Centre for Efficiency and Productivity Analysis

Working Paper Series
No. 03/2005

Title

Constructing Internationally Comparable Real Income Aggregates by Combining Sparse Benchmark Data with Annual National Accounts Data. A State-Space Approach

Authors
Alicia N. Rambaldi
D.S. Prasada Rao & Howard E. Doran

Date: April 2005

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

Constructing Internationally Comparable Real Income Aggregates by Combining Sparse Benchmark Data with Annual National Accounts Data. A State-Space Approach

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

Howard E. Doran
School of Economics
University of New England
Armidale, Australia

Revised version of the paper prepared for the:
WORKSHOP ON FRONTIERS IN BENCHMARKING TECHNIQUES AND THEIR APPLICATION TO OFFICIAL STATISTICS
Organised by OECD and EUROSTAT
Luxembourg, 7 & 8 April 2005
Plateau du Kirchberg

Abstract

The importance of availability of comparable real income aggregates and their components to applied economic research is highlighted by the popularity of the Penn World Tables. Any methodology designed to achieve such a task requires the combination of data from several sources. The first is purchasing power parities (PPP) data available from the International Comparisons Project roughly every five years since the 1970s. The second is national level data on a range of variables that explain the behaviour of the ratio of PPP to market exchange rates. The final source of data is the national accounts publications of different countries which include estimates of gross domestic product and various price deflators. In this paper we present a method to construct a consistent panel of comparable real incomes by specifying the problem in state-space form. We present our completed work as well as briefly indicate our work in progress.

JEL Classification: C53, C33

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

The authors gratefully acknowledge the support of the Australian Research Council (DP0557606) to pursue research on this topic.

Constructing Internationally Comparable Real Income Aggregates by Combining Sparse Benchmark Data with Annual National Accounts Data. A State-Space Approach

Alicia N. Rambaldi, D.S. Prasada Rao and Howard E. Doran

1. Introduction

Econometric studies of growth, catch up and convergence are heavily reliant on internationally comparable time series of gross domestic product and per capita incomes, which are expressed in a common currency unit and adjusted for price differences across countries and over time. The Penn World Tables (PWT) have been the main source of such data for over two decades. The PWT data are based on the Purchasing Power Parities (PPP) compiled under the auspices of the International Comparison Program (ICP) known as benchmark PPPs. These data are reliable in that the benchmarking exercise is conducted in a given year across a number of participating countries, using a common basket of commodities. However, benchmarking exercises are conducted roughly every five years (since the 1970s) and the number of countries participating in the exercise has varied. The first few benchmark exercises were limited to a handful of countries, although the participation has substantially increased over the three decades. For the current phase of the ICP, in 2005-2006, a large number of countries (around 150) are participating. Thus, the problem is one of extrapolating the benchmark information over time and across non-participating countries to construct a large panel.

The current method for the construction of time series of PPPs, PWT, for a large number of countries is a two-step method. The PWT are constructed by: (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; and (ii) extrapolation to non-benchmark years. The second step combines the information from step (i) with national accounts data to produce the tables.

The main objective of this paper is to propose a methodology that will allow the joint use of *all* benchmark PPP data with data from the other two sources for purposes of extrapolations and projections. The methodology makes *full and efficient* use of all the information available and obtains optimal predictors of PPPs for all the countries and time periods, as well as making possible the derivation of standard errors associated with the PPPs thereby providing measures of errors in predictions for various macroeconomic aggregates.

The paper proposes the use of a state-space formulation that can generate predictions for non-participating countries in different benchmark years and at the same time provide projections of PPPs that are consistent with country-specific temporal movements in prices. As an illustration, we develop a fairly general econometric model that allows for cross-sectional correlations through an appropriately specified spatially correlated error structure. The feasibility and performance of the method is demonstrated using the state-space formulation of this model on data from 23 OECD countries.

2. Combining Economic Theory with Available Data

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

$$R_i = \frac{PPP_i}{ER_i} \tag{1}$$

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 and 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-tradeable sector and on the nature and extent of any barriers to free trade (Clague, 1988; Bergstrand, 1991, 1996; Ahmad, 1996).

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:

$$R_i = f(X_1, X_2, X_3, \ldots X_k) + e_i \tag{2}$$

where e_i is a random disturbance with specific distributional characteristics.

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:

$$PPP_{i,t} = PPP_{i,t-1} \times \frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}} \tag{3}$$

From (3) the movement of the national price level over time is then given by:

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

This can be used in conjunction with the prediction model in equation (2).

Equations (2) to (4) clearly demonstrate the type of data needed for the construction of a panel of PPPs over time and across countries. It is evident that the basic data requirements consist of: (i) PPPs and exchange rates for countries from the ICP on the LHS of equation (2); (ii) data on a number of explanatory variables to explain the ratio R_i – such data available from national sources; and (iii) data on GDP deflators of all the countries needed in equation (4) – these data are available from the national accounts of countries. Thus, construction of a consistent panel of PPPs requires efficient use of information drawn from a variety of sources, an exercise in combining benchmark data.

The next section develops an econometric model to combine the three sources of data and the national price level literature to obtain a panel of PPPs.

