tilde{\boldsymbol{\beta}}_s + \ln(ER_{it}) + \hat{\mathbf{u}}_t^f) - \hat{p}_{1t}$ is the prediction obtained using the estimates from the algorithm described in Steps 1-4 above; $\hat{\mathbf{u}}_t^f = \tilde{\phi}\mathbf{W}_t\hat{\mathbf{u}}_t^*$; $\tilde{\beta}_{0t}$, $\tilde{\boldsymbol{\beta}}_s$, $\tilde{\phi}$, and $\tilde{\sigma}_u^2$ are the estimates obtained from the algorithm in Steps 1-4.

[20]This is expected as the iterations over values of $\boldsymbol{\theta}$ simply update the ordinary least squares estimate of $\boldsymbol{\beta}$ to a maximum likelihood estimate through iterations of the GLS estimator.

ii) Benchmark years

$$\mathbf{y}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\hat{\mathbf{p}}_t^f \\ \tilde{\mathbf{p}}_t \end{bmatrix}, \ \mathbf{Z}_t = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_{np} \\ \mathbf{S}_p \end{bmatrix}, \ \boldsymbol{\zeta}_t = \begin{bmatrix} 0 \\ \mathbf{S}_{np}\mathbf{v}_t \\ \mathbf{S}_p\xi_t \end{bmatrix} \tag{29}$$

$$\hat{\mathbf{H}}_t = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\sigma}_u^2 \boldsymbol{S}_{np}\tilde{\boldsymbol{\Omega}}_t\boldsymbol{S}_{np}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\sigma}_\xi^2 \boldsymbol{S}_p\boldsymbol{V}_t\boldsymbol{S}_p' \end{bmatrix} \tag{30}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \boldsymbol{\eta}_t \tag{31}$$

where $\tilde{\sigma}_\xi^2$ is the estimate obtained from the algorithm in Steps 1-4, and $\hat{\mathbf{Q}}_t$=$\tilde{\sigma}_\eta^2\mathbf{V}_t$.

***Smoothing Algorithm*:**

This step estimates the unknown state vector using the state-space in (26) and (31) where all system parameters have now been replaced by estimates. The Kalman Filter (forward filter) and Kalman smoother (backward filter) algorithms [21] are used to obtain the model's predictions of the smoothed state vector α_{it} (for all i and t), $\mathbf{a}_t$, and its covariance matrix, $\hat{P}_t$.

The completed panel provides predictions of $\alpha_{it} = \ln(PPP_{it})$. To obtain predictions of PPP_{it} we reverse the natural log transformation which provides a median unbiased estimate:

$$\hat{PPP}_{it} = e^{a_{it}} \tag{32}$$

where a_{it} is the corresponding Kalman smoothed element.

The standard errors for the predicted *PPPs* are computed as follows[22]:

$$se(\hat{PPP}_{it}) = \sqrt{e^{2a_{it}}e^{P_{ii,t}}(e^{P_{ii,t}} - 1)} \tag{33}$$

where $P_{ii,t}$ is the *ith* diagonal element of the estimated smoothed covariance of the state vector, P_t.

[21] The interested reader will find the equations of the Kalman filter algorithm in Section A.6 of the Appendix and the equations of the Kalman smoother used in this paper in Appendix B.

[22] The standard errors are computed under the assumption of log-normality of the predictions.

5.3 Dealing with newly formed countries

There are 22 countries in our data set that were not designated as separate countries over the complete period. For these countries data are available only from 1990 when they came into existence as separate entities after the dissolution of USSR and former Yugoslavia. To be able to retain the time-space consistency and reference-country invariance properties for the majority of countries, the estimation is conducted in two steps, both use the RRD method. In step 1, RRD is implemented for 159 countries that have full data coverage. In this step all the parameters of the model are estimated. Given the parameters estimated in Step 1, RRD method is implemented for all the 181 countries for the period 1990 to 2013. See Appendix A of the UQICD User Guide (Rao et al. [2015]).

6 Empirical Results

UQICD Version 2.1.0 covers 181 countries from all geographical regions of the world. The database provides data for the years 1971 to 2012. Users can download necessary series for countries and years of their choice using the friendly interface developed for UQICD. In order to facilitate analysis based on country groupings, UQICD makes it possible to download data for selected country groupings. Regional groupings available on UQICD are based on the World Bank World Development Indicators classification. Figure 1 shows the download screen which provides the user with all the downloading options

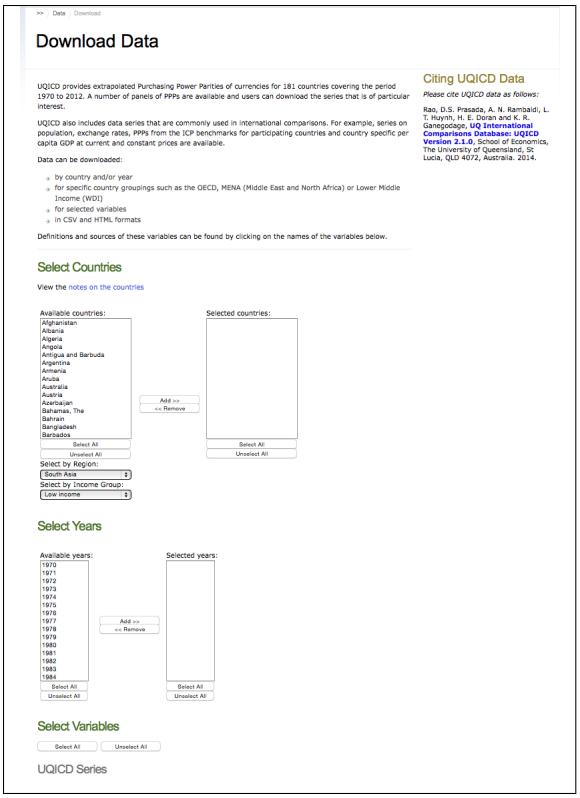

Figure 1: ScreenShoot Download Page UQICD

6.1 Series available

The main series available from UQICD relate to PPPs and real expenditures at the GDP level and for the three main components: Consumption; Investment; and Government Expenditure. A special feature of UQICD is the availability of charts showing different series over time for all the countries in the database. In addition to the panels of PPPs available from the UQICD, a number of standard series of interest from secondary sources are made available to the users to facilitate basic analyses as well as the construction of a number of series that can be derived from the series that can be downloaded from UQICD. Figure 2 shows some of the series that can be downloaded from the website.

