EXTRAPOLATION OF PURCHASING POWER PARITIES USING MULTIPLE BENCHMARKS AND AUXILIARY INFORMATION: A NEW APPROACH

BY D. S. PRASADA RAO,* ALICIA RAMBALDI AND HOWARD DORAN

University of Queensland

The paper presents an econometric framework for the construction of a consistent panel of purchasing power parities (PPPs) which makes it possible to combine all the PPP benchmark data from various phases of the International Comparison Program with the data on national price movements in the form of implicit deflators from national accounts. The method improves upon the current practice used in the construction of the Penn World Tables (PWT), and similar tables produced by the World Bank which tend to be anchored on a selected benchmark. The econometric formulation is based on a regression model for the national price levels where the disturbances are assumed to be heteroskedastic and spatially correlated across countries. The regression model along with data on country specific price movements are combined using a state–space formulation and optimum predictions of PPPs are obtained. As a property of the method presented in the paper, we show that the resulting PPP predictions are weighted averages of extrapolations of PPPs from different benchmarks—thus the method provides a formal approach which has a simple intuitive interpretation. The smoothed PPP predictions (and standard errors) obtained through the state–space are produced for both ICP-participating and non-participating countries and non-benchmark years. A complete tableau of PPPs for 141 countries spanning the period 1970 to 2005 is compiled using the method. Results for some selected countries are presented and the new series are compared and contrasted with the currently available PWT series. Extrapolated series for the remaining countries are available from the authors upon request.

1. INTRODUCTION

In the current global economic climate there is an increasing demand for internationally comparable data on major economic aggregates such as gross domestic product (GDP), private and government consumption, and gross fixed capital formation. Over the last four decades, there has been a consensus that market exchange rates are not suitable for converting economic aggregate data from different countries expressed in respective national currency units.[1] Instead,

[1]For a detailed discussion of the issues relating to the use of exchange rates, the reader is referred to Kravis *et al.* (1982) as well as the ICP Handbook available on the World Bank website. In addition, the most recent publication from the Asian Development Bank on the 2005 comparisons in the Asia Pacific (Asian Development Bank, 2007) also provides an in-depth discussion on the use of exchange rates and purchasing power parities.

purchasing power parities (PPPs) of currencies which measure price level differences across countries are considered more appropriate for converting nominal aggregates into real terms.[2] PPP-converted real per capita incomes are used in influential publications like the World Development Indicators of the World Bank (World Bank, various years) and the Human Development Report (UNDP 2006) which publishes values of the Human Development Index (HDI) for all countries in the world. The PPPs are also used in a variety of areas, including: the study of global and regional inequality (Milanovic, 2002); measurement of regional and global poverty using international poverty lines like $1/day and $2/day (regularly published in the Word Development Indicators, World Bank); the study of catch-up and convergence in real incomes (Sala-i-Martin, 2002; Barro and Sala-i-Martin, 2004; Durlauf *et al.*, 2005) and issues surrounding carbon emissions and climate change (Castles and Hendersen, 2003; McKibbin and Stegman, 2005). Macroeconomic analyses and growth and convergence studies rely on panel data on PPPs with a good coverage of countries and period.

The principal source of information on PPPs for the whole economy has been the International Comparison Program (ICP). The data-intensive nature of the ICP and the need to undertake complex price surveys has meant that the ICP could be conducted roughly once every five years. The project began as a research project at the University of Pennsylvania which focused on 10 countries for the benchmark year 1970; the ICP concluded the 2005 round early in 2008,[3] with an impressive coverage and 146 participating countries.[4] However, the coverage for the intermediate benchmark years has not been that comprehensive. The last ICP round with a global coverage was the 1993/1996 round though the comparisons for this benchmark year can best be described as regional. Details of the country participation in different benchmarks are presented in the Appendix, Table A1.[5] Given the sparseness of the panel of PPPs available from the ICP, researchers rely heavily on PPPs and real income data available in the form of Penn World Tables (PWT) (PWT 6.2 are available at: http://pwt.econ.upenn.edu[6]) and from alternative sources such as the World Development Indicators (World Bank, various issues) and the Maddison series (GGDC, 2008).

The basic approach used in generating a panel of PPPs covering a large number of countries spanning a long period involves two major steps. In the first step, PPPs for the participating countries from a given benchmark year are extrapolated to yield PPPs for non-participating countries; and in the second step, these PPPs for the benchmark year are extrapolated using country-specific GDP deflators from national accounts data to produce the tables. The actual implementation of these

---

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

[3]See www.worldbank.org/data/icp for a copy of the final report of ICP 2005.

[4]Details of the history of the ICP and its coverage are well documented in the recent report of the Asian Development Bank (http://adb.org/Documents/Reports/ICP-Purchasing-Power-Expenditures/default.asp).

[5]Details presented in the Appendix are for only those countries included in our study.

[6]The popularity of the PWT is reflected in the vast number of citations made to Summers and Heston (1991). PWT version 6.3 was released after the work on the paper had been completed. Therefore, we use only PWT 6.2 in this paper.

steps in the construction of PWT 6.2 can be found in Heston *et al.* (2006).[7] Similarly, the currently available World Bank series on PPP-converted real per capital income are anchored solely on the latest benchmark year, 2005.

From this approach we observe that such extrapolations based on a single benchmark tend to ignore a wealth of information on PPPs contained in the remaining benchmarks. In fact each benchmark can be used in generating a separate panel of PPPs and these panels can differ significantly due to the inherent time–space inconsistency and the divergence between the implicit growth rates derived from the country-specific GDP deflators and the price movements implied in the PPPs compiled from different benchmarks. An obvious solution to this problem is to construct a single panel of PPPs derived as a weighted average of the different panels derived based on different benchmarks. This approach was described as the consis-tentization approach advocated in Summers and Heston (1991).[8] While the use of a weighted average is an attractive solution to the problem, as it makes use of information from all the benchmarks, the mechanism of selecting the weights accorded to extrapolations from different benchmarks is not clear.

In this paper we provide a formal framework for implementing this strategy whereby PPP data from several benchmarks can be used in conjunction with data on national price movements implicit in the national GDP deflators. The panel we seek to construct consists of countries that may or may not have participated in the benchmarks. In identifying the framework, we trace the essential elements involved in the process. The first element is the extrapolation of PPPs observed for countries participating in a given benchmark year to the non-participating countries. Next, we consider the process of updating PPPs in a given year using national level price deflators. The last element represents the optimal[9] predictors of unob-served PPPs for non-participating countries and non-benchmark years.

The econometric approach pursued in this paper combines all these elements and executes them in a single step. The approach proposed in the paper makes use of a state–space formulation which is designed to generate predictions of PPPs, along with their standard errors, over time and across countries that are broadly consistent with benchmark data on PPPs and observed country-specific price movements.

The main objective of this paper is to outline this approach and show[10] that the optimal predictions of unobserved PPPs are indeed "weighted" averages of PPPs obtained by extrapolating PPPs from benchmark years to non-benchmark years. The weights are endogenously determined, which eliminates the need to use guesswork in assigning weights to different benchmarks available in the dataset.

---

[7]The actual approach for PWT is more complicated as extrapolations are undertaken for the Consumption, Investment, and Government components separately and then combined using the Geary–Khamis aggregation method. The discussion is limited to one aggregate as the focus of the paper is on PPPs only at the GDP level. Other series which focus only on GDP include the World Bank's series published in the World Development Indicators (WDI, various issues) and the Maddison series (Maddison, 1995, 2007).

[8]A recent paper by Hill (2007) also revisits the problem of consistentization. However, this paper uses a different stochastic model and does not address the issue of extrapolation to non-benchmark countries.

[9]Optimal here is used in a statistical sense, implying that the predictions derived from the procedure possess minimum mean squared errors—a criterion commonly used in producing forecasts.

[10]We provide a rigorous analytical derivation of this result.

The paper is then mainly devoted to a description of the regression methodology used for the purpose of extrapolating *PPP*s to non-benchmark countries in a given year and the procedure used in generating optimal predictions of the unobserved *PPP*s which are used in filling the incomplete tableau of *PPP*s implied by the ICP benchmark data. The results presented here refer to a specific version of the more general model proposed by Rao, Rambaldi, and Doran (2008) (RRD). These results resemble closely to those approaches that have been discussed above.

The paper implements the new econometric methodology on a dataset that covers 141 countries over the period 1970 to 2005. There are 110 of the included countries that participated in the 2005 round of the ICP. A panel of PPPs is constructed for all countries in the sample. The sensitivity of the extrapolated PPPs is examined through a comparison of the panels with and without the 2005 benchmark data. The results show an encouraging degree of robustness to the inclusion of the 2005 data.

The paper is organized as follows. Section 2 briefly describes the econometric method proposed in the paper, and the basic assumptions underlying the stochastic specification are discussed in detail. Section 3 is devoted to a formal description of our approach and then presents, in Section 3.3, an analytical proof establishing that the predictions generated from our model are weighted averages of the tableaus generated through extrapolations from each of the benchmarks. We discuss the nature of the weights accorded to different benchmarks. Section 4 describes the data used in the empirical exercise. In this section we discuss the variables included in the basic regression model and highlight the specification of the spatial autocorrelation matrix used in modeling disturbances in the regression model. In Section 5 we present the optimal predictors of PPPs generated using our methodology and discuss the implications. Due to limitations on the length of the paper, we focus on a selected set of countries across the developed and developing world, but a complete set of results generated using our method are available from the authors upon request. The extrapolated PPPs are presented along with estimates from PWT 6.2. In Section 6 we offer a few concluding remarks.

2. The New Approach

The approach outlined in this paper is designed to take into account all the available information from all the benchmarks in the process of finding optimal predictors of the missing elements of the panel of PPPs. The steps and components used in the new approach are quite similar to those used in the PWT or by the World Bank but the novelty of the approach is in combining all these steps to produce predictions of PPPs. These steps are discussed below.

The main variable of interest is denoted by $p_{it} = \ln(PPP_{it})$ for country $i = 1, 2, \ldots, N$ and time $t = 1, 2, \ldots T$, where PPP_{it} represents the purchasing power parity of currency of country i, expressed in terms of the currency of a reference country.[11] In this paper the U.S. is used as the reference country for the purposes of illustration and the results presented here are invariant to the choice of

[11]Throughout this paper, we use italics for scalars and bold letters to represent vectors and matrices. Also note that PPP is used as an abbreviation for purchasing power parity of a currency.

reference country. This means that if the PPP data were expressed relative to another reference country, say the U.K., then the extrapolations generated would be identical to those derived using the U.S. as the reference currency after rebasing the results.[12]

2.1. *PPPs Compiled by the ICP for Various Benchmarks*

The first source of data used in our method is the ICP which provides PPPs for different benchmarks and for the countries participating in these ICP benchmarks. The main difference between our approach and the previous approaches is that we consider the ICP PPPs to be observations on *unknown* PPPs with a measurement error incorporated. For a given country i participating in the benchmark year t, we let $\tilde{p}_{it} = \ln\left(\widetilde{PPP_{it}}\right)$ be the logarithm of the ICP benchmark observation on the PPP for country i. Then, the ICP benchmark observations are related to the true, but unknown, PPPs through the following equation:

$$(1) \qquad \tilde{p}_{it} = p_{it} + \xi_{it}$$

where $p_{it} = \ln(PPP_{it})$ with PPP_{it} representing the true but unknown PPP for country i in period t; and ξ_{it} is a random error accounting for measurement error with $E(\xi_{it}) = 0$. Several implications of (1) are presented and discussed below.

While equation (1) simply means that the true PPPs are observed with error, it can allow considerable flexibility into our model through the magnitude of ξ_{it}. If we believe that PPPs within the ICP are measured without error and therefore the expectation is that any extrapolation method should track the ICP based PPPs (i.e. extrapolated PPPs should be exactly equal to the ICP PPPs) then it is possible to impose this in our model by making $Var(\xi_{it})$ equal to zero or for practical purposes choose a value close to zero.[13]

We now turn to the specification of the variance–covariance terms involving the disturbance term, ξ_{it} for different countries i for a given time period t. The basic assumption is that the PPP of the currency of each country is expressed relative to a numeraire (here we consider this as unspecified) and its variance (reciprocal of reliability) is directly related to per capital income.[14] Therefore, we postulate that

$$(2) \qquad Var\left(p_i^{(R)}\right) = \sigma_i^2 \text{ and } \sigma_i^2 = \sigma^2 \times \frac{1}{GDP_i}$$

---

[12]This basically means that the *PPP*s generated from our method will satisfy the same property of base invariance exhibited by the *PPP*s from the ICP. A mathematical proof of this property is provided in appendix 3 of RRD.

[13]The flexibility accorded by this specification allows us to compile extrapolations where the benchmark *PPP*s are preserved or whether the extrapolations can deviate from the benchmarks but preserve observed movements in relative price levels in different countries. This is an important property of the method proposed here.

[14]We use the nominal measure of per capita income obtained by converting per capita incomes using exchange rates. Expressing these in $U.S. using market exchange rates will accentuate the difference between developed and developing countries and thus provide us with a suitable measurement of the desired effect.

where $p_i^{(R)}$ represents the *PPP* of i in terms of a numeraire country, R; and GDP_i is the nominal gross domestic product per capita of country i derived using exchange rates.[15] Now we consider $\tilde{p}_{it}$ which represents the *PPP*s expressed using the U.S. as the numeraire country. Without loss of generality we denote the reference country U.S. as country 1. Then we have

$$\tilde{p}_{it} = p_{it}^{(R)} - p_{1t}^{(R)} \text{ and, therefore, } \tilde{p}_{1t} = 0^{[16]}.$$

Further, the variance structure for the *PPP*s expressed relative to country 1 as the numeraire is given below. By definition, variance attached to the *PPP* of the reference country is equal to zero. We have

(3) $Var(\tilde{p}_{1t}) = 0;$
$Var(\tilde{p}_{it}) = Var(p_{it}^{(R)} - p_{1t}^{(R)}) = Var(p_{it}^{(R)}) - Var(p_{1t}^{(R)}) = \sigma_i^2 + \sigma_1^2;$ and
$Covar(\tilde{p}_{it}\tilde{p}_{jt}) = \sigma_1^2.$

Basically this means that the standard error (square root of variance) associated with *PPP* for country i in period t incorporates the uncertainty associated with the *PPP* for country 1 which is used as the reference currency. This leads to a general covariance matrix of the form

$$Var(\tilde{\mathbf{p}}_t) = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_{1t}^2 + diag(\sigma_{2t}^2, \sigma_{3t}^2 \dots \sigma_{Mt}^2) \end{bmatrix}.$$

This formulation of the variance structure provides for a simple interpretation of the standard errors usually reported. It is important to note that the standard error associated with the PPP of the currency of a given country explicitly incorporates the uncertainty associated with the PPP of the reference country. This formulation of the covariance structure is also an essential element in ensuring that the methodology proposed here is invariant to the choice of the numeraire country (see RRD for technical details and a formal proof of this statement[17]).