2.1 Econometric formulation of the problem

A random variable $r_{it} = \ln(PPP_{it} / ER_{it})$ is considered for each country *i* (*i* =1,2, …, N) and year *t* (*t* = 1,2, …, T) where PPPs and exchange rates are all measured relative to the currency of a reference currency (US is used as the reference country in the empirical illustration reported here). By definition, $r_{it} \equiv 0$ for the reference country[1], but it is otherwise observed with error. We wish to produce a panel of predictions of r_{it} (denoted $\hat{r}_{it}$) accompanied by standard errors which optimally uses all relevant available data, 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

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

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

2.2 Assumptions

(i) There is a linear relationship[2] $\mathbf{r}_t = \mathbf{X}_t^* \boldsymbol{\beta} + \mathbf{e}_t$

where,

$\mathbf{X}_t^*$ (N × K) is observed and $\boldsymbol{\beta}$ (K × 1) is an unknown parameter vector.

(ii) Because of the time-series/cross-section nature of r_{it}, we assume that it is characterized by both autocorrelation and spatial correlation. We adopt a simple model for $\boldsymbol{e}_t$, to capture these effects, as follows

[1] The USA is the customary choice.
[2] Specification of this model including the choice of regressors draws heavily from the literature on explaining national price levels (see (Kravis and Lipsey 1983 and 1986; Clague, 1988; Bergstrand, 1996, and Ahmad (1996)).

$$\mathbf{e}_t = \rho \mathbf{e}_{t-1} + \mathbf{u}_t \tag{5}$$

where,

$|\rho| \leq 1$ is unknown;

$\mathbf{u}_t$ is normally distributed with

$E(\mathbf{u}_t) = 0, \qquad E(\mathbf{u}_t \mathbf{u}_t') = \sigma^2 \mathbf{\Omega}, \qquad \mathrm{E}(\mathbf{u}_t \mathbf{u}_{t-s}') = 0$

with

$$\mathbf{\Omega}^{-1} = (\mathbf{I} - \phi \mathbf{W})(\mathbf{I} - \phi \mathbf{W})'.$$

Here σ^2 and ϕ are unknown parameters, and $\mathbf{W}$ (N $\times$N) is a known matrix which is determined by contiguity relationships between countries. We assume W has been "row normalized" (for example, rows adding to 1), and ($\mathbf{I}$ - $\phi\mathbf{W}$) is positive definite. These assumptions imply that $\phi < 1$.

2.3 Observations

While r_{it} is never observed, relevant observations are available to enable its estimation.

(i) Causal or conditioning variables $X^*_{it,j}$ (j = 1,2, …, K) are observed in all countries and all years.

(ii) For all years, a variable $\mathbf{g}^*_t$, can be observed from National Accounts. We call $\mathbf{g}^*_t$ the observed growth rate vector[3]. We recognise that there is some measurement error in the National Accounts and assume that $\mathbf{g}^*_t$ is not identical to $\mathbf{g}_t = \mathbf{r}_t - \mathbf{r}_{t-1}$.

---

[3] The growth rate is $u_{it} = \dfrac{R_{it} - R_{i,t-1}}{R_{i,t-1}}$, where $R_{it} = PPP_{it} / ER_{it}$. Then $\dfrac{R_{it}}{R_{i,t-1}} = 1 + u_{it}$. Taking logarithms, $r_{it} - r_{i,t-1} = \ln(1+u_{it}) \approx u_{it}$ assuming μ_{it} <<1. Thus, $\mathbf{g}_t$ is approximately equal to the growth-rate vector.

(iii) In "benchmark" years, a known subset N_t of the countries participate in benchmarking. The benchmark r_{it}^* is taken to be an approximation to the unobserved r_{it}. We denote the N_t – vector of benchmarks by $\mathbf{r}_t^*$.

(iv) The reference country, $i=N$, must satisfy the constraint $\hat{r}_{Nt} = r_{Nt} \equiv 0$ for all t[4].

Thus,

For all years,

$$\mathbf{g}_t^* = \mathbf{g}_t + \boldsymbol{\xi}_{1t} = (\mathbf{X}_t^* - \mathbf{X}_{t-1}^*)\boldsymbol{\beta} + (\mathbf{e}_t - \mathbf{e}_{t-1}) + \boldsymbol{\xi}_{1t} \qquad (6)$$

and for benchmark years there is a known $N_t \times N$ selection matrix $\mathbf{S}_t$ which selects the participating countries and relates $\mathbf{r}_t^*$ and $\mathbf{r}_t$ by

$$\mathbf{r}_t^* = \mathbf{S}_t \mathbf{r}_t + \boldsymbol{\xi}_{2t} \qquad (7)$$

$\boldsymbol{\xi}_{st}$, s=1,2 are the measurement error of the growth rate and benchmark respectively, taken to be normally distributed. A crucial assumption is that the variances of $\xi_{s,it}$ are inversely proportional to the level of development, measured here by per capita GDP.

Thus,

$$E(\xi_{s,it}) = 0, \;\; E(\xi_{s,it}^2) = \sigma_s^2 V_{ii,t} \;\;\; E(\xi_{s,it}\xi_{s,jt}) = 0 \;\; (j \neq i).$$

where $V_{ii,t}$ is the inverse of per capita GDP of country i in year t and σ_s^2 are unknown constants of proportionality.

We now present a state-space formulation of the model specified above.