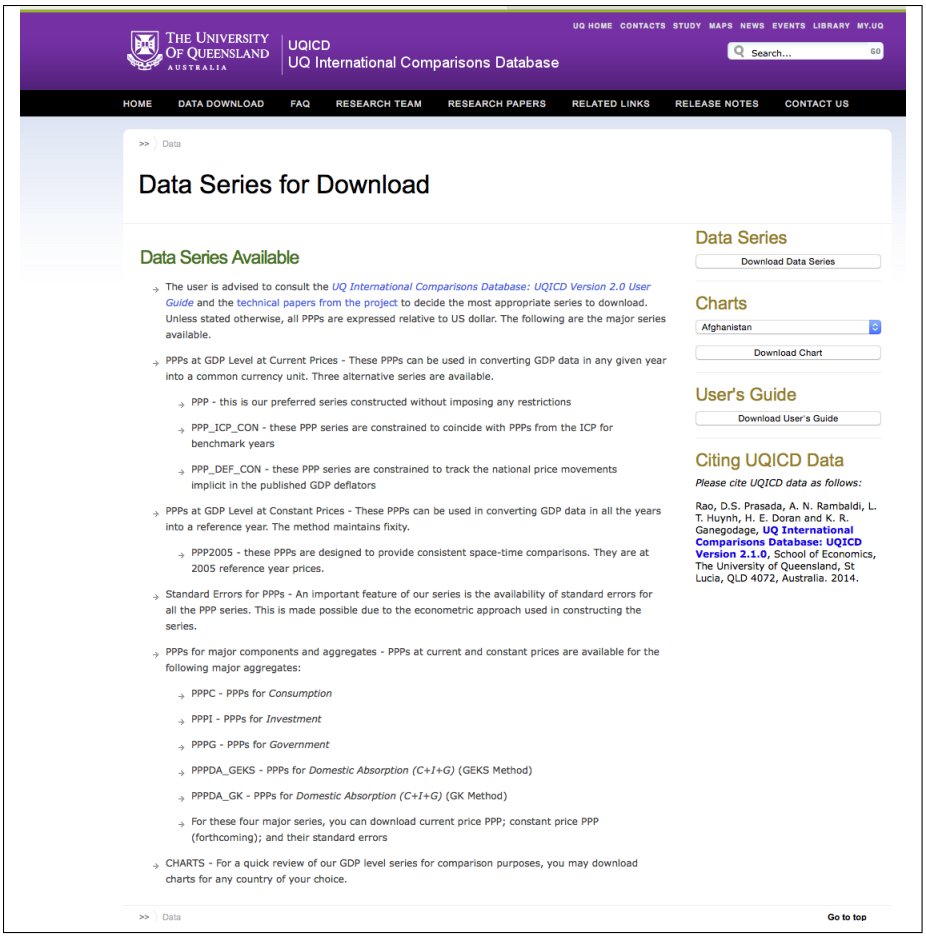

Figure 2: ScreenShoot Download Page UQICD

6.1.1 Purchasing Power Parities (PPPs) of currencies

In UQICD, all the PPPs are expressed relative to the US dollar. For example if PPP of Indian rupee is 15 rupees, this means that 15 Indian rupees have the same purchasing power as one US dollar. The UQICD PPPs provided are at current prices and provide parities of currencies with respect to one unit of reference currency in a given year. PPPs at current prices can be used in making real expenditure comparisons across countries in a specific year for which the PPPs refer to. For example, if PPP for AUS dollar is 1.50 in 2010, this means that expenditure in Australian dollars can be converted into US dollars in 2010 at the rate of 1.50 AUD for US dollar.

The PPPs are used to construct panels of real incomes and components (consumption, government and investment) in current and constant US dollar which are available from download from the UQICD website.

6.1.2 PPP series at current prices

These are the PPPs that can be used, in any given year, for converting national accounts aggregates into a common currency unit (US dollar in this case) which can be compared across countries. As PPPs refer to the aggregate under consideration, users must select the PPP series that is most appropriate for their analysis or application.

GDP Level: Three alternative series are available.

1. ***PPP*** - This is our recommended series for users at the GDP level. This series is constructed without imposing any constraints to track ICP benchmarks or the domestic price deflators.

2. ***PPP_ICP_CON*** - This is the series users would use if they place complete confidence on PPPs from the ICP benchmarks. These series ensure that ICP benchmarks are respected.

3. ***PPP_DEF_CON*** - This is the series constrained to track GDP price deflators published by the national statistical offices.

PPP series for GDP Components: Only one series is available for each of the components. An addition series is created for *Domestic Absorption (DA)* which is the sum of Consumption, Investment and Government, i.e. C+I+G. As DA is the sum of three components, aggregation is involved in this process. Two series based on the Geary-Khamis (GK) and Gini-Elteto-Koves-Szulc (GEKS) methods are available for DA. Thus a total of five series are available for the components.

1. ***PPPC*** – Unconstrained extrapolated series for the Consumption aggregate

2. ***PPPI*** – Unconstrained extrapolated series for the Investment aggregate

3. ***PPPG*** – Unconstrained extrapolated series for the Government expenditure aggregate

4. ***PPPDA_GEKS*** – Unconstrained extrapolated series for Domestic Absorption using GEKS aggregation method

5. ***PPPDA_GK*** – Unconstrained extrapolated series for Domestic absorption using GK method

The methodology used in the estimation of PPPs for components and details of the GK and GEKS aggregation methods are provided in Appendix B.