2.2. *Extrapolation to Non-Benchmark Countries: A Regression Model of National Price Levels*

The second component of our methodology is the same as the one that is used in the construction of the Penn World Tables (Summers and Heston, 1991; Heston *et al.*, 2006) as well as the approach followed by the World Bank in the extrapolation of the benchmark PPPs for countries not participating in the benchmark. Our approach draws on the extensive literature on the theories that purport to explain the deviation of PPPs from the exchange rates. National price levels are defined as the ratios of PPPs to exchange rates:

[15]The use of exchange rates avoids possible endogeneity problems arising out of the use of PPPs in converting GDP in national currency units. Further, we suppress the time subscript for purposes of exposition, but use it in the actual empirical implementation of the method.

[16]This means that PPP for the reference country, the U.S., is equal to 1 and therefore ln(*PPP*) is equal to zero for the reference country.

[17]Due to the technical nature of the proof we have decided not to include it in this paper.

$$(4) \qquad R_{it} = \frac{PPP_{it}}{ER_{it}}$$

where *PPP* and *ER* respectively denote purchasing power parities and exchange rates for national currencies. For example, if the *PPP* and *ER* for Japan, with respect to one U.S. dollar, are 155 and 80 yen respectively, then the price level in Japan is 1.94, indicating that prices in Japan are roughly double those in the United States. A value of this ratio greater than one implies national price levels in excess of international levels and vice versa. A stylized fact on national price levels is that the ratio in (4) is typically well below unity for countries with low to middle levels of per capita income and typically around unity for richer countries. The recently released *PPP* estimates for the 2005 benchmark from the World Bank,[18] Asian Development Bank,[19] and others clearly support this phenomenon.

Specification of the regression models inevitably relies on the basic purchasing power parity theory.[20] The general theories of national price levels are anchored mainly on the share of tradable and non-tradables in the gross domestic product. While Kravis and Lipsey (1983, 1986) pioneered work on this, Clague (1988) and Bergstrand (1991, 1996) provide an analytical framework and theoretical models to establish and identify the main characteristics and the associated socio-economic variables that can potentially explain the observed national price levels. Some of the explanations also rely on productivity level differences between developed and developing countries which in turn offer support for the inclusion of education related variables. We draw heavily on Clague (1988) in the selection of variables for inclusion in our regression models. The variables recommended for use on the basis of these theories include: real income levels; resource endowments; foreign trade ratios; openness of the economy; education levels; dependence on agriculture and related variables. Ahmad (1996) makes use of a similar set of variables in the specification of a regression equation for extrapolation. His work includes black market exchange rate as a ratio to official exchange rate.

In this paper we essentially follow the same approach in extrapolating PPPs for non-participating countries in different benchmarks. The distinguishing feature of the approach here is that information from all the benchmarks is used in calibrating the regression model. Consequently, the regression model is specified as a panel data model. The model is mixed log-linear. Following the theories explaining national price levels, we postulate the model as:

$$(5) \qquad r_{it} = \beta_{0t} + \mathbf{x}'_{it}\boldsymbol{\beta}_s + u_{it}$$

where $r_{it} = \ln(PPP_{it}/ER_{it})$ is the logarithm of the national price level for country i in period t; $\mathbf{x}'_{it}$ a set of conditioning variables; β_{0t} intercept parameter; $\boldsymbol{\beta}_s$ a vector of slope parameters; and u_{it} a random disturbance with specific distributional characteristics. In this model, the slope parameters are assumed to be invariant over time and space. However, the intercept term is time varying, suggesting that the

[18]http://www.worldbank.org/data/icp and follow the links to the final report on the global comparisons.
[19]http://adb.org/Documents/Reports/ICP-Purchasing-Power-Expenditures/default.asp
[20]Officer (1982) provides an excellent review of literature on this topic.

temporal fixed effects are explicitly modeled whereas the country-specific fixed effects are implicitly modeled through the inclusion of several dummy variables which represent various trade and currency groupings to which the countries belong. We discuss several specific features of the variables included in the model and the distributional assumptions made regarding the disturbance term.

Dependent Variable

In this study we cover all the benchmark data from 1975 and include ICP PPPs for the years 1975, 1980, 1985, 1990, 1993, 1996, 1999, 2002, and 2005. Of all these benchmarks, the 2005 benchmark has the highest coverage, with 146 participating countries; 1975 has only 13 countries. Several benchmarks, including 1990, 1993, 1999 and 2002, have a coverage limited to only OECD and EU member countries.

Explanatory or Regressor Variables

We have two types of dummy variables and several continuous regressor variables included in the model on the basis of the theory of national price levels discussed above.

The first type of dummy variables is time dependent and each of these dummy variables assumes a different intercept term for each benchmark year. Thus inclusion of these variables is similar to modeling time specific fixed effects, taking into account the fact that benchmark comparisons are available only for selected years.

In the place of dummy variables that are traditionally representing country-specific fixed effects, we make use of dummy variables to account for the membership of countries in different currency unions or trade agreements. For example, we include D-euro to represent a dummy variable for countries which have used the euro since 1999 and D_scucar—a dummy for strict currency union in Caribbean. We also include a dummy variable, D_usd, for countries with currencies either pegged to the U.S. dollar for substantial amounts of time or expected to move closer to the U.S. dollar. As the dependent variable is the ratio of PPP to the exchange rate, ER, it is important to model effects on exchange rates.

Inclusion of the time and country-group specific dummy variables is necessary within the context of a panel approach to the model where data from all the benchmarks are used.[21] If we were to estimate the model using a single benchmark, then several of these dummy variables could be dropped from the model.

The third set of explanatory variables included may be considered essentially as the drivers of the national price level model. However, there are some innovations introduced. Instead of using a real per capita income as in Clague (1988), Atlas income as in Ahmad (1996), or real income as the World Bank did in its recent extrapolation work, we use proxies for the level of development, thus reducing the possibility of endogeneity in the regressor variables. We use *Life*, life

[21]The inclusion of time dummy shifts for benchmark years is a sufficient condition for invariance of the method to the reference country (see appendix 2 of RRD).

expectancy at birth; *agedep*, age dependency ratio;[22] *Radpccn*, radios per capita; and *Phones*, number of telephone mainlines per 1000 people. We make use of *Expg*, exports of goods and service as a percentage of GDP; *Manufimp and Manufexp*, manufacturing exports and imports as a percentage of GDP, respectively, of exports and imports; and *Tradeg*, trade as a percentage of GDP as variables that represent openness of the economy and the share of manufacturing in exports and imports. Finally, we include variables that reflect the traditional and low productivity sector in the economy by a group of variables including *Advag*, value added of agriculture as a percentage of GDP; *Rurpop*, share of rural population in total; and *Tractorpw*, agricultural machinery measured using tractors per agricultural worker. Level of education is measured using *Literate*, a measure of literacy rates defined as population aged 15 and over which is literate per 1000 population, and *Secenr*, the number of secondary enrollments per 1000 population. The variables used in our study are similar to those conventionally used. For example, Ahmad (1996) uses OPEN for openness; AGR for agricultural value added as a proportion of GDP; and ENROL, secondary school enrollment ratio. A similar correspondence can be made with some of the variables used by Clague (1988). It is possible to speculate on the expected signs of the variables included in the model.[23] We note that it is possible to obtain signs contrary to expectations due to the presence of possible multicollinearity in the regressor variables. As the main function of the model in (5) is for the prediction of the national price level for non-benchmark countries, the overall fit of the model is more important than the signs of the individual coefficients.[24]

Predictions Using the Model

If estimates of β_{0t} and $\boldsymbol{\beta}_s$ are available, model (5) can provide a prediction of the variable of interest, $\ln(PPP_{it})$, consistent with the price level theory:

$$(6) \qquad \hat{p}_{it} = \hat{\beta}_{0t} + \mathbf{x}'_{it}\hat{\boldsymbol{\beta}}_s + \ln(ER_{it}).$$

Thus, (6) states that price level theory provides a prediction, $\hat{p}_{it}$, and hence of $\ln(PPP_{it})$.

We note here that it is possible to obtain predictions of PPPs for all countries and all time periods using (6) provided we have observations on the regressor variables. *However, in this paper we make use of the regression model only to generate predictions for non-benchmark countries in the benchmark years. No predictions are generated using (6) for non-benchmark years.*[25]

At this stage, we are in a position to produce PPPs for all the countries for the years when the ICP benchmarks are conducted. The next step is to extrapolate the PPPs to the non-benchmark years and thereby complete the panel of PPPs.

[22]See Appendix Table A2 for a detailed description of the variables included in the model. We also indicate the source from which data are drawn for the empirical part of the paper.
[23]We refer the reader to Clague (1988) for an excellent discussion on the expected signs based on theory. These are not repeated here.
[24]We revert to this when we present the results of the estimated model.
[25]A more general approach that makes use of predictions for all the countries in all years (both benchmark and non-benchmark years) is described in RRD.

2.3. *Extrapolation to Non-Benchmark Years: Use of National Growth Rates and Implicit GDP Deflators*

A major consideration in the construction of a panel of PPPs, and real incomes, is that the growth rates in real income obtained using PPPs should be the same or close to the national growth rates observed and reported by the respective national statistical offices. This is considered an important property to be satisfied by the extrapolated PPPs. The currently available series from PWT, the World Bank's World Development Indicators, and the Maddison series all adhere to this important principle. Thus, the national growth rates and the price movements implicit in such growth rates which are referred to as GDP deflators offer crucial information for updating PPPs.

The temporal movements in PPPs are, therefore, governed by a simple relationship presented in equation (7). It is easy to see that equation (7) is a simple identity if PPPs were the price of a single commodity. However, in the case of PPPs at the GDP level, the GDP is treated as a composite commodity.

$$(7) \qquad PPP_{i,t} = PPP_{i,t-1} \times \frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}}$$

where $GDPDef_{i,[t-1,t]}$ denotes the GDP deflator showing price movements from period $t-1$ to t in country i.[26] Equation (7) simply provides a mechanism for updating PPPs using movements in the GDP deflator of the country concerned, and therefore provides a definition for the growth rate of PPP_{it}.

The main source of data on deflators is the national accounts published by countries, generally on an annual basis. In the econometric specification of our model, we include a random disturbance term that makes it possible for the extrapolated PPPs to deviate from the PPPs implied by the growth rates in GDP deflators.

Taking the logarithm of (7) and accounting for the measurement error:

$$(8) \qquad p_{it} = p_{i,t-1} + c_{it} + \eta_{it}$$

where,

$c_{it} = \ln\left(\frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}} \right)$; and η_{it} is a random error accounting for measurement error in the growth rates. We assume that η_{it} is a random variable with mean equal to zero and, further, that errors in the updating equation, η_{it}, are uncorrelated with errors in the PPP equation with $E(\eta_{it}\xi_{it}) = 0$.

In a vector form where PPPs for all the countries are stacked in a column, equation (8) can be written as:

---

[26]The U.S. GDP deflator is used in the updating equation reflecting the choice of the U.S. as the reference country. However, the method described here is invariant to the choice of the numeraire country and the relative PPPs will remain the same even if some other country, say India, is used as the reference country.

$$(9) \qquad \mathbf{p}_t = \mathbf{p}_{t-1} + \mathbf{c}_t + \boldsymbol{\eta}_t$$

where $\mathbf{p}_t$ is the unobserved state vector and $\mathbf{c}_t$ is the observed growth rate of $\mathbf{p}_t$.

Now we turn to the stochastic specification of the error term in equation (9). The error, η_{it}, is assumed to have a mean equal to zero and a covariance matrix $E(\eta_t \eta_t') \equiv \mathbf{Q}_t$ where elements of $\mathbf{Q}_t$ are defined on the basis that the GDP deflators have variance inversely proportional to the level of development of the country as measured by per capita GDP. If country 1 is selected as the reference country, then the $\mathbf{Q}$ matrix, denoted by $\mathbf{Q}_t^{(1)}$ is defined as $\mathbf{Q}_t^{(1)} = \sigma_\eta^2 \mathbf{V}_t$ where,

$$\mathbf{V}_t = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_{1t}^2 \mathbf{j}\mathbf{j}' + diag\left(\sigma_{2t}^2, \ldots, \sigma_{Nt}^2\right) \end{bmatrix},$$

σ_{it}^2 is the variance of the i-th country price deflator's growth rates, and $\mathbf{j}$ is a column of ones.

A few remarks on equation (9) and the flexibility it offers in the treatment of national growth rates and implicit price movements is useful here. First, we note that the approach used in the PWT, the World Bank, and the Maddison series is a special case of (9) where η_{it} is set to zero or its variance set close to zero (mean is equal to zero by assumption). In this case the temporal movements in PPPs are identical to the relative movements in national GDP deflators, and hence the implied growth rates in income are also maintained. Second, we note that it is possible that the national deflators are also subject to measurement errors and growth rates are subject to errors. It is possible to build such uncertainty associated with the growth rates into the extrapolation methodology presented here. We assume that the reliability of the growth rates is again inversely proportional to the per capita income of the country. Again the model is flexible and if we decide to impose the growth rates exactly then we can set the variance of η_{it} close to zero. Equation (9) is then satisfied as an exact equality. In the absence of any preference one may leave the magnitudes of the variances to be determined by the data.

2.4. *Exact Restriction from the Choice of the Reference Country*

The definition of PPP requires a choice of reference country. For the reference country, PPP is defined to have a value of one for all time periods.[27] Thus, we know the value of the variable of interest for the reference country for all time periods. As the U.S. is taken as the reference country and denoted as the first country in the list, it then follows that for all t

$$(10) \qquad p_{USA,t} \equiv p_{1t} = 0.$$

We ensure that this restriction is imposed in our model for all the time periods. In addition, in RRD it was formally shown that the empirical results emanating from

---

[27]PPPs between currencies of two countries are invariant to the choice of the base country. In the current study, we use the U.S. dollar as the reference currency which, in turn, gives equation (7). The method used in this paper is invariant to the choice of the reference currency.

an alternative choice of the reference country would be identical to those (in relative terms) obtained when the U.S. is used as the reference country.

2.5. *Spatially Autocorrelated Error Structure*

A major innovation and a departure from the standard practice in the extrapolation of *PPP*s to non-participating countries in the benchmark year is the use of spatially autocorrelated structures in the model that describes the national price levels. Recalling the model described in equation (5) in Section 2.2, we have

$$r_{it} = \beta_{0t} + \mathbf{x}'_{it}\boldsymbol{\beta}_s + u_{it}.$$

In the models currently used in the construction of PWT and the World Bank's extrapolations, the disturbance terms for different countries are assumed to have the same variance for all the countries and are assumed to be uncorrelated across countries. In the approach described here, we relax both of these. As discussed earlier, the variance attached to a particular country which reflects the reliability of its *PPP* from the benchmark is inversely proportional to the per capita income.