2.4 A state space representation

We define an unobservable "state vector" α_t by

$$\alpha_t = [\mathbf{e}_t', \mathbf{e}_{t-1}']' \qquad (8)$$

Thus, from (5)

[4] This is because both PPP_{it} and ER_{it} are measured relative to the currency of the reference country.

$$\boldsymbol{\alpha}_t = \mathbf{D}\,\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \qquad (9)$$

where,

$\mathbf{D} = \rho\mathbf{I}_{2N}$ and $\boldsymbol{\eta}_t = (\mathbf{u}'_t, \mathbf{u}'_{t-1})'$. Denoting $E(\boldsymbol{\eta}_t\boldsymbol{\eta}'_t) = \sigma^2\mathbf{Q}_t$ and $E(\boldsymbol{\alpha}_t\boldsymbol{\alpha}'_t) = \sigma^2\mathbf{P}_t$

$$\mathbf{Q}_t = \mathbf{I}_2 \otimes \boldsymbol{\Omega}_t \qquad (10)$$

and

$$\boldsymbol{\alpha}_t \sim N(\mathbf{0}, \sigma^2\mathbf{P}_t)$$

Furthermore, for all t, there exists an observed vector $\mathbf{y}_t$ satisfying

$$\mathbf{y}_t = \mathbf{Z}_t\boldsymbol{\alpha}_t + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\xi}_t \qquad (11)$$

where $E(\boldsymbol{\xi}_t) = 0$ and $E(\boldsymbol{\xi}_t\boldsymbol{\xi}'_t) = \lambda\mathbf{H}_t$. The measurement equation includes an exact constraint to insure $\hat{r}_{it} = 0$ when i is the reference country.

The matrices $\mathbf{y}_t$, $\mathbf{Z}_t$, $\mathbf{X}_t$ and $\mathbf{H}_t$ are defined differently for the benchmark years, the years after the benchmark and for the remaining non-benchmark years, as follows:

(i) <u>Non-benchmark years</u>

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{j}_2 \\ \mathbf{S}_{N-1}\mathbf{g}_t \end{bmatrix}$$

$$\mathbf{Z}_t = \begin{Bmatrix} (\mathbf{I}_2 \otimes \nu'_{RC}) \\ \mathbf{S}_{N-1}[\mathbf{I}_N, -\mathbf{I}_N] \end{Bmatrix}$$

$$\mathbf{X}_t = \begin{Bmatrix} (\mathbf{I}_2 \otimes \nu'_{RC})[\mathbf{X}_t^{*\prime}, \mathbf{X}_{t-1}^{*}{}']' \\ \mathbf{S}_{N-1}(\mathbf{X}_t^* - \mathbf{X}_{t-1}^*) \end{Bmatrix}$$

$$\mathbf{H}_t = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mu(\mathbf{S}_{N-1}\mathbf{V}_t\mathbf{S}'_{N-1}) \end{bmatrix}$$

where,

$\mathbf{j}_2 = [0, 0]'$ is an augmentation term to satisfy the reference country constraint

$\mathbf{S}_{N-1}$ is (N-1) × N and selects all but the reference country, and

ν'_{RC} is a selection vector for the reference country

$\mathbf{I}_M$ is an identity matrix of dimension *M*, and

$\mathbf{X}_t^*$ is the matrix of observed conditioning variables

$\mathbf{V}_t$ is diagonal with elements $V_{ii,t}$

$\mu = \sigma_2^2 / \sigma_1^2$

$\lambda = \sigma_1^2 / \sigma^2$

These definitions simply express the fact that $\mathbf{g}_t$, the observed growth rates from National Accounts are subject to some measurement error proportional to the inverse of per capita GDP of country *i* in year *t*. The row dimension of all matrices is N_{1t} = N+1.

(ii) Benchmark years

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{j}_2 \\ \mathbf{r}_t^* \\ \mathbf{S}_{N-1}\mathbf{g}_t \end{pmatrix}$$

$$\mathbf{Z}_t = \begin{Bmatrix} (I_2 \otimes \nu'_{RC}) \\ \mathbf{S}_t[\mathbf{I}_N, \mathbf{0}] \\ \mathbf{S}_{N-1}[\mathbf{I}_N, -\mathbf{I}_N] \end{Bmatrix}$$

$$\mathbf{X}_t = \begin{Bmatrix} (\mathbf{I}_2 \otimes \nu'_{RC})[\mathbf{X}_t^{*\prime}, \mathbf{X}_{t-1}^{*}{}'] ' \\ \mathbf{S}_t \mathbf{X}_t^* \\ \mathbf{S}_{N-1}(\mathbf{X}_t^* - \mathbf{X}_{t-1}^*) \end{Bmatrix}$$

$$\mathbf{H}_t = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_t\mathbf{V}_t\mathbf{S}'_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu(\mathbf{S}_{N-1}\mathbf{V}_t\mathbf{S}'_{N-1}) \end{bmatrix}$$

As above, these definitions reflect the fact that in benchmark years, the growth rate information is augmented by approximations to $\mathbf{r}_t$, given by (7). The row dimension of all matrices is $N_{1t} = N_t + N+1$.

