Econometric Estimation and Aggregation of PPP Panels for Components of GDP

Alicia Rambaldi (University of Queensland, Australia)

Linh Huynh (University of Queensland, Australia)

Prasada Rao (University of Queensland, Australia)

Paper Prepared for the IARIW 33rd General Conference

Rotterdam, the Netherlands, August 24-30, 2014

Session 8C

Time: Friday, August 29, Afternoon

Econometric Estimation and Aggregation of PPP Panels for Components of GDP

L. T. Huynh, A. N. Rambaldi and D. S. P. Rao

August 8, 2014

School of Economics, The University of Queensland. St Lucia, QLD 4072. Australia.

Abstract

This paper is a part of a larger project on the construction of panels of PPPs undertaken by Rao, Rambaldi and Doran at the University of Queensland. The project aims to develop a coherent econometric framework to extrapolate PPPs using information on PPPs from the benchmark data from the International Comparison Program as well as the information on price deflators from national sources. In the first stage of the project, the econometric framework has been developed and implemented to extrapolate PPPs at the GDP level. In the second stage, as a natural progression, the method is extended to produce extrapolated PPPs for the three major components of GDP, consumption (C), investment (I) and government (G). The contributions of the paper are in three areas. The first is on the specification of structural models to explain the price level of the components of GDP. There are no ready-made results on the structural determinants of the price level for C, G and I. Therefore, we bring in elements of the macroeconomic literature to define the economic models for C, I and G. Through these economic models, groups of variables are identified and included in the econometric estimation of the price level for each component. The second contribution is on econometric methodology. Here we propose to use a bootstrap estimation approach to incorporate the statistical uncertainty associated with the estimation of a subset of the parameters of the model that was ignored by the RRD method. The paper makes use of the GEKS as well as the GK methods to obtain PPPs for domestic absorption (DA) for each country and time period by aggregating the estimated PPPs of each component. The methodology is implemented in generating panels of PPPs for C, G and I for 181 countries covering the period 1970 to 2010. Using the recently released ICP 2011 benchmark results PPPs for GDP level are extrapolated for the period 1970-2012. A set of experimental calculations and extrapolations are presented which include/exclude the 2005 and 2011 benchmark results.

1 Introduction

This paper describes an econometric framework to create a balanced panel across countries and time periods by extrapolating the purchasing power parities (PPP) produced by the International Comparison Program (ICP) for GDP and its the components (consumption (C), investment (I) and government (G)). The ICP PPPs are available for only benchmark countries in benchmark years, and thus extrapolation of the PPPs for components of GDP to non-benchmark countries and years are needed. This paper deals with the extrapolation of the components, as well as a method to aggregate the estimates to construct PPPs for GDP. The approach is an extension of the method proposed by Rao et al. (2010b,a)-RRD Method to construct PPPs for GDP at current prices.

The contributions of the paper are in three areas. The first is on the specification of structural models to explain the price level of the components of GDP. There are no ready-made results on the structural determinants of the price level for C, G and I. Therefore, we bring in elements of the macroeconomic literature to define the economic models for C, I and G. Through these economic models, groups of variables are identified and included in the econometric estimation of the price levels for each component. The second contribution is the development of a bootstrap estimation approach to incorporate the statistical uncertainty associated with the estimation of a subset of the parameters of the model that was ignored by the RRD method. The third aspect is the use of the GEKS as well as the GK methods to obtain PPPs for GDP for each country and time period by aggregating the estimated PPPs of each component. Empirical results are presented in the form of panels of PPPs for C, G and I and aggregated domestic absorption (DA) and gross domestic product (GDP) for 181 countries covering the period 1970 to 2010. As the recent ICP benchmark (2011) results have been released, we also present the extrapolated PPPs for GDP level obtained using the RRD method for the period 1970 to 2012. We present a set of results for a few selected countries from the forthcoming release of UQICD `http://uqicd.economics.uq.edu.au`.

Typically the extrapolation of ICP PPPs involves two stages. In the first stage, PPPs in a given benchmark year are extrapolated to non-participating countries. In the second stage, PPPs for both participating and non-participating countries are extrapolated to non-benchmark years. However, different approaches have been used in the past with each using its own methodology and producing different sets of results with different properties. These methods include the Reduced Information method and various regression methods (REG, PWT all versions up to and including PWT7.0[1], RRD). The reduced information method (Ahmad (1980), Ahmad (1988)) makes quick estimates for non-participating countries PPPs collecting prices for a reduced sample of carefully selected items, then making ICP type calculations for GDP and a small number of its components. Apart from the ICP aggregation methods, this approach is concerned more with price collection than estimation. The main regression based methods include the REG method (Ahmad (1996)), the earlier PWT method (Heston et al. (2012)) and the RRD method which are, respectively, the methodologies used by the World Bank, the PWT (versions 5, 6 and 7), and the UQICD. Among the regression methods, only the PWT method has used the regression method to estimate PPPs for the components C, I and G. Other methods mainly estimate PPPs at the GDP level. Still, these methods are mentioned here as a starting point for the development of new techniques to construct complete PPP panels for each of the GDP components.

PWT7 is based on benchmark data from the 2005 ICP Benchmark. Apart from being a source of complete GDP PPP panel like the World Bank and the UQICD, PWT7 is the only existing source of complete PPP panels for C, I and G. The basic data consist of price and expenditure data (from national accounts) for basic headings of consumption, government expenditures, investment, and exports and imports in local currency units are aggregated to obtain PPPs for Consumption (C), Government expenditures (G), Investment (I) and Net Foreign Balance (NFB). The PPP construction of the PWT (versions 5-7) essentially involved three main steps: (i) aggregation of PPPs at basic heading level and aggregate component level for the most current ICP benchmark for participating countries, (ii) extrapolation of these PPPs for nonparticipating countries for the benchmark year, (iii) extrapolation of price levels obtained in step (i) and (ii) over time and aggregation of them into GDP price levels for all countries and years. Readers who are interested in the PWT method may refer to the PWT technical documentation in Heston et al. (2012).

[1]The PWT Version 8.0 uses an approach to extrapolation of PPPs that differs significantly from the earlier versions of PWT including PWT 7.0. PWT 8.0 discontinued the practice of extrapolating PPPs for countries not participating in the benchmark years of ICP. Extrapolation of PPPs for countries which have have participated in two benchmark years, say s and t ($> s$), is based on an weighted average of PPPs from the two benchmarks with weights that depend on the distance from the two benchmarks. For country j and a period $t*$, the interpolated value is: $PPP_j^{t*} = PPP_j^s \times \frac{(t-t^*)}{(t-s)} + PPP_j^s \times \frac{(t*-s)}{(t-s)}$ for $s \leq t^* \leq t$. For countries which have participated in only one benchmark, then PPPs are extrapolated forwards and backwards using movements in deflators relative to the reference country. PWT 8.0 in principle does not provide PPPs for counties that have never participated in ICP. As ICP 2005 covered 146 countries, this is not a major issue.

2 The General RRD methodology for the Construction of Panels of PPPs at Current Prices

The RRD method by Rao, Rambaldi and Doran (Rao et al. (2010b,a)) is an econometric based approach to the construction of panels of PPPs and real incomes. The method improves upon the PWT and the REG methods as it improves on (i) time-space inconsistency of the data produced from different benchmarks and (ii) standard errors for the predictions. By using a single step state space econometric framework, all the available information of PPPs for countries over time is combined efficiently.

2.1 Description of the method

The econometric problem is one of signal extraction. That is, there are a number of sources of "noisy" information that can be combined to extract the signal. A state-space (SS) is a suitable representation for this type of problems. At any time period t the N countries can be placed in one of three groups when t is an ICP benchmark year or in two groups otherwise. In an ICP year, the groups are: the reference country (without loss of generality this is set to be the first country), the non-participating countries and the participating countries. In a non-ICP benchmark year there are only two groups: the reference country and all others.

The mapping is from what is observed or measured *with some error* at time t to a vector of true but unobserved PPPs to be estimated. It is convenient to work with log transformations and thus, at each t the vector of log PPPs, at current prices, (for the N countries) is denoted by $p_t = ln(PPP_t)$, with elements $p_{it} = ln(PPP_{it})$ for $i = 1, ..., N$. The objective is to estimate p_t for all N countries and $t = 1, ..., T$ time periods to generate a complete panel. The mapping equations (known as a observation and transition equations in the state-space literature) are given in equations (1) and (13). The rest of the sub-section presents the economic and econometric framework that leads to these two sets of equations. Equation (1) simply links the *observed* information and noise to the latent p_t. Equation (13) provides the law of motion of p_t over time, which is derived from index theory and is the established updating approach used by PWT and Maddison (2007).

$$y_t{=}Z_t p_t{+}\zeta_t \tag{1}$$

In an ICP benchmark year the mapping is as follows,

$$y_t{=}\begin{bmatrix}0\\ \hat{p}_t\\ \tilde{p}_t\end{bmatrix};\ Z_t{=}\begin{bmatrix}S_1\\ S_{np}\\ S_p\end{bmatrix};\ \zeta_t{=}\begin{bmatrix}0\\ S_{np}v_t\\ S_p\xi_t\end{bmatrix} \tag{2}$$

where,

y_t is a vector of *observed* information

$\tilde{p}_t$ is a vector of log transformations of the ICP PPP benchmarks for participating countries, $P\tilde{P}P_t$.

$\hat{p}_t$ is a vector of log PPP regression predictions for non-participating countries. The predictions are based on a model of the log of price levels ($\ln(PPP_{it}/XR_{it})$), some details provided below;

The first element of y_t is zero as that is the observation for the reference country $p_{1t} = ln(PPP_{1t})$ which is a constraint in the system

Z_t is a partitioned selection matrix with components which select the reference country (country 1), S_1, the non-participating countries, S_{np}, and the participating ICP countries, S_p; and

ζ_t is a random vector capturing the uncertainty arising from each set of sources of observed values of PPP_{it}. The first row is zero as it represents the reference country constraint. The non-participating countries have error v_t, and the ICP measures have error ξ_t. The variance-covariance matrix of ζ_t is then given by,

$$E\left(\zeta_t\zeta_t'\right) \equiv H_t = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sigma_u^2 S_{np}\Omega_t S_{np}' & 0 \\ 0 & 0 & \sigma_\xi^2 S_p V_t S_p' \end{bmatrix} \tag{3}$$

In a non-benchmark years there are no observations from the ICP, thus the only observations are those produced by the predictions from the price level model and the constraint,

$$y_t = \begin{bmatrix} 0 \\ S_{np}\hat{p}_t \end{bmatrix};\ Z_t = \begin{bmatrix} S_1 \\ S_{np} \end{bmatrix};\ \zeta_t = \begin{bmatrix} 0 \\ S_{np}v_t \end{bmatrix} \tag{4}$$

The components of this mapping are derived from the following theoretical considerations,

1. The observed PPPs from the ICP, in the benchmark years, are related to the true PPPs through the following equation:

$$\tilde{p}_{it} = p_{it} + \xi_{it} \tag{5}$$

where ξ_{it} is a random error accounting for measurement error with the properties:

$$E(\xi_{it}) = 0;\ E(\xi_{it}^2) = \sigma_\xi^2 V_{it} \tag{6}$$

The measurement error variance-covariance is of the form

$$V_t = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_{1t}^2 jj' + diag(\sigma_{2t}^2, \ldots, \sigma_{Nt}^2) \end{bmatrix}$$

where j is a vector of 1's and σ_{it}^2 is the variance of the PPP from the ICP benchmark for country i in period t. Here σ_{1t}^2 is the variance of the reference country (country 1). In the empirical implementation of the method, σ_{it}^2 is assumed to be inversely related to the GDP of country i in period t [2].

2. The numerical value of the PPP for the reference/numeraire country, 1, is set at 1. Thus

$$p_{1,t} = 0;\ t = 1, 2, \ldots, T \tag{7}$$

3. The key element of the approach is the regression model used in extrapolating PPPs to non-participating countries using PPP data from the ICP benchmarks. The regression model draws on the literature on the explanation of national price levels (Kravis and Lipsey (1983); Clague (1988) and Bergstrand (1991, 1996)). A linear model in logarithms of price levels is postulated as below:

$$r_{it} = ln(PPP_{it}/XR_{it}) = \beta_{0t} + \mathbf{x'}_{it}\beta_s + e_{it} \tag{8}$$

for all $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$

Deviating from the usual assumptions on the disturbance term, we assume that errors in (8) are spatially autocorrelated . The following specification is used

$$e_t = \phi W_t e_t + u_t \tag{9}$$

[2] In order to avoid circularity, GDP in $US adjusted by market exchange rates is used in the estimation process.

where $|\phi| < 1$ and $W_t(N \times N)$ is a spatial weights matrix and $u_{it} \sim N(0, \sigma_u^2)$. The term spatial in the present contexts refers to socio-economic distance rather than the traditional geographical distance. It follows that $E(u_t u_t')$ is proportional to $\Omega_t = (I - \phi W_t)^{-1}(I - \phi W_t)^{-1'}$. If estimates of parameters in (8) are available, then predictions of PPPs consistent with price level theory can be generated for any country in any period. These are given by:

$$\hat{p}_{it} = \hat{\beta}_{0t} + \mathbf{x'}_{it}\hat{\beta}_s + ln(XR_{it}) + \hat{\phi}W_t\hat{e}_t \tag{10}$$

Inspection of equation (8) shows it is possible to obtain estimates of the parameters by using the unbalanced panel available through using as dependent variable $\hat{r}_t = \hat{p}_t - ln(XR_t)$. However, these predictions can be improved.

Using this as a set of starting predictions, RRD embeds a re-written version of equation (10) in a re-writing (1) as follows,

$$y_t = Z_t p_t + B_t X_t \theta + \zeta_t \tag{11}$$

where, θ, a function of β_{0t} and β_s and B_t a mapping matrix to non-participating countries. Upon convergence of the estimation algorithm (which involves the Kalman filter algorithm) $\theta \to 0$ and (11) reduces to (1), which is then used by the Kalman filter and Smoothing algorithm to produce estimates of the latent vector p_t and an associated mean squared prediction error matrix. The point to note here is that unlike the PWT and other extrapolation methods, this approach generates predictions for all the cells (time periods and countries). However, it is trivial to limit the regression based PPPs, $\hat{p}_t$, (through Z_t and B_t) to be used by the model's predictor to only those countries and years when no ICP benchmark observations, $\tilde{p}_t$, are available.

The identification of p_t from the above mapping requires information on how PPPs evolve over time. The updating of PPPs from period $t-1$ to t is through the GDP deflators in the country concerned and in the reference country. Thus,

$$PPP_{i,t} = PPP_{i,t-1} \times \frac{GDPDef_{i,[t-1,t]}}{GDPDef_{1,[t-1,t]}} \tag{12}$$

Taking logarithms on both sides of (12), and assuming the updating equation (12) holds on average due to measurement error, we have

$$p_{it} = p_{i,t-1} + c_{it} + \eta_{it} \tag{13}$$

where $c_{it} = ln(\frac{GDPDef_{i,[t-1,t]}}{GDPDef_{US,[t-1,t]}})$; and $\eta_{it} \sim N(0, \sigma_\eta^2)$ is random error accounting for measurement error in the growth rates. Equation (13) is commonly used in constructing panels of PPPs including the PWT and in the construction of the Maddison series[3]. The variance covariance matrix of η_{it} is assumed to be similar to the matrix in equation (5).

As the current problem is one of finding predictions for the vectors of PPPs from a variety of sources of noisy information through the ICP benchmarks; regression predictions and, finally, the updating equation in (13), a state-space (SS) representation is suitable for these kinds of problems and the approach proposed formulates all the information in equations (1) to (13) in the form of a set of observation and transition equations on the state vector p_t which is the vector of unknown $ln(PPP_t)$. Under Gaussian assumptions, the Kalman filter and Smoother predictor of the conditional mean, $\widetilde{p_{it}}$, conditional on information available at time t, is a minimum squared error predictor of the state vector, p_t[4]. The panel of PPPs is the obtained by,

[3] Maddison (2007) presents series that are extrapolated from the 1990 benchmark year.

[4] Technical details and equations for the Kalman Filter and Smoother are provided in Appendix A.6 and Appendix B of Rao et al. (2010b).

$$\widetilde{PPP_{it}} = \exp(\widetilde{p_{it}})\; i = 1, ..., N \text{ and } t = 1, ..., T \tag{14}$$

where the wide "$\sim$" is used to denote the RRD estimates of the log of PPP, $\widetilde{p_{it}}$, and corresponding smoothed estimated PPPs, $\widetilde{PPP_{it}}$

2.2 Analytical properties

In order to provide a better appreciation of the features of RRD, a number of analytical results are presented here. In particular, these properties demonstrate the flexibility of the method and show how it provides intuitively meaningful predictions under specific scenarios. The following properties are stated without proofs but complete proofs are provided in the referenced materials.

1. The predicted PPPs are weighted sums of all the available information

   Using the results from Koopman and Harvey (2003) we can write the estimate of the PPPs at period t, $\widetilde{p}_t$, as a weighted sum of information immediately closest to the time period t, with the highest weight at t and decreasing weights as j is further away in each direction from t. The weights w_{jT} depend on the benchmark information, regression information and measurement error structures attached to that information.

$$\widetilde{p}_t = \sum_{j=1}^{T} \omega_{jT} y_j \tag{15}$$

   The size and shape of the weights depends on the time period. The form of the weights are shown in Appendix A. For example, if the sample goes up to 2012, for t =2011, the highest weight will be from the 2011 information with a discounted memory back to the start of the sample; although in practice most of the non-zero weights might be from the immediate past years. Some weight, although likely to be small, will be from the 2012 information. The weights sum to one.