We assume that the disturbance terms in (5) are spatially correlated across countries. In this regard it is useful to rewrite the disturbance term in (5) as:

$$u_{it} = r_{it} - \beta_{0t} - \mathbf{x}'_{it}\boldsymbol{\beta}_s.$$

Essentially the disturbance term represents the magnitude by which the observed price level for a country differs from the value expected from the model. If the disturbance is positive, the observed price level exceeds the expected value and vice versa. For a given period, we can write the regression model as:

$$\mathbf{r}_t = \beta_{0t} \times \mathbf{j} + \mathbf{x}'_t\boldsymbol{\beta}_s + \mathbf{u}_t$$

where $\mathbf{u}_t$ is a $N \times 1$ vector of disturbances.

We assume that the elements of the disturbance vector $\mathbf{u}_t$ are spatially auto-correlated with the following structure:

$$(11) \qquad\qquad \mathbf{u}_t = \phi\mathbf{W}_t + \mathbf{e}_t$$

with $|\phi| < 1$[28] and $\mathbf{e}_t$ is vector of disturbance with a mean equal to $\mathbf{0}$, and they are uncorrelated across countries/observations. Following Anselin (1988) and standard spatial correlation literature, the matrix $\mathbf{W}_t$ is a spatial weights matrix that drives the underlying spatial structure of the disturbances. That is, the rows of $\mathbf{W}_t$ add up to one and the diagonal elements are zero. It follows that the covariance matrix of the disturbances in (5), for a given t, is given by:

$$(12) \qquad\qquad \mathrm{E}(\mathbf{u}_t\mathbf{u}'_t) = \sigma^2_u(\mathbf{I} - \phi\mathbf{W}_t)^{-1}(\mathbf{I} - \phi\mathbf{W}_t)^{-1\prime} = \sigma^2_u\boldsymbol{\Omega}_t.$$

If the matrix $\mathbf{W}_t$ is known, then one can use standard approaches like the generalized least squares or maximum likelihood methods to estimate the param-

[28]In most economic applications ϕ is likely to be positive.

eters of the model. We need to specify the elements of the $\mathbf{W}_t$ matrix. In this paper we make use of a simple specification which assumes that the disturbances of two countries tend to behave similarly (for example, the national price levels above the expected value) if the two countries have strong trade relationships.[29] The model for national price levels in (5) is anchored on the purchasing power parity theory, which implies that price levels get transmitted through trade and ultimately equalized if there is free trade between all the countries. In the actual implementation, which is further discussed in Section 4 dealing with data construction, the weights in the $\mathbf{W}_t$ matrix are proportional to bilateral trade flows.

This completes the description of the basic model and its salient features. Now we turn to the overall econometric approach used in constructing the panel of PPPs.

3. A State–Space Approach to the Extrapolation Problem

The basic approach followed in this paper is to make use of the various components of the extrapolation work described in equations (1) to (12) in generating predicted purchasing power parities to complete the required panel. Given the nature of the problem where we have information on how PPPs are generated through the price level equation and the ICP benchmarks and a model to extend it over time using relative movements in price levels, a state–space (SS) representation is highly suitable for this type of problem.

There are three elements necessary to provide a state–space representation. First, we need to identify a state vector for which we wish to obtain optimal predictions over time. Second, we need a model that describes the observations on the state vector through the *observation equation*. In our case the national price level model in (5) and the ICP benchmark information equation (1) together provide the observation equation. Finally, we need a *transition equation* which provides a mechanism that describes how the state vector moves from one period to the next. Our equation (9), which explains how PPPs are related to national price movements, provides the basis for the transition equation.

3.1. *State–Space Representation of the Model*

State Vector

The state vector in our model is the vector $\mathbf{p}_t$, the vector of logarithms of PPPs of currencies of different countries. *Our aim is to obtain optimal predictors of $\mathbf{p}_t$ for all the time periods in the analysis.*

Observation Equations

As previously discussed, there are two noisy sources of observations of the state vector, $\mathbf{p}_t$. First we have measurements of the state vector in the form of $\tilde{\mathbf{p}}_t$

---

[29]A common assumption is to use the assumption of geographic contiguity, which translates to the assumption that the price levels around their mean tend to behave similarly for countries which are geographically contiguous. However, we find little theoretical support for such a model. In a more recent paper, Rambaldi *et al*. (2010) construct a measure of economic distance which includes geographical, trade, and other historically relevant indicators.

which are based on PPPs from the ICP. This information is described in equation (1). The second set of measurements, $\hat{\mathbf{p}}_t$, come from predictions for the regression model for those countries that have not participated in the ICP benchmarks and *this information is restricted to only benchmark years*. We note here that this is a special case of the more general framework described in RRD where regression predictions are used in benchmark as well as non-benchmark years. The main reason for restricting the use of predictions to only benchmark years is to relate this procedure to the intuitive notion that ultimately the panel of predicted PPPs must be a weighted average of available PPPs in benchmark years, that is, observed PPPs from the ICP benchmarks for participating countries and regression predictions for non-participating countries.

In writing the observation equations, we wish to relate the conditioning variables to the state variable of interest, $\mathbf{p}_t$, by formulating the problem in state space form. The equations (1) and (6) are necessary to obtain the "observation equations" of the state–space in equation (13):

$$(13) \qquad \mathbf{y}_t = \mathbf{Z}_t \mathbf{p}_t + \mathbf{B}_t \mathbf{X}_t \boldsymbol{\theta} + \boldsymbol{\zeta}_t$$

for only those years t which represent ICP benchmark years.[30] In equation (13), we use the following notation:

$\mathbf{y}_t = [0 \quad \hat{\mathbf{p}}_t \quad \tilde{\mathbf{p}}_t]'$ is a vector of "observations" of the state vector, i.e. reference country value,[31] regression predictions, and benchmark observations.

$\mathbf{Z}_t$ is a known permutation matrix.

$\mathbf{B}_t$ is a selection matrix that maps $\mathbf{X}_t \boldsymbol{\theta}$ to the *PPP* regression predictions for non-participating countries.

$\mathbf{X}_t$ is a matrix of observable socio-economic variables.

$\mathbf{p}_t$ is the unobserved state vector.

$\boldsymbol{\theta}$ is a vector of parameters to be estimated and is a known form of the vector $\hat{\boldsymbol{\beta}}$ in (5).

$\boldsymbol{\zeta}_t$ is an error with $E(\boldsymbol{\zeta}_t) = 0$ and $E(\boldsymbol{\zeta}_t \boldsymbol{\zeta}_t') \equiv \mathbf{H}_t$.

$\mathbf{H}_t = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 \mathbf{S}_{np} \boldsymbol{\Omega}_t \mathbf{S}_{np}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\xi^2 \mathbf{S}_p \mathbf{V}_t \mathbf{S}_p' \end{bmatrix}$ is partition to account for the variance of the

reference country, and the variance–covariance structure associated with the $\hat{\mathbf{p}}_t$ and $\tilde{\mathbf{p}}_t$ respectively.

S_{np} and S_p are selection matrices for non-participating and participating countries, respectively.

Though equation (13) looks complex, it has a fairly simple interpretation. The first point to note is that these observation equations are defined and used only for ICP benchmark years. Second, the selection matrix $\mathbf{Z}_t$ is a permutation matrix that separates countries into participating and non-participating countries in a benchmark year. The covariance matrix associated with the errors in benchmark PPPs is given by the last diagonal partition in the $\mathbf{H}_t$ matrix. Finally, for those countries

[30]Details of the derivation of (13) are provided in RRD.
[31]Without loss of generality the reference country is the first.

that have not participated in the benchmark, the regression prediction, $\hat{\mathbf{p}}_t$ is derived using the regression model in (6), making use of the estimated β coefficients. The selection matrix $\mathbf{B}_t$ selects all those non-participating countries, and the subscript t indicates that the composition of the non-participating countries is different for each benchmark year.

Transition Equations

The transition equations show the evolution of the state vector over time. In the construction of PPPs, equation (7) defines the temporal movements as explained in Section 2.3. Taking natural logarithms and accounting for measurement error in the growth rates, the temporal evolution is given by equation (9).

Thus, equations (13) and (9) define the state–space model that produces the complete panel of PPPs, given estimates of the unknown parameters.

3.2. *Estimation and Smoothing*

Given estimates of equations (13) and (9), and the distribution of the initial value of the state vector, $\mathbf{p}_0$, and under the assumption of normality of the disturbances and the initial state vector, the Kalman filter computes the conditional mean (based on the information available at time t), $\breve{\mathbf{p}}_t$, and corresponding covariance matrix, $\mathbf{\Psi}_t$, of the distribution of $\mathbf{p}_t$.

We have two types of predictions that can be constructed using the state–space approach and Kalman filters. The first set, referred to as the filtered predictions, are derived making use of the information that is available up to a given point in time. These predictions are given by the *updating equations* given below. Once the Kalman filter is run through all time periods, an optimal prediction for period T, $\breve{\mathbf{p}}_T$, is derived. Given, $\breve{\mathbf{p}}_T$, $\mathbf{\Psi}_t$ and $\mathbf{\Psi}_{t|t-1}$, $t = 1, \ldots, T$, the Kalman smoother provides predictions $\mathbf{p}^*_{t|T}$, $t = 1, \ldots, T$ of the state vector based on all the information available. For each period the Kalman smoother for period t is denoted by $\mathbf{p}^*_{t|T}$ as given below.

Kalman Filter Equations

The forward filter is conceptually composed of two sets of equations as follows:

Prediction Equations

$$(14) \qquad \mathbf{p}_{t|t-1} = \mathbf{p}_{t-1} + \mathbf{c}_t$$
$$\mathbf{\Psi}_{t|t-1} = \mathbf{\Psi}_{t-1} + \hat{\mathbf{Q}}_t.$$

Updating Equations

$$(15) \qquad \mathbf{F}_t = \mathbf{Z}_t \mathbf{\Psi}_{t|t-1} \mathbf{Z}'_t + \hat{\mathbf{H}}_t$$
$$\mathbf{p}_t = \mathbf{p}_{t|t-1} + \mathbf{\Psi}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \left(\mathbf{y}_t - \mathbf{B}_t \mathbf{X}_t \hat{\boldsymbol{\theta}} - \mathbf{Z}'_t \mathbf{p}_{t|t-1} \right) = \mathbf{p}_{t|t-1} + \mathbf{\Psi}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \hat{\boldsymbol{\zeta}}_t$$
$$\mathbf{\Psi}_t = \mathbf{\Psi}_{t|t-1} - \mathbf{\Psi}_{t|t-1} \mathbf{Z}'_t \mathbf{F}_t^{-1} \mathbf{Z}_t \mathbf{\Psi}_{t|t-1}.$$

Kalman Smoother Equations

The backward filter is given by the following set of equations:

(16)
$$\mathbf{p}^{*}_{t|T} = \breve{\mathbf{p}}_{t} + \hat{\boldsymbol{\Psi}}_{t}(\mathbf{p}^{*}_{t+1|T} - \mathbf{c}_{t+1} - \breve{\mathbf{p}}_{t})$$

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

$$\hat{\boldsymbol{\Psi}}_{t} = \boldsymbol{\Psi}_{t}\boldsymbol{\Psi}^{-1}_{t+1|t}$$

where,

$\breve{\mathbf{p}}_{t|t-1}$ is the Kalman filter estimate conditional on the information up to $t-1$.

$\breve{\mathbf{p}}_{t}$ is the Kalman filter estimate of the state vector at time t.

$\boldsymbol{\Psi}_{t}$ is the Kalman filter unconditional covariance of the state vector.

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

$\mathbf{p}^{*}_{t|T}$ is the Kalman smoothed estimate of the state vector.

$\boldsymbol{\Psi}^{*}_{t|T}$ is the Kalman smoothed estimate of the covariance of the state vector.

Estimation of Unknown Parameters

There are unknown parameters in the covariance structure, $\mathbf{Q}_{t}$, $\mathbf{H}_{t}$ as well as the vector of parameters in the mean of the observation equation (associated with the regression part), $\boldsymbol{\theta} = (\hat{\boldsymbol{\beta}}_{I} - \boldsymbol{\beta})$, where $\hat{\boldsymbol{\beta}}_{I}$ is an estimate of the vector of parameters in equation (5) obtained by using least squares.

To obtain the estimates of the unknown parameters in the state–space model we use an iterative algorithm that estimates the parameters by maximum likelihood and updates $\hat{\mathbf{p}}_{t}$ in (13) at each iteration. The log likelihood function for the unknown parameters is given by:

$$\log L\left(\phi, \sigma^{2}_{\eta}, \sigma^{2}_{u}, \sigma^{2}_{\xi}, \boldsymbol{\theta}\right) = -\frac{TN}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{F}^{-1}_{t}| - \frac{1}{2}\sum_{t=1}^{T}\hat{\boldsymbol{\zeta}}'\mathbf{F}^{-1}_{t}\hat{\boldsymbol{\zeta}}_{t}.$$

To obtain a value of the log likelihood, we numerically maximize this function over σ^{2}_{η}, σ^{2}_{u}, σ^{2}_{ξ}, and ϕ and obtain an estimate, $\hat{\boldsymbol{\theta}}(\phi, \sigma^{2}_{\eta}, \sigma^{2}_{u}, \sigma^{2}_{\zeta})$, by regressing a set of "innovations" $\mathbf{y}^{**}_{t}$ on the "innovations" $\mathbf{X}^{**}_{t}$, where these "innovations" are obtained by running the Kalman filter separately for $\mathbf{y}_{t}$ and each column of $\mathbf{X}_{t}$ (the procedure follows Harvey, 1990, pp. 130–3). For the interested reader, the completed algorithm is presented in detail in RRD.

We note that under normality of the disturbances, the conditional distribution of the observation vector $\mathbf{y}_{t}$ is given directly by the Kalman filter (we refer the reader to Harvey (1990) for details). By writing the log likelihood in prediction error decomposition form, a pass through the Kalman filter allows the computation of a value of the likelihood function. Given the parameter estimates, the sample is then run through the equations of the Kalman filter and Kalman smoother to obtain the model's predicted p_{it} (for all i and t), p^{*}_{it}, and associated standard errors.

In our study, we propose that the Kalman smoother, p^{*}_{it}, is the preferred predictor for filling the time–space tableau of PPPs. The Kalman filter prediction provides a prediction of $\ln(PPP_{it}) = p_{it}$ which makes use of all the information

available up to and including information available in period t. However, the Kalman smoother provides a predicted value of p_{it} based on the available information for all the time periods $t = 1,2. \ldots , T$. This is one of the goals set for the study, i.e., to derive predictions for PPPs based on all the available information from all the benchmarks and relative price movements data available from national sources.

Since the state vector $\mathbf{p}_t$ is in fact the vector of ln (PPP_{it}) and our interest is in PPPs, we make the following transformation to derive the predicted PPPs.