(iii) First year after a Benchmark

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{j}_2 \\ \mathbf{r}^*_{t-1} \\ \mathbf{S}_{N-1}\mathbf{g}_t \end{pmatrix}$$

$$\mathbf{Z}_t = \begin{Bmatrix} (I_2 \otimes \nu'_{RC}) \\ \mathbf{S}_{t-1}[\mathbf{I}_N, \mathbf{0}] \\ \mathbf{S}_{N-1}[\mathbf{I}_N, -\mathbf{I}_N] \end{Bmatrix}$$

$$\mathbf{X}_t = \begin{Bmatrix} (\mathbf{I}_2 \otimes \nu'_{RC})[\mathbf{X}^{*\prime}_t, \mathbf{X}^{*\prime}_{t-1}]' \\ \mathbf{S}_{t-1}\mathbf{X}^*_{t-1} \\ \mathbf{S}_{N-1}(\mathbf{X}^*_t - \mathbf{X}^*_{t-1}) \end{Bmatrix}$$

$$\mathbf{H}_t = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{t-1}\mathbf{V}_{t-1}\mathbf{S}'_{t-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu(\mathbf{S}_{N-1}\mathbf{V}_t\mathbf{S}'_{N-1}) \end{bmatrix}$$

These definitions recognise that the year following a benchmark year cannot be treated as a regular non-benchmark year given that the state vector involves $\mathbf{e}_t$ and $\mathbf{e}_{t-1}$. Here also $N_{1t} = N_t + N+1$.

Equations (9) and (11) are the "transition" and "observation" equations of a conventional state-space system. Conditional on the unknown parameters ρ, ϕ, β, σ^2, σ_1^2 and σ_2^2, optimal MSE estimates $\hat{\boldsymbol{\alpha}}_t$ of the state vector $\boldsymbol{\alpha}_t$ can be obtained using the Kalman Filter (see Harvey 1990, 100-110 and 130-133).

3. Estimation

For ease of reference, we will set down the recursive equations of the Kalman Filter, generally using Harvey's (1981, 1990) notation. At this stage we are assuming that ρ, ϕ, β, σ^2, σ_1^2 and σ_2^2 are known which in turn implies that $\mathbf{Q}$ and $\mathbf{H}_t$ are known. Starting with the covariance matrix $\mathbf{P}_{t-1}$ the '*covariance cycle'* is given as follows.

$$\mathbf{P}_{t|t-1} = \mathbf{D}\,\mathbf{P}_{t-1}\,\mathbf{D}' + \mathbf{Q} \tag{12}$$

$$\mathbf{F}_t^* = \mathbf{Z}_t\left(\sigma^2\,\mathbf{P}_{t|t-1}\right)\mathbf{Z}_t{}' + \sigma_1^2\,\mathbf{H}_t \tag{13}$$

For later convenience, we define

$$\mathbf{F}_t = \mathbf{F}_t^* / \sigma^2$$

$$\lambda = \sigma_1^2 / \sigma^2 \tag{14}$$

Then σ^2 can be cancelled from (13) to yield

$$\mathbf{F}_t = \mathbf{Z}_t\ \mathbf{P}_{t|t-1}\mathbf{Z}_t' + \lambda\,\mathbf{H}_t \tag{15}$$

Finally, the cycle is completed by

$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\,\mathbf{Z}_t'\mathbf{F}_t^{-1}\,\mathbf{Z}_t\,\mathbf{P}_{t|t-1} \tag{16}$$

Thus the 'covariance cycle' moves from $\mathbf{P}_{t-1}$ to $\mathbf{P}_t$ in the sequence:

$$\mathbf{P}_{t-1} \Rightarrow \mathbf{P}_{t|t-1} \Rightarrow \mathbf{F}_t \Rightarrow \mathbf{P}_t$$

as given in (12), (15) and (16).

The '*state-vector cycle*' starts with $\hat{\boldsymbol{\alpha}}_{t-1}$ and updates as follows:

$$\hat{\boldsymbol{\alpha}}_{t|t-1} = \mathbf{D}\,\hat{\boldsymbol{\alpha}}_{t-1}$$

$$\boldsymbol{\upsilon}_t = \mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta} - \mathbf{Z}_t\,\hat{\boldsymbol{\alpha}}_{t|t-1}\ ,$$

where υ_t is the prediction error with covariance matrix $\sigma^2\,\mathbf{F}_t$. The prediction error is used to obtain $\hat{\boldsymbol{\alpha}}_t$ by

$$\hat{\boldsymbol{\alpha}}_t = \hat{\boldsymbol{\alpha}}_{t|t-1} + \mathbf{K}_t\,\boldsymbol{\upsilon}_t \tag{17}$$

where $\mathbf{K}_t$, known as the Kalman gain, is given by

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\,\mathbf{Z}_t\,\mathbf{F}_t^{-1}. \qquad (18)$$

Thus the 'state vector cycle' updates $\hat{\boldsymbol{\alpha}}_{t-1}$ in the sequence $\hat{\boldsymbol{\alpha}}_{t-1} \Rightarrow \hat{\boldsymbol{\alpha}}_{t|t-1} \Rightarrow \boldsymbol{\upsilon}_t \Rightarrow \hat{\boldsymbol{\alpha}}_t$.