   As presented in Appendix A, the adjustment provided by the weights is from information about the movement of PPPs between benchmarks *after the deflator movement has been incorporate*. This information include national accounts data, ICP benchmarks and the influence of movements in other trading partners which is brought into the weights through the cross-sectional correlation information gather by the price level regression used by UQICD.

2. The predicted PPPs are "weighted averages" of benchmark year only extrapolations

   Suppose there are $M + 1$ benchmark years. If regression based predictions are used to extrapolate PPPs to non-participating countries *only in benchmark years* and then use the implicit price deflators to extrapolate from one year to the next, then it is possible to construct a panel of extrapolated PPPs for each of the benchmark years. In this case, an obviously intuitive approach is to make use of an average of these $M + 1$ panels of PPPs. An important property of the RRD approach is that, in this case, the predictions $\widetilde{p}_t$ can be shown to be a weighted average of the $M + 1$ panels of PPPs, where the weights are determined by the diagonal elements of the 'Kalman Gain' matrices, which represent the gain in information provided by an additional benchmark. The weights can be interpreted as reflecting the reliability of the $j - th$ benchmark. The proof of this important property is presented in Rao et al. (2010b).

3. Invariance of the Predicted PPPs to the Choice of the Reference Country

   The relative purchasing powers of currencies of countries should, in principle, be invariant to the choice of the reference country. It can be shown that RRD satisfy this important invariance property. The proof of this property is quite involved and it is presented in Appendix A of Rao et al. (2010a).

4. Constraining the model to track PPPs for countries participating in the benchmarks

   As the ICP is the main source of PPPs for countries participating in different benchmarks and given that respective PPPs are determined using price data collected from extensive price surveys, one may consider it necessary that the econometric method proposed should generate predicted PPPs that are identical to PPPs for the countries participating in different ICP benchmarks. In RRD this can be achieved by simply setting the variance of the disturbance term in (5) to be equal to zero. In this case a particular property of Kalman filter predictions is that the predicted PPPs ($\widehat{PPP_{it}}$) will be identical to the ICP benchmark, $P\tilde{P}P_{it}$, [5] when t is a benchmark year.

5. Constraining the model to preserve movements in the Implicit GDP Deflator

   In the currently available PWT and the Maddison series, growth rates in real GDP and movements in the implicit price deflators are preserved. As the GDP deflator data are provided by the countries and given that such deflators are compiled using extensive country-specific data, it is often considered more important that the predicted PPPs preserve the observed growth rates implicit in the GDP deflator. This essential feature can be guaranteed in RRD by simply stipulating the variance of the error in the updating equation (13) be zero. It is trivial to show that the national level movements in prices are preserved using the formulae for the fixed interval Kalman Smoother[6].

We note here that it is not possible to simultaneously constrain the predictors to track the benchmark PPPs as well as the national movements in GDP deflators. One has to choose either one or none of these restrictions when generating panels of extrapolated PPPs. The recommended approach is to simply use unconstrained equations and thereby not impose either of the restrictions described above.

2.3 Improvement to the standard error estimation by bootstrapping

State space based approaches are very popular in the literature for describing the dynamic evolution of macroeconomic and financial time series since they allow the implementation of the Kalman filter and smoothing algorithms which deliver minimum mean squared error estimates when the linear state space models have known hyperparameters. Hyperparameters (also labelled 'other parameters' in the literature) are typically associated with the variance-covariance of the measurement equation, (1), and the transition equation, (13), of the state-space representation. In RRD these are ϕ, σ_u^2, σ_ξ^2 (associated with the measurement equation, see (3) and definitions) and σ_η^2, associated with the transition equation. The filters and smoothers also deliver measures of uncertainty associated with the estimates which are the prediction mean squared errors (PMSE). Standard errors for the predicted PPPs in (14) can be computed under log-normal assumptions as follows,

$$SE(\widetilde{\widehat{PPP_{it}}}) = \sqrt{exp(2 \times \widetilde{p_{it}})exp(\hat{\Psi}_{ii,t})exp(\hat{\Psi}_{ii,t} - 1)} \tag{16}$$

where $\hat{\Psi}_{ii,t}$ is the ith diagonal element of the estimated smoothed covariance of the state vector. Both $\widetilde{p_{it}}$ and $\hat{\Psi}_{it}$ obtained from the Kalman Filter and Smoother are dependent on the hyperparameters ϕ, σ_u^2, σ_ξ^2 and σ_η^2 which are unknown and therefore also estimated (for details of estimation of the hyperparameters as well as the state vector and its mean squared error covariance please see Durbin and Koopman (2012)). However, in equation (16) these hyperparameters are treated as given since the formula does not account for the uncertainty associated with estimating them. As a result, these standard errors might underestimate the true PMSE of the state (i.e. estimates PPPs) Ansley and Kohn (1986), Hamilton (1986), Durbin and Koopman (2000), Quenneville and Singh (2000).

[5] This result follows from the work of Doran (1992).

[6] The proof of this property is provided in Appendix B of Rao et al. (2010a).

There are four main methods in the literature to incorporate the hyperparameter's uncertainty into the standard errors of the states. First are the Bayesian methods which generate distributions of the states and, therefore, incorporate the hyperparameter uncertainty naturally. Please refer to Carter and Kohn (1994) and Durbin and Koopman (2012). The drawback of this method is that it is computationally complicated in large models, as well dependent on the assumptions about the conditional distribution of the hyperparameters and states Quenneville and Singh (2000). The second approach, as proposed by Ansley and Kohn (1986) and Kass and Steffey (1989), is to use the asymptotic distribution or second order approximation of the hyperparameter estimator. However, the asymptotic distribution can be a poor approximation to the finite sample distribution when the sample size is not large enough and the second order approximation can be computationally demanding. The third approach by Quenneville and Singh (2000) reduces the biases in the prediction mean squared error (PMSE) by approximating the posterior distribution of the hyperparameters using a Bayesian method before computing the PMSE of the states using the Monte Carlo integration of the approximated distribution. However, this approach is only applicable to the local level model, and can be computationally demanding in more general contexts. The final method, which has been put forward by authors like Pfeffermann and Tiller (2005) and Rodriguez and Ruiz (2012) is to use bootstrap procedures to compute the PMSE together with the Kalman filter. This is the method used here since bootstrap procedures have the advantage over the first three methods of being computationally simple and are robust against misspecification of the error distribution. The method is also based on the Monte Carlo integration of the distribution of the hyperparameter, but approximate this distribution by a bootstrap distribution instead of an asymptotic or posterior distribution. Appendix B presents a summary of the approach[7].

3 Construction of PPPs for components – C, G and I

One of the objectives of this paper is to present our recent research into the extrapolation of PPPs for the aggregate components of GDP: Private Consumption (C), Government Expenditure (G), and Gross Capital Formation (I). In order to do the extrapolation, economic models and econometric method are needed. The econometric method chosen to estimate PPPs for GDP components is the RRD method described in the previous sections. This section is devoted to discussing the economic models for each component together with the data constraints, which altogether produce the resulting estimates of PPPs for each component.

There are no ready-made results on the structural determinants of price level for C, G and I. Therefore, we wish to bring in elements of the macroeconomic literature to define the economic models for C, I and G. Through these economic models, the groups of variables that are specific in explaining price level for each component will be identified below.

3.1 Private Consumption (C)

Compared to investment, government expenditure and net export, consumption is the largest components of GDP. On average, individual consumption constitutes 69 percent of GDP International Comparison Program (2005). These personal expenditures fall under one of the following categories: durable goods, non-durable goods, and services. Examples include food, rent, jewelry, gasoline, and medical expenses but does not include the purchase of new housing. Hence, consumption involves both tradables (goods) and nontradables (services) like GDP, therefore, the structural determinants of consumption price level should be similar to those of GDP. Again, there is no ready-made model in macroeconomics to account for price level of private consumption, but if we look into the theoretical reasonings of the structural determinants of national price level, we can see that these are also applied to consumption price levels.

[7]The results in this draft are still those based on the RRD method using equation (16)

First of all, consumption goods include both tradables and nontradables, therefore, the productivity differential model of Balassa is relevant for consumption price. Just like in the case of GDP, the law of one price holds for tradables so prices for traded goods are similar between countries, but prices of nontradable goods and services will be different due to different productivity levels in the tradable sector of countries. A rich country with high productivity level will pay higher wages to the tradable sector labour than poor countries whose productivity are lower. Even though international productivity differences are smaller for non-traded sector, the low wages established in poor countries in the low-productivity traded goods industries will apply also to the not-so-low productivity nontraded goods industries. The consequences will be rich countries having higher consumption price levels; or income per capita is also a structural determinant of private consumption price. Apart from per-capita income, other long-run structural factors that might also influence the consumption price levels are resource abundance, the degrees of openness, international tourism, country size, foreign trade ratios and trade balance. These judgements follow those in Rao et al. (2010b) as for price level in GDP level.

While the structural determinants are similar between consumption price level and GDP price level, the magnitude of influence of those determinants on the price levels might be different between the two. Under the expenditure approach, GDP is the sum of consumption, investment, government expenditure and net exports. Government spending and investment involves both tradables and nontradables, however, net exports only concern tradables. As a matter of fact, the proportion of nontradables in consumption will be larger than that in total GDP. By the productivity differential hypothesis, the positive correlation between consumption price level and per capita income will be higher than the correlation between national price level and income per person. With similar reasoning, resource abundance, international tourism, country size, the degrees of openness and trade balance also have stronger effects on consumption price than the national price level.

In conclusion, in constructing economic model to explain consumption price level, the set of variables, X_t, to be included in the regression component of RRD (see (8) and (11)) should include all groups of variables that explain the national price level.

3.2 Government Expenditure (G)

Government expenditure contains government consumption on final goods and services and gross government investment. Examples of government consumption spending includes salaries of public servants to produce and provide services to the public, such as public school education, health care, defense, justice, general administration, and the protection of the environment. Gross investment by the government consists of spending for fixed assets that directly benefit the public, such as highway construction, or that assist government agencies in their production activities, such as purchases of military hardware. It does not include any transfer payments, such as social security or unemployment benefits Burstein et al. (2004a). Therefore, government expenditure mainly consists of salary payments to government employees and purchase of tradable goods like machinery and equipments or military weapons.

From macroeconomic theory we know the salary payment or wage rates are determined by the marginal productivity of labour. As a result, labour high-income countries with high labour productivity will earn higher wage than their counterpart in low-income countries, which postulates a positive relationship between wage rates and national average income. The price of capital goods like equipments and military hardware are, on the other hand, seems to be negatively correlated with per capita income (as per discussion in the previous subsection). Therefore, the relationship between overall price level of government expenditure (which is the combination of wage rates and capital goods price level) with per capita income depends on the proportion of service (employment) and tradable goods purchased. It is also found that the volume of military spending is positively correlated with the national price level Bergstrand (1996), hence positively correlated with government expenditure price level.

In summary, an economic model explaining government expenditure price would ideally include variables explaining wage rates and capital-goods price; which are variables measuring labour productivity, average income, proportion of service and goods purchased by the government, volume of military spending and investment rates of the governments.

3.3 Gross Capital Formation (I)

Investment or Gross Fixed Capital Formation together with government expenditure and net exports only take up about a third of GDP, though there are exceptions like China. Investment measures expenditures, which mostly comprise purchases of equipment and construction services and distributions services (wholesaling, retailing, and transportation) are much less important for investment than for consumption (International Comparison Program (2005), Burstein et al. (2004b)). Examples include business investment in equipment, construction of a new mine, purchase of software, purchase of machinery and equipment for a factory or spending by households (not government) on new houses. One point to note is that investment in this context does not include exchanges of existing assets or purchases of financial products. Buying financial products is actually classified as 'saving', as opposed to investment.

From the two main categories of investment: equipment purchase and construction, it can be inferred that investment involves both tradable goods and nontradable services like GDP and consumption. However, while consumption contains in itself higher proportion of nontradables, investments mostly involves tradable capital goods as the import content of investment much larger than that of consumption Burstein et al. (2004b). It is agreed that services prices are lower in low-income countries, but it is controversial whether equipment or capital-goods prices are the same across nations.

Hsieh and Klenow (2007) claim that the absolute price of capital goods is no higher in poor countries than in rich countries. Their study, which uses data from the Penn World Tables, produces positive and mostly significant results suggesting, if anything, higher investment price in rich countries. The author explains that the high relative price of investment in poor countries is due to the low price of consumption goods in those countries since poor countries have low efficiency in producing investment goods and need to produce consumer goods to trade for them Hsieh and Klenow (2007). This result is exactly what is predicted by the Balassa-Samuelson hypothesis. On the contrary, common views and other empirical evidence seems to suggest the opposites. Alfaro and Ahmed (2010) use highly disaggregated data on trade in capital goods to study differences in the price of capital across countries and find that the price of imported capital goods is negatively and significantly correlated with the income of the importing country. This finding explains why in poor countries, the relative prices of capital to consumption goods are observed to be higher Alfaro and Ahmed (2010).

Several hypotheses have been proposed to explain why tradable capital goods are actually more expensive in poor countries. The first reason might be the measurement problems in the PWT and ICP price data set, especially in regards to developing countries. RRD also acknowledges this problem by incorporating measurement errors of the ICP into their econometric model, assuming that the variance of errors are inversely related income per capita (see equation (5) and footnote (2). The second possible reason is price discrimination, which means producers set their selling prices of the same goods higher for poorer countries. Price discrimination has long been present in the literature Mertens and Ginsburgh (1985), Verboven (1996), Ayres and Siegelman (1995), which speculate that it might be profitable for firms to charge higher prices to groups of consumers that have a lower average reservation price if the variance of reservation prices within the group is sufficiently large. Within the context of traded capital goods, a vendor that knows this might rationally charge higher prices to all of its customers in poor countries. The third possible reason is transaction costs. For many developing countries, high tariffs or other form of capital control would likely drive up the price of imported capital goods. Besides, higher costs for poor countries are associated with searching for and negotiating (directly or indirectly) foreign purchases, as well as the volume of trade. Low-income countries might also be paying more for capital goods shipped in smaller quantities. Alfaro and Ahmed (2010).

Other factors beside income that are documented to affect capital goods prices are investment rates or growth Alfaro and Ahmed (2010). For example, in a research using a data set for capital-goods and equipment prices covering the 1870–1950 period for 11 OECD countries, the authors have argued that relative capital-goods prices are strongly negatively correlated with investment rates (Collins and Williamson (2001)).

From the discussion above, there are several groups of variables that should ideally be included in the economic model explaining investment price. These are variables that measure the proportion of equipment purchase (tradable capital-goods) to construction service (non-tradables), income per capita, transaction costs (e.g. capital control, volume of trade), investment rates and growth.

3.4 Data constraint and choices of variables

Data for the PPP extrapolation of C, G and I are the ICP benchmark PPPs for the components, the socio-economic data for each country and the bilateral trade data required to compute the spatial weights matrix (see equation (9)).

The benchmark PPP data for Consumption, Government expenditure and Investment were collected from two different sources for the 11 benchmarks. For 1970, 1973, 1975, 1980, 1985 and 2005, benchmark PPP data for the components were collected from ICP and the remaining years of 1990, 1993, 1996, 1999, 2002 were obtained from Eurostat-OECD. Several features of the PPP data are noteworthy. The number of countries vary over benchmarks. The first benchmark (1970) covered only 13 countries, while the most recent (2005) benchmark represents truly global comparisons with 146 countries. Another related point worth noting is the fact that PPPs for all the benchmarks prior to 1990 were based on the GK method and PPPs for the more recent years are all based on the GEKS method of aggregation.

The socio-economic data and the already computed spatial weight matrix are both obtained from the UQICD database. In this database there are socio-economic variables, variables representing productivity level, the degree of openness of the economy, national resource, trade balance, currency and trade agreements.

The spatial weight matrix W_t (in equation (9)) used in modeling the spatial error structure is proportional to trade closeness as measured by bilateral trade flows (see Rambaldi et al. (2010)).

The dimensions of the extrapolation were largely determined by data availability. A number of countries were excluded because of missing data and the time frame 1970-2010 was likewise chosen because of poor data availability prior to 1970. As a result, the complete PPP panels for C, G and I will be for 181 countries and 40 periods (year 1970 data are used for computation of growth rates so results are only for 1971 to 2010).

Socio-Economic Variables (forming $\boldsymbol{x}_{it}$ in equation (8)) included the regression are chosen based on the determinants of the national price level as well as the structural economic determinants of price level for each component discussed above, and, the availability of our data. Details of the variables chosen for each component will be discussed in the next subsections.

Explanatory variables for private consumption

To construct the economic model to explain consumption price level, the set of variables should include all the variables that explain the national price level. They are per-capita income, national resource, the degrees of openness, international tourism, country size, foreign trade ratios and trade balance. Per-capita income exchange rate adjusted for each country is used to construct the matrix V_t (see equation (6)). Per-capita income cannot be used directly as an explanatory variable since there will be an endogeneity problem. To overcome this difficulty, variables representing productivity, which are in accordance with the productivity differential model of Balassa in explaining national price level and are proxies for income per-capita are chosen.

The procedure to select variables to explain C, G and I price levels are similar. First, given available data and the theoretical structural determinants discussed above, the largest possible set of variables are chosen for each component. Then, subsets of these variables are selected by statistical fittings in order to maximize the adjusted R-square in the initial run using 491 benchmark observations to obtain an initial estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^0$ by regressing r_{it} on $\boldsymbol{x}'_{it}$. Once the regression is calibrated, a first set of predictions of PPPs is obtained to start the state-space based estimation (see equation (10)). Estimates of these initial regressions are presented in Appendix C.