Standard Errors for selected PPP Series

SE(PPP) - Standard errors for unconstrained PPP series at GDP level

SE(PPP_ICP_CON); ***SE(PPP_DEF_CON)***; ***SE(PPPC)***; ***SE(PPPI)***; and ***SE(PPPG)*** are also available to accompany the corresponding PPP series

6.1.3 GDP Level and Aggregates (C, G and I) at Current and Constant Prices

In order to make comparisons over time and space, it is necessary to have GDP, Private Consumption, Investment and Government Expenditures series which are expressed at constant prices. For example if the objective is to compare real per capita income of India in 2012 with real per capita income of the US in 1995, then there is a need to compare these aggregates at constant price series. In compiling PPP series at constant prices, it is necessary to fix the reference country and also a reference year. In the UQICD series at constant prices, United States is the reference country and the year 2005 is the reference year. Suppose PPP for India in 2000 at constant 2005 prices is 12 rupees, this means that 12 rupees in Indian in 2000 have the same purchasing power as one US dollar in 2005.

UQICD Version 2.1.0 provides **Constant Prices Real GDP and Components**. Details of the procedure and the derivations involved are shown in Appendix C of Rao et al. [2015].

***CRGDP_PC*-** these are consistent space-time comparable **incomes per capita** in USD of 2005 prices

***CRGDPC_PC*-** these are consistent space-time comparable **household consumption expenditures per capita** USD of 2005 prices

***CRGDPG_PC*-** these are consistent space-time comparable general **government expenditures per capita** USD of 2005 prices

***CRGDPI_PC*-** these are consistent space-time comparable **gross capital formation per capita** USD of 2005 prices

UQICD also provides **Current Prices Real GDP.** These are obtained by dividing GDP per capita in local currency units by our unrestricted estimates of PPPs.

RGDP_PC - these are comparable **income per capita** in USD - current prices

6.1.4 Derived Series

The user can derive a number of other series using the series provided by UQICD. A few useful series are described below.

Gross domestic product Let GDP_{it} represent the gross domestic product of country i in period t expressed in local currency units in current prices. The GDP_{it} can be derived as:

$$GDP_{it} = GDPPCAP_LCU_CURRENT \times pop$$

Nominal and Real GDP at current prices Let XR_{it} and PPP_{it} represent exchange rate and PPP of currency of country i in period M in current prices. Then *nominal* and *real* GDP can be obtained as:

$$\text{Nominal GDP} = NGDP_{it} = \frac{GDP_{it}}{XR_{it}}$$

$$\text{REAL GDP} = RGDP_{it} = \frac{GDP_{it}}{PPP_{it}}$$

PPP at constant 2005 year prices Let PPP_{it}^{2005} represent the PPP of currency of country i in period t expressed in 2005 prices. Using UQICD constant price real GDP, denoted by $CRGDP_{it}^{2005}$, the PPP is obtained as follows:

$$PPP_{it}^{2005} = \frac{GDP_{it}}{CRGDP_{it}^{2005}}$$

UQICD provides an estimate of $CRGDP_{it}^{2005}$ per capita which is labelled ***CRGDP_PC*** (see Appendix C of the User Guide for details)

Real GDP at constant prices with year of your choice In certain instances, users may wish to use a year different from 2005 to express constant price series. This can be easily achieved using the series available on UQICD. Suppose real GDP at constant year τ prices where τ is different from 2005. Let $PPP_{US,\tau}^{2005}$ represent PPP for US dollar in year τ at constant 2005 prices (obtained from PPP_{it}^{2005} described before), then PPPs at constant year τ prices are given by:

$$PPP_{it}^{\tau} = \frac{PPP_{it}^{2005}}{PPP_{US,\tau}^{2005}}$$

To illustrate this, take the example where the constant price PPPs are shifted from 2005 year prices to 2000. This can be simply achieved by:

$$PPP_{it}^{2000} = \frac{PPP_{it}^{2005}}{PPP_{US,2000}^{2005}}$$

The constant year τ prices GDP for country i in period t ,$CRGDP_{it}^{\tau}$, is given by:

$$CRGDP_{it}^{\tau} = \frac{GDP_{it}}{PPP_{it}^{\tau}}$$

For example, the constant price real GDP in year 2000 prices can be obtained as:

$$CRGDP_{it}^{2000} = \frac{GDP_{it}}{PPP_{it}^{2000}}$$

This section provided examples of how derived series can be constructed using series available from UQICD. For example, all the series illustrated above refer to GDP aggregate. However, such series can be derived for consumption, investment, government expenditure and domestic absorption.

6.2 Illustration

The website also provides graphs of current prices PPP figures for GDP (with a two-standard error band) and its components (C,G and I), as well as a measure of PPP for Domestic Absorption for each country. Each country's benchmark data from ICP participation is also shown. The statistical uncertainty is very large for countries that have not participated in earlier ICP comparision, this is illustrated by China. India participated in earlier rounds; however, a large gap exists between 1985 and 2005, which leads to increased standard errors attached to that period. For established developed countries, the uncertainty is much smaller as demonstrated by Australia for which estimated standard errors are very much smaller.