Because the N_{1t} dimensional prediction error $\boldsymbol{\upsilon}_t$ has distribution $\boldsymbol{\upsilon}_t \sim N(\mathbf{0}, \sigma^2 \mathbf{F}_t)$, the log of the likelihood function can be written as $\mathbf{L} = \sum_{t=1}^{T} L_t$, where

$$L_t = -\frac{N_{1t}}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma^2 \mathbf{F}_t\right| - \frac{1}{2}\boldsymbol{\upsilon}_t{}'(\sigma^2\mathbf{F}_t)^{-1}\boldsymbol{\upsilon}_t$$

Thus,

$$L = -\frac{1}{2}\left[\ln(2\pi) + \ln\sigma^2\right]\sum_{t=1}^{T} N_{1t} - \frac{1}{2}\sum_{t=1}^{T}\ln\left|\mathbf{F}_t\right| - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\boldsymbol{\upsilon}_t'\mathbf{F}_t^{-1}\boldsymbol{\upsilon}_t\ .$$

It is quite simple to derive a concentrated form of the likelihood function as

$$L_c = -\frac{1}{2}\left[1 + \ln(2\pi) + \ln\hat{\sigma}^2\right]\sum_{t=1}^{T} N_{1t} - \frac{1}{2}\sum_{t=1}^{T}\ln\left|\mathbf{F}_t\right| \qquad (19)$$

where

$$\hat{\sigma}^2 = \sum_{t=1}^{T}\boldsymbol{\upsilon}_t\mathbf{F}_t^{-1}\boldsymbol{\upsilon}_t \Big/ \sum_{t=1}^{T} N_{1t} \qquad (20)$$

The parameters ρ, φ, μ and λ are hyperparameters, which are bounded between 0 and 1 in this case, are obtained by numerical maximization of L_c. Estimates of $\boldsymbol{\beta}$ are obtained at every iteration by a conditional GLS (see Harvey, 1990: 130-133). A final pass of the Filter yields $\hat{\sigma}^2$ and L_c conditional on $\hat{\boldsymbol{\beta}}, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\mu},$ and $\hat{\rho}$. The Kalman Filter and smoother are then run to obtain the sequences ($\hat{\boldsymbol{\alpha}}_t$, $\mathbf{P}_t$) for t = 1,2, . . ., T.

The standard errors for the predicted PPP are computed as follows:

$$P\hat{P}P_{it} = \exp(\hat{r}_{it}) \times ER_{it} \qquad (21)$$

$$\begin{aligned} SE(P\hat{P}P_{it}) &= \sqrt{\operatorname{var}(\hat{r}_{it})} \times P\hat{P}P_{it} \\ &= \sqrt{\hat{P}_{ii,t}} \times P\hat{P}P_{it} \end{aligned} \qquad (22)$$

where,

$\hat{P}_{ii,t}$ is the *ith* diagonal element of the estimated covariance of the state vector, $\hat{\boldsymbol{\alpha}}$.

Equation (22) is obtained using the definition of the variance of a function and a Taylor's Expansion.

4. An Illustration

We present a small illustration of the method using OECD data. These data can be easily accessed through the OECD and World Bank sites. Several of the countries in the OECD were participants in the ICP project since its first benchmark year. We include 23 countries in this illustration, they are: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Japan, (S.) Korea, Mexico, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Turkey, the United Kingdom, and the United States as the reference country.

The data for the empirical example are for the period 1970 to 2000, annual, and we discuss next the *dependent*, *explanatory*, and *covariance related variables*.

Dependent Variable

Benchmark PPP information, GDP Deflators, and exchange rates were collected from the OECD site and the World Bank's Stars data set. Benchmark years were: 1975, 1980, 1985, 1990, 1995 and 1999. All countries in this sample with the exception of Hungary (did not participate in 1975 and 1990) participated in all the benchmarks.

The data are expressed in the following currencies:

COUNTRY	**ABBREVIATION**	**CURRENCY**
Australia	AUS	Australian dollar
Austria	AUT	Euros (1999 ATS euro)
Canada	CAN	Canadian dollar
Belgium	BEL	Euros (1999 BEF euro)
Denmark	DNK	Danish kroner
Finland	FIN	Euros (1999 FIM euro)

France	FRA	Euros (1999 FRF euro)
Germany	DEU	Euros (1999 DEM euro)
Greece	GRC	Euros (2001 GRD euro)
Hungary	HUN	Forint
Italy	ITA	Euros (1999 ITL euro)
Japan	JPN	Yen
S. Korea	KOR	South Korean Won
Mexico	MEX	Mexican pesos
Netherlands	NLD	Euros (1999 NLG euro)
New Zealand	NZL	New Zealand dollar
Norway	NOR	Norwegian kroner
Portugal	PRT	Euros (1999 PTE euro)
Spain	ESP	Euros (1999 ESP euro)
Sweden	SWE	Swedish kronor
Turkey	TUR	Turkish lire
Great Britain	GBR	Pounds sterling

Explanatory Variables

The following variables were included as explanatory variables in the model:

Euro Dummy: Takes the value of 1 from 1993 onwards for the countries that joined the euro currency by 2000.

FDI%: Foreign direct investment, net inflows (% of GDP)

LE: Life Expectancy in years

SERV%: Services, value added (% of GDP)

OPEN%: Trade (% of GDP)

$\mathbf{CPI}_{it}/\mathbf{CPI}_{US,t}$, for $i = 1, \ldots, N$

Labour Productivity = (Population × per capita GDP)/ Labour Force

The choice of conditioning variables is based on national price level theory and data availability. The data were obtained from the OECD site and from various issues of the World Development Indicators.

The model used here is adapted from those used in the literature to suit the nature and scope of the current study. In particular, since the model is only illustrative and is applied to only OECD countries, a variable like education has not been included. The effect of productivity differentials on national price levels is captured through the inclusion of a labour productivity measure. The model is a first approximation and further work and refinements are planned for the next stage of this project.