A set of 24 variables that are expected to capture country-specific episodes that may influence the price level, variables that capture trade or monetary agreements, variables representing productivity, national resource, degree of openness and trade balance have been selected. The model is specified with time fixed effects. The initial regression (using available benchmark data) produces an adjusted R-square of 72.14%.

Explanatory variables for Government expenditure

The same procedure of variable selection for Private Consumption is used for Government Expenditure. First, the theoretical discussion by Bergstrand (1996) suggests that government expenditure price would ideally include variables explaining wage rates and capital-goods price; which are variables measuring labour productivity, average income, proportion of service and goods purchased by the government, volume of military spending and investment rates of the governments. However, we did not have data on the share of services to government expenditure, or the volume of military spending and investment rates. As a result, a set of variables representing productivity and average income are selected together with economic variables which include measures of trade balance and degree of openness. The initial model with the highest adjusted R-square of 79.79%. See Appendix C.

Explanatory variables for Gross Capital Formation

Among the three components, Gross Capital Formation is the most difficult one to model given our dataset at this stage. From a theoretical perspective, the group of variables that should ideally be included in the economic model explaining investment price are: variables that measure the proportion of equipment purchase (tradable capital-goods) to contraction services (non-tradables), income per capita, transaction costs (e.g. capital control, volume of trade), investment rates and growth. Given available data, a set variables which represent income per capita, transaction cost (capital control), and trade volume are used. The adjusted R-square for initial regression is only 60.56%, which is lowest among the three.

The benchmark PPPs of Gross Capital Formation is found highly correlated with market exchange rate (with correlation coefficient of 0.95). This reflect the fact that investment goods are mostly tradables. However, we cannot use exchange rate as an explanatory variable given it is in the denominator of the dependent variable. Hopefully the explanatory power of the regression will be improved when we can include variables that measure the tradables-to-nontradables ratio in investment, investment rates and growth.

3.5 Estimation results

For a discussion of the estimation results, we have chosen a set of countries that represent both developed (the UK, the Netherlands) and developing countries (South Africa, Brazil, Mexico, India, Kenya and China) for all the three components. With each graph, our estimation of the Price levels (PPP/ER) for each country in the 40 years period from 1970 to 2010 are presented together with their standard errors and compared against the corresponding estimates from PWT 7.1. While there are some common features across the results for Private Consumption, Government Expenditure and Gross Investment; there are also different points among them that are worth noting.

The first common feature among the estimated figures for the three components is that our estimates track the ICP benchmark closer than the PWT 7.1. Also, in comparison with PWT 7.1, though our predictions are different, they mostly follow the same trend across the 40 years period. From the graphs, we can also see that our estimates and those produced by PWT 7.1 are generally closer in developed countries like the UK, the Netherlands, than in developing countries like Kenya; and generally closer in the second half of the time periods (1990-2010) than in the first half (1970-1990). The second common feature across the results is in the standard errors (SE) of the estimates. SE are generally smaller close to the benchmark; smaller towards the end of the estimation period and smaller in developed than developing countries. These facts about the SE might reflect the quality and availability of data since with the more recent benchmarks, we have more data from the ICP with their improvements in benchmark PPPs computation.

Between the three components, our estimates for Private Consumption are closest to those from the PWT 7.1. Most of the time, our estimates for Investment are higher while those for Government Expenditure are lower than theirs. Standard errors for Gross Investment seem to be largest among the three, the reason might lie in the fact that some variables like tradables-to-nontradables ratio in investment, investment rates and growth haven't been included in the regression as suggested by economic theory, due to data constraints.

[Figures 1-12 here]

4 Aggregation of components

The econometric approach to extrapolation of PPPs for the components generates panels of PPPs for Individual Consumption (C), Investment (I) and Government Expenditure (G). Then domestic absorption, DA, and gross domestic product, GDP, are given by:

$$DA = C + G + I; \text{ and } GDP = C + G + I + (X - M)$$

where X and M denote exports and imports.

The PPPs for C, I and G form the price data and expenditure data from national accounts are the source of weights for aggregation. let p_{ij} and e_{ij} represent respectively the PPP and expenditure in national currency units for aggregate i ($= 1,2,3$ or C,G,I) and country j $(=1,2,\ldots,M)$. We can define implicit quantity as: $q_{ij} = e_{ij}/p_{ij}$. These price, expenditure and quantity data can be aggregated to generate PPPs for GDP.

Two aggregation methods are considered. The first is the Gini-Elteto-Koves-Szulc (GEKS) method and the second is the Geary-Khamis (G-K) method. Diewert (2013) provides a description of these two methods and their relative merits. The GEKS method is used here as it is the recommended aggregation method for the ICP and it is known to be relative free from Grechenkron effect. The G-K method is also used as it is the aggregation method used in all the versions of PWT including PWT 8.0. The G-K method possesses additivity property which is useful in considering national accounts in real terms.

4.1 The GEKS Method

The GEKS method provides transitive PPPs from the matrix of binary comparisons between all pairs of countries obtained using the Fisher binary index or any other suitable binary index satisfying country-reversal test. The PPP for the currency of country k with country j as base is given by:

$$PPP_{jk} = \prod_{l=1}^{M} \left[F_{jl} \times F_{lk}\right]^{1/M} \text{ where } F_{jl} = \left[\frac{\sum_{i=1}^{N} p_{il}q_{ij}}{\sum_{i=1}^{N} p_{il}q_{ij}} \times \frac{\sum_{i=1}^{N} p_{il}q_{il}}{\sum_{i=1}^{N} p_{ij}q_{il}}\right]$$

The GEKS PPPs are transitive and base invariant.

4.2 The Geary-Khamis Method

The Geary-Khamis method due to Geary (1958) and Khamis (1972) provides PPPs from a system of simultaneous equations that relate PPPs to international average prices of commodities. Let denote the international average price of commodity i (=*1,2* and *3*). Then the GK system is defined through the following equations.

$$P_i = \frac{\sum_{j=1}^{M} p_{ij} q_{ij} / PPP_j}{\sum_{j=1}^{M} q_{ij}} \; i = 1, 2, \ldots N$$

$$PPP_j = \frac{\sum_{j=1}^{N} p_{ij} q_{ij}}{\sum_{j=1}^{N} P_i q_{ij}} \; i = 1, 2, \ldots M$$

The G-K PPPs are computed by solving this system of equations iteratively. Khamis (1972) shows the existence and uniqueness of solutions to this system of equations.

The G-K method was the main aggregation method for the ICP in the early rounds from 1970 to 1985 and has been replaced by GEKS in the 1993, 2005 and 2011 rounds of the ICP. The G-K method produces additively consistent international comparisons. However, it is known to exhibit Grechenkron effect and tend to overstate the real expenditures of low income countries. The bias induced by G-K method was discussed by Dowrick and Akmal (2005).

4.3 Domestic absorption estimates

Using the methods described, we have been able to produce estimates of PPPs for domestic absorption (DA). Examples of results are shown in the plots below. For each countries, five series are graphed together: the DA PPPs computed using GK method (GK), the DA PPPs aggregated using GEKS method (GEKS), the PPPs for GDP produced by RRD method (UQICD GDP) (see next section), PPPs for GDP from PWT7.1 (PWT7.1) and the ICP benchmark(PPP-ICP). All are expressed in price levels.

As can be seen, all the series follow the same trends in both developed and developing countries. The GK and the GEKS are very close to each other and the UQICD GDP series.

[Figures 13-16 here]

5 UQICD Version 2.0

The UQICD is a database generated by the project team. The members include D.S. P Rao, A. N. Rambaldi and H.E. Doran as principal researchers, L.T. Hyunh as PhD student (working on the components estimation and bootstrap standard errors), K. R. Ganegodage as database manager, and L. Brough as website designer. The database is available to researchers via a dedicated website `http://uqicd.economics.uq.edu.au`.

UQICD Version 2.0 will be providing complete panels of PPPs for GDP levels at current and constant prices (the methodology for the construction of PPPs at constant prices is available from Rao and Rambadi (2013) and Huynh et al. (2014)). The panel provides data for 181 countries and the period 1970 - 2012. In addition, PPPs for the components of GDP are available for 181 countries and the period 1970-2010 based on the work presented in Section 4.

5.1 The Addition of the data from 2011 ICP

The ICP has recently released the 2011 benchmark results for GDP level. We have incorporated these results and constructed a panel for 181 countries using the RRD method. We present results for the selected group of countries with specific reference to the expected "alignment" between the last two consecutive rounds, 2005 and 2011 ICP rounds, measured by applying each country's deflator movement with respect to that of the reference country.

We start the discussion by recalling that in RRD the relative deflators movement is used as the law of motion of the true log of PPPs; however the ICP benchmarks, regression predictions and deflators are assumed to be measured with error and these measurement error uncertainty are combined to weight the observed ICP information (please see Appendix A). As a result, the predicted PPP for a benchmark year might be different from that implied applying only the movement in the GDP deflators. This is a major advantage of the RRD method as it combines information from all the sources.

Figures 17-20 present a number of estimates for four pairs of countries, two European (UK and Netherlands), two Latin American (Mexico and Brazil), two African (Kenya and South Africa) and two Asian (China and India). The estimates provided are the RRD based estimates for the period 1970 to 2012 of: a) the price level and two standard errors bound when all available benchmark information is used (ICP from 1970 to 2011 as well as OECD benchmarks). These are labelled **AllBench** and **2SE(AllBench)** (these are in black); b) the price level and two standard errors bound when the recent 2011 ICP round information is not used. These are labelled **Predict2011** and **2SE(Predict2011)** (these are in blue); c) the price level when the 2005 ICP round data are ignored but all other ICP and OECD information is used. This is labelled **Ignore2005** (this is red). The first conclusion to draw is that there is not a single consistent pattern about the "alignment" of the 2005 and 2011 ICP rounds.

Consider European countries first, Figure 17. For the UK we find that the ICP 2011 PPP was above the predicted value, but at the upper end of the **2SE(Predict2011)**. When using 2011 information but ignoring the 2005, the predicted 2005 value is below the actual value (see **Ignore2005** estimate). In the case of the Netherlands, the **Predict2011**, **AllBench** are identical to the realised value of ICP 2011. The **Ignore2005** series shows the predicted 2005 value slightly below the realised.

Consider Mexico and Brazil depicted in Figure 18 next. Mexico shows perfect alignment of both rounds and the **SE(Predict2011**) bound to be slightly wider than **2SE(AllBench)** as one would expect. Brazil on the other hand show the **Predict2011** predicted a higher value for 2011 than that realised. The realised value is at the lower bound of the **SE(Predict2011)** two standard error bound. The **Ignore2005** series indicates the prediction for 2005 is lower than the realised ICP 2005 value.

The African countries are presented in Figure 19. Kenya's result is similar to that of Brazil in that **Predict2011** is higher than the realised value and **Ignore2005** is lower than the realised value. However, the deviation from the 2011 ICP realised value is much smaller than that of Brazil and the **2SE(Predic2011)** much wider. In the case of South Africa the 2005 ICP realised value is consistent with the prediction and the prediction of 2011 is marginally above the realised value. Like Kenya, the **2SE(Predict2011)** two standard error bound is very wide.

The two largest Asian economies, China and India are presented in Figure 20. In the case of China, the predicted values for 2005 (when using, 2011 or both 2005 and 2011) are identical to the ICP realised value. The 2011 prediction is higher than the realised value; however within the two standard error bound. The realised value is on the lower bound. The two standard error bound for **Predict2011** is relatively wide. However, we note that the misalignment of the 2011 realised value is much smaller than that of the case of Brazil. For India, the story is similar, in that the 2011 prediction is higher than that realised value which is close to the two standard error lower bound. Also, the 2005 prediction, when incorporating 2011 information, is slightly below the realised value. However, similar to China, the difference is much less than that observed for Brazil.

[Figures 17-20 here]

5.2 The New Countries

A new addition to UQICD is 22 countries that were newly formed after the break up of the Soviet block. The estimated series for the Russian Federation and Serbia are presented in Figure 21. For Russia, the 2008 OECD and 2011 ICP seem to indicate price levels are just below 0.6, while the price level was around 0.5 in 2005. For Serbia the 2005 value is just above 0.4 and the 2011 value is 0.5. Thus, the trend in both countries is upwards indicating the purchasing power parity is slowly converging to the market exchange rate.

[Figure 21 here]

6 Conclusions

In this paper we have presented an overview of the results from the second stage of our project on the construction of consistent panels of PPPs for GDP and its components. Structural models to explain the difference in price levels for the consumption (C), investment (I) and government (G) components of GDP are developed and tested. The extrapolation method developed for PPPs at the GDP level is implemented for the components. A new method for the computation of standard errors is proposed. The extrapolated PPPs for the components C, G and I are then aggregated used GEKS and Geary-Khamis methods to produce panels of PPPs for domestic absorption (DA). The paper makes use of the recently released PPP data from ICP 2011 benchmark in compiling extrapolated PPPs for 181 countries and the period 1970 to 2012. Recognising the commentary on the apparent inconsistency between extrapolations from 2005 to 2011 and the 2011 benchmark results, the paper presents experimental computations and extrapolations under three scenarios: (i) include both 2005 and 2011 benchmarks; (ii) include only 2005 benchmark; and (iii) include only 2011 benchmark and a comparison of the extrapolations are presented for selected countries. These results do not suggest any consistent patterns. We conclude the most reasonable approach is to include all the currently available data in the process of extrapolation using the RRD method which provides a consistent time-space extrapolation by combining individual countries' ICP benchmarks information and GDP deflator movements with time and cross-sectionally weighted information from national accounts.

References

Ahmad, S. (1980). Approaches to purchasing power parity and real product comparison using shortcuts and reduced information. Number 418. World Bank Staff Working Paper.

Ahmad, S. (1988). International real income comparison. In *World Comparisons of Incomes, Prices and Product.*

Ahmad, S. (1996). *Regression Estimates of per Capita GDP Based on Purchasing Power Parities*, volume International Comparisons of Prices, Output and Productivity: Contributions to Economic Analysis Series, chapter II, pages 237–264. Elsevier.

Alfaro, L. and Ahmed, F. Z. (2010). The price of capital: Evidence from trade data. Technical report, SSRN eLibrary.

Ansley, C. F. and Kohn, R. (1986). Prediction mean squared error for state space models with estimated parameters. *Biometrika*, 73(2):467–473.

Ayres, I. and Siegelman, P. (1995). Race and gender discrimination in bargaining for a new car. *The American Economic Review*, 85(3):304–321.

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

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

Burstein, A., Neves, J. C., and Rebelo, S. (2004a). Investment prices and exchange rates: Some basic facts. National Bureau of Economic Research. NBER Working Papers 10238.

Burstein, A., Neves, J. C., and Rebelo, S. (2004b). Investment prices and exchange rates: Some basic facts. NBER Working Papers 10238, National Bureau of Economic Research, Inc.

Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):pp.541–553.

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

Collins, W. J. and Williamson, J. G. (2001). Capital-goods prices and investment, 1870-1950. *The Journal of Economic History*, 61:59–94.

Diewert, W. E. (2013). *Methods of Aggregation above the Basic Heading Level within Regions*, volume Measuring the Real Size of the World Economy: The Framework, Methodology, and Results of the International Comparison Program-ICP, chapter 5, pages 121–167. World Bank.

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

Dowrick, S. and Akmal, M. (2005). Contradictory trends in global income inequality: A tale of two biases. *The Review of Income and Wealth*, 51(2):201–229.

Durbin, J. and Koopman, S. (2000). Time series analysis of non-gaussian observations based on state space models from both classical and bayesian perspectives. *Journal of the Royal Statistical Society Series B*, 62:3–56.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods*. Oxford Statistical Science Series. Oxford University Press, 2nd edition.

Geary, R. (1958). A note on the comparison of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society. Series A*, 121(Part I):97–99.

Hamilton, J. (1986). A standard error for the estimated state vector of a state-space model. *Journal of Econometrics*, 33:387–397.

Heston, A., Summers, R., and Aten, B. (2012). Penn World Table version 7.1.

Hsieh, C. T. and Klenow, P. J. (2007). Relative prices and relative prosperity. *American Economic Review*, 97(3):562–585.

Huynh, L. T., Rambaldi, A. N., and Rao, D. P. (2014). A new approach to inter-country comparisons of real income and expenditures: The UQICD panel of PPPs. In *Workshop on Inter-Country and Intra-Country Comparisons of Prices and Standards of Living*.

International Comparison Program (2005). Global purchasing power parities and real expenditures.

Khamis, S. (1972). A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society. Series A*, 135(Part I):96–121.

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

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

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

Mertens, Y. and Ginsburgh, V. (1985). Product differentiation and price discrimination in the european community: The case of automobiles. *Journal of Industrial Economics*, 34(2):151–166.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap approximation to prediction MSE for state-space models with estimated parameters. *Journal of Time Series Analysis*, 26:893–916.

Quenneville, B. and Singh, A. (2000). Bayesian prediction mean squared error for state space models with estimated parameters. *Journal of Time Series Analysis*, 21:219–236.

Rambaldi, A., Rao, D., and Ganegodage, K. (2010). *Modelling Spatially Correlated Error Structures in the Time-Space Extrapolation of Purchasing Power Parities*, volume Price Index Numbers in Time and Space of *Contributions to Statistics*, chapter Part II, pages 63–96. Springer-Verlag, Berlin, Germany.

Rao, D., Rambaldi, A., and Doran, H. (2010a). An econometric approach to the construction of complete panels of purchasing power parities: analytical properties and empirical results. In *The Econometric Society World Congress, 17-21 August, 2010. Shanghai.*

Rao, D. P., Rambaldi, A. N., and Doran, H. E. (2010b). Extrapolation of purchasing power parities using multiple benchmarks and auxiliary information: A new approach. *Review of Income and Wealth*, 56(Suppl 1):S59–S98.