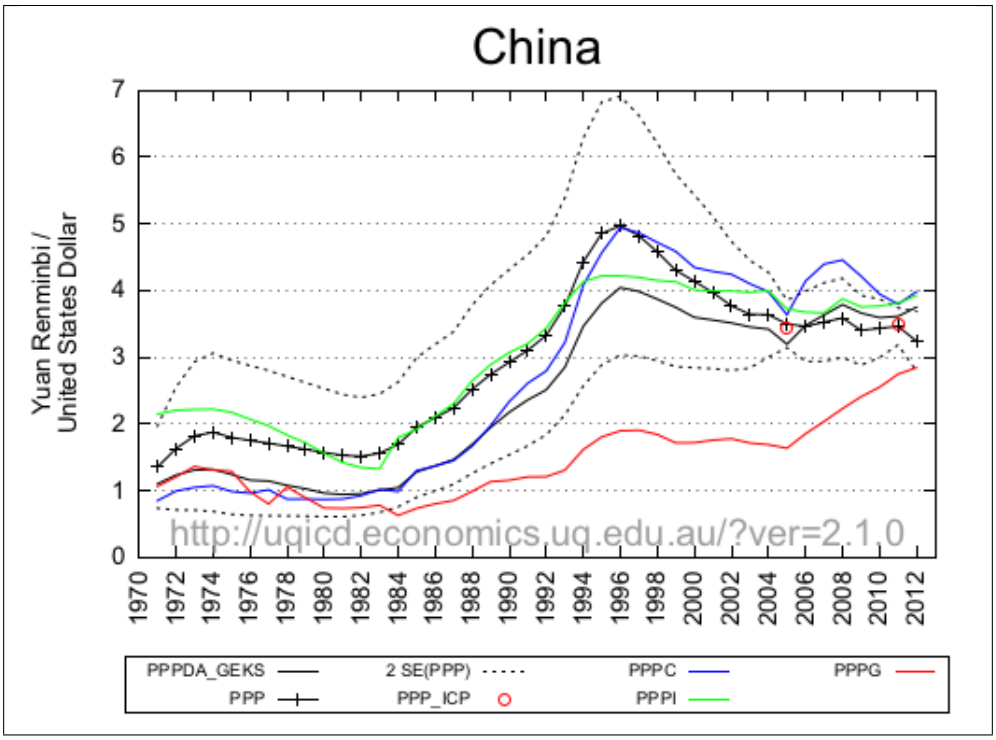

Figure 3: UQICD - v2 Website Plot - China

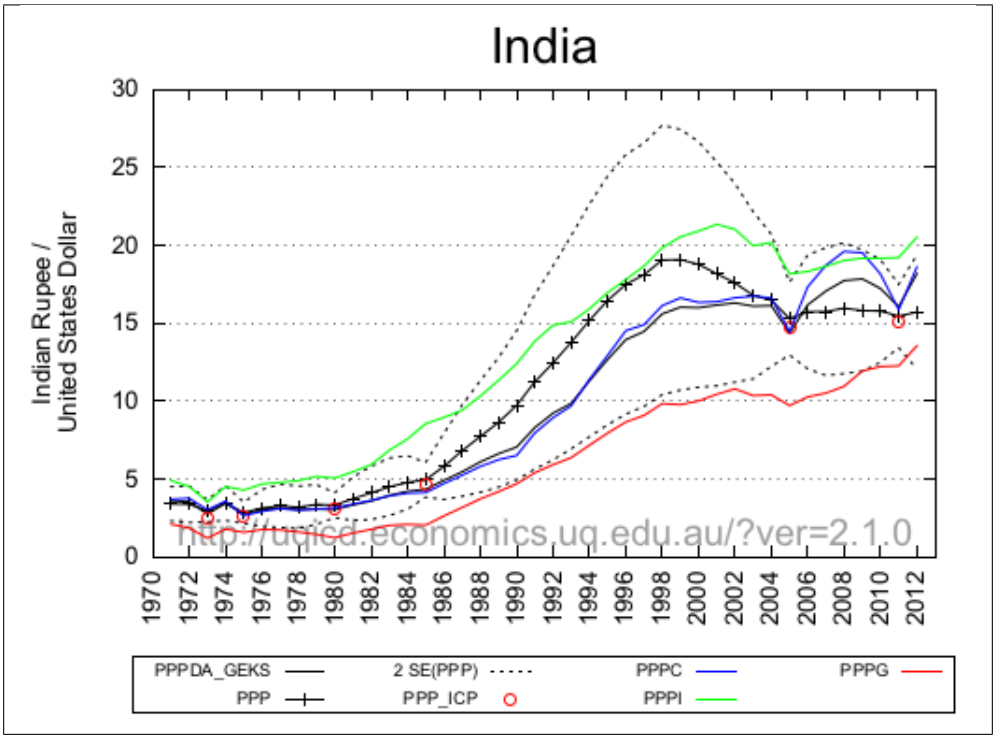

Figure 4: UQICD - v2 Website Plot - India

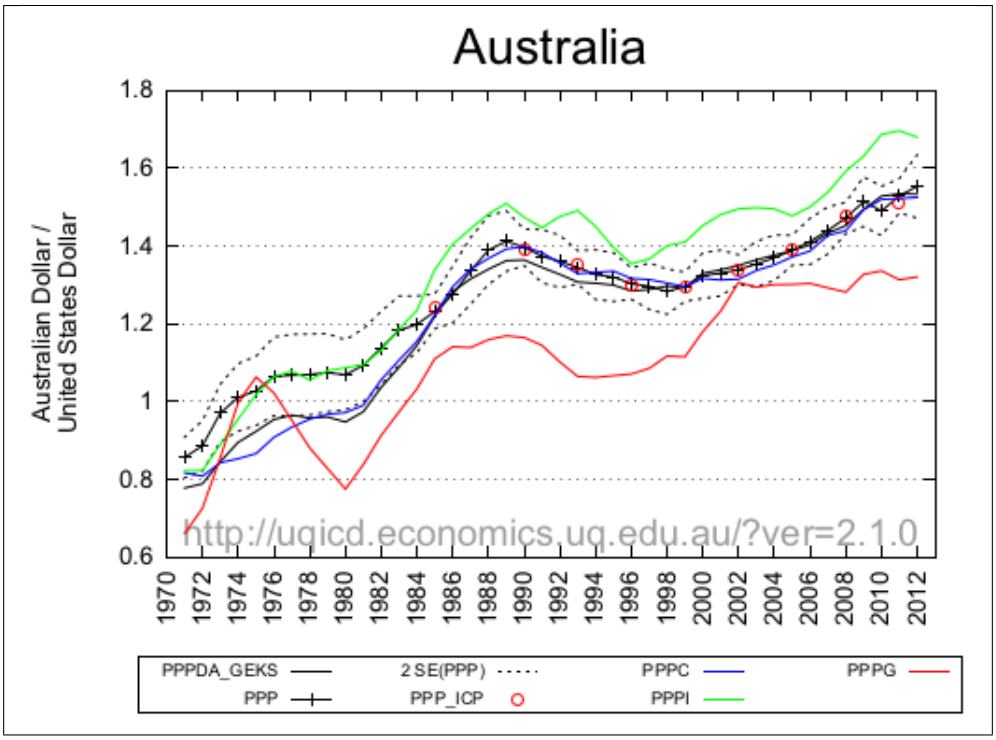

Figure 5: UQICD - v2 Website Plot - Australia

7 Conclusions

The econometric methodology suggested in the paper for the construction of a consistent panel of purchasing power parities represents a major step forward as it advocates a transparent and coherent approach. The approach suggested is designed to make use of all the principal and auxiliary information available for the purpose of extrapolation of the International Comparison Program (ICP) benchmarks. Existing approaches to the construction of panels of PPPs are not based on an econometric approach. Economic theory is placed within an econometric model and expressed as a state-space model as the problem of estimating PPPs is one of signal extraction. The paper demonstrates that the new approach is flexible in that it can be used to consider a number of scenarios including restrictions on some variance parameters to generate extrapolations that track the observed ICP PPPs in benchmark years or the implied price movements over time for individual countries. A proof of the invariance of the resulting predictions to the choice of reference country is provided. Further, this is the first available approach to producing not only a panel of PPPs, but also associated standard errors that can be incorporated into any further modelling using these

estimates.

The methodology proposed has been implemented through a dedicated website which provides PPPs for GDP and its components as well as space-time consistent constant price measures of real GDP, household consumption expenditures, general government expenditures and gross capital formation in US dollars. The current prices PPPs and standard errors for India, China and Australia are presented to illustrate the results obtained.

References

ADB Asian Development Bank. Purchasing power parities and real expenditures. 2005 international comparison program in asia and the pacific., 2007. URL `https://www.adb.org/publications/purchasing-power-parities-and-real-expenditures-2007`.

R.J. Barro and X. Sala-i Martin. *Economic Growth*. MIT Press, Cambridge, MA, second edition, 2004.

J.H. Bergstrand. Structural determinants of real exchange rates and national price levels: Some empirical evidence. *American Economic Review*, 81(1):325–334, 1991.

J.H. Bergstrand. *Productivity, Factor Endowments, Military Expenditures, and National Price Levels*, volume International Comparisons of Prices, Output and Productivity: Contributions to Economic Analysis Series, chapter II, pages 297–317. Elsevier Science Publishers, 1996.