Covariance related Variables

a) *Measuring spatial correlation*

A contiguity matrix was constructed using volumes of bilateral trade in 1990. This is the matrix **W** (see Section 2.).

b) *Capturing accuracy of benchmark data collection and National Accounts' computation of the national price level.*

As stated we assume that the accuracy of a PPP benchmark and the growth rate on the national price level is inversely related to a country's GDP per capita (measured in constant US$ of 1995).

4.1 Model Estimates

In Section three we showed the Kalman Filter cycle to obtain a value of the concentrated likelihood function by rewriting three of the five original hyper-parameters into two ratios $\lambda = \sigma_1^2 / \sigma^2$ and $\mu = \sigma_2^2 / \sigma_1^2$. The main benefit is to obtain a specification where all hyper-parameters are bounded above by one which highly simplifies the search for starting values. We ran the alternative specifications (ie assuming $\sigma_1^2 > \sigma_2^2$ or $\sigma_1^2 < \sigma_2^2$) and found that in all cases the estimated values for λ were consistent with $\sigma^2 \approx \sigma_1^2$. Table 1 presents a summary of the estimation results that form the basis for the predictions presented in Figures 1, 3 and Table 2. The numerical optimisation worked well and the estimates of the spatial and autocorrelated parameters were fairly robust over all possible alternative specifications of λ and μ.

The regression fits the data well. Running a simple pooled regression of the benchmark data over the sample period (160 observations) yields a R^2 of 0.62, with parameters estimates close to those obtained by our method (when they are significant). Other relevant variables could be included in the regression and any future work will further explore alternative regression specifications.

4.2 Predictions of PPP and National Price Level

We only present the results for two countries as an illustration of the method, they are Australia and Turkey[5]. We present graphical results for Australia (see Figures 1, 2 and 3) and a table of results for Turkey. Due to the hyperinflation suffered by Turkey during the sample period, it is difficult to capture the results in a graphical form. Further, Figures 1 and 2 compare the predictions for PPP of the method under complete and incomplete benchmark information, as well as to the PWT6.1 values. Predictions in Figure 1 are based on the use of all available benchmark information. In contrast Figure 2 assumes that Australia only participated in the 1999 benchmark exercise and therefore the results show how the model performs when predictions are formed primarily from the observation of national account's growth rates and the spatial covariance structure. Figures 1 and 2 also illustrate how prediction intervals widen considerably when no benchmark information is available.

Australia floated its exchange rate in 1983. This can be observed in Figure 3 where both our predicted ratio and the PWT6.1 are presented. During the fixed exchange rate period it is widely accepted that the Australian dollar was over- valued. Note that the price level ratio is hovering around one since the floating of the exchange rate. This is the expected result, consistent with the purchasing power parity theory and the theory of national price level.

Table 2 presents the results for Turkey. It is clear that the predictions of our model are consistent and track the observed benchmark information closely, even during the periods of hyperinflation. We believe this result provides a strong indication that our modelling approach is performing well.

[5] Full results for all countries in the sample are available from the authors.

Overall, the results for all countries in the sample are similar to those presented above for Australia and Turkey. That is, the PPP predictions are close to benchmark observations and consistent with the known historical facts of the individual countries.

5. Conclusions

The main objective of the paper is to demonstrate how a state-space approach can be employed in the estimation of a panel of purchasing power parities necessary for constructing a consistent set of internationally comparable real income aggregates. The methodology described here successfully combines data drawn from a number of national and international sources in estimating PPPs. It offers several improvements over the existing PWT approach, which is the only source of such data at the present time. These improvements include a method that: (i) can make use of all the PPP data from the ICP for all the benchmark years since 1970; (ii) can provide optimal predictors for PPPs for ICP-non-participating countries and for non-benchmark years; (iii) produces PPPs that are consistent with observed movements in prices in different countries; and (iv) provides standard errors associated for the PPPs and, therefore, for the estimates of real per capita incomes. To achieve these objectives the paper proposes the use of an econometric model with errors that are spatially correlated cross-sectionally and autocorrelated temporally. The econometric model is re-formulated in a state-space form and estimated using Kalman filtering techniques. The new methodology is applied to an illustrative data set of 23 OECD countries for the period 1970 to 2000. The results from the illustrative application demonstrate the feasibility of using the model for consistent space-time extrapolation. Our results show how prediction intervals widen considerably during non-benchmark years and when only a limited number of benchmark data are used. Further research focusing on refinements to the model specification is currently in progress.

References

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

Bergstrand, J.H, (1991), “Structural determinants of real exchange rates and national price levels”, *American Economic Review*, 81, 325-334.

Bergstrand, J.H. (1996), “ Productivity, Factor Endowments, Military Expenditures, and National Price Levels” in *International Comparisons of Prices, Ouput and Productivity*, D.S. Prasada Rao and J. Salazar-Carrillo (eds.), Elsevier Science Publishers B.V. North Holland.

Clague, C.K.(1988) “Explanations of National Price Levels,” in *World Comparison of Incomes, Prices and Product*, J. Salazar-Carrillo and D.S. Prasada Rao (eds.), Elsevier Science Publishers B.V, North Holland.

Harvey, A. C. (1990), Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge Univ. Press. Cambridge.

Heston, A., R. Summers and B. Aten, Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

Kravis, Irving B., and R E. Lipsey (1983) “Toward an Explanation of National Price Levels,” Princeton Studies in International Finance, No 52 Princeton, N.J.: Princeton University, International Finance Section.