Rao, D. S. P. and Rambadi, A. N. (2013). A state-space and constrained GEKS approach to the construction of panels of real incomes at current and constant prices. In *International Comparisons Conference (PPP/ICP). Princeton University.*

Rodriguez, A. and Ruiz, E. (2012). Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters. *Computational Statistics & Data Analysis*, 56:62–74.

Verboven, F. (1996). International price discrimination in the european car market. *RAND Journal of Economics*, 27(2):240–286.

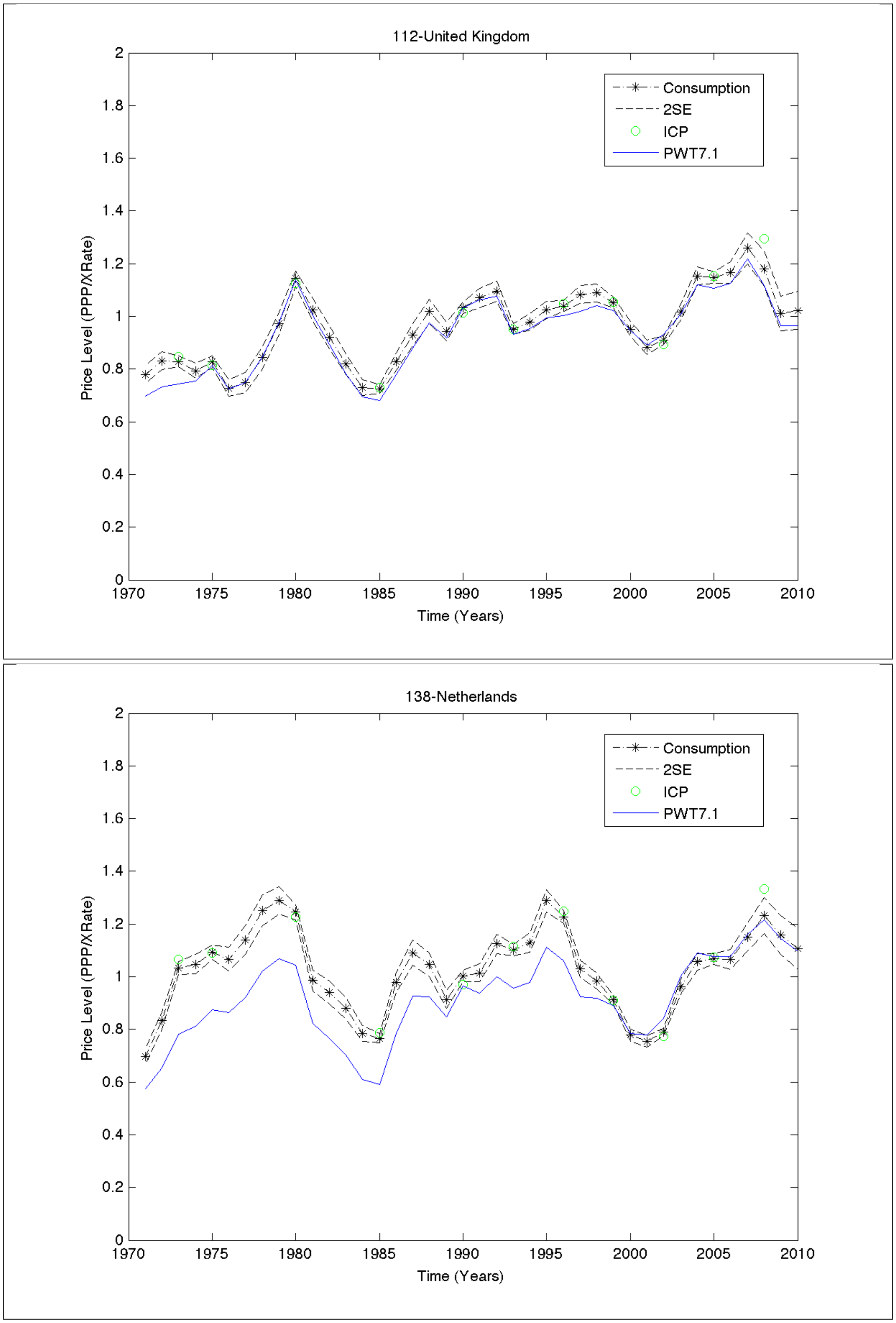


Figure 1: Private Consumption Price Level (PPP/ER) - European Countries

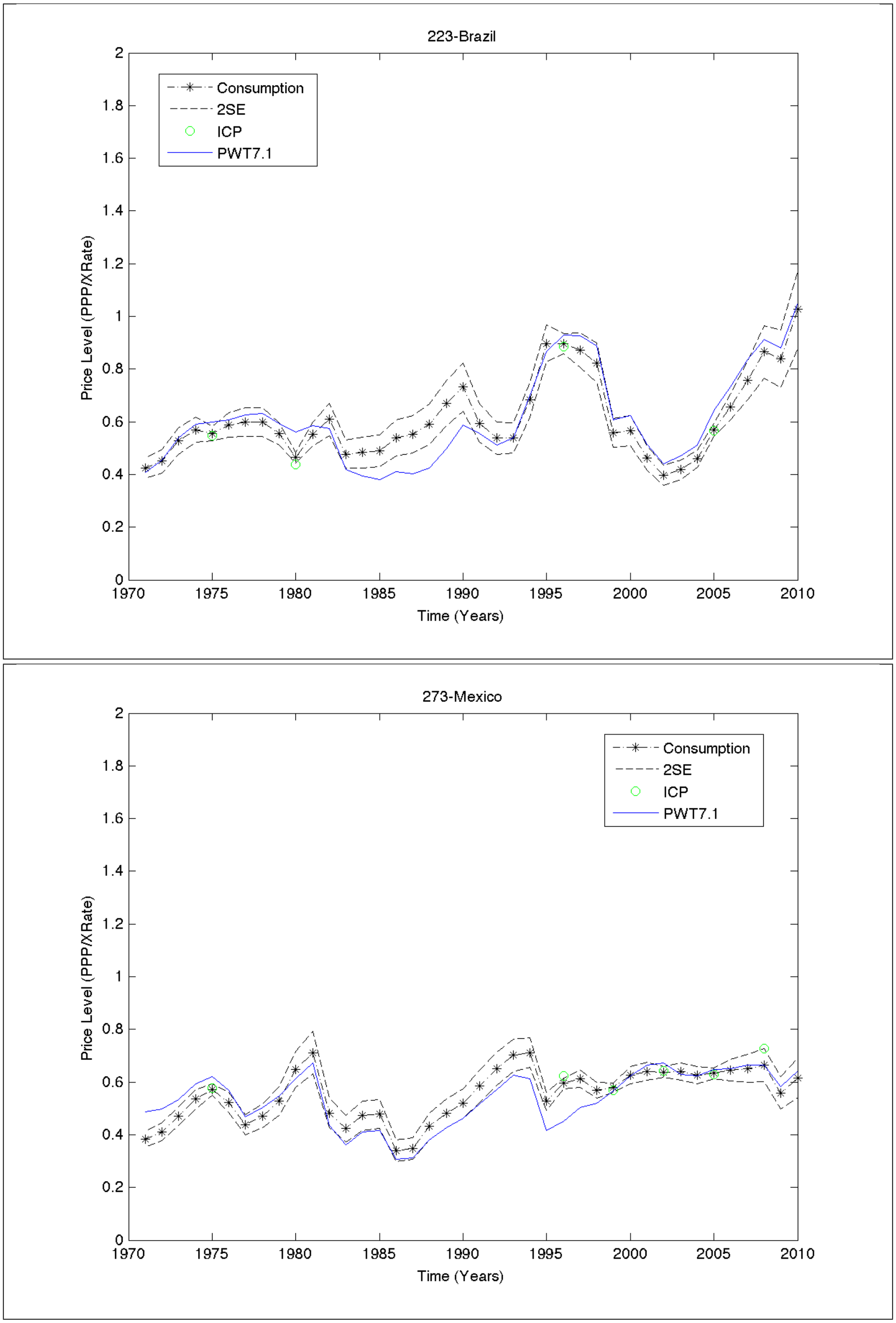
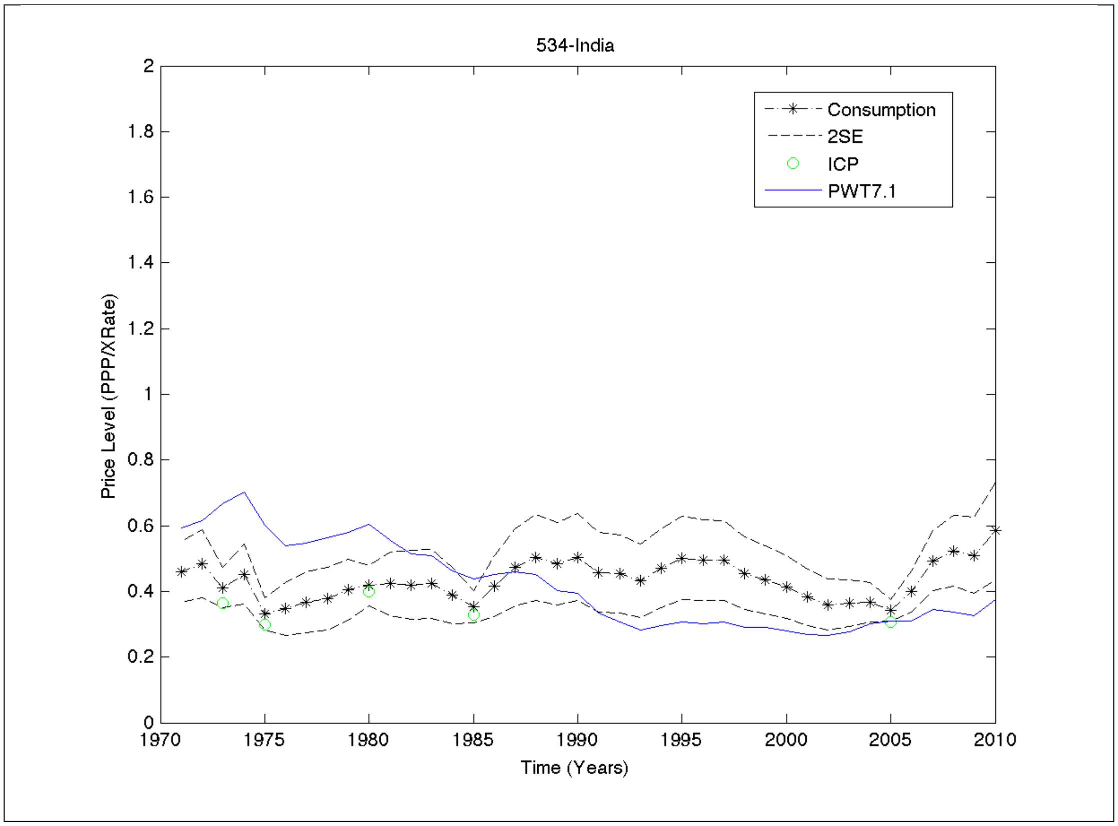


Figure 2: Private Consumption Price Level (PPP/ER) - LA Countries



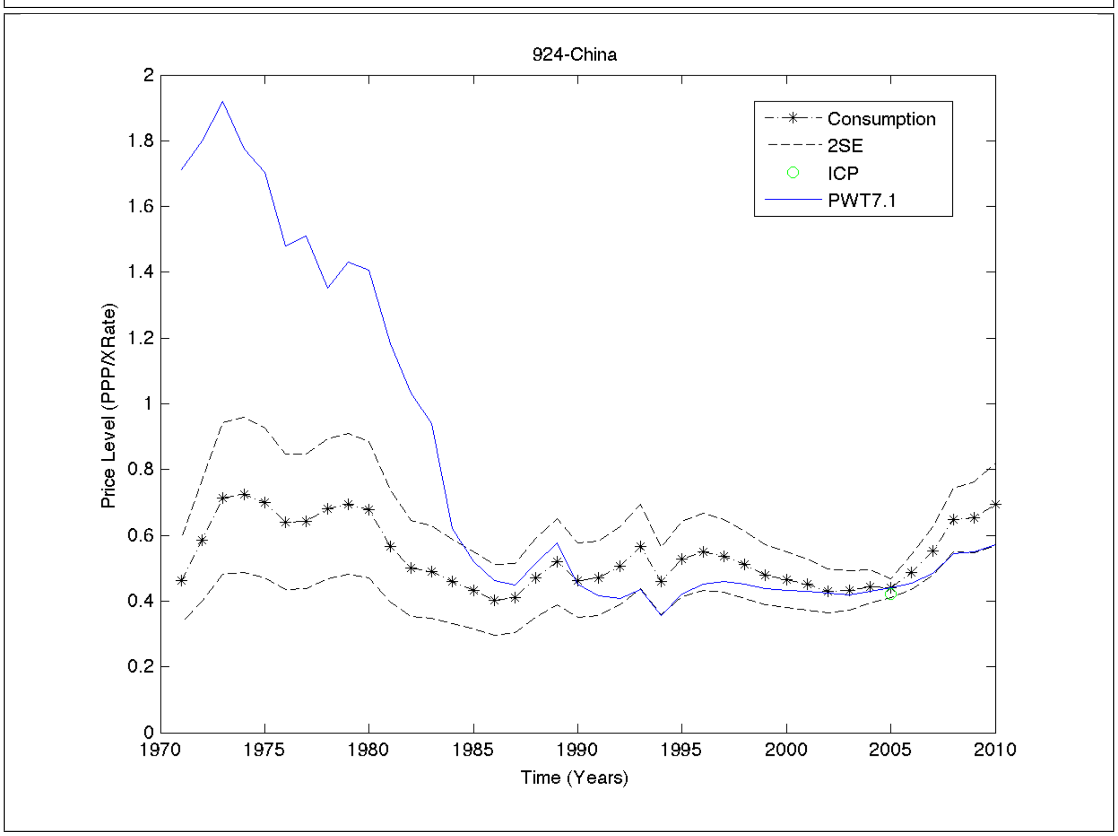


Figure 3: Private Consumption Price Level (PPP/ER) for Asian Countries

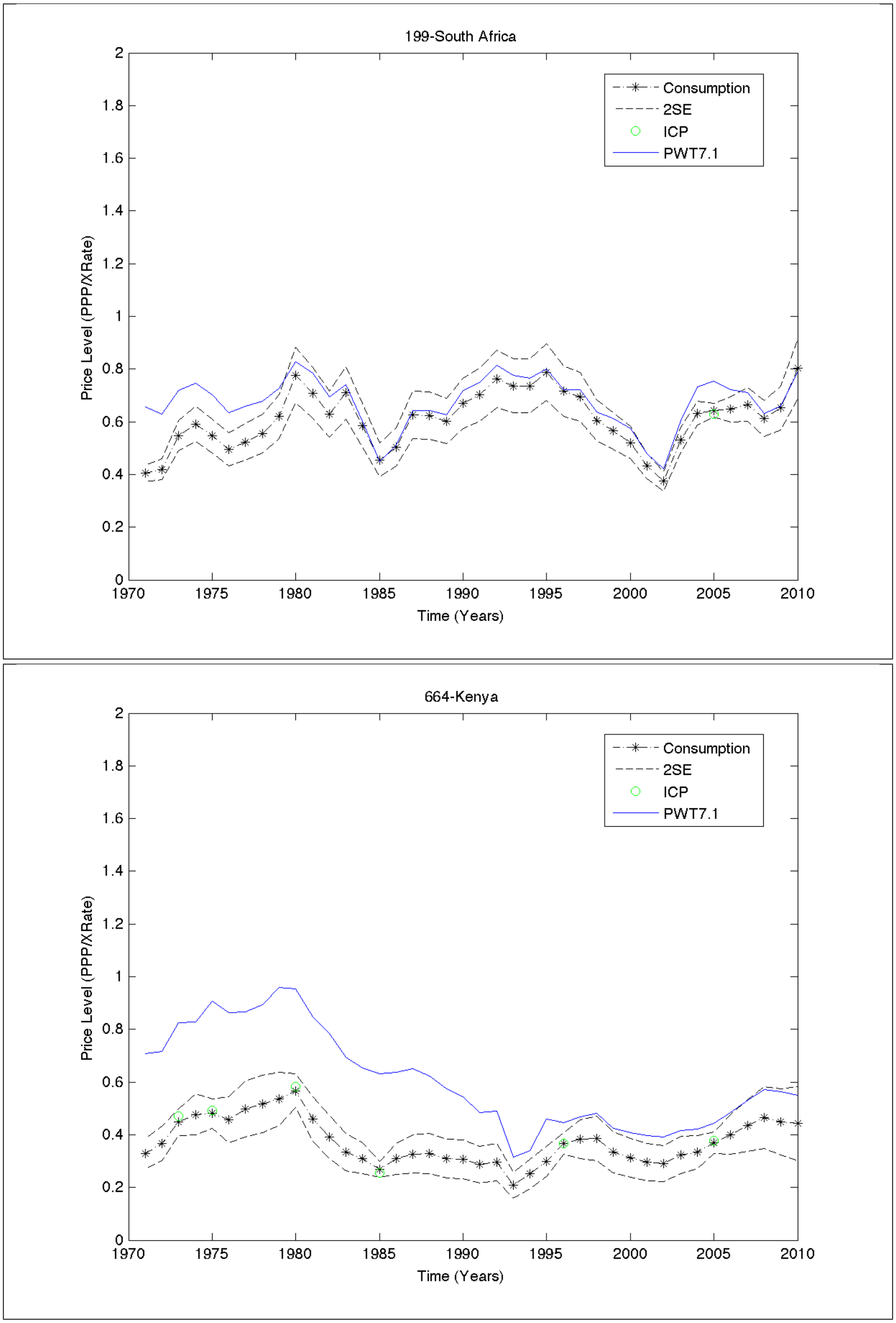
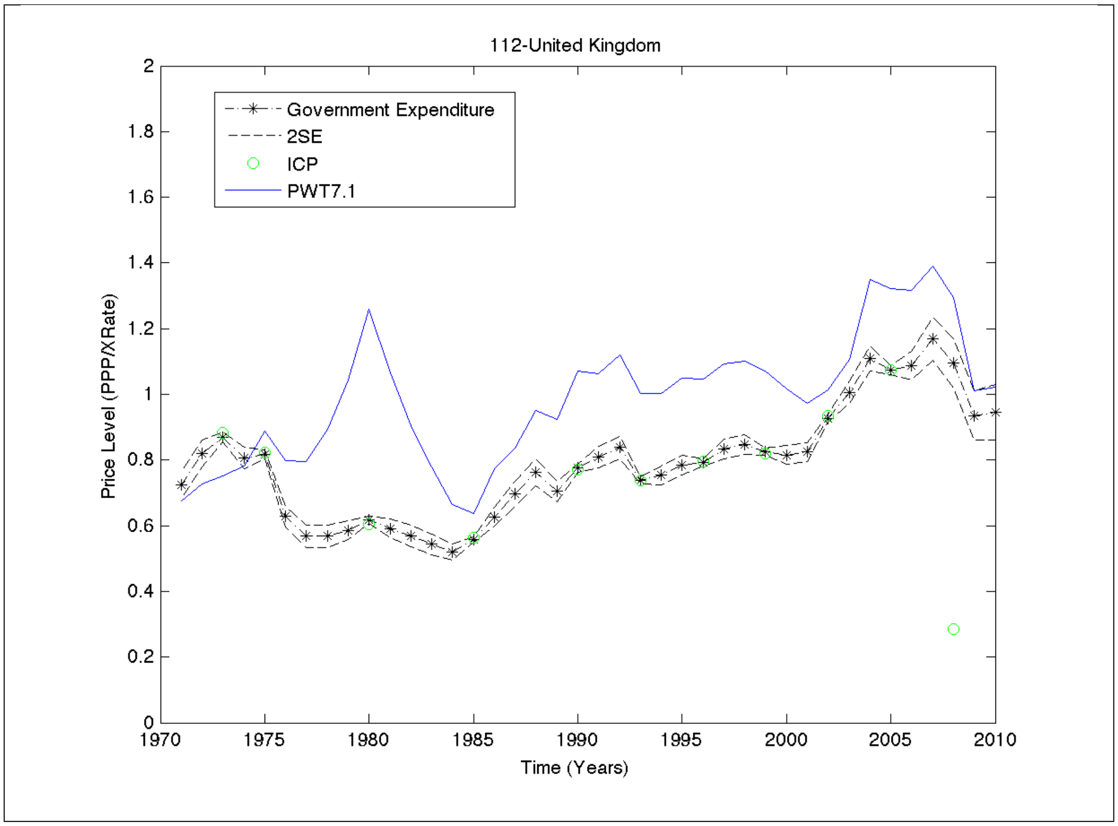