A prediction of PPP_{it} is given by:

(17) $$P\hat{P}P_{it} = e^{p_{it}^*}$$

where p_{it}^* is the corresponding Kalman smoothed element.[32]

The standard errors for the predicted PPPs are computed as follows:[33]

(18) $$se\left(P\hat{P}P_{it}\right) = \sqrt{e^{2p_{it}^*}e^{\psi_{it,t}^*}\left(e^{\psi_{it,t}^*}-1\right)}$$

where
$\psi_{ii,t}^*$ is the i-th diagonal element of the estimated smoothed covariance of the state vector, $\mathbf{\Psi}_t^*$.

We note here that the standard error associated with a predicted PPP reflects the uncertainty associated with the PPP of the currency of a given country when the United States (country 1) is used as the reference country. Thus the standard error includes the standard error associated with the PPP of the U.S. dollar. Therefore, the standard errors should be carefully interpreted.

3.3. *Main Properties of the Proposed Smoothing Method*

The econometric approach described is a versatile smoothing approach with a number of useful properties. First, the predictions of PPPs generated for any country at a given point of time are based on all the information on all the benchmarks, GDP deflators, and data on socio-economic variables available for the purpose of extrapolation. Further, the predictions obtained are optimal in a statistical sense (this is a property of the Kalman filter and smoother). Second, the PPPs generated are by construction equal to 1 for the reference country. The predicted PPPs that form the final panel are invariant to the choice of the reference country. For example, if Australia is used instead of the U.S. as the reference, the relative PPPs (PPP of a currency relative to another currency) will remain unchanged. Third, the Kalman filtered predictions can be smoothed to produce

[32]It may be noted here that the estimator of PPP is biased as the exponential function is a non-linear transformation. It is possible to make a correction for the bias in the case where p_{it}^* is normally distributed. However, no such correction is applied in this paper which is consistent with the current practice of the World Bank and the PWT. In our method the "σ^2" in the correction formula is time varying in general. When the smoother that preserves growth rates is used, the value converges to a constant. In the later case the bias correction factor for China is of the order of 1.016. In general, we do not expect these factors to make a significant difference in practice for the majority of the countries.
[33]The standard errors are computed under the assumption of the lognormality of the predictions.

predictions p_{it}^* that preserve the national price movements available in the form of GDP deflators. In this case the estimation of the parameters and the forward filter, the Kalman filter, account for all sources of error; however, the backward filter, the Kalman smoother, is restricted to preserving national price movements.[34] Fourth, if the ICP benchmarks are required to be preserved, the estimation, filtering, and smoothing can be constrained to produce predictions that are identical to the ICP observed value, $\tilde{p}_{it}$, in benchmark years. PPPs for currencies of the ICP participating countries are determined using price data collected from extensive price surveys, thus one may consider it necessary that the extrapolated PPPs track these benchmark PPPs accurately. This is achieved by setting to zero the variance of ξ_{it} in equation (1).

However, we focus on the main property of our approach. The model we use in this paper represents a simple case of the RRD smoothing method, where benchmark PPPs and extrapolations to non-participating countries are used only in ICP benchmark years, but no regression information is introduced for years in between benchmark years. The panel of PPP estimates produced is a "weighted average" of the extrapolations from different benchmarks. Furthermore, the weights are not arbitrary, but derived from the covariance properties of the model resulting in optimal predictions for PPPs. We present an illustration of the results obtained using this form of the algorithm in Section 5. An algebraic proof of the statement is provided in the Appendix. The special case discussed at the end of the Appendix shows that the extrapolations derived under the absence of spatial autocorrelation using the Kalman filter are weighted averages of extrapolations from different benchmarks with weights proportional to the variances attached to the country-specific PPPs in the benchmarks.

3.4. *Computational Aspects*

As expected, developing the algorithms and writing program code for the estimation of parameters of the model and for generating the optimal predictions using the Kalman filter and smoother has been a challenge. The program code is developed in GAUSS and runs into 9000 lines.[35] The code has been thoroughly tested before using it to generate predictions of PPPs and their standard errors. As the code is specifically written for the purpose of extrapolation of PPPs, the code is not made available in the public domain.[36] However, a complete set of results and predictions are available from the authors upon request.

4. Data Compilation and Data Construction

This section describes the dataset used in this study. The dataset covers 141 countries over the years 1970 to 2005. Table A1 lists the 141 countries included in

---

[34]When the restricted smoother is used, the standard errors converge to a constant value for all time periods (see RRD, appendix 1 for the form of the smoother equations and derivation of the standard error).

[35]The code was largely developed and tested by Alicia Rambaldi though Howard Doran had contributed to the development in the initial stages of the process.

[36]The code is not a generic code that could be used by different users. It is not like a code for the computation of inverse of a matrix which can then be used by any one interested in inverting a matrix. Our code is problem specific and therefore considered confidential for commercial purposes.

the study. This table also lists the currency of each country and the years each country has participated in the ICP benchmark comparisons. The empirical analysis in this paper includes PPP data from the ICP 2005 round. Given that the 2005 PPPs were only recently released, we have been able to include only 110 of the 146 countries that participated in the round. We will be adding the remaining countries pending availability of the required socio-economic variables. Table A2 gives definitions and sources of the variables used in the study, while Table A3 provides some basic descriptive statistics of the variables.

4.1. *PPP Data*

The state variable in the state space model is $\ln(PPP_{it})$, and observed values (which define the dependent variable in the measurement equation) are obtained from all the benchmarks conducted so far. Thus ICP PPP data are drawn from the early benchmarks of 1975, 1980, and 1985 as well as from more recent benchmark information for the years 1990, 1993, 1996, 1999, 2002, and 2005. Several features of the PPP data are noteworthy. The first benchmark covered 13 countries. The 1980, 1985, and the recent 2005 benchmarks represent truly global comparisons with PPPs computed using data for all the participating countries. For the years from 1990 to 2002, data are essentially from the OECD and EU comparisons, with the exception of 1996.[37] The 1996 benchmark year again is a global comparison with PPPs for countries from all the regions of the world. However, the 1996 benchmark may be considered weaker than the 1980, 1985, and 2005 benchmark comparisons as no systematic linking of regional PPPs was undertaken. In terms of reliability, one would consider the 1996 benchmark PPPs to be less reliable. Another related point of interest is the fact that PPPs for all the benchmarks prior to 1990 were based on the Geary–Khamis method (Geary, 1958; Khamis, 1972), and PPPs for the more recent years are all based on the EKS method (Elteto and Koves, 1964; Szulc, 1964) of aggregation.[38] In the current empirical analysis, we have not made any adjustments to the PPP data, but making the series comparable through the use of the same aggregation methodology is a part of our ongoing research program.

4.2. *Socio-Economic Variables Included in the Regression*

We have discussed in Section 2.2 the rationale for inclusion of a set of socio-economic variables used in modeling the national price levels. In particular, the reader will note that the regression model here concerns a panel rather than a cross-section for a given benchmark year. Therefore a range of dummy variables to model the time and country-specific fixed effects are also included. In Table A2 we provide a description of the socio-economic variables included in the regressions.

---

[37]We are indebted to Ms Francette Koechlin (OECD) for providing ICP benchmark data for these years. PPPs for those countries which joined in the Euro zone, the pre-Euro domestic currencies were converted using the *1999 Irrevocable Conversion Rates* (source: European Central Bank, 1998). The irrevocable conversion rate of the drachma vis-à-vis the euro was set at GRD 340.750 (source: http://www.bankofgreece.gr/en/euro).

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

During the stage of data compilation, a variety of sources have been used. For example, if data were not available from standard sources like the World Development Indicators, then use is made of the CIA Factbook (CIA, 2008). When this approach failed, a simple imputation procedure using values from a similar country were used. When there are missing entries in time series data for a variable in a country, simple statistical extrapolations were used.[39]

4.3. *Covariance Variables*

Measuring Spatial Correlation

The spatial weights matrix, $\mathbf{W}_t$, used in modeling spatial autocorrelation is derived using the volume of bilateral trade flows. Each row in $\mathbf{W}_t$ is a country, and the entry corresponding to each column, *j*, is the proportion of country *i*'s total trade across the 141 countries corresponding to country *j*. Thus, the rows of $\mathbf{W}_t$ add up to one, the diagonal elements are zero, and off diagonal element are trade weights. These spatial weights matrices were constructed for the years 1970, 1975, 1980, 1985, 1990, 1995, 2000, and 2005. Thus, we assume that the relative weights of partner countries do not change within each of the five year periods. These matrices are compiled using data from Rose and IMF Trade Directions.

Accuracy of Benchmarks and National Accounts' Growth Rates

As discussed in Section 2, we postulate that the quality of the PPPs as well as estimates of implicit deflators are directly related to the per capita income, indicating that poorer countries will have large variances (and, therefore, lower quality) associated with the PPPs and deflators. The model specification allows for the modeling of accuracy of benchmark PPPs and national growth rates (equations (1) and (8)). We assume that the measurement errors in both cases have variances that are inversely proportional to the per capita GDP expressed in U.S. dollars. This means that countries with higher per capita incomes are expected to have more reliable data, as reflected by lower variances associated with them.[40]

5. EMPIRICAL RESULTS

The empirical implementation of the econometric methodology proposed in the paper has generated a lot of output, including detailed tables of PPPs for each country included in the analysis. Keeping the length of the paper as a guide, we had to be selective in what is presented in this section.[41] At the outset we present the estimates of the parameters of the models used in extrapolating PPP, that is the regression parameters from the model of national price levels as well as the

[39]Information on country and variable specific details of the imputation methods used is available from the authors.

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

[41]Detailed results are available from the authors. Country specific PPPs generated from the state–space methodology are also available for interested readers.

TABLE 1
Estimates of Parameters from Different Model Specifications

Variable	Regression No Spatial Errors (Panel 1)[a]		State–Space with Spatial Errors (Panel 2)		State–Space Without Spatial Errors (Panel 3)		State–Space 2005 Not a benchmark Year Spatial Errors (Panel 4)[b]		State–Space 2005 Not a Benchmark Year No Spatial Errors (Panel 5)[b]	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Intercept	-0.199	0.063	-3.684	2.337	-4.277	1.976	-0.622	1.589	-1.593	2.177
dum75_79	-0.128	0.238	0.702	0.602	0.543	0.376	-1.047	0.405	-0.501	0.436
dum80_84	-0.621	0.239	0.539	0.595	0.481	0.375	-2.751	0.401	-2.343	0.435
dum85_89	-0.282	0.244	0.579	0.593	-0.085	0.371	-2.883	0.398	-2.890	0.431
dum90_92	-0.451	0.244	-0.091	0.589	-0.100	0.362	-2.871	0.395	-2.961	0.420
dum93_95	-0.445	0.241	-0.721	0.587	-0.443	0.362	-3.154	0.392	-3.312	0.420
dum96_98	-0.692	0.246	-0.841	0.589	-1.558	0.363	-4.107	0.394	-4.389	0.422
dum99_01	-0.786	0.238	-1.553	0.595	-0.893	0.368	-3.646	0.399	-3.773	0.428
dum02_04[c]	-0.540	0.237	-1.824	0.606	-1.071	0.374	-3.803	0.407	-3.981	0.435
dum05	-0.770	0.221	-2.133	0.618	-2.526	0.379				
D_anz	0.016	0.080	-2.647	0.792	-2.566	0.756	-2.456	0.513	-2.272	0.801
D_asean	-0.029	0.155	-0.285	0.489	-0.472	0.406	-0.236	0.315	-0.480	0.429
D_cac	0.101	0.116	0.693	0.544	0.511	0.441	0.716	0.352	0.241	0.467
D_cafrica	0.118	0.094	1.676	0.569	2.195	0.485	2.143	0.365	2.686	0.509
D_eafrica	0.092	0.045	1.688	0.501	1.941	0.421	2.141	0.323	2.641	0.444
D_euro	0.045	0.073	0.076	0.408	-0.149	0.346	0.049	0.266	-0.163	0.368
D_mena	-0.081	0.082	-0.234	0.380	-0.247	0.319	-0.296	0.248	-0.209	0.341
D_mercsr	-0.243	0.086	-0.570	0.547	-0.659	0.443	-0.189	0.356	-0.492	0.472
D_nafta	0.066	0.122	-1.492	0.713	-2.044	0.580	-1.574	0.464	-2.110	0.616
D_safrica	0.228	0.148	1.537	0.571	1.918	0.488	1.373	0.365	1.732	0.510
D_scucar	0.632	0.206	-0.871	0.542	-0.072	0.465	-1.160	0.355	-0.238	0.495
D_spr	0.073	0.069	1.506	0.526	1.741	0.426	2.073	0.343	2.054	0.455
D_usd			0.593	0.292	0.639	0.251	0.305	0.191	0.345	0.268

D_wafrica	0.256	0.089	1.327	0.481	1.757	0.403	1.774	0.310	2.423	0.426
AGEDEP	0.365	0.174	-0.426	0.958	-1.235	0.840	-0.314	0.630	-1.168	0.891
AGVAGUN	-0.009	0.002	-0.016	0.011	-0.022	0.010	-0.017	0.007	-0.021	0.011
TRACTORPW	0.094	0.061	2.039	0.478	2.216	0.414	2.386	0.329	2.575	0.461
LABPOP	-0.003	0.003	0.037	0.018	0.027	0.016	0.023	0.012	0.010	0.017
LIFE	-0.006	0.004	-0.004	0.022	0.025	0.019	0.007	0.015	0.030	0.020
LITERATE	2.14E-04	1.40E-04	0.001	0.001	-2.00E-04	0.001	3.01E-04	4.55E-04	-1.14E-04	0.001
NTRVAG2	-0.004	0.003	0.021	0.012	0.010	0.010	0.009	0.008	0.002	0.011
EXPG	-0.002	0.003	-0.017	0.012	-0.026	0.011	-0.012	0.008	-0.022	0.012
PHONES	0.001	1.78E-04	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
RADPCCN	5.00E-06	7.00E-06	-6.80E-05	4.18E-05	-6.46E-05	3.84E-05	-9.78E-05	2.97E-05	-7.65E-05	4.46E-05
RURPOP	-0.004	0.001	-0.009	0.007	3.34E-04	0.006	-0.003	0.005	0.006	0.007
SECENR	3.30E-05	5.40E-05	-2.99E-04	3.20E-04	-2.59E-04	2.79E-04	4.49E-06	2.22E-04	-7.12E-06	3.15E-04
TRADEGUN	-2.09E-04	0.002	0.006	0.006	0.013	0.006	0.003	0.004	0.011	0.006
MANUFEXP	-2.41E-04	0.001	0.008	0.004	0.008	0.004	0.012	0.003	0.011	0.004
MANUFIMP	0.003	0.001	0.006	0.007	0.016	0.006	0.008	0.005	0.017	0.007
R^2	0.737									
logL			-6.75e+06		-8.90e+06		-4.53e+03		-4.54e+03	
Benchmarks	449		449		449		339		339	
σ_η^2 (Growth Rates)			10.000	0.010	8.000	0.007	3.000	0.015	8.000	0.015
σ_u^2 (Regression)			8.000	0.131	6.000	0.180	3.000	0.015	6.000	0.015
σ_ζ^2 (Benchmarks)			3.000	0.004	2.000	0.002	0.500	0.005	2.000	0.005
ϕ (Spatial Auto)			0.500	0.026	0	—	0.500	0.015	0	—

Notes:
(a) Initial regression predictions are produced from this model to start the algorithm.
(b) A regression similar to that in Panel 1 with data from benchmark years 1975–2002 is used to produce the initial predictions for the algorithm.
(c) dum02_05 for Panels 4 and 5.