Kravis, Irving B., and R E. Lipsey (1986), “The Assessment of National Price Levels,” Paper presented at Eastern Economic Association Meetings, Philadelphia, April.

Summers, R. and A. Heston (1988), "Comparing International Comparisons", in *World Comparisons of Incomes, Prices and Product*, (eds.) Salazar-Carrillo and D.S. Prasada Rao, Contributions to Economic Analysis Series, North-Holland.

Summers, R. and A. Heston (1991), “The Penn World Tables (Mark 5): An expanded set of international comparisons, 1950-88”, *Quarterly Journal of Economics*, 2, 1-45.

Table 1. Estimated Parameters

HYPERPARAMETERS		
Parameters	Estimates	
$\hat{\lambda}$	1.0000	
$\hat{\mu}$	5.8890e-003	
$\hat{\phi}$	7.1269e-001	
$\hat{\rho}$	5.8049e-001	
L_c	3.1965	
Constants of Proportionality		
$\hat{\sigma}^2$	2.3033e-002	
$\hat{\sigma}_1^2$	2.3033e-002	
$\hat{\sigma}_2^2$	1.3564e-004	
Regression Parameters		
Regressor	Estimates	Standard Error
Intercept	-9.6460e-001(**)	1.4569e-001
Euro Dummy	1.0318e-001(**)	4.2208e-002
FDI%	1.3768e-003	3.5938e-003
LE%	3.1581e-003(**)	1.0253e-003
SERV%	5.6478e-004	2.3803e-003
OPEN%	3.1757e-004	4.6479e-004
Labour Productivity	1.7881e-005(**)	9.6717e-007
CPI_i/CPI_{us}	5.2888e-002(**)	1.2525e-002