Figure 4: Private Consumption Price Level (PPP/ER) for African Countries



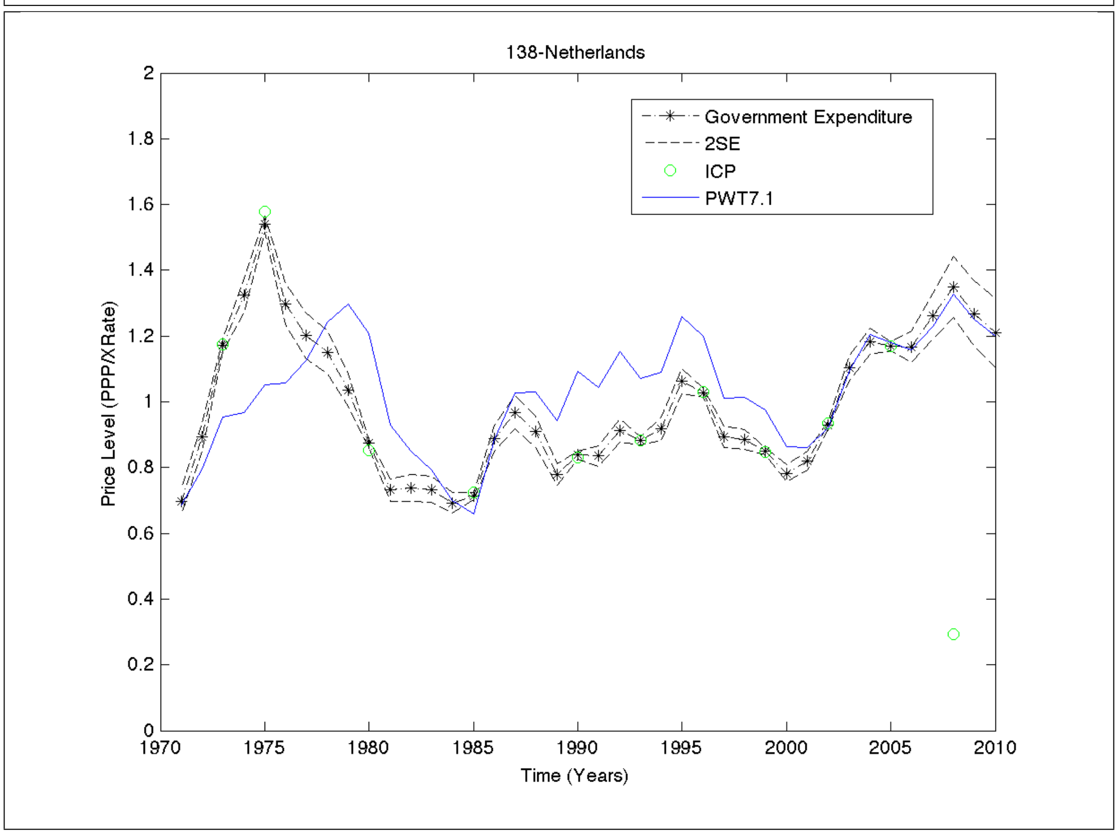
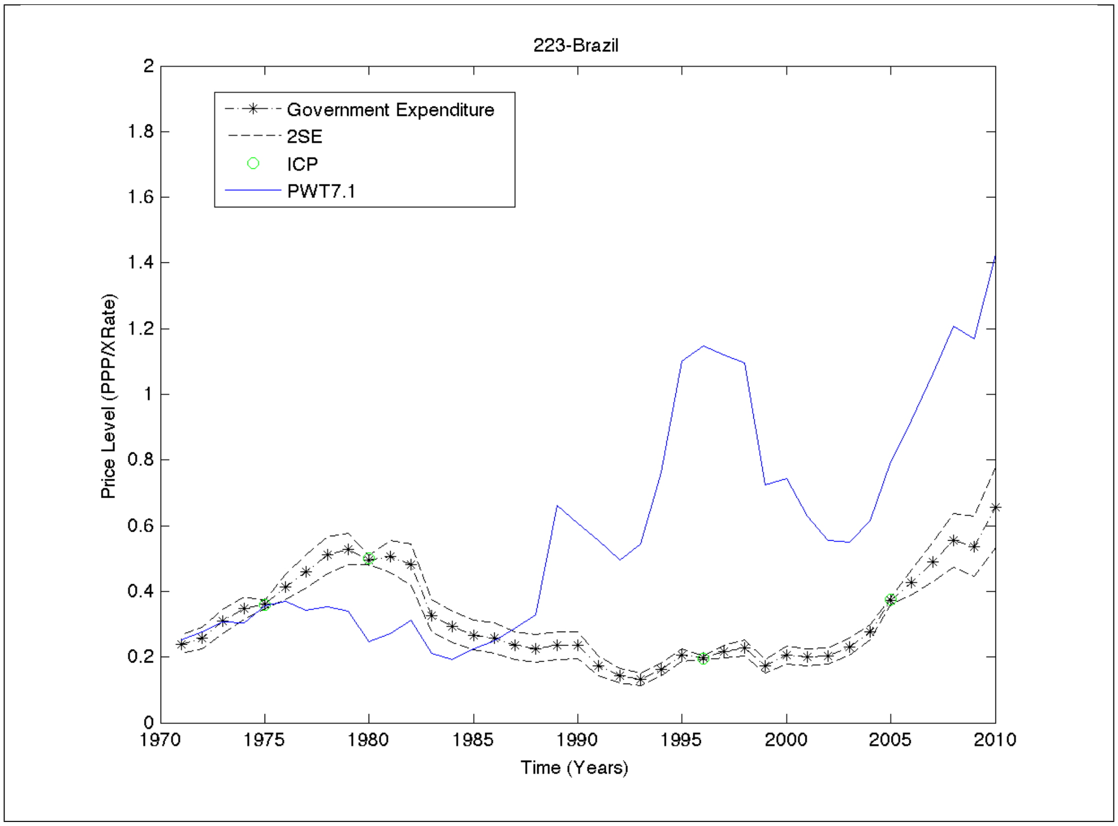