D. Chotikapanich, W.E. Griffiths, and D.S. Prasada Rao. Estimating and combining national income distributions using limited data. *Journal of Business and Economic Statistics*, 25: 97–109, 2007.

C. Clague. *Explanations of National Price Levels*, volume World Comparisons of Incomes, Prices and Product: Contributions to Economic Analysis Series, chapter III, pages 237–262. Elsevier, North Holland, 1988.

Angus Deaton and Alan Heston. Understanding ppps and ppp-based national accounts. *American Economic Journal: Macroeconomics*, 2(4):1–35, 2010.

H.E. Doran. Constraining kalman filter and smoothing estimates to satisfy time-varying restrictions. *Review of Economics and Statistics*, 74(3):568–572, 1992.

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

S.N. Durlauf, P. Johnson, and J. Temple. *Handbook of Economic Growth*, volume 1, chapter Growth Econometrics, pages 555–677. Elsevier, 2005.

Robert C. Feenstra, Robert Inklaar, and Marcel P. Timmer. The next generation of the Penn World Table. *American Economic Review*, 105(10):3150–82., 2015. doi: 10.1257/aer.20130954.

A. Harvey. *Journal of the Royal Statistical Society Series B*, 62:34–35, 2000.

A. C. Harvey, T. M. Trimbur, and H. K. Van Dijk. Trends and cycles in economic time series: A Bayesian approach. *Journal of Econometrics*, 140(2):618–649, October 2007. URL `http://ideas.repec.org/a/eee/econom/v140y2007i2p618-649.html`.

A.C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University, Cambridge, 1989.

Alan Heston, Robert Summers, and B. Aten. Penn world table version 6.2, 2006. URL `http://pwt.econ.upenn.edu/php_site/pwt_index.php`.

G. Koop and H.K. van Dijk. Testing for integration using evolving trend and seasonal models: A bayesian approach. *Journal of Econometrics*, 97:261–91, 2000.

S.J. Koopman and A.C. Harvey. Computing observation weights for signal extraction and filtering. *Journal of Economic Dynamics and Control*, 27:1317–1333, 2003.

I.B. Kravis and R.E. Lipsey. *Toward an Explanation of National Price Levels*. Princeton University Press, Princeton, 1983.

J.P. LeSage and R.K.Pace. *Introduction to Spatial Econometrics*. Chapman & Hall/CRC. Taylor & Francis Group. Boca Raton, FL. USA., 2009.

A. Maddison. *Monitoring the World Economy, 1820-1992*. Development Centre of the Organisation for Economic Co-operation and Development, Paris and Washington, D.C., 1995.

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

B. Milanovic. True world income distribution, 1988 and 1993: First calculation based on household surveys alone. *The Economic Journal*, 112:51–92, 2002.

L. Pritchett. Divergence, big time. *Journal of Economic Perspectives*, 11:3–17, 1997.

D. Quah. Empirics for growth and distribution: Stratification, polarization, and convergence clubs. *Journal of Economic Growth*, 2:27–59, 1997.