(**) Statistically Significant at the 5%

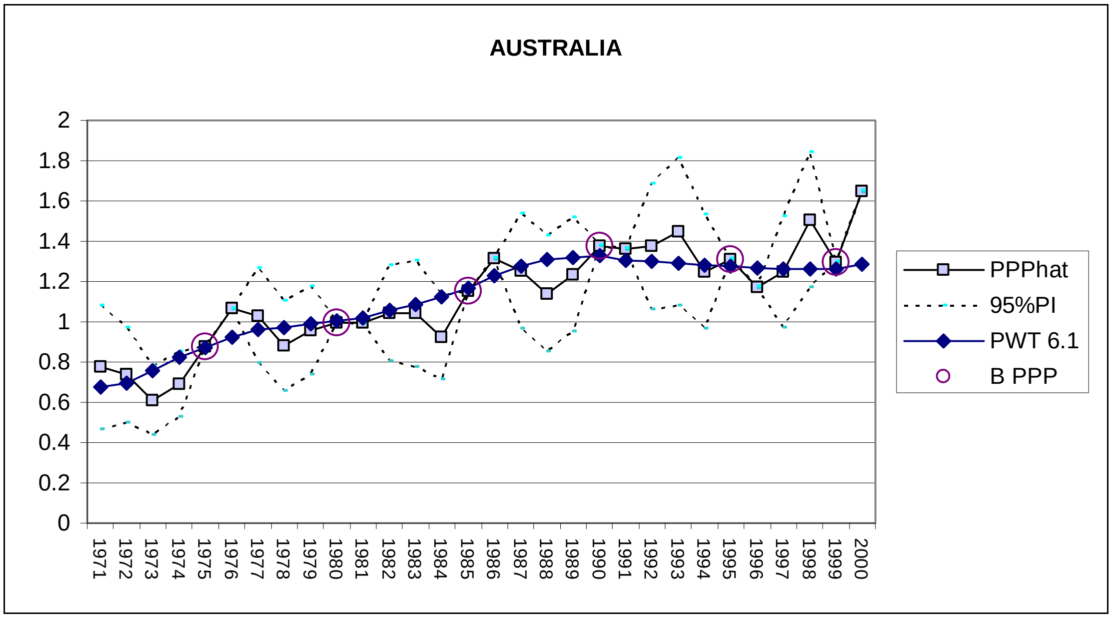

Figure 1. All Benchmark information assumed to be known

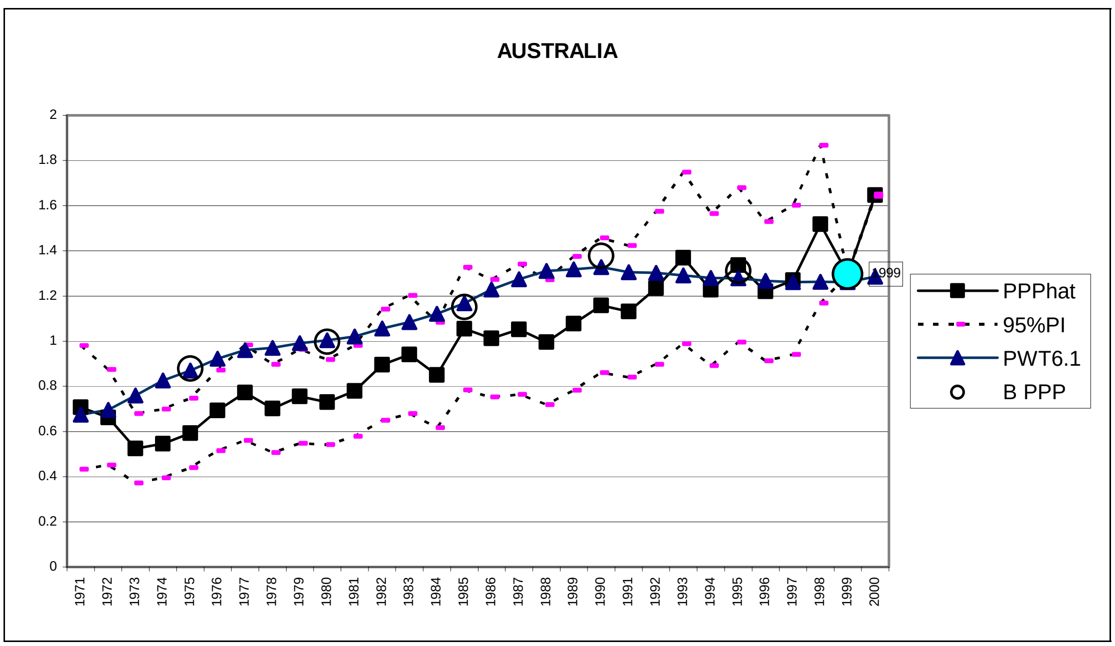

Figure 2. Only 1999 Benchmark information used

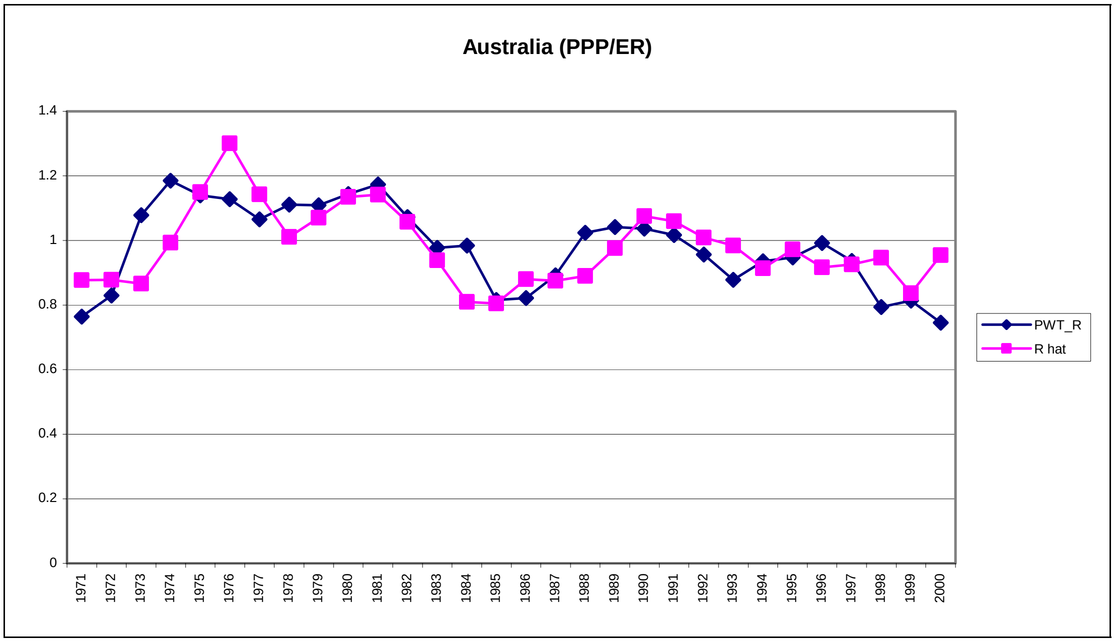

Figure 3. Comparison of Estimated Price Level Ratio

Table 2. Predicted PPPs and Standard Errors with Complete Information, Turkey

Exchange Rate	Benchmark PPP	Predicted log of PPP/ER	Predicted PPP	Standard Error of Predicted PPP
14.917		-0.3975	10.0	2.06
14.15		-0.5672	8.0	1.34
14.15		-0.4768	8.8	1.29
13.927		-0.4166	9.2	1.13
14.442	10.598	-0.3095	10.6	0.04
16.053		-0.1185	14.3	0.07
18.002		-0.3287	13.0	1.56
24.282		-0.3503	17.1	2.28
31.078		-0.4906	19.0	2.29
76.038	51.468	-0.3894	51.5	0.18
111.219		0.2647	144.9	0.70
162.553		-0.1107	145.5	17.48
225.457		-0.3913	152.5	20.32
366.678		-0.5013	222.1	26.68
521.983	211.067	-0.9049	211.2	0.68
674.512		-0.3644	468.5	2.12
857.216		-0.4215	562.4	67.54
1422.35		-0.2097	1153.3	153.75
2121.68		-0.3014	1569.5	188.51
2608.64	1539.871	-0.5267	1540.6	4.60
4171.82		0.3686	6031.1	25.55
6872.42		0.0789	7436.5	893.14
10984.6		-0.1728	9241.5	1232.01
29608.7		-0.0547	28033.3	3366.84
45845.1	22886.97	-0.6940	22902.7	66.38
81404.9		0.4349	125756.4	509.09
151865		0.1375	174245.9	20354.64
260724		-0.2851	196046.9	22901.25
418783	191772.5	-0.7806	191863.5	542.40
625218		0.0033	627297.7	2474.53