Figure 5: Government Expenditure Price Level (PPP/ER) - European Countries



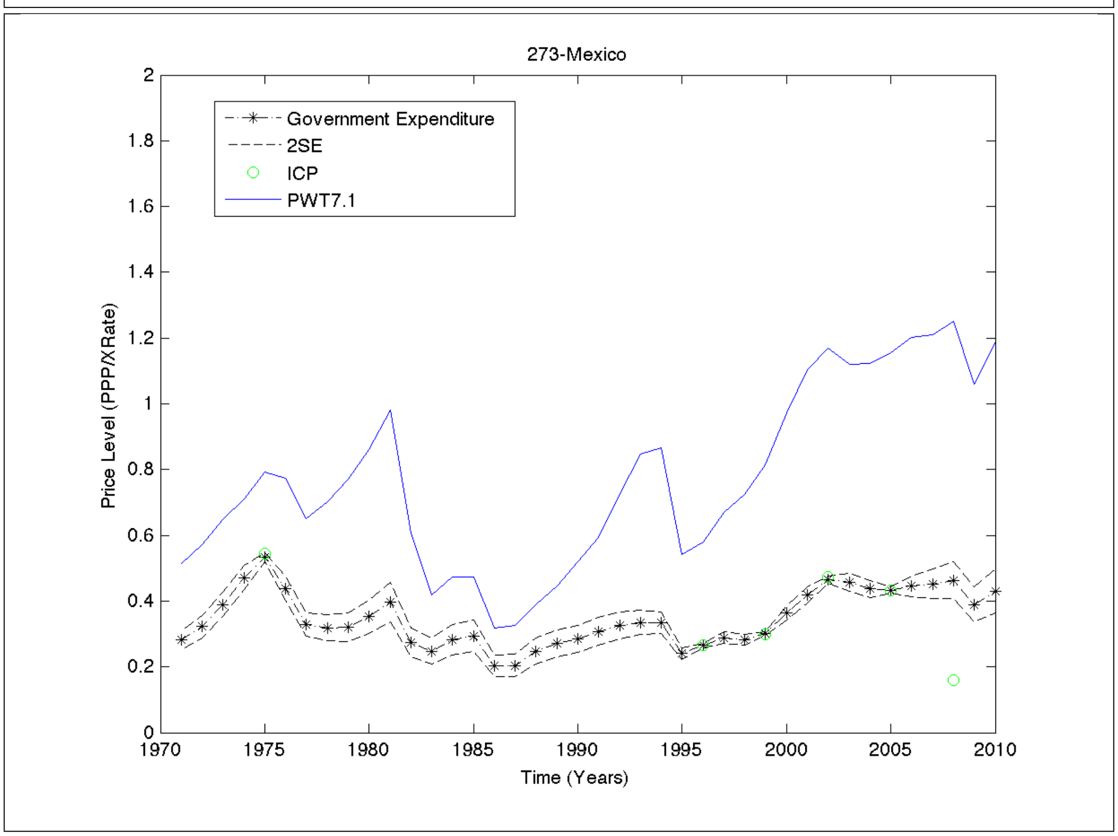


Figure 6: Government ExpenditurePrice Level (PPP/ER) - LA Countries

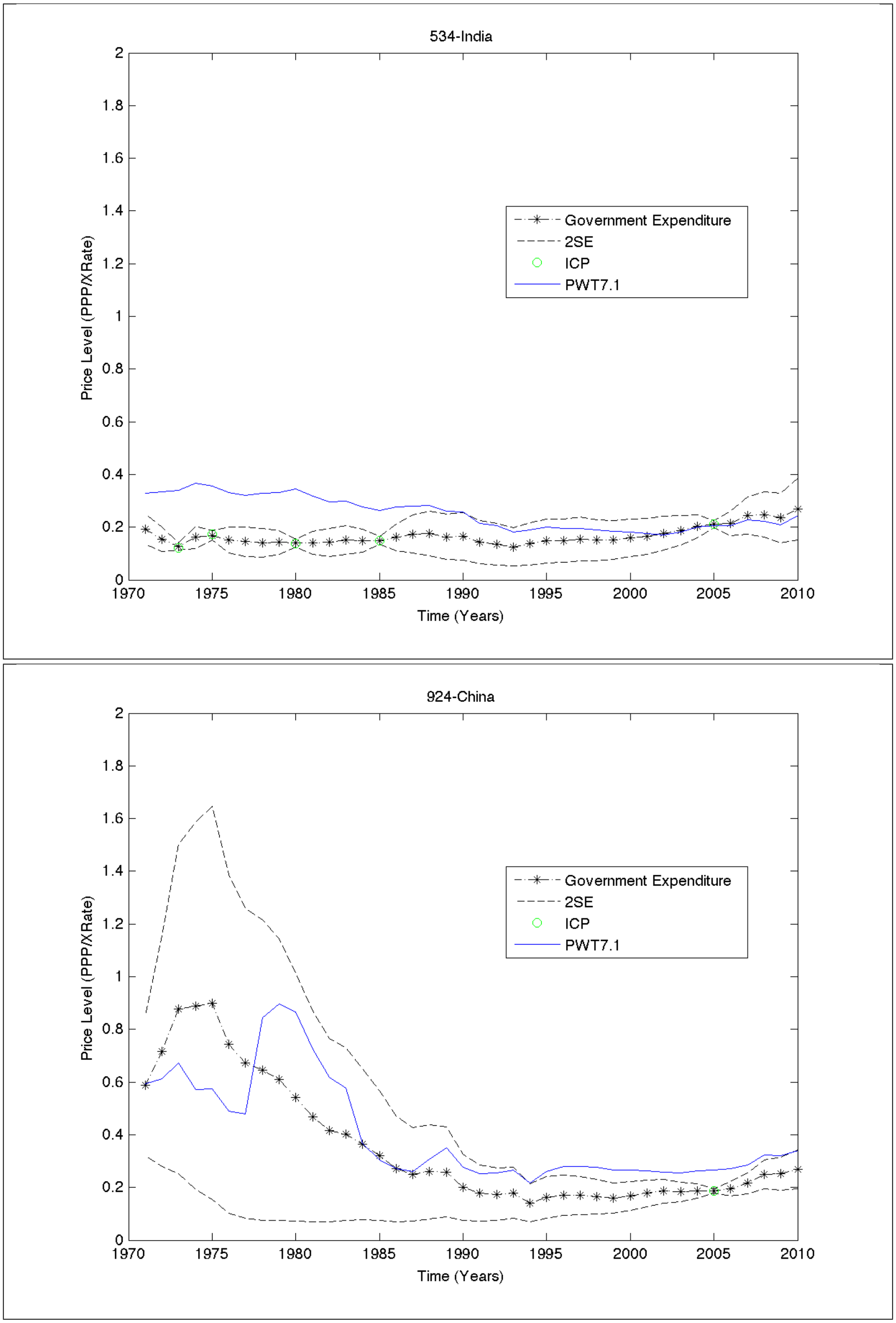


Figure 7: Government Expenditure Price Level (PPP/ER) for Asian Countries

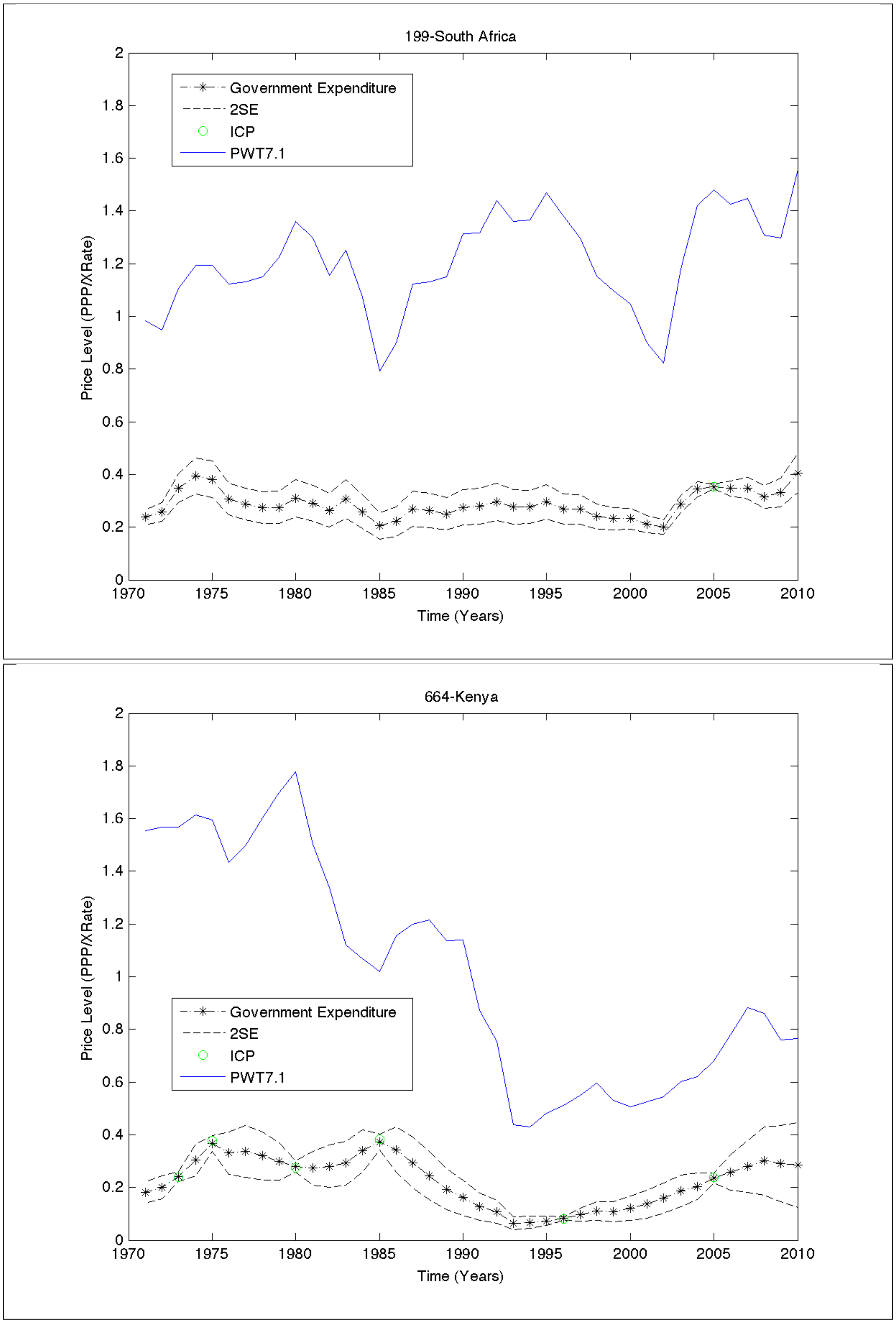
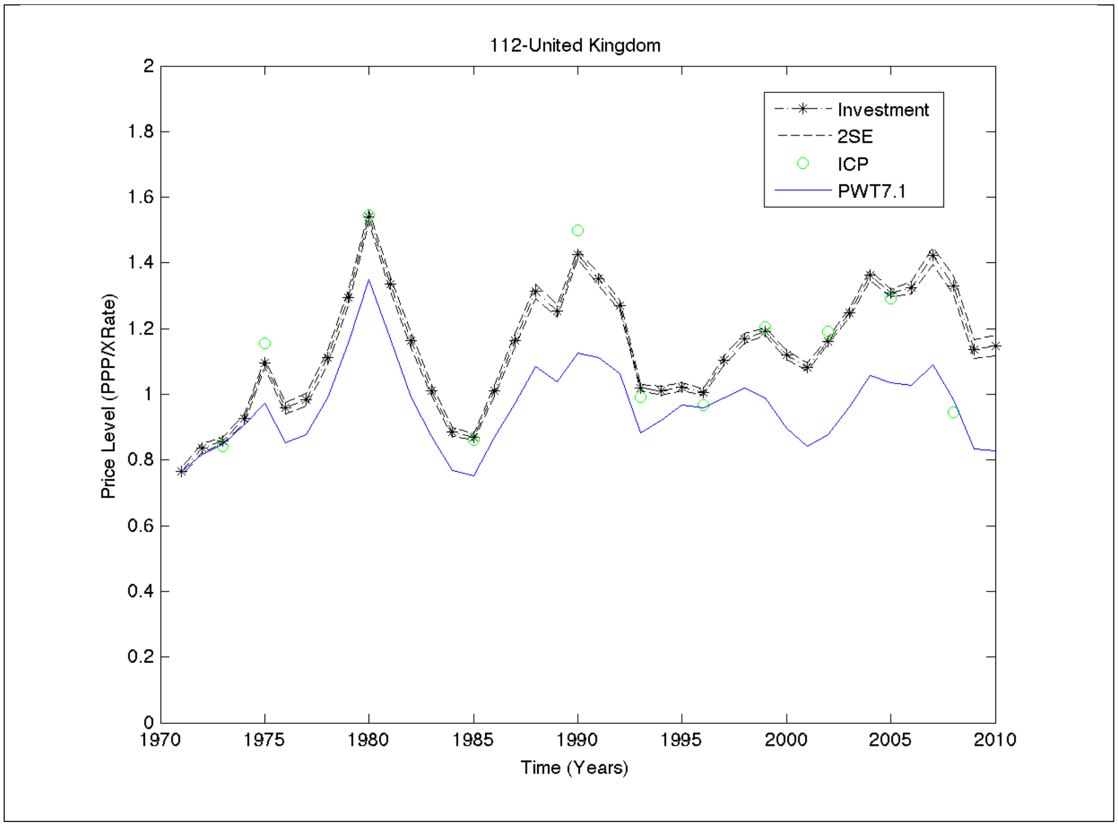


Figure 8: Government Expenditure Price Level (PPP/ER) for African Countries



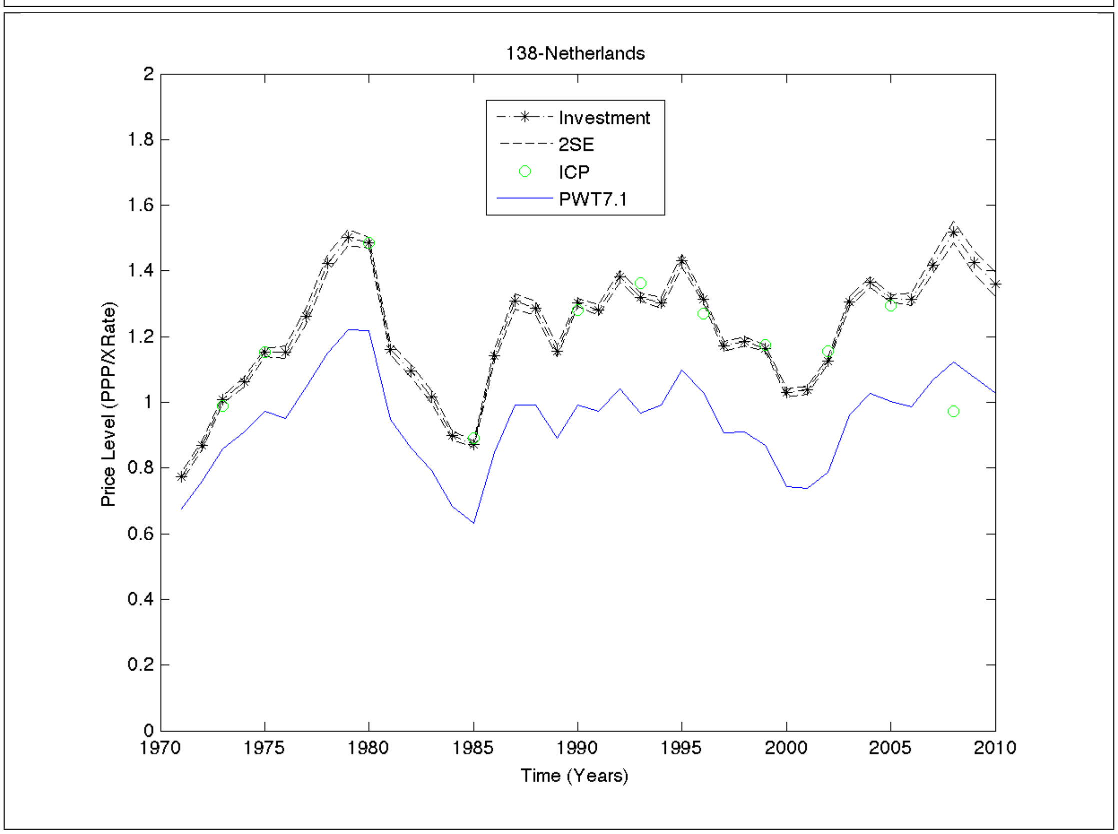


Figure 9: Investment Price Level (PPP/ER) - European Countries

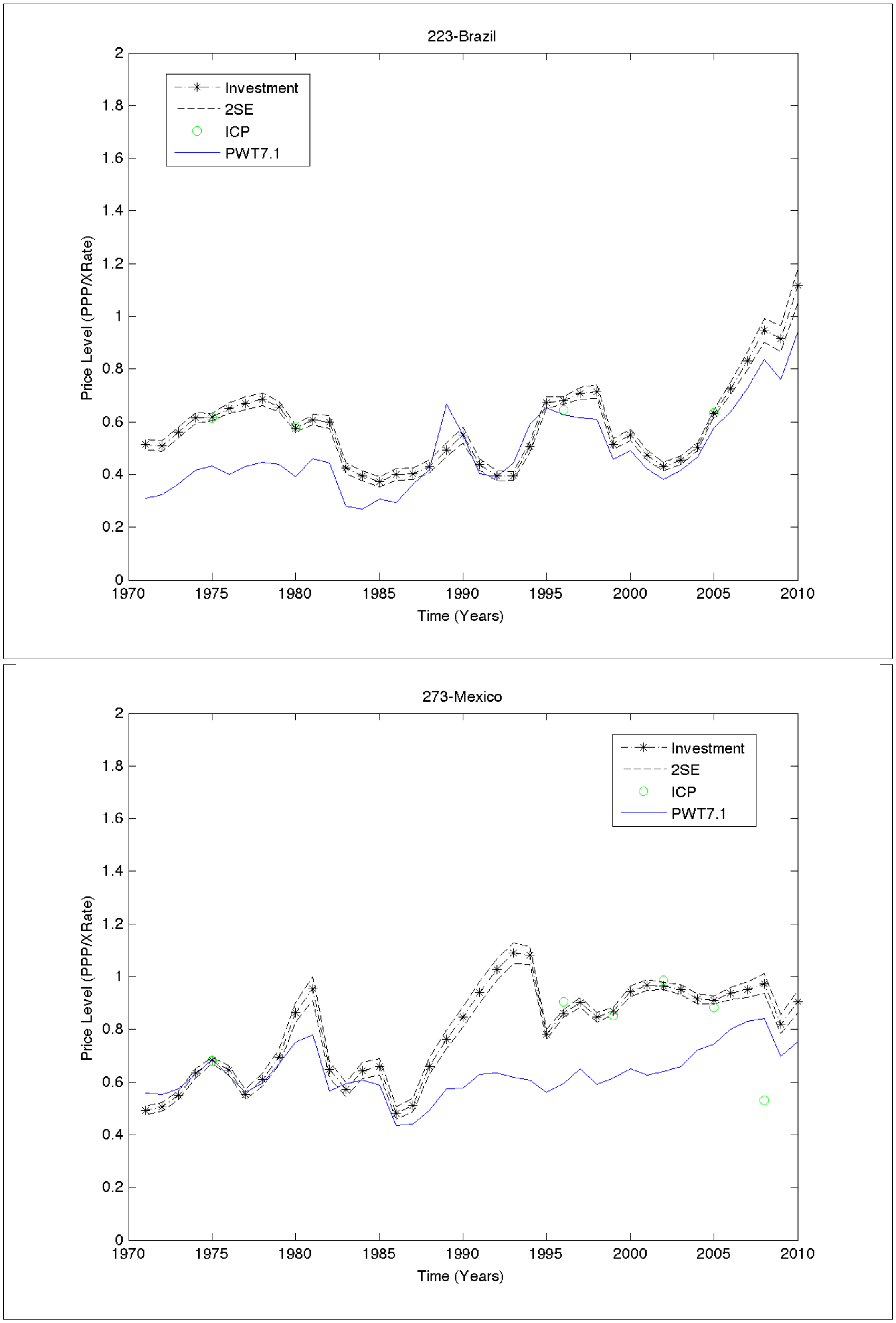
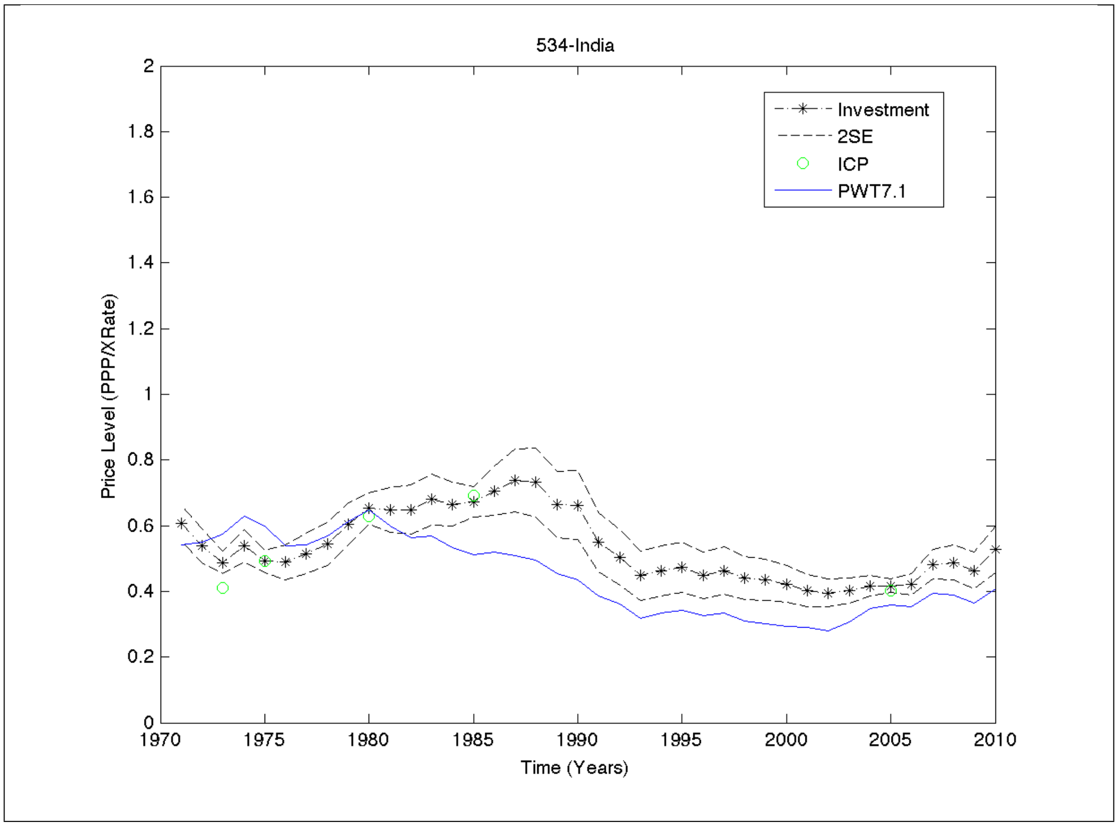