D. S. Prasada Rao, A. N. Rambadi, K. R. Ganegodage, L. T. Huynh, and H. E. Doran. *UQICD v2 User Guide*. The University of Queensland, https://ideas.repec.org/p/qld/uq2004/534.html, version 2.1.0 edition, February 2015. URL `https://ideas.repec.org/p/qld/uq2004/534.html`.

D.S. Prasada Rao, A. N. Rambaldi, and H.E. Doran. Extrapolation of purchasing power parities using multiple benchmarks and auxiliary information: A new approach. *The Review of Income and Wealth*, 56(S1):59–98, 2010.

X. Sala-i Martin. The world distribution of income: Falling poverty and ... convergence, period. *Quarterly Journal of Economics*, 121:351–97, 2006.

R. Summers and A. Heston. A new set of international comparisons of real product and price levels estimates for 130 countries: 1950-1985. *Review of Income and Wealth*, 34(1): 1–25, 1988.

R. Summers and A. Heston. The penn world table (mark 5): An expanded set of international comparisons, 1950-1988. *The Quarterly Journal of Economics*, 106(2):327–368, May 1991.

H. Theil. The development of international inequality, 1960-1985. *Journal of Econometrics*, 42:145–155, 1989.

UNDP. Statistics of the human development report, various. URL `http://hdr.undp.org/en/statistics/`.

UQICD Website, February 2015. URL `http://uqicd.economics.uq.edu.au`.

A. The Invariance of the (Kalman Filter) Predictions to the Choice of Reference Country - Proposition 2.

A.1 Notation and Conventions

Without loss of generality we will take two reference countries as countries 1 and 2, and denote the $ln(PPP_t)$ relative to the two bases as $\boldsymbol{\alpha}_t^{(1)}$ and $\boldsymbol{\alpha}_t^{(2)}$. Other consequent notation will usually be obvious, making definition unnecessary.

By definition

$$\boldsymbol{\alpha}_t^{(2)} = \boldsymbol{\alpha}_t^{(1)} - \alpha_{2t}^{(1)} \tag{34}$$

Also $\alpha_{2t}^{(2)} \equiv \alpha_{1t}^{(1)} \equiv 0$.

Because the p_{it} is always zero for the base country, we will remove it from the Kalman filter cycle, and re-define $\boldsymbol{\alpha}_t^{(1)}$ and $\boldsymbol{\alpha}_t^{(2)}$ as the $N-1$ vectors $\boldsymbol{\alpha}_t^{(1)} = [\alpha_{2t}^{(1)}, \alpha_{3t}^{(1)}, ..., \alpha_{Nt}^{(1)}]'$ and $\boldsymbol{\alpha}_t^{(2)} = [\alpha_{1t}^{(2)}, \alpha_{3t}^{(2)}, ..., \alpha_{Nt}^{(2)}]'$. It follows from (34) that

$$\boldsymbol{\alpha}_t^{(2)} = \boldsymbol{A}\boldsymbol{\alpha}_t^{(1)} \tag{35}$$

where $\boldsymbol{A}$ is a non-stochastic, non-singular $(N-1) \times (N-1)$ matrix given by

$$\mathbf{A} = \begin{bmatrix} -1 & \mathbf{0}'_{N-2} \\ -\boldsymbol{j}_{N-2} & \boldsymbol{I}_{N-2} \end{bmatrix} \tag{36}$$

$\mathbf{j}_{N-2}$ is a vector of ones and $\mathbf{0}'_{N-2}$ a (row) vector of zeros.

Denoting the Kalman filter estimates obtained by using observations relative to the two base countries by $\hat{\boldsymbol{\alpha}}_t^{(1)}$and $\hat{\boldsymbol{\alpha}}_t^{(2)}$, the invariance property holds if it can be established that

$$\hat{\boldsymbol{\alpha}}_t^{(2)} = \boldsymbol{A}\hat{\boldsymbol{\alpha}}_t^{(1)} \tag{37}$$

A.2 Regression Estimates

a) *Benchmark years*

Estimates of β_{ot} and β_s are obtained by regressing benchmark observations $\tilde{\mathbf{p}}_t$ on the conditioning variables $\mathbf{x}_t = [x_{1t}, x_{2t}, ..., x_{N_1 t}]'$ where we have taken countries $i = 1, 2, \ldots, N_1$, as the participating countries.

Now, by definition,

$$\tilde{\mathbf{p}}_t^{(2)} = \tilde{\mathbf{p}}_t^{(1)} - \tilde{p}_{2t}^{(1)}\mathbf{j}_{N-2} \tag{38}$$

That is, the dependent variable $\tilde{\mathbf{p}}_t^{(2)}$ is obtained by subtracting the same number $\tilde{p}_{2t}^{(1)}$ from each observation in $\tilde{\mathbf{p}}_t^{(1)}$. Because the regressors $\mathbf{X}_t$ do not change when the base country is changed from 1 to 2, by standard regression theory

$$\hat{\beta}_{0t}^{(2)} = \hat{\beta}_{0t}^{(1)} - \tilde{p}_{2t}^{(1)} \tag{39}$$

$$\hat{\boldsymbol{\beta}}_s^{(2)} = \hat{\boldsymbol{\beta}}_s^{(1)} = \hat{\boldsymbol{\beta}}_s$$

That is, intercepts change but slopes are invariant. It follows that for non-participating countries

$$\hat{\mathbf{p}}_t^{(2)} = \hat{\mathbf{p}}_t^{(1)} - \hat{p}_{2t}^{(1)}\mathbf{j}$$

Thus, defining the "observation vector" $\mathbf{y}_t$ by $\mathbf{y}_t = [\tilde{\mathbf{p}}_t, \hat{\mathbf{p}}_t]'$ and discarding the base country observation (as it is always zero) we have

$$\mathbf{y}_t^{(2)} = \mathbf{A}\mathbf{y}_t^{(1)} \tag{40}$$

b) *Non-benchmark years*

Here the observation is the regression prediction $\hat{\beta}_o^{(i)}\mathbf{j}_N + \mathbf{X}_t\hat{\boldsymbol{\beta}}_s$ $(i = 1, 2)$. We now adjust the observation by subtracting the base country prediction from all predictions. This ensures the base country observation is zero, and the value of the intercept is irrelevant.