5.2. *Predicted PPP Series for Selected Countries from the State–Space Model*

In this section we illustrate the econometric methodology proposed in this paper by presenting predicted *PPP* series derived from the estimated state–space model described in Section 3 under the different scenarios indicated above. We present these results through tables and figures for the five countries chosen to illustrate the results. The countries chosen are Australia, China, India, Nigeria, and Honduras. Australia is an OECD country and has participated in benchmark comparisons since 1985. Results for Australia are representative of results for the case of a developed country that has consistently participated in most of the global as well as OECD comparisons; it will illustrate the case when all sources of available information (national accounts and benchmark data) seem to provide a consistent picture. China participated in a benchmark comparison for the first time in 2005. India had participated in earlier benchmarks; however, it had not participated since 1985 and has again participated in the 2005 round. Nigeria had participated in the earlier comparisons, 1975, 1980, 1985, and 1996, and has participated in the 2005 round. Honduras had participated in the 1980 comparison and it is one of the countries in the sample that did not participate in the 2005 round.

The results presented in the tables show the extrapolated PPPs for the period 1971 to 2005 along with the benchmark PPPs where available and also the PWT 6.2 extrapolations. As this version of PWT was produced without the 2005 benchmark data, only extrapolations based on the models in Panels 4 and 5 are strictly comparable to the PWT extrapolations. Standard errors for the predicted PPPs are also presented. The post-fix "GRC" denotes the use of the constrained Kalman smoother so that the series preserves the implied growth rates in the GDP deflators reported by each country through national accounts. The figures show the extrapolated PPPs for each version of the model when the smoothing is constrained to preserve growth rates.

Australia

Table 2 and Figure 1 present the results for Australia. It is clear from these results that all sources of information for Australia are consistent and therefore, the resulting PPP series are virtually identical independently of which model is used. That is, the spatial error does not change the predictions from the non-spatial model, and the constrained smoother induces only marginal corrections to the values of earlier benchmarks. Given the frequency of participation of Australia in international comparisons, the addition of the 2005 ICP benchmark reduces the standard error (from AUD 0.04 to AUD 0.02 using the models with spatial errors) as expected; however, the out-of-sample prediction of the 2005 *PPP* (from models in Panels 4 and 5) is identical to the ICP value of AUD 1.390. This result holds for most countries in the OECD that have participated in all comparisons since 1990. Unfortunately, there are only 23 countries in this category. It is important to note the role of the ICP in the reduction of uncertainty in the form of standard errors. This result highlights the importance of participation in the international comparison exercises.

TABLE 2
PPP Estimates for Australia, 1971–2005

Year	ER	ICP	Spatial Model Panel 2	SE	Spatial Model GRC Panel 2	Non-Spatial Model[a] Panel 3	SE	Spatial Model 2005 Not a Benchmark[a] Panel 4	SE	Non-Spatial Model 2005 Not a Benchmark[a] Panel 5	SE	PWT 6.2
1971	0.883		0.580	0.038	0.712	0.577	0.034	0.535	0.021	0.562	0.033	0.771
1972	0.839		0.611	0.041	0.745	0.608	0.037	0.568	0.022	0.595	0.036	0.789
1973	0.703		0.679	0.045	0.821	0.676	0.040	0.636	0.024	0.663	0.040	0.838
1974	0.697		0.732	0.047	0.878	0.730	0.042	0.691	0.025	0.717	0.042	0.916
1975	0.764		0.768	0.047	0.914	0.767	0.042	0.732	0.025	0.755	0.041	0.973
1976	0.818		0.822	0.052	0.963	0.821	0.047	0.787	0.028	0.810	0.046	1.026
1977	0.902		0.850	0.054	0.980	0.850	0.048	0.818	0.029	0.838	0.048	1.041
1978	0.874		0.876	0.054	0.995	0.876	0.048	0.847	0.029	0.865	0.048	1.053
1979	0.895		0.904	0.052	1.010	0.904	0.047	0.878	0.028	0.894	0.046	1.062
1980	0.878		0.922	0.052	1.015	0.922	0.042	0.899	0.025	0.912	0.042	1.054
1981	0.870		0.971	0.050	1.037	0.972	0.045	0.953	0.027	0.964	0.044	1.068
1982	0.986		1.043	0.051	1.077	1.043	0.046	1.030	0.027	1.037	0.045	1.108
1983	1.110		1.116	0.049	1.117	1.117	0.043	1.108	0.026	1.112	0.043	1.150
1984	1.140		1.162	0.040	1.129	1.163	0.035	1.160	0.021	1.160	0.035	1.185
1985	1.430	1.240	1.226	0.021	1.157	1.227	0.017	1.231	0.009	1.227	0.017	1.236
1986	1.500		1.276	0.041	1.213	1.276	0.036	1.279	0.021	1.276	0.036	1.307
1987	1.430		1.333	0.049	1.276	1.333	0.044	1.335	0.026	1.332	0.043	1.353
1988	1.280		1.394	0.051	1.344	1.394	0.045	1.396	0.027	1.393	0.045	1.392
1989	1.260		1.407	0.044	1.366	1.406	0.038	1.407	0.023	1.406	0.038	1.414
1990	1.280	1.389	1.392	0.022	1.361	1.391	0.018	1.391	0.009	1.391	0.018	1.416
1991	1.280		1.373	0.039	1.339	1.372	0.034	1.373	0.020	1.372	0.034	1.406
1992	1.360		1.362	0.038	1.325	1.361	0.033	1.363	0.020	1.361	0.033	1.402
1993	1.470	1.350	1.345	0.021	1.306	1.345	0.017	1.347	0.009	1.345	0.017	1.393
1994	1.370		1.324	0.036	1.295	1.323	0.032	1.326	0.019	1.323	0.032	1.378
1995	1.350		1.318	0.035	1.298	1.317	0.031	1.318	0.019	1.317	0.031	1.374
1996	1.280	1.299	1.302	0.019	1.291	1.300	0.016	1.301	0.008	1.300	0.016	1.363
1997	1.350		1.297	0.034	1.286	1.296	0.030	1.298	0.018	1.296	0.030	1.366
1998	1.590		1.285	0.033	1.273	1.284	0.029	1.286	0.017	1.284	0.029	1.377
1999	1.550	1.297	1.294	0.018	1.281	1.293	0.015	1.296	0.008	1.293	0.015	1.374
2000	1.720		1.323	0.033	1.314	1.322	0.029	1.324	0.018	1.322	0.029	1.399
2001	1.930		1.324	0.033	1.319	1.323	0.029	1.325	0.017	1.323	0.029	1.398
2002	1.840	1.337	1.337	0.018	1.336	1.335	0.015	1.337	0.008	1.335	0.016	1.404
2003	1.540		1.349	0.033	1.349	1.348	0.029	1.350	0.022	1.348	0.036	1.389
2004	1.360		1.366	0.033	1.366	1.365	0.029	1.367	0.030	1.365	0.049	1.386
2005	1.309	1.390	1.389	0.019	1.389	1.388	0.016	1.390	0.036	1.388	0.060	

Note: (a) Growth Rate Constrained series appears in Figure 1.

Australia
Implied Growth Rates Preserved
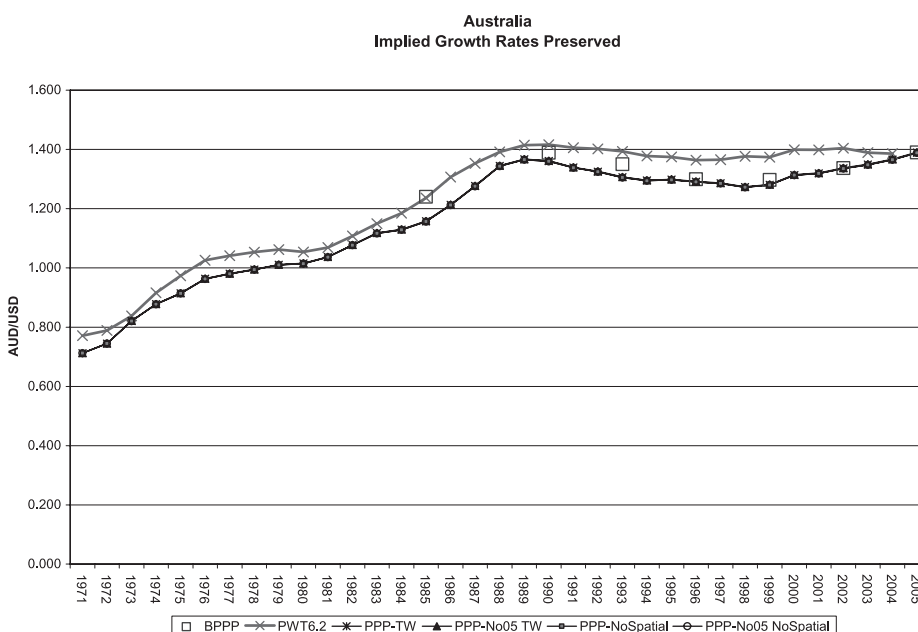
Figure 1. Australia. Extrapolated PPPs using models in Panels 2–5 and the constrained smoother

Notes: BPPP represent PPPs in benchmark years from the ICP; PWT 6.2 denotes PPPs from Version 6.2 of the Penn World Tables; PPP-TW are the predicted PPPs from the model in Panel (2) of Table 1 (includes trade weighted spatial errors and the 2005 ICP benchmark data); PPP-No05TW are the predicted PPPs from the model in Panel (4) of Table 1 (includes trade weighted spatial errors but 2005 is treated as a non-benchmark year); PPP-NoSpatial is the same as PPP-TW except that no adjustments are made for the presence of spatial autocorrelation (predictions from the model in Panel (3) of Table 1); and PPP-No05NoSpatial is the model without spatial autocorrelation, and the 2005 benchmark data are not included in the analysis (predictions from Panel (5) in Table 1). These results are useful in assessing the robustness of predictions from our model with respect to the inclusion of the 2005 benchmark data.

China

A large number of countries in the sample participated in an ICP comparison for the first time in 2005; China is one of them. The results for China are presented in Table 3 and Figure 2. We first concentrate on the extrapolated PPPs from the model estimated without the 2005 benchmark data. These series are comparable to PWT 6.2. As no benchmark data existed for China, the constructed series relies entirely on the predictions from the price level regression. It is in this case that we can evaluate the role of the spatial error structure. Through both the figure and the table it is clear that the inclusion of the spatial structure has made a substantial difference to the predicted PPP series. The prediction for China from PWT 6.2 for 2004 was Yuan 2.14; however, given the ICP benchmark estimate of Yuan 3.45 for 2005, the 2004 prediction should have been in the order of Yuan 3.3 to be on target. Our model (with spatial errors) predicted the 2004 value to be Yuan 2.90 and Yuan 2.98 for 2005. The approach presented in this paper is a simpler form of the more general framework we are

TABLE 3

PPP Estimates for China, 1971–2005

Year	ER	ICP	Spatial Model Panel 2	SE	Spatial Model GRC Panel 2	Non-Spatial Model[a] Panel 3	SE	Spatial Model 2005 Not a Benchmark[a] Panel 4	SE	Non-Spatial Model 2005 Not a Benchmark[a] Panel 5	SE	PWT 6.2
1971	2.460		0.795	0.273	3.086	0.818	0.248	0.927	0.171	0.953	0.295	1.808
1972	2.250		0.829	0.398	2.960	0.857	0.358	0.954	0.239	0.958	0.414	1.733
1973	1.990		0.850	0.492	2.808	0.883	0.440	0.968	0.288	0.950	0.499	1.654
1974	1.960		0.845	0.553	2.581	0.882	0.493	0.952	0.320	0.912	0.551	1.473
1975	1.860		0.821	0.579	2.331	0.860	0.516	0.915	0.335	0.858	0.571	1.347
1976	1.940		0.850	0.641	2.200	0.887	0.566	0.929	0.368	0.837	0.613	1.216
1977	1.860		0.881	0.690	2.091	0.916	0.605	0.946	0.398	0.822	0.646	1.172
1978	1.680		0.902	0.717	1.980	0.936	0.626	0.953	0.420	0.801	0.663	1.134
1979	1.550		0.929	0.736	1.893	0.961	0.641	0.966	0.441	0.787	0.677	1.055
1980	1.500		0.948	0.739	1.801	0.977	0.642	0.972	0.456	0.768	0.680	0.997
1981	1.700		0.968	0.743	1.684	0.991	0.642	0.977	0.472	0.747	0.687	0.937
1982	1.890		0.990	0.741	1.585	1.006	0.636	0.984	0.488	0.730	0.691	0.900
1983	1.980		1.040	0.751	1.541	1.050	0.642	1.020	0.516	0.736	0.711	0.888
1984	2.320		1.127	0.781	1.557	1.131	0.666	1.092	0.560	0.769	0.754	0.878
1985	2.940		1.282	0.849	1.664	1.280	0.722	1.230	0.638	0.846	0.839	0.926
1986	3.450		1.359	0.863	1.703	1.349	0.732	1.288	0.678	0.868	0.877	0.952
1987	3.720		1.435	0.870	1.741	1.417	0.737	1.344	0.717	0.889	0.911	0.981
1988	3.720		1.600	0.924	1.887	1.574	0.781	1.485	0.800	0.964	1.000	1.101
1989	3.770		1.725	0.942	1.979	1.689	0.795	1.584	0.862	1.012	1.060	1.172
1990	4.780		1.804	0.923	2.013	1.759	0.779	1.641	0.899	1.030	1.089	1.142
1991	5.320		1.903	0.915	2.076	1.847	0.771	1.715	0.951	1.059	1.136	1.161
1992	5.510		2.047	0.921	2.190	1.980	0.775	1.830	1.025	1.113	1.208	1.235
1993	5.760		2.367	0.995	2.490	2.282	0.837	2.103	1.186	1.261	1.382	1.475
1994	8.620		2.839	1.119	2.942	2.727	0.940	2.503	1.427	1.482	1.648	1.755
1995	8.350		3.209	1.180	3.279	3.071	0.989	2.810	1.616	1.644	1.850	1.975
1996	8.310		3.393	1.157	3.425	3.237	0.969	2.953	1.711	1.709	1.943	2.074
1997	8.290		3.404	1.075	3.419	3.254	0.904	2.957	1.729	1.701	1.961	2.083
1998	8.280		3.352	0.971	3.353	3.211	0.820	2.906	1.714	1.662	1.940	2.059
1999	8.280		3.274	0.860	3.264	3.143	0.728	2.835	1.685	1.612	1.903	1.997
2000	8.280		3.268	0.771	3.260	3.152	0.657	2.826	1.695	1.602	1.919	1.963
2001	8.280		3.253	0.676	3.248	3.151	0.579	2.810	1.699	1.588	1.927	1.960
2002	8.280		3.210	0.569	3.211	3.123	0.489	2.771	1.688	1.562	1.917	1.941
2003	8.280		3.229	0.469	3.229	3.170	0.406	2.787	1.712	1.571	1.954	1.977
2004	8.280		3.364	0.362	3.364	3.330	0.314	2.903	1.798	1.637	2.062	1.977
2005	8.194	3.450	3.448	0.179	3.448	3.439	0.146	2.975	1.856	1.677	2.137	2.145

Note: (a) Growth Rate Constrained series appears in Figure 2.