Figure 10: Investment Price Level (PPP/ER) - LA Countries



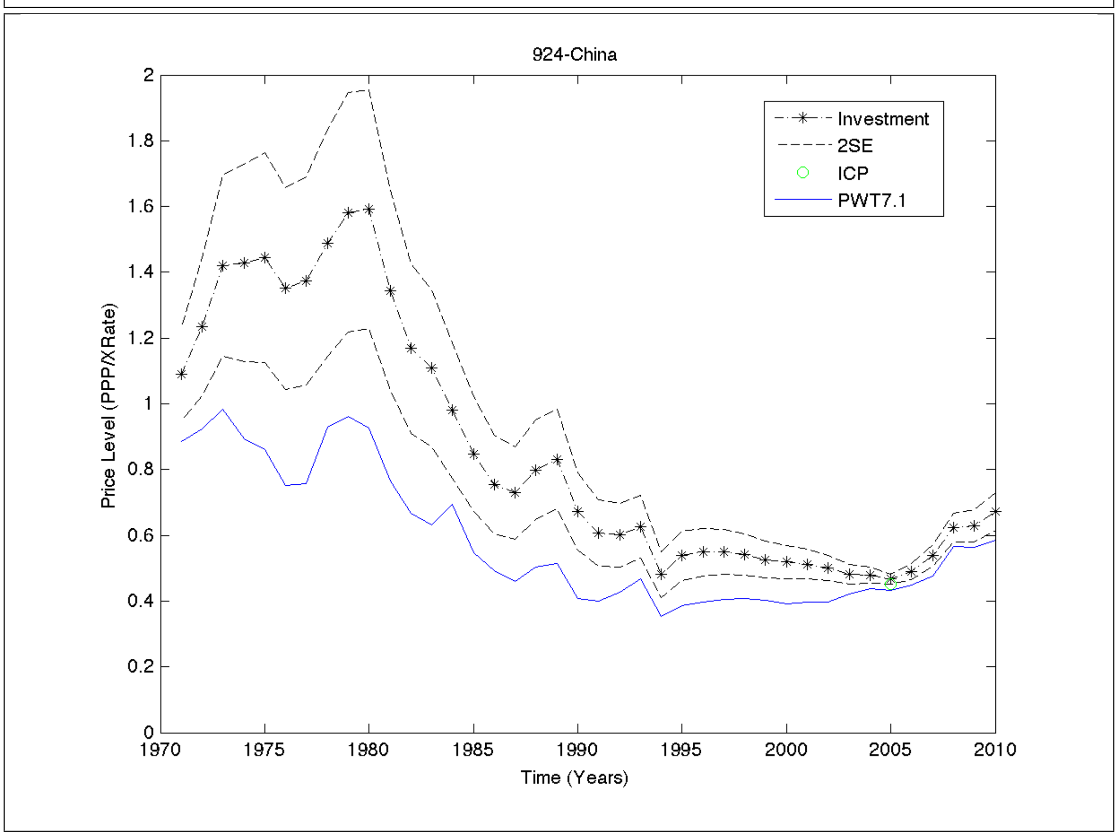
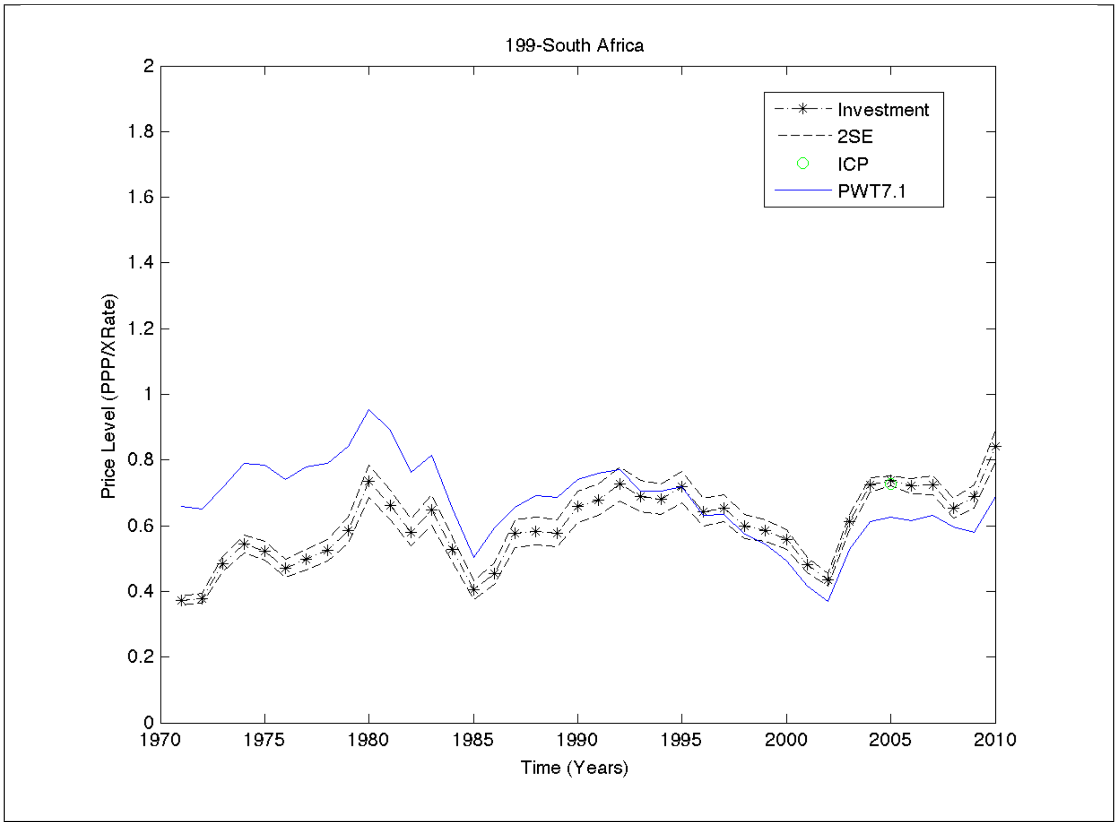


Figure 11: Investment Price Level (PPP/ER) for Asian Countries



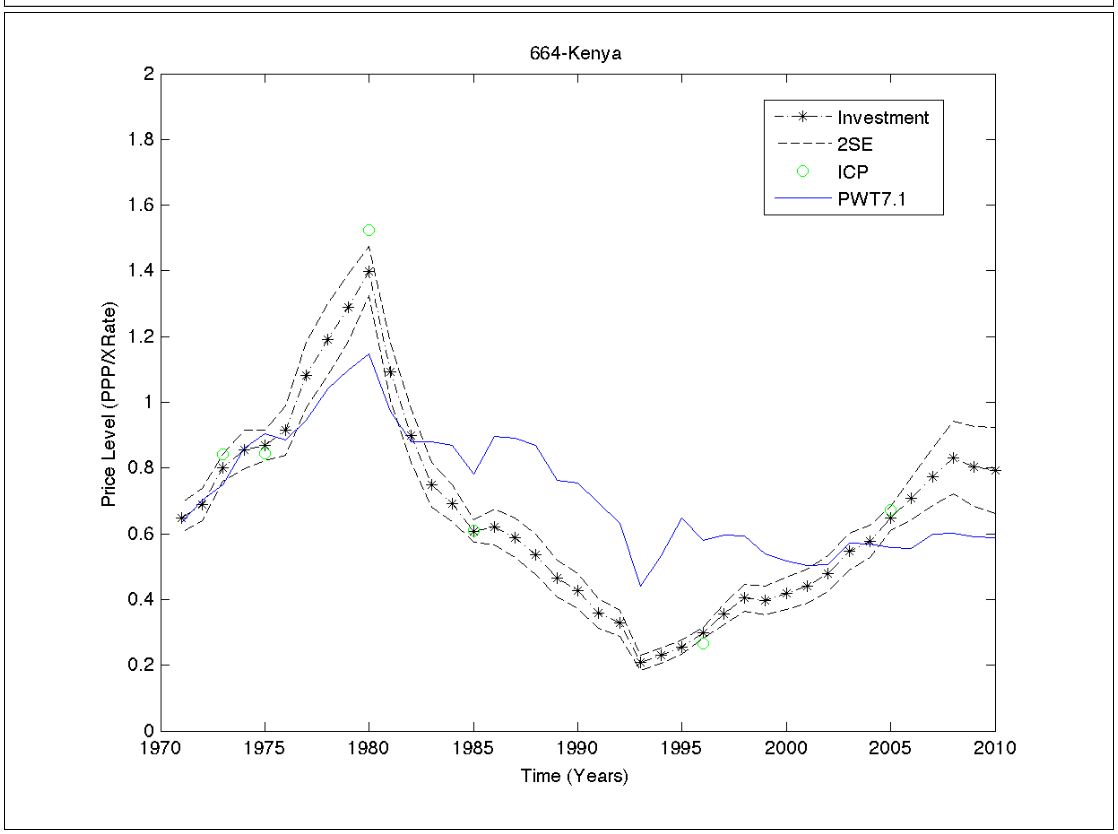


Figure 12: Investment Price Level (PPP/ER) for African Countries

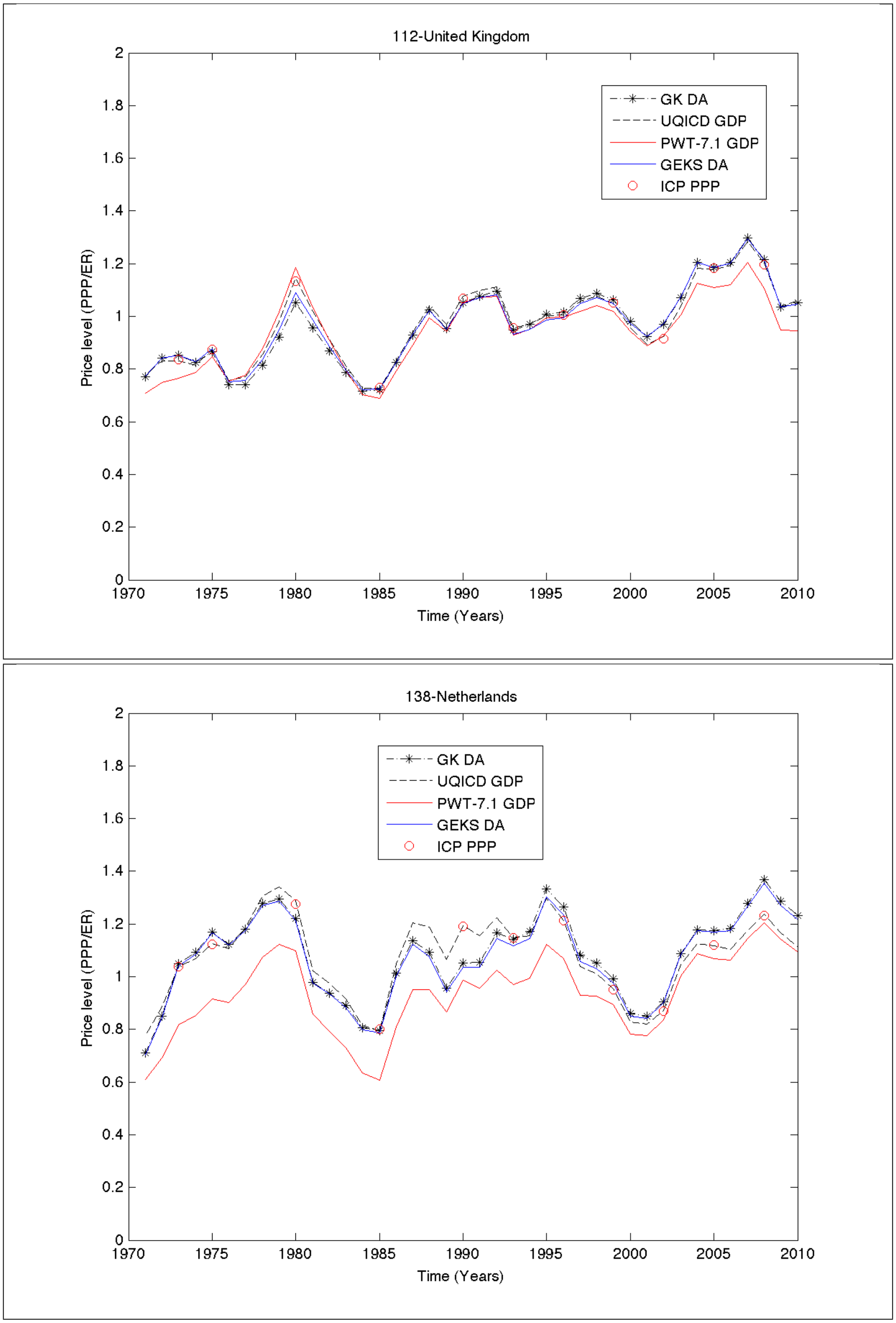


Figure 13: DA (PPP/ER) - European Countries

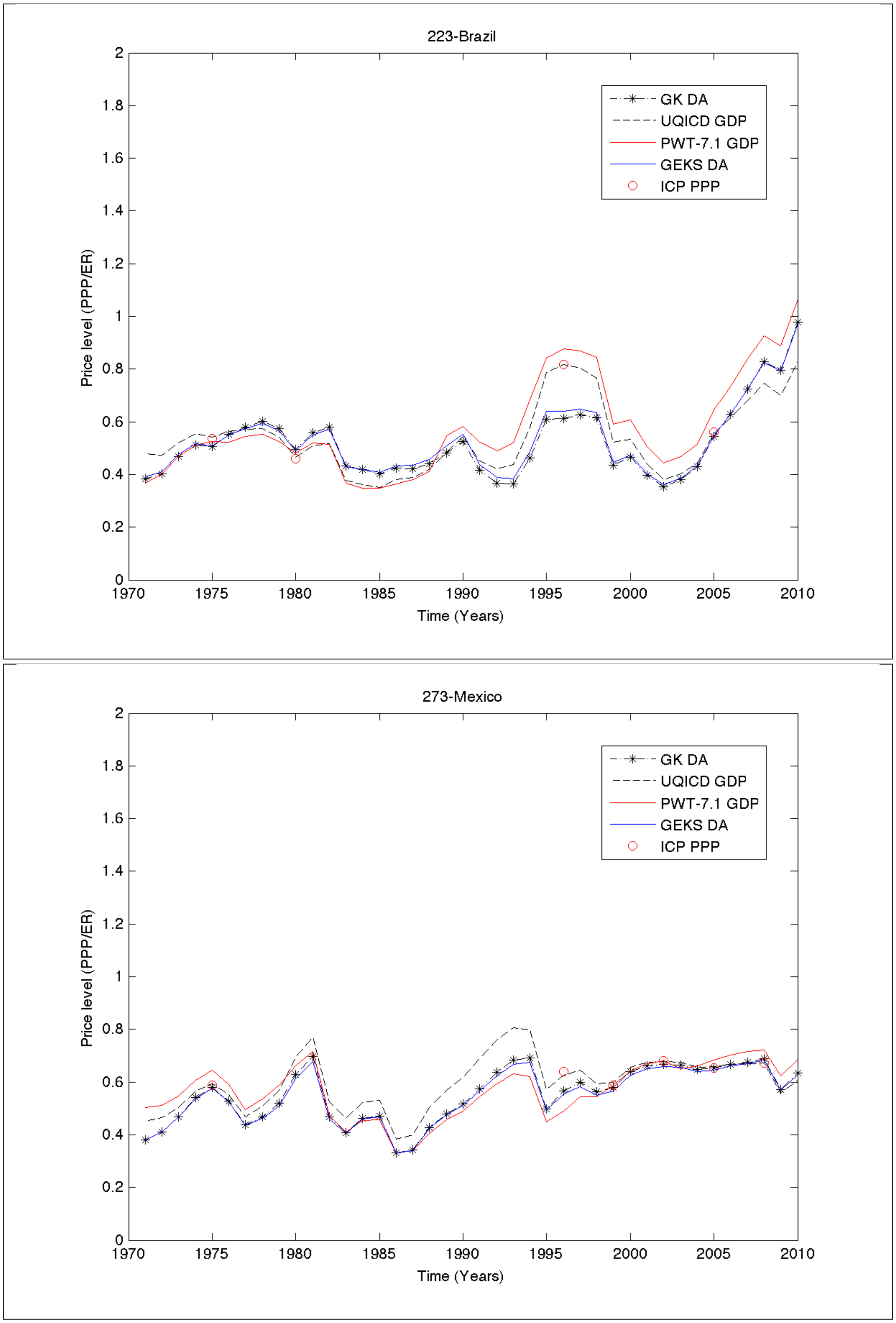
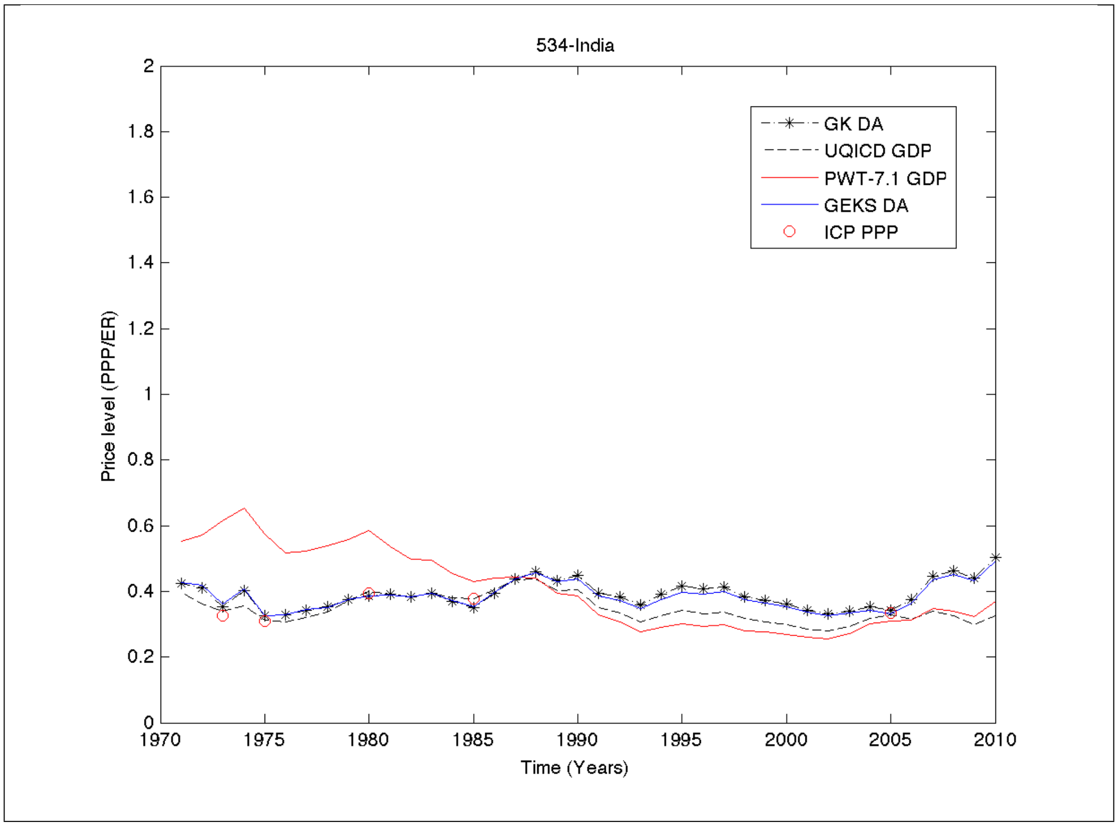