Then,

$$y_{it}^{(1)} = (x'_{it} - x'_{1t})\hat{\boldsymbol{\beta}}_s; \qquad y_{it}^{(2)} = (x'_{it} - x'_{2t})\hat{\boldsymbol{\beta}}_s$$
$$(x'_{it} - x'_{1t})\hat{\beta}_s - (x'_{2t} - x'_{1t})\hat{\boldsymbol{\beta}}_s = y_{it}^{(1)} - y_{2t}^{(1)}$$

Thus,

$$\mathbf{y}_t^{(2)} = \mathbf{A}\mathbf{y}_t^{(1)} \tag{41}$$

It follows from (40) and (41) that for both benchmark and non-benchmark years, the fundamental transformation $\mathbf{y}_t^{(2)} = \mathbf{A}\mathbf{y}_t^{(1)}$ holds.

A.3 The covariance of the measurement error

The measurement error in the benchmark *PPPs* and growth rates are assumed to have a covariance proportional to the form[23]:

$$\mathbf{V}_t = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_{1t}^2 \boldsymbol{j}\boldsymbol{j}' + diag(\sigma_{2t}^2, ..., \sigma_{Nt}^2) \end{bmatrix} \tag{42}$$

where, σ_{it}^2 is the variance of country i at time t and σ_{1t}^2 is the variance of the reference country.

Let $\mathbf{V}_t^{(1)}$ the $(N-1) \times (N-1)$ matrix obtained by ignoring the first row and column of $\mathbf{V}_t$,

$$\mathbf{V}_t^{(1)} = \sigma_\eta^2[\sigma_{1t}^2 \mathbf{j}\mathbf{j}' + diag(\sigma_{2t}^2, ..., \sigma_{Nt}^2)]$$

Then,

$$\mathbf{A}\mathbf{V}_t^{(1)}\mathbf{A}' = \sigma_\eta^2[\sigma_{2t}^2 \mathbf{j}\mathbf{j}' + diag(\sigma_{1t}^2, \sigma_{3t}^2, ..., \sigma_{Nt}^2)] = \mathbf{V}_t^{(2)}$$

[23] See RRD for a formal derivation of $\mathbf{V}_t$ from the definition of PPP.

A.4 The observation equation

The fundamental observation equation used in the method is

$\mathbf{y}_t = \boldsymbol{\alpha}_t + \boldsymbol{\zeta}_t, \qquad E(\boldsymbol{\zeta}_t \boldsymbol{\zeta}_t') = \mathbf{H}_t$

where $\mathbf{y}_t$ is an observation of the unobserved state; $\boldsymbol{\alpha}_t$ and $\boldsymbol{\zeta}_t$ is an observation error.

Because $\boldsymbol{\alpha}_t^{(2)} = \mathbf{A}\boldsymbol{\alpha}_t^{(1)}$ by definition and $\mathbf{y}_t^{(2)} = \mathbf{A}\mathbf{y}_t^{(1)}$ by regression properties and construction (see previous sections) it follows that $\boldsymbol{\zeta}_t^{(2)} = \mathbf{A}\boldsymbol{\zeta}_t^{(1)}$. And thus because $\boldsymbol{A}$ is non-stochastic,

$$\mathbf{H}_t^{(2)} = \mathbf{A}\mathbf{H}_t^{(1)}\mathbf{A}' \tag{43}$$

This is the fundamental result that enables us to prove invariance.

A.5 The transition equation

The transition equation used is of the form

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \boldsymbol{\eta}_t \tag{44}$$

where,

c_t is the observed growth rate of $\boldsymbol{\alpha}_t$; $\boldsymbol{\eta}_t$ is an error with $E(\boldsymbol{\eta}_t) = 0$ and $E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') \equiv \mathbf{Q}_t{=}\sigma_\eta^2 \mathbf{V}_t$.

By defining $\mathbf{V}_t$ as in (42), it follows that,

$$\mathbf{Q}_t^{(2)} = \mathbf{A}\mathbf{Q}_t^{(1)}\mathbf{A}' \tag{45}$$

A.6 Proposition 2 Proof

For the reader's reference the Kalman filter equations, are given by

Prediction Equations

$\hat{\boldsymbol{\alpha}}_{t|t-1} = \hat{\boldsymbol{\alpha}}_{t-1} + \mathbf{c}_t$; $\mathbf{P}_{t|t-1} = \mathbf{P}_{t-1} + \mathbf{Q}_t$

Updating Equations

$\mathbf{F}_t = \mathbf{P}_{t|t-1} + \mathbf{H}_t$; $\hat{\boldsymbol{\alpha}}_t = \hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{F}_t^{-1}(\mathbf{y}_t - \hat{\boldsymbol{\alpha}}_{t|t-1})$; $\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{F}_t^{-1}\mathbf{P}_{t|t-1}$

Assume,

$$\hat{\boldsymbol{\alpha}}_{t-1}^{(2)} = \mathbf{A}\hat{\boldsymbol{\alpha}}_{t-1}^{(1)} \tag{46}$$

from which it follows (because $\boldsymbol{A}$ is non-stochastic) that

$$\mathbf{P}_{t-1}^{(2)} = \mathbf{A}\mathbf{P}_{t-1}^{(1)}\mathbf{A}' \tag{47}$$