China
Implied Growth Rates Preserved

Figure 2. China. Extrapolated PPPs using models in Panels 2–5 and the constrained smoother
  Note: Legends used here are the same as those used in Figure 1.

proposing for the construction of panels of PPP, and the reader is referred to RRD for the performance of this model when regression predictions are used for all years (benchmark and non-benchmark years). In this case, the predictions of our model are Yuan 3.10 and Yuan 3.18 for 2004 and 2005, respectively. We note that the predictions produced by the model in Panel 5 (without spatial errors and 2005 ICP data) are extremely off target, Yuan 1.67 for 2005.

Turning to the models that included the ICP 2005 benchmark data, we note that the inclusion of the spatial error structure has only a marginal effect on the prediction of 2005, as the model without spatial autocorrelation predicts Yuan 3.44 and Yuan 3.45 for 2005 without and with spatial errors, respectively. This is an expected result given that this is an in-sample prediction.

India

India participated in the global comparisons of 1980, 1985, and 2005. The time lapsed between 1985 and 2005 is considerable, and thus the task of the PWT as well as our model (without the 2005 benchmark data) is considerable. The PPP predictions from our model when 2005 is not known and there are no spatial errors is again extremely off target and simply unacceptable. The predictions are substantially improved by the inclusion of the spatial error. The PWT 6.2 prediction for 2003 is Rupee 8.14, our spatial model predicts Rupee 10.3. To be consistent
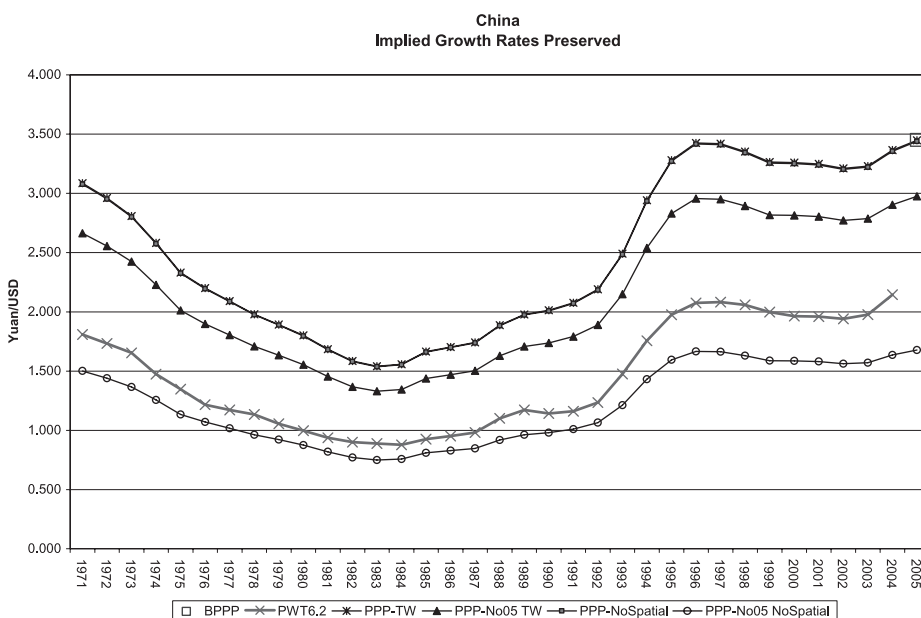
India
Implied Growth Rates Preserved
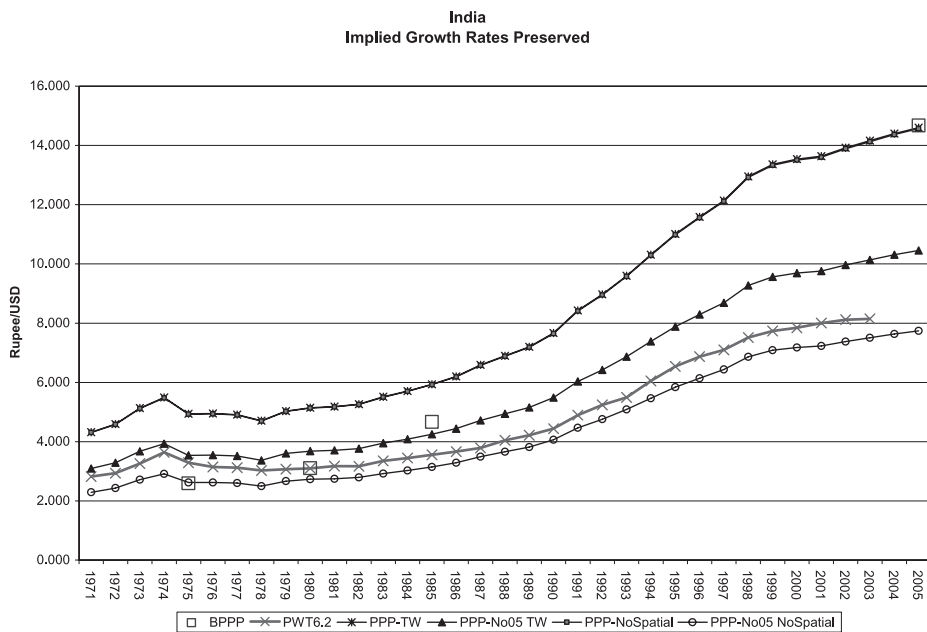
Figure 3. India. Extrapolated PPPs using models in Panels 2–5 and the constrained smoother
Note: Legends used here are the same as those used in Figure 1.

with the 2005 ICP value (Rupee 14.67), the 2003 prediction should have been in the order of Rupee 13.80.[45]

The predictions from the models that include the 2005 ICP data (with and without spatial errors) produce a marginally lower prediction than that of the ICP with values of Rupee 14.60 for 2005. The PPP series presented in Figure 3 and the fourth and sixth column of Table 4 are useful to observe the degree of inconsistency between the ICP benchmark observations and national accounts implied movement in prices. When the smoother is not constrained to preserve growth rates, the series obtained is that in Table 4, fourth column. In this case the PPP predictions are closer to the ICP benchmarks of 1975, 1980, and 1985 than the predictions obtained when the smoother is constrained to maintained implicit price movements (sixth column of Table 4 and Figure 3). For example, the ICP benchmark for 1985 is Rupee 4.667, the predicted value when growth rates are allowed to deviate from published sources is Rupee 4.621, while the predicted value from the series consistent with nationally reported price movements is Rupee 5.94.

Nigeria

Nigeria participated in the 1980, 1985, 1996, and 2005 ICP benchmarks. The predicted PPPs are in Tables 5 and Figure 4. Relying heavily on the historical ICP

---

[45]As in the case of China, it is also the case for India that when we use our general framework (see RRD and Rambaldi *et al.*, 2010) the predictions are much closer to the ICP value.

TABLE 4

PPP Estimates for India, 1971–2005

Year	ER	ICP	Spatial Model Panel 2	SE	Spatial Model GRC Panel 2	Non-Spatial Model[a] Panel 3	SE	Spatial Model 2005 Not a Benchmark[a] Panel 4	SE	Non-Spatial Model 2005 Not a Benchmark[a] Panel 5	SE	PWT 6.2
1971	7.490		2.231	0.441	4.318	2.230	0.393	2.050	0.223	2.090	0.369	2.820
1972	7.590		2.388	0.578	4.593	2.386	0.512	2.238	0.290	2.273	0.488	2.933
1973	7.740		2.682	0.656	5.129	2.681	0.578	2.565	0.330	2.596	0.560	3.257
1974	8.100		2.889	0.594	5.490	2.887	0.521	2.818	0.299	2.841	0.512	3.641
1975	8.380	2.594	2.611	0.274	4.938	2.610	0.225	2.596	0.113	2.608	0.225	3.287
1976	8.960		2.693	0.552	4.948	2.692	0.482	2.679	0.282	2.689	0.482	3.144
1977	8.740		2.749	0.658	4.913	2.747	0.578	2.736	0.340	2.744	0.577	3.119
1978	8.190		2.704	0.646	4.706	2.702	0.568	2.693	0.335	2.700	0.567	3.027
1979	8.130		2.976	0.599	5.031	2.973	0.524	2.965	0.307	2.971	0.524	3.067
1980	7.860	3.104	3.127	0.321	5.143	3.124	0.263	3.118	0.133	3.122	0.263	3.099
1981	8.660		3.319	0.650	5.184	3.316	0.569	3.312	0.334	3.307	0.567	3.175
1982	9.460		3.549	0.809	5.264	3.544	0.710	3.544	0.420	3.528	0.707	3.166
1983	10.100		3.906	0.876	5.513	3.900	0.770	3.904	0.456	3.875	0.765	3.352
1984	11.400		4.242	0.797	5.708	4.235	0.696	4.243	0.410	4.201	0.691	3.444
1985	12.400	4.667	4.621	0.447	5.937	4.613	0.366	4.626	0.185	4.567	0.365	3.553
1986	12.600		4.891	0.979	6.200	4.830	0.844	4.803	0.510	4.596	0.833	3.660
1987	13.000		5.267	1.365	6.590	5.147	1.170	5.076	0.729	4.710	1.149	3.789
1988	13.900		5.581	1.664	6.899	5.401	1.413	5.285	0.904	4.766	1.389	4.047
1989	16.200		5.893	1.918	7.202	5.649	1.613	5.488	1.062	4.815	1.594	4.215
1990	17.500		6.340	2.182	7.665	6.022	1.817	5.809	1.236	4.964	1.814	4.435
1991	22.700		7.064	2.534	8.430	6.672	2.097	6.353	1.470	5.306	2.122	4.893
1992	25.900		7.611	2.788	8.969	7.150	2.294	6.721	1.667	5.492	2.367	5.239
1993	30.500		8.246	3.031	9.598	7.704	2.481	7.154	1.880	5.722	2.628	5.489
1994	31.400		8.998	3.294	10.311	8.386	2.690	7.655	2.121	6.020	2.935	6.048
1995	32.400		9.752	3.508	11.011	9.068	2.860	8.144	2.361	6.304	3.234	6.539
1996	35.400		10.406	3.633	11.587	9.654	2.957	8.538	2.573	6.510	3.490	6.868
1997	36.300		11.029	3.712	12.139	10.286	3.041	8.908	2.789	6.742	3.781	7.095
1998	41.300		11.898	3.809	12.953	11.154	3.140	9.466	3.067	7.115	4.153	7.510
1999	43.100		12.402	3.723	13.363	11.684	3.087	9.728	3.248	7.264	4.394	7.734
2000	44.900		12.697	3.530	13.539	12.063	2.955	9.800	3.374	7.297	4.583	7.845
2001	47.200		12.912	3.246	13.634	12.368	2.742	9.813	3.475	7.287	4.737	8.003
2002	48.600		13.309	2.915	13.920	12.850	2.483	9.964	3.621	7.378	4.953	8.116
2003	46.600		13.757	2.513	14.161	13.438	2.162	10.135	3.781	7.506	5.208	8.146
2004	45.300		14.203	1.943	14.401	14.025	1.679	10.308	3.937	7.633	5.459	
2005	44.272	14.670	14.599	0.969	14.599	14.563	0.789	10.449	4.077	7.738	5.687	

Note: (a) Growth Rate Constrained series appears in Figure 3.

TABLE 5

PPP Estimates for Nigeria, 1971–2005

Year	ER	ICP	Spatial Model Panel 2	SE	Spatial Model GRC Panel 2	Non-Spatial Model[a] Panel 3	SE	Spatial Model 2005 Not a Benchmark[a] Panel 4	SE	Non-Spatial Model 2005 Not a Benchmark[a] Panel 5	SE	PWT 6.2	SE
1971	0.713		0.349	0.063	0.506	0.351	0.057	0.401	0.041	0.397	0.064	0.551	0.064
1972	0.658		0.343	0.080	0.499	0.345	0.072	0.388	0.049	0.387	0.081	0.540	0.081
1973	0.658		0.340	0.090	0.498	0.343	0.081	0.380	0.054	0.381	0.090	0.562	0.090
1974	0.630		0.445	0.126	0.658	0.450	0.112	0.492	0.074	0.497	0.124	0.651	0.124
1975	0.616		0.499	0.143	0.742	0.506	0.128	0.545	0.082	0.554	0.140	0.797	0.140
1976	0.627		0.541	0.153	0.803	0.548	0.136	0.581	0.086	0.589	0.147	0.887	0.147
1977	0.645		0.565	0.150	0.835	0.570	0.134	0.597	0.083	0.603	0.142	0.863	0.142
1978	0.635		0.604	0.141	0.889	0.607	0.125	0.626	0.076	0.631	0.130	0.937	0.130
1979	0.604		0.624	0.114	0.915	0.625	0.100	0.636	0.060	0.639	0.102	1.085	0.102
1980	0.547	0.643	0.645	0.058	0.943	0.645	0.047	0.646	0.024	0.648	0.047	1.210	0.047
1981	0.618		0.691	0.129	1.002	0.691	0.113	0.692	0.067	0.693	0.113	1.087	0.113
1982	0.673		0.675	0.152	0.969	0.674	0.134	0.674	0.079	0.676	0.134	1.086	0.134
1983	0.724		0.760	0.176	1.083	0.759	0.155	0.759	0.092	0.761	0.155	1.222	0.155
1984	0.767		0.863	0.170	1.221	0.861	0.149	0.859	0.087	0.862	0.149	1.356	0.149
1985	0.894	0.860	0.874	0.090	1.228	0.871	0.073	0.869	0.037	0.872	0.073	1.354	0.073
1986	1.750		0.894	0.190	1.183	0.891	0.166	0.893	0.097	0.895	0.167	1.388	0.167
1987	4.020		1.390	0.382	1.728	1.385	0.334	1.394	0.198	1.397	0.337	1.891	0.337
1988	4.540		1.729	0.538	2.028	1.722	0.470	1.741	0.279	1.744	0.476	2.279	0.476
1989	7.360		2.544	0.842	2.821	2.533	0.735	2.572	0.438	2.575	0.747	2.752	0.747
1990	8.040		2.768	0.934	2.910	2.754	0.815	2.809	0.488	2.810	0.832	2.916	0.832
1991	9.910		3.399	1.143	3.379	3.384	0.997	3.455	0.598	3.452	1.018	3.491	1.018
1992	17.300		6.453	2.087	6.065	6.428	1.823	6.571	1.096	6.557	1.861	5.698	1.861
1993	22.100		10.189	3.040	9.050	10.154	2.659	10.394	1.605	10.357	2.714	7.220	2.714
1994	22.000		13.567	3.537	11.324	13.543	3.098	13.839	1.869	13.762	3.152	9.319	3.152
1995	21.900		22.073	4.464	17.309	22.070	3.899	22.514	2.334	22.341	3.957	19.146	3.957
1996	21.900	32.539	31.536	3.103	23.253	31.583	2.543	32.163	1.312	31.853	2.598	25.925	2.598
1997	21.900		30.191	6.001	23.182	30.011	5.214	31.848	3.470	31.181	5.820	25.910	5.820
1998	21.900		27.082	6.795	21.655	26.718	5.889	29.552	4.410	28.601	7.282	24.183	7.282
1999	92.300		28.778	8.062	23.969	28.175	6.938	32.496	5.896	31.085	9.675	26.616	9.675
2000	102.000		37.521	11.009	32.411	36.704	9.467	43.662	9.145	41.533	15.058	37.479	15.058
2001	111.000		39.124	11.371	35.047	38.240	9.771	46.912	10.995	44.378	18.183	40.613	18.183
2002	121.000		38.509	10.451	35.782	37.608	8.974	47.592	12.252	44.768	20.358	41.683	20.358
2003	129.000		44.459	10.631	42.371	43.722	9.185	56.355	15.707	53.012	26.405	49.188	26.405
2004	133.000		50.694	9.475	49.509	50.190	8.197	65.849	19.644	61.943	33.401	58.771	33.401
2005	131.274	60.000	60.386	5.611	60.386	60.176	4.567	80.315	25.424	75.551	43.707		43.707

Note: (a) Growth Rate Constrained series appears in Figure 4.