Figure 14: DA (PPP/ER) - LA Countries



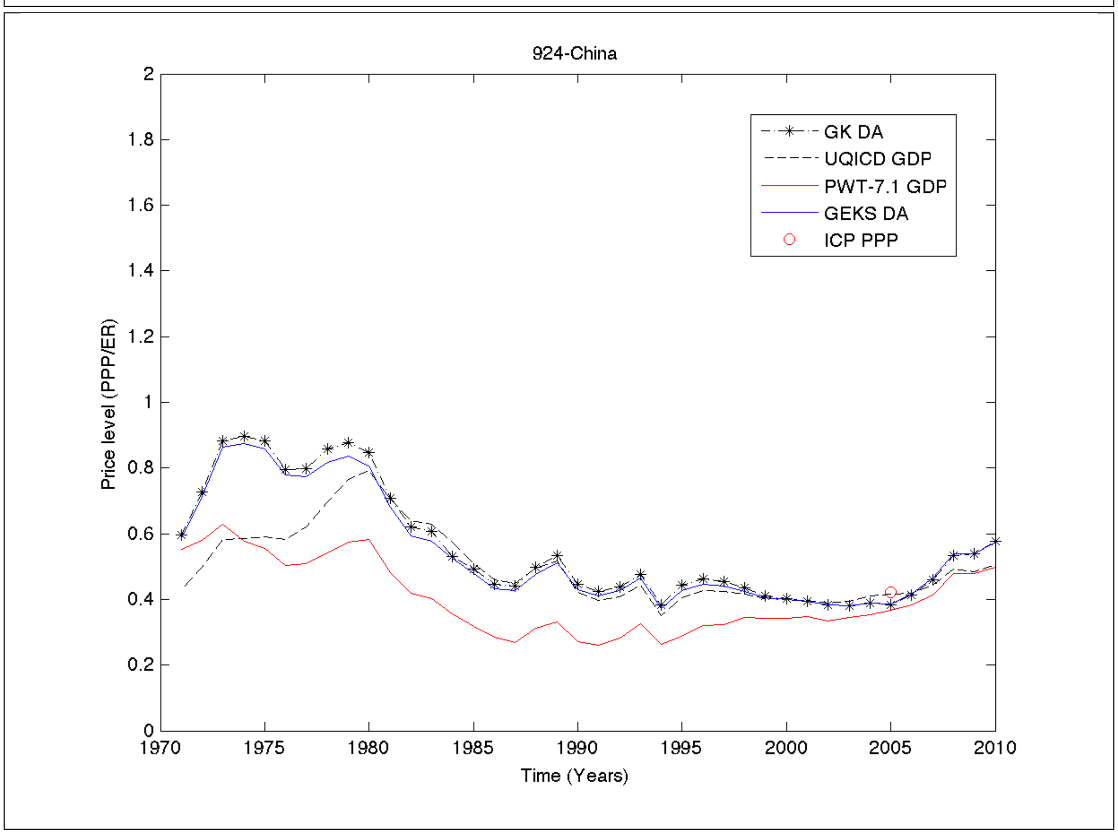


Figure 15: DA (PPP/ER) for Asian Countries

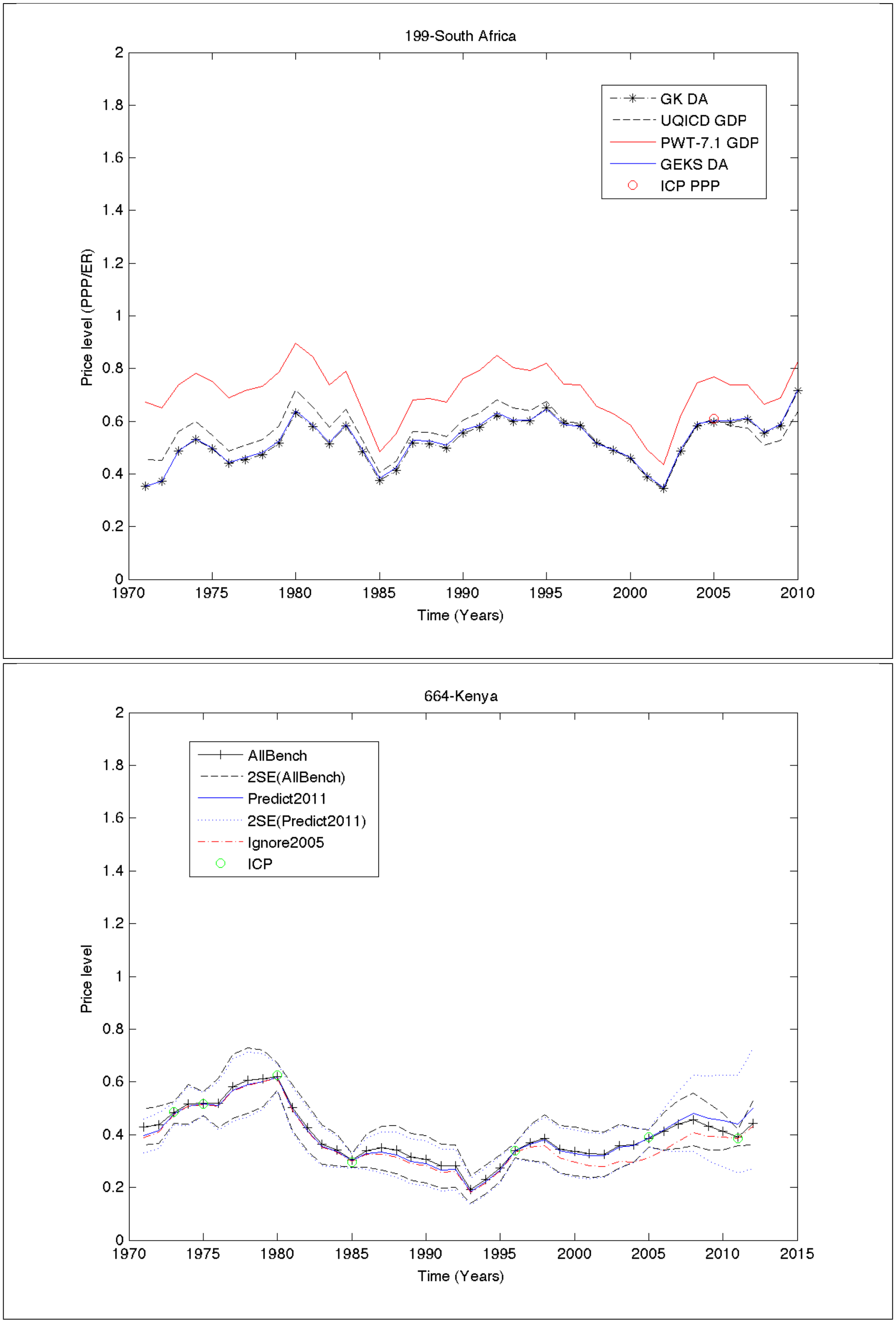
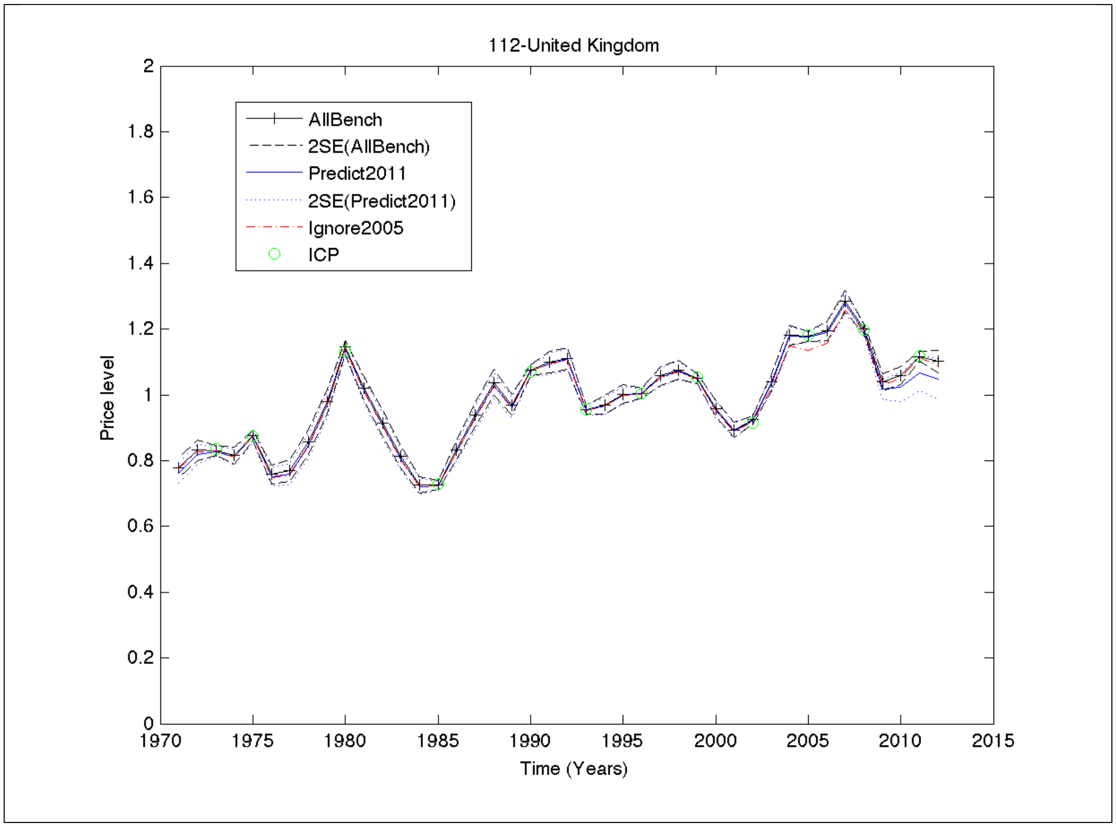


Figure 16: DA (PPP/ER) for African Countries



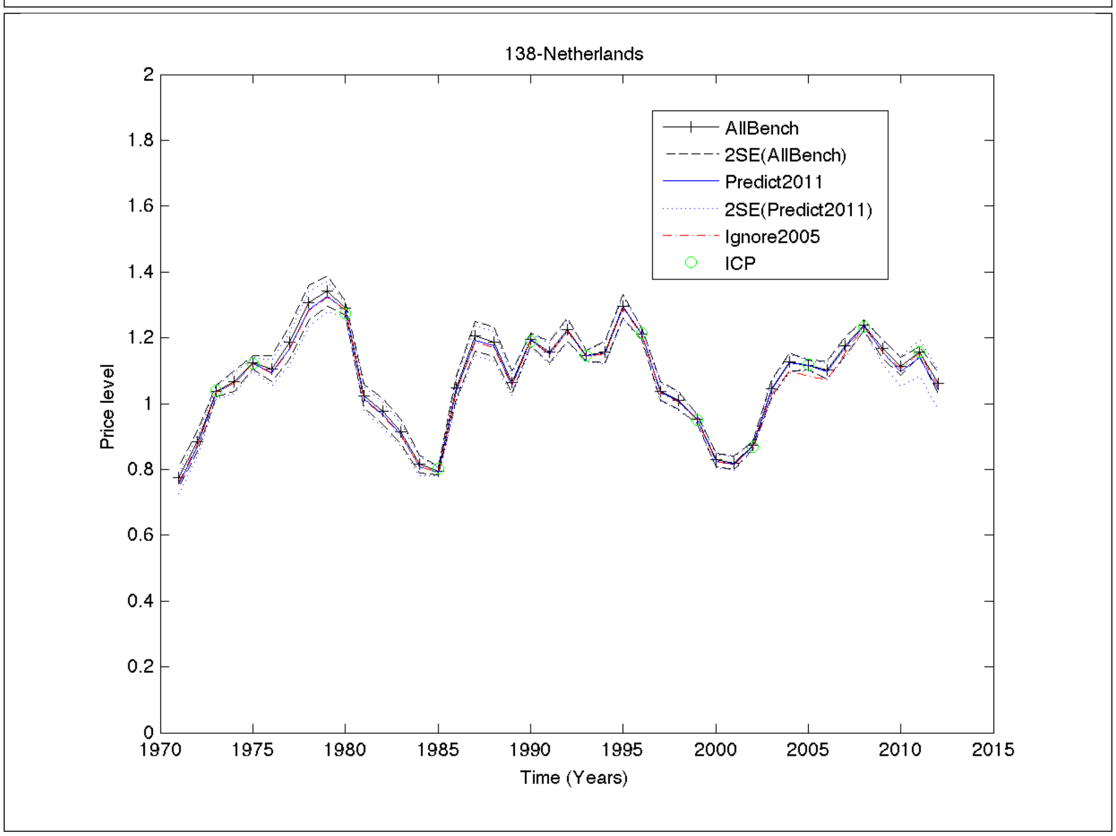
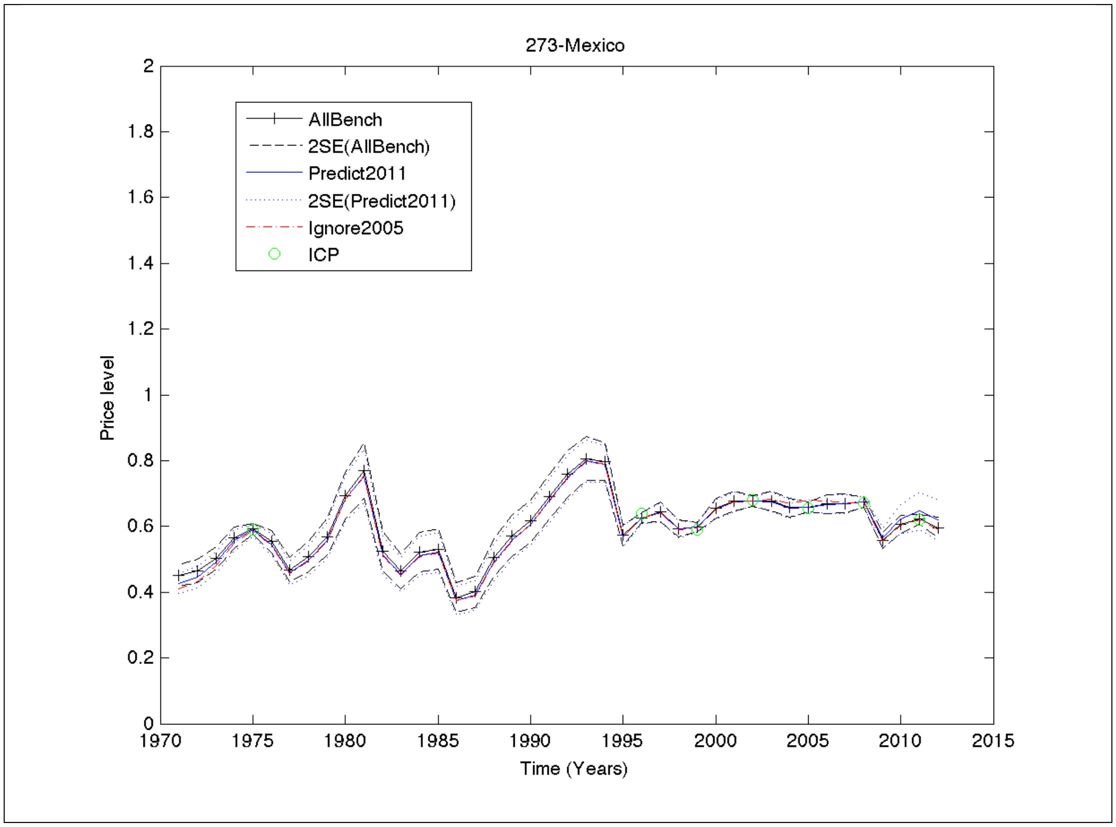


Figure 17: UQICD - GDP Level - European Countries



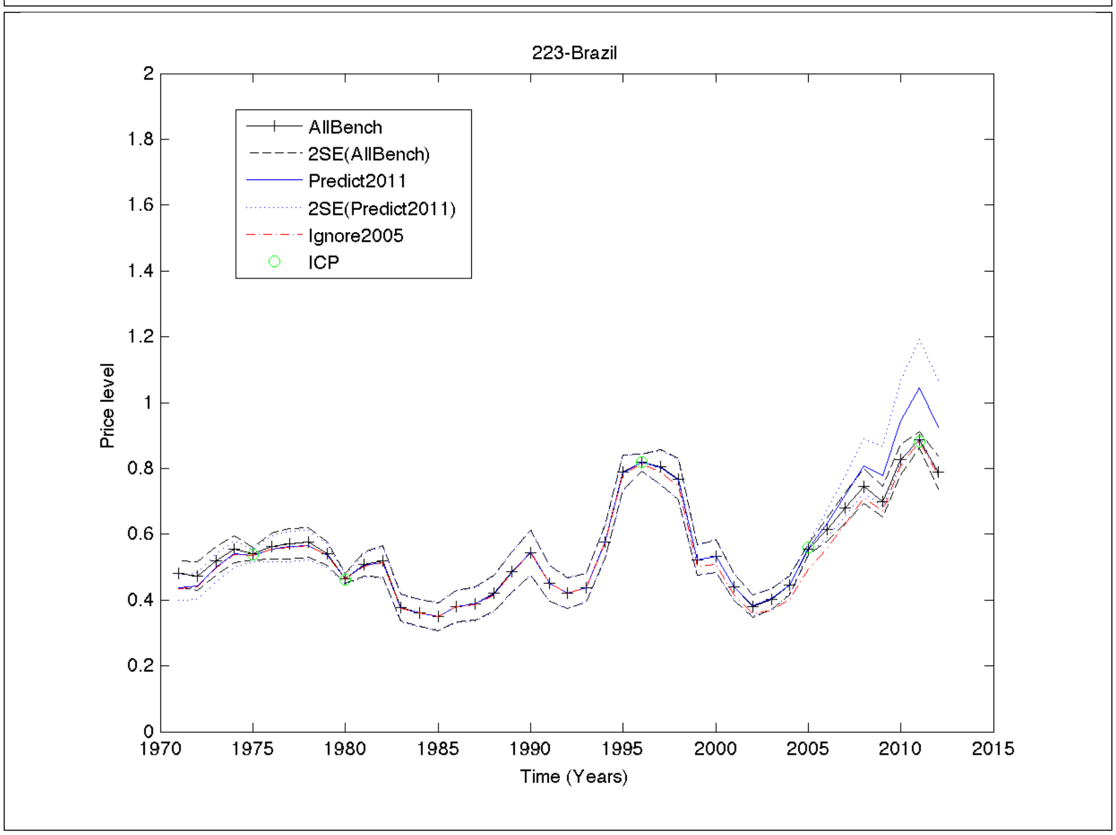