Following the Kalman filter covariance cycle

$$\mathbf{P}_{t|t-1}^{(2)} = \mathbf{P}_{t-1}^{(2)} + \mathbf{Q}_t^{(2)} = \mathbf{A}\mathbf{P}_{t-1}^{(1)}\mathbf{A}' + \mathbf{A}\mathbf{Q}_t^{(1)}\mathbf{A}' \text{ (by(45))}$$

$$\mathbf{P}_{t|t-1}^{(2)} = \mathbf{A}\mathbf{P}_{t|t-1}^{(1)}\mathbf{A}' \tag{48}$$

$$\mathbf{F}_t^{(2)} = \mathbf{P}_{t|t-1}^{(2)} + \mathbf{H}_{t|t-1}^{(2)} = \mathbf{A}\mathbf{P}_{t|t-1}^{(1)}\mathbf{A}' + \mathbf{A}\mathbf{H}_t^{(1)}\mathbf{A}' \text{ (by (42))}$$

$$\mathbf{F}_t^{(2)} = \mathbf{A}\mathbf{F}_t^{(1)}\mathbf{A}' \tag{49}$$

The updating equation for $\hat{\boldsymbol{\alpha}}_t^{(2)}$ is

$$\hat{\boldsymbol{\alpha}}_t^{(2)} = \hat{\boldsymbol{\alpha}}_{t-1}^{(2)} + \mathbf{P}_{t|t-1}^{(2)}\left(\mathbf{F}_t^{(2)}\right)^{-1}(\mathbf{y}_t^{(2)} - \mathbf{c}_t^{(2)} - \hat{\boldsymbol{\alpha}}_{t-1}^{(2)})$$

Substituting using 43, 46, 47 and 40,

$$\hat{\boldsymbol{\alpha}}_t^{(2)} = \mathbf{A}\hat{\boldsymbol{\alpha}}_{t-1}^{(1)} + \mathbf{A}\mathbf{P}_{t|t-1}^{(1)}\mathbf{A}'\left(\mathbf{A}\mathbf{F}_t^{(1)}\mathbf{A}'\right)^{-1}(\mathbf{A}\mathbf{y}_t^{(1)} - \mathbf{A}\mathbf{c}_t^{(1)} - \mathbf{A}\hat{\boldsymbol{\alpha}}_{t-1}^{(1)})$$

$$= \mathbf{A}[\hat{\boldsymbol{\alpha}}_{t-1}^{(1)} + \mathbf{P}_{t|t-1}^{(1)}\left(\mathbf{F}_t^{(1)}\right)^{-1}(\mathbf{y}_t^{(1)} - \mathbf{c}_t^{(1)} - \hat{\boldsymbol{\alpha}}_{t-1}^{(1)})] \text{ (because } \boldsymbol{A}\text{is non-singular)} \tag{50}$$

From the definition of $\boldsymbol{c}_t$ following equation (7), it is clear that $\boldsymbol{c}_t^{(2)} = \boldsymbol{A}\boldsymbol{c}_t^{(1)}$.

Thus,

$$\hat{\boldsymbol{\alpha}}_t^{(2)} = \mathbf{A}\hat{\boldsymbol{\alpha}}_t^{(1)} \tag{51}$$

It follows by induction that if the estimation is commenced when (51) holds, invariance will be true for all subsequent years.

B. Preserving Movements in Implicit GDP Deflators through the Smoothing Filter

In this appendix we show that using a fixed interval smoother with $\sigma_\eta^2 = 0$, the resulting smoothed estimates of the state vector, $\mathbf{a}_{t|\mathrm{T}}$, preserve the movement in the implicit price deflator and the covariance matrix of the smoothed estimate equals the Kalman filter estimate of the covariance at time T for all t.

The equations of a fixed interval smoother are,

$$\mathbf{a}_{t|T} = \hat{\boldsymbol{\alpha}}_t + \mathbf{P}_t^*(\mathbf{a}_{t+1|T} - \mathbf{c}_{t+1} - \hat{\boldsymbol{\alpha}}_t) \tag{52}$$

$$\mathbf{P}_{t|T} = \mathbf{P}_t + \mathbf{P}_t^*(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{P}_t^{*'} \tag{53}$$

$$\mathbf{P}_t^* = \mathbf{P}_t\mathbf{P}_{t+1|t}^{-1} \tag{54}$$

where, $\hat{\boldsymbol{\alpha}}_t$ is the Kalman filter estimate of the state vector; $\mathbf{P}_t$ is the Kalman filter unconditional covariance of the state vector ; $\mathbf{P}_{t+1|t}$ is the Kalman filter conditional covariance of the state vector; $\mathbf{a}_{t|T}$ is the Kalman smoothed estimate of the state vector; $\mathbf{P}_{t|T}$ is the covariance of $\mathbf{a}_{t|T}$.

Now, if $\sigma_\eta^2 = 0$, $\mathbf{P}_{t+1|t} = \mathbf{P}_t$, which from (54) implies $\mathbf{P}_t^* = \mathbf{I}_N$. Therefore, $\mathbf{a}_{t|\mathrm{T}} = \mathbf{a}_{t+1|\mathrm{T}} - \mathbf{c}_{t+1}$, or $\mathbf{a}_{t+1|\mathrm{T}} = \mathbf{a}_{t|\mathrm{T}} + \mathbf{c}_{t+1}$. That is, smoothed estimates, $\mathbf{a}_{t|\mathrm{T}}$ preserve the movement in the implicit price deflator.

Now consider the covariance matrix in (53). Since, $\mathbf{P}_{t+1|t} = \mathbf{P}_t$ and $\mathbf{P}_t^* = \mathbf{I}_N$ we have, $\mathbf{P}_{t|\mathrm{T}} = \mathbf{P}_{t+1|\mathrm{T}}$. Thus, $\mathbf{P}_{t|\mathrm{T}}$ is constant with respect to t and $\mathbf{P}_{t|\mathrm{T}} = \mathbf{P}_{T|T} = \mathbf{P}_{T|T}$ for all t.