Nigeria
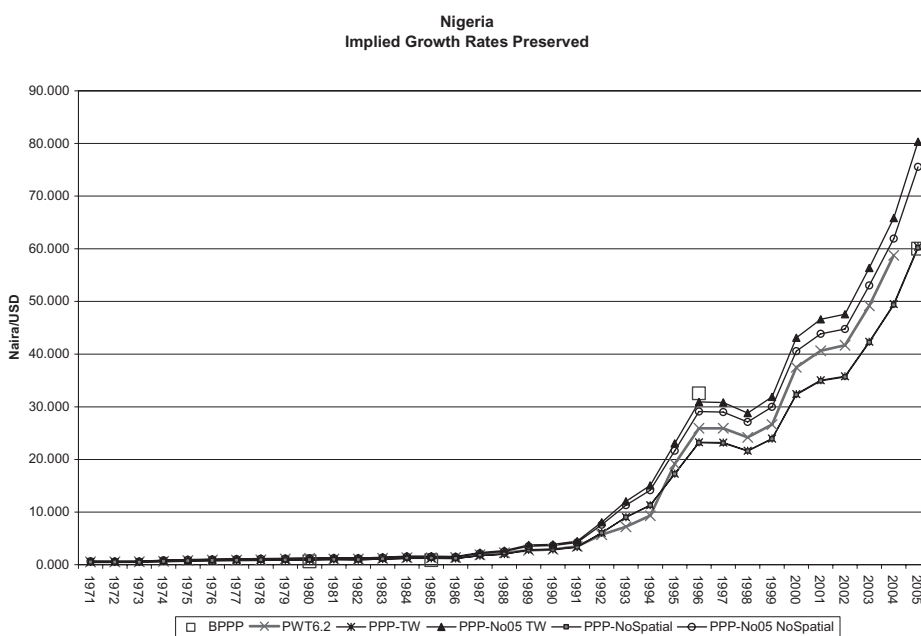**Implied Growth Rates Preserved**

Figure 4. Nigeria. Extrapolated PPPs using models in Panels 2–5 and the constrained smoother
Note: Legends used here are the same as those used in Figure 1.

benchmarks and published growth rates, the model that does not incorporate the 2005 ICP, predicts Naira 80.3 when spatial errors are include and Naira 75.55 when no spatial error are used. The PWT 6.2 prediction for 2004 is Naira 58.77, which would have implied Naira 68.8 in 2005. The ICP 2005 is Naira 60.00,[46] which implies a 2004 value of Naira 50.10. Our prediction of 2005, when 2005 ICP is incorporated is Naira 60.39.

The combination of the earlier benchmarks and reported GDP deflator for Nigeria strongly indicated a considerably higher value than that produced by the ICP for 2005. This is a consistent pattern from PWT as well as our model.

Honduras

Honduras is one of the countries that did not participate in the 2005 round of the ICP. Since no Central American country participated in this benchmark comparison, the available information for the region is only that from socio-economic variables and GDP deflators. Table 6 and Figure 5 present the results. The predictions based on the models without spatial errors are significantly different in both cases, that is, when 2005 ICP data are in the model as well as when these are not in the model (see Table 6). The inclusion of the spatial error results in predictions that are in the neighborhood of Lempira 8.00 when 2005 ICP is not known

---

[46]The complete model that includes regression predictions in non-benchmark years predicted Naira 66.00 which is higher but much closer to the value ICP 2005 value (see RRD).

TABLE 6

PPP Estimates for Honduras, 1971–2005

Year	ER	ICP	Spatial Model Panel 2	SE	Spatial Model GRC Panel 2	Non-Spatial Model[a] Panel 3	SE	Spatial Model 2005 Not a Benchmark[a] Panel 4	SE	Non-Spatial Model 2005 Not a Benchmark[a] Panel 5	SE	PWT 6.2
1971	2.000		0.809	0.114	1.032	0.803	0.101	0.800	0.063	0.801	0.101	0.991
1972	2.000		0.827	0.149	1.028	0.821	0.132	0.820	0.080	0.820	0.132	0.975
1973	2.000		0.845	0.170	1.026	0.840	0.151	0.841	0.091	0.840	0.151	0.979
1974	2.000		0.913	0.194	1.080	0.908	0.171	0.911	0.104	0.908	0.172	1.064
1975	2.000		0.911	0.195	1.050	0.906	0.171	0.912	0.104	0.908	0.172	1.026
1976	2.000		0.965	0.201	1.077	0.960	0.177	0.966	0.107	0.961	0.177	0.973
1977	2.000		1.049	0.203	1.137	1.044	0.179	1.050	0.108	1.044	0.179	0.983
1978	2.000		1.059	0.181	1.116	1.054	0.159	1.061	0.095	1.054	0.159	1.006
1979	2.000		1.121	0.151	1.148	1.116	0.132	1.123	0.078	1.115	0.132	1.069
1980	2.000	1.202	1.197	0.080	1.192	1.192	0.065	1.199	0.033	1.190	0.065	1.075
1981	2.000		1.192	0.169	1.168	1.162	0.144	1.192	0.087	1.152	0.143	1.091
1982	2.000		1.194	0.229	1.149	1.139	0.192	1.192	0.120	1.120	0.189	1.053
1983	2.000		1.249	0.291	1.183	1.165	0.238	1.245	0.152	1.136	0.233	1.105
1984	2.000		1.264	0.338	1.178	1.152	0.271	1.257	0.177	1.115	0.263	1.099
1985	2.000		1.311	0.390	1.203	1.168	0.305	1.301	0.203	1.122	0.294	1.132
1986	2.000		1.326	0.435	1.223	1.158	0.333	1.311	0.225	1.104	0.319	1.081
1987	2.000		1.320	0.469	1.224	1.130	0.351	1.301	0.240	1.072	0.335	1.105
1988	2.000		1.353	0.515	1.260	1.135	0.377	1.329	0.262	1.070	0.358	1.120
1989	2.000		1.390	0.561	1.300	1.144	0.402	1.360	0.283	1.072	0.380	1.172
1990	4.110		1.614	0.689	1.516	1.302	0.482	1.575	0.345	1.213	0.454	1.378
1991	5.320		1.958	0.882	1.846	1.550	0.604	1.905	0.438	1.439	0.567	1.677
1992	5.500		2.080	0.982	1.969	1.616	0.659	2.018	0.484	1.495	0.618	1.820
1993	6.470		2.300	1.133	2.187	1.756	0.745	2.227	0.555	1.619	0.697	2.137
1994	8.410		2.896	1.491	2.760	2.174	0.962	2.797	0.725	2.003	0.900	2.673
1995	9.470		3.536	1.895	3.378	2.610	1.200	3.407	0.915	2.403	1.123	3.189
1996	11.700		4.254	2.367	4.075	3.089	1.471	4.091	1.134	2.841	1.377	3.895
1997	13.000		5.081	2.936	4.901	3.648	1.801	4.897	1.403	3.371	1.696	4.815
1998	13.400		5.571	3.335	5.411	3.956	2.019	5.382	1.588	3.672	1.913	5.324
1999	14.200		6.086	3.772	5.951	4.272	2.252	5.892	1.791	3.984	2.148	5.978
2000	14.800		6.483	4.164	6.386	4.506	2.457	6.286	1.968	4.231	2.363	6.302
2001	15.500		6.783	4.506	6.732	4.669	2.628	6.588	2.120	4.413	2.548	6.719
2002	16.400		7.033	4.823	7.033	4.795	2.780	6.841	2.260	4.562	2.718	6.916
2003	17.300		7.421	5.264	7.421	5.024	3.006	7.218	2.450	4.814	2.966	7.337
2004	18.200		7.844	5.742	7.844	5.275	3.251	7.630	2.657	5.089	3.237	7.986
2005	19.000		8.373	6.313	8.373	5.592	3.543	8.144	2.905	5.431	3.561	

Note: (a) Growth Rate Constrained series appears in Figure 5.

Honduras
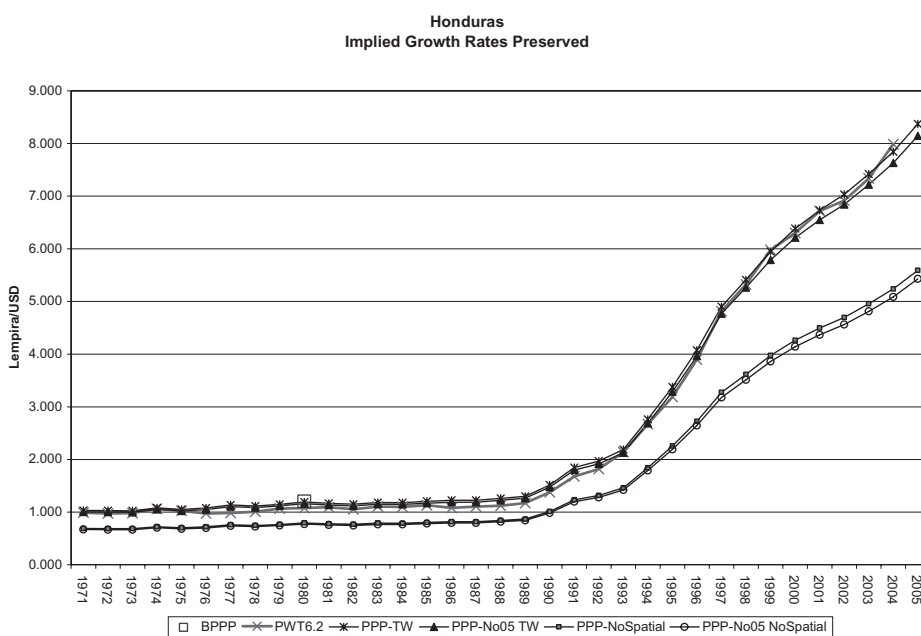**Implied Growth Rates Preserved**

Figure 5. Honduras. Extrapolated PPPs using models in Panels 2–5 and the constrained smoother
Note: Legends used here are the same as those used in Figure 1.

and higher when the 2005 ICP benchmark is included in the model.[47] These estimates are in line with the PWT 6.2 predictions for Honduras.

6. Conclusions

The econometric methodology suggested in the paper for the construction of a consistent panel of purchasing power parities represents a significant attempt to provide a clear and coherent approach since the first attempt by Summers and Heston (1988) to achieve consistentization between benchmarks and national growth rates. The approach proposed in this paper offers a formal econometric solution to the problem raised, and the method discussed here proposes a flexible method of extrapolating PPPs under different types of constraints including, firstly, tracking benchmark PPPs, and secondly, tracking national growth rates. The method described here makes use of PPP data from all the benchmarks which is a superior approach to that in the current practice of basing extrapolations on data from a single benchmark. Further, the method ensures that the extrapolations derived are weighted averages of the extrapolations from different benchmarks derived using movements in national deflators. The paper provides an analytical proof of this property of the extrapolation method.

[47]The estimates are closer to Lempira 10.00 and standard errors lower if regression predictions are used in non-benchmark years (see RRD).

In the empirical part of the paper we present extrapolations based on our methodology and compare and contrast them with the published PPPs from the Penn World Table Version 6.2. The methodology proposed is applied to a large dataset covering 141 countries and a 35-year period, 1970 to 2005, for generating predictions. In this paper we also examine the performance of the new methodology by producing extrapolations with and without the 2005 benchmark year. We find that when spatial autocorrelation is modeled and used, the extrapolated PPPs without the 2005 benchmark are quite meaningful. Given the large size of the panel of PPPs, the paper presents results for a selected set of countries, including China, India, Australia, Honduras, and Nigeria, to examine the plausibility of the extrapolations. The results based on the new econometric methodology are very encouraging and perform quite well compared to the existing extrapolations available.

The econometric approach enunciated in this paper based on a state–space formulation of the basic information is shown to provide a viable alternative to the present practice of basing the extrapolations of PPPs on a single benchmark year. A major advantage with our approach is the flexibility it offers and its ability to make use of all the information available at any given point of time. The method also provides standard errors associated with the extrapolated PPPs.

APPENDIX: KALMAN FILTER PREDICTIONS WITH NO REGRESSION INFORMATION IN NON-BENCHMARK YEARS IS A WEIGHTED SUM OF OBSERVED BENCHMARKS

We present the equations of the Kalman filter to assist the presentation.

(A.1)
$$\breve{\mathbf{p}}_{t|t-1} = \breve{\mathbf{p}}_{t-1} + \mathbf{c}_t$$

(A.2)
$$\boldsymbol{\Psi}_{t|t-1} = \boldsymbol{\Psi}_{t-1} + \hat{\mathbf{Q}}$$

(A.3)
$$\breve{\mathbf{p}}_t = \breve{\mathbf{p}}_{t|t-1} + \boldsymbol{\Psi}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \left(\mathbf{y}_t - \mathbf{B}_t \mathbf{X}_t \hat{\boldsymbol{\theta}} - \mathbf{Z}_t' \breve{\mathbf{p}}_{t|t-1} \right)$$

(A.4)
$$\boldsymbol{\Psi}_t = \boldsymbol{\Psi}_{t|t-1} - \boldsymbol{\Psi}_{t|t-1} \mathbf{Z}_t' \mathbf{F}_t^{-1} \boldsymbol{\Psi}_{t|t-1}$$

(A.5)
$$\mathbf{F}_t = \mathbf{Z}_t \boldsymbol{\Psi}_{t|t-1} \mathbf{Z}_t' + \hat{\mathbf{H}}_t$$

where,

$\hat{\mathbf{Q}}$, $\hat{\mathbf{H}}_t$, $\hat{\boldsymbol{\theta}}$ are estimates.

$\breve{\mathbf{p}}_t$ is the Kalman filter estimate of the state vector.

$\boldsymbol{\Psi}_t$ is the Kalman filter estimate of the unconditional covariance of the state vector.

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

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

Let $\breve{\mathbf{p}}_T$ be the Kalman filter estimate of $\mathbf{p}_T$ and $\breve{\mathbf{p}}_{T,j}$, $j = 0, 1, \ldots, M$ be the $M + 1$ different estimates of $\mathbf{p}_T$ obtained by applying growth rates to the benchmark

observations until time $t = T$. Further, we define $\mathbf{G}(i)$, the *Kalman gain*[48] at $t = t(i)$ which in our case takes the form:

$$(A.6) \qquad \mathbf{G}(i) = \begin{cases} \boldsymbol{\Psi}_{t|t-1}\mathbf{F}_t^{-1} & \text{for } i > 0 \\ \mathbf{I} & \text{for } i = 0 \end{cases}.$$

Proposition

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

$$(A.7) \qquad \breve{\mathbf{p}}_T = \sum_{i=0}^{M} \Upsilon_i^{(M)} \ddot{\mathbf{p}}_{T,i}$$

where the weights $\Upsilon_i^{(M)}$ are defined as

$$(A.8) \qquad \Upsilon_i^{(M)} = \begin{cases} \left[\displaystyle\prod_{j=1}^{M-i} (\mathbf{I} - \mathbf{G}(M-j+1)) \right] \mathbf{G}(i) & \text{for } i = 0, 1, \ldots, M-1 \\ \mathbf{G}(i) & i = M \end{cases}.$$

Lemma

The $\Upsilon_i^{(M)}$ defined in (A.8) are the product of positive definite (pd) matrices and

$$(A.9) \qquad \sum_{i=0}^{M} \Upsilon_i^{(M)} = \mathbf{I}_N.$$

Proof of Lemma

In (A.2) $\boldsymbol{\Psi}_{t-1}$ is positive semidefinite (psd) or pd and $\mathbf{Q}_t$ is positive definite (pd). Therefore, $\boldsymbol{\Psi}_{t|t-1}$ is pd for all t. Also, by definition $\mathbf{F}_t$ in (A.5) must be pd as $\mathbf{H}_t$ is pd. Thus, $\mathbf{G}(i)$ is the product of pd matrices for all i.

Also, post-multiplying (A.5) by $\mathbf{F}_t^{-1}$, we have

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

Therefore, $\mathbf{I}_N - \mathbf{G}(i) = \hat{\mathbf{H}}_t\mathbf{F}_t^{-1}$, and is also the product of pd matrices for all i. Thus, it follows that by (A.8) $\Upsilon_i^{(M)}$ is the product of pd matrices.

We will now establish that for $\Upsilon_i^{(M)}$ defined by (A.8), (A.9) holds. The proof will proceed by induction and we note that the form of $\Upsilon_i^{(M)}$ in (A.8) implies that:

$$(A.10) \qquad \Upsilon_i^{(M)} = [\mathbf{I} - \mathbf{G}(M)]\Upsilon_i^{(M-1)}.$$

We will now assume that (A.9) is true for $M - 1$. That is,

[48]See Harvey (1990, p. 110).

(A.11)
$$\sum_{i=0}^{M-1} \Upsilon_i^{(M-1)} = \mathbf{I}_N.$$

Then from (A.10) and (A.9)

$$\sum_{i=0}^{M} \Upsilon_i^{(M)} = \sum_{i=0}^{M-1} \Upsilon_i^{(M)} + \Upsilon_M^{(M)}$$
$$= [\mathbf{I} - \mathbf{G}(M)] \sum_{i=0}^{M-1} \Upsilon_i^{(M-1)} + \mathbf{G}(M),$$

and so by the assumption (A.11)

$$\sum_{i=0}^{M} \Upsilon_i^{M} = \mathbf{I}_N.$$

Therefore if (A.9) is true for $M - 1$, it is also true for M.

Now, set $M = 1$

$$\sum_{i=0}^{M} \Upsilon_i^{(M)} = \Upsilon_0^{(1)} + \Upsilon_1^{(1)}.$$

From (A.8) and (A.6)

(A.12)
$$\Upsilon_o^{(1)} = (\mathbf{I} - \mathbf{G}(1)), \quad \Upsilon_1^{(1)} = \mathbf{G}(1).$$

Therefore, (A.9) is true for $M = 1$ and so, by induction,

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

Proof of Proposition

In order to ease the notational burden, we will prove (A.7), first for the case $\mathrm{T} = t(M)$ and then extend to the case $\mathrm{T} > t(M)$.

Assume (A.7) and (A.8) are true for $\mathrm{T} = t(M - 1)$.

That is,

(A.13)
$$\breve{\mathbf{p}}_{t(M-1)} = \sum_{i=0}^{M-1} \Upsilon_i^{(M-1)} \ddot{\mathbf{p}}_{(M-1),i}.$$

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

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

The Kalman filter updating formula (see (A.3)) gives:

(A.14) $$\breve{\mathbf{p}}_{t(M)} = \left(\breve{\mathbf{p}}_{t(M-1)} + \overline{\mathbf{c}}\right) + \mathbf{G}(M)\left[\mathbf{y}(M) - \left(\breve{\mathbf{p}}_{t(M-1)} - \overline{\mathbf{c}}\right)\right]$$

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

Thus,

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

Now, by assumption (A.13)

$$
\begin{aligned}
\breve{\mathbf{p}}_{t(M-1)} + \overline{\mathbf{c}} &= \sum_{i=0}^{M-1} \Upsilon_i^{(M-1)} \ddot{\mathbf{p}}_{(M-1),i} + \overline{\mathbf{c}} \\
&= \sum_{i=0}^{M-1} \Upsilon_i^{(M-1)} \left(\ddot{\mathbf{p}}_{(M-1),i} + \overline{\mathbf{c}}\right) \quad \text{(by A.11)} \\
&= \sum_{i=0}^{M-1} \Upsilon_i^{(M-1)} \ddot{\mathbf{p}}_{t(M),i}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\breve{\mathbf{p}}_{t(M)} &= \sum_{i=0}^{M-1} [\mathbf{I} - \mathbf{G}(M)]\Upsilon_i^{(M-1)} \ddot{\mathbf{p}}_{M,i} + \mathbf{G}(M)\breve{\mathbf{p}}_{t(M)|M} \\
&= \sum_{i=0}^{M} \Upsilon_i^{(M)} \ddot{\mathbf{p}}_{t(M),i}.
\end{aligned}
$$

And so if (A.7) and (A.8) are true for $t(M-1)$, then they are also true for $t(M)$.

Now set $M = 1$. This implies two benchmark years, at $t(0) = 0$ and $t(1)$. By definition, $\breve{\mathbf{p}}_{t(1),0} = \breve{\mathbf{p}}_0 = \mathbf{y}(0)$, and $\breve{\mathbf{p}}_{t(1),1} = \mathbf{y}(1)$.

Then, using the Kalman updating formula,

$$
\begin{aligned}
\breve{\mathbf{p}}_{t(1)} &= [\mathbf{I} - \mathbf{G}(1)](\breve{\mathbf{p}}_0 + \overline{\mathbf{c}}) + \mathbf{G}(1)\mathbf{y}(1) \\
&= [\mathbf{I} - \mathbf{G}(1)]\ddot{\mathbf{p}}_{t(1),0} + \mathbf{G}(1)\ddot{\mathbf{p}}_{t(1),1} \\
&= \Upsilon_0^{(1)}\ddot{\mathbf{p}}_{t(1),0} + \Upsilon_1^{(1)}\ddot{\mathbf{p}}_{t(1),1} \quad \text{(by (A.8)).}
\end{aligned}
$$

Thus (A.7) and (A.8) hold for $M = 1$, and hence, by induction, for all M.

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

$$
\begin{aligned}
\breve{\mathbf{p}}_T &= \breve{\mathbf{p}}_{t(M)} + \overline{\mathbf{c}} \\
&= \sum_{i=0}^{M} \Upsilon_i^{(M)} \ddot{\mathbf{p}}_{t(M),i} + \overline{\mathbf{c}} \\
&= \sum_{i=0}^{M} \Upsilon_i^{(M)} \left(\ddot{\mathbf{p}}_{t(M),i} + \overline{\mathbf{c}}\right)
\end{aligned}
$$

$$\breve{\mathbf{p}}_T = \sum_{i=0}^{M} \Upsilon_i^{(M)} \ddot{\mathbf{p}}_{T,i}.$$

Special Case

If the elements of η_t and ξ_t are contemporaneously uncorrelated (that is, $\mathbf{Q}_t$ and $\mathbf{H}_t$ are diagonal) it is easily shown that the $\Upsilon_i^{(M)}$ are diagonal and positive definite for all $i = 1, \ldots, M$, provided $\Psi_0 \neq \mathbf{0}$.

Suppose that $\breve{p}_{jT}$ and $\vec{p}_{jT,i}$ are the Kalman filter and benchmark estimates (from the i-th benchmark) of the PPP of country j at time $t = T > t(M)$. Denote by $\upsilon_{jj,i}^{(M)}$ the j-th diagonal element of $\Upsilon_i^{(M)}$. It then follows that

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

Furthermore, because $\Upsilon_i^{(M)}$ is pd, and from (A.9), it follows that $\upsilon_{jj,i}^{(M)} > 0$ and $\sum_{i=0}^{M} \upsilon_{jj,i}^{(M)} = 1$.

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

References

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

Anselin, L., *Spatial Econometrics: Methods and Models*, Kluwer Academic, Boston, MA, 1988.

Asian Development Bank, "Purchasing Power Parities and Real Expenditures," 2005 International Comparison Program in Asia and the Pacific, Economics and Research Department, Manila, 2007 (available at: www.adb.org/Documents/Reports/ICP-Purchasing-Power-Expenditures/default. asp).

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

Barro, R. J. and X. X. Sala-i-Martin, *Economic Growth*, 2nd edn, MIT Press, Cambridge, MA, 2004.

Bergstrand, J. H., "Structural Determinants of Real Exchange Rates and National Price Levels: Some Empirical Evidence," *American Economic Review*, 81, 325–34, 1991.

———, "Productivity, Factor Endowments, Military Expenditures, and National Price Levels," in P. D. S. Rao and J. Salazar-Carrillo (eds), *International Comparisons of Prices, Output and Productivity*, Elsevier Science, North Holland, 1996.

Castles, I. and D. Hendersen, "IPPC Issues: A Swag of Documents," 2003 (available at: http:// www.lavoisier.com.au/papers/articles/IPPCissues.html; accessed December 14, 2007).

CIA, *The CIA World Factbook*, The Central Intelligence Agency, Langley, VA, 2008 (available at: https://www.cia.gov/library/publications/the-world-factbook/).

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

Durlauf, S. N., P. Johnson, and J. Temple, "Growth Econometrics," in P. Aghion and S. N. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1, Elsevier, Amsterdam, 555–677, 2005.

Elteto, O. and P. Koves, "On an Index Computation Problem in International Comparisons" (in Hungarian), *Statiztikai Szemle*, 42, 507–18, 1964.

European Central Bank, "Determination of the Euro Conversion Rates," Press Release, December 31, 1998 (available at: http://www.ecb.eu/press/pr/date/1998/html/pr981231_2.en.html).

FAO, *AGROSTAT*, Food and Agriculture Organization, 2006 (available at: http://faostat.fao.org/ default.aspx).

Geary, R. C., "A Note on the Comparison of Exchange Rates and Purchasing Power between Countries," *Journal of the Royal Statistical Society A*, 121, 97–9, 1958.

GGDC, 2008 (available at: http://www.ggdc.net/databases/other.htm).

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

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

Hill, R. J., "Convergence or Divergence: How to Get the Answer You Want," School of Economics Discussion Paper 2007/06, University of New South Wales, Sydney, 2007.

IMF, *International Financial Statistics*, International Monetary Fund, 2007.

Khamis, S. H., "A New System of Index Numbers for National and International Purposes," *Journal of the Royal Statistical Society A*, 135, 96–121, 1972.

Kravis, I. B. and R. E. Lipsey, *Toward an Explanation of National Price Levels, Princeton Studies in International Finance, No. 52*, Princeton University Press, Princeton, 1983.

———, "The Assessment of National Price Levels," Paper presented to Eastern Economic Association Meetings, Philadelphia, April 1986.

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

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

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

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

Milanovic, B., "True World Income Distribution, 1988 and 1993: First Calculation Based on Household Surveys Alone," *Economic Journal*, 112, 51–92, 2002.

Officer, L., *Purchasing Power Parity and Exchange Rates*, JAI Press, 1982.

Rambaldi, A. N., D. S. P. Rao, and K. R. Ganegodage, "Modelling Spatially Correlated Error Structures in the Time-Space Extrapolation of Purchasing Power Parities," in L. Biggeri and G. Ferrari (eds), *Price Indexes in Time and Space, Methods and Practice*, Springer, Heidelberg, 63–96, 2010.

Rao, D. S. P., A. N. Rambaldi, and H. E. Doran, "A Method to Construct World Tables of Purchasing Power Parities and Real Incomes Based on Multiple Benchmarks and Auxiliary Information: Analytical and Empirical Results," CEPA Working Paper Series, WP05/2008, University of Queensland, Brisbane, 2008.

Rose, A. K., "Bilateral Trade Data Set," 2004 (available at: http://www.cepr.org/data/#IM_bilateral_trade_observations; accessed March 2010).

Sala-i-Martin, X. X., "15 Years of New Growth Economics: What Have We Learnt?" Unpublished Manuscript, 2002.

Summers, R. and A. Heston, "A New Set of International Comparisons of Real Product and Price level Estimates for 130 Countries, 1950–1985," *Review of Income and Wealth*, 34, 1–25, 1988.

———, "The Penn World Tables (Mark 5): An Expanded Set of International Comparisons, 1950–88," *Quarterly Journal of Economics*, 2(1), 1–45, 1991.

Szulc, B., "Index Numbers of Multilateral Regional Comparisons" (in Polish), *Przeglad Statistyczny*, 11, 239–54, 1964.

UN, *National Accounts Main Aggregates Database*, United Nations, 2007 (available at: http://unstats.un.org/unsd/snaama/; accessed June 2007).

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

World Bank, *World Development Indicators*, various years (available at: http://go.worldbank.org/U0FSM7AQ40; accessed March 2010).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table A1. List of Countries Used in the Study.
Table A2. Definitions and Sources of Variables.
Table A3. Descriptive Statistics for Variables.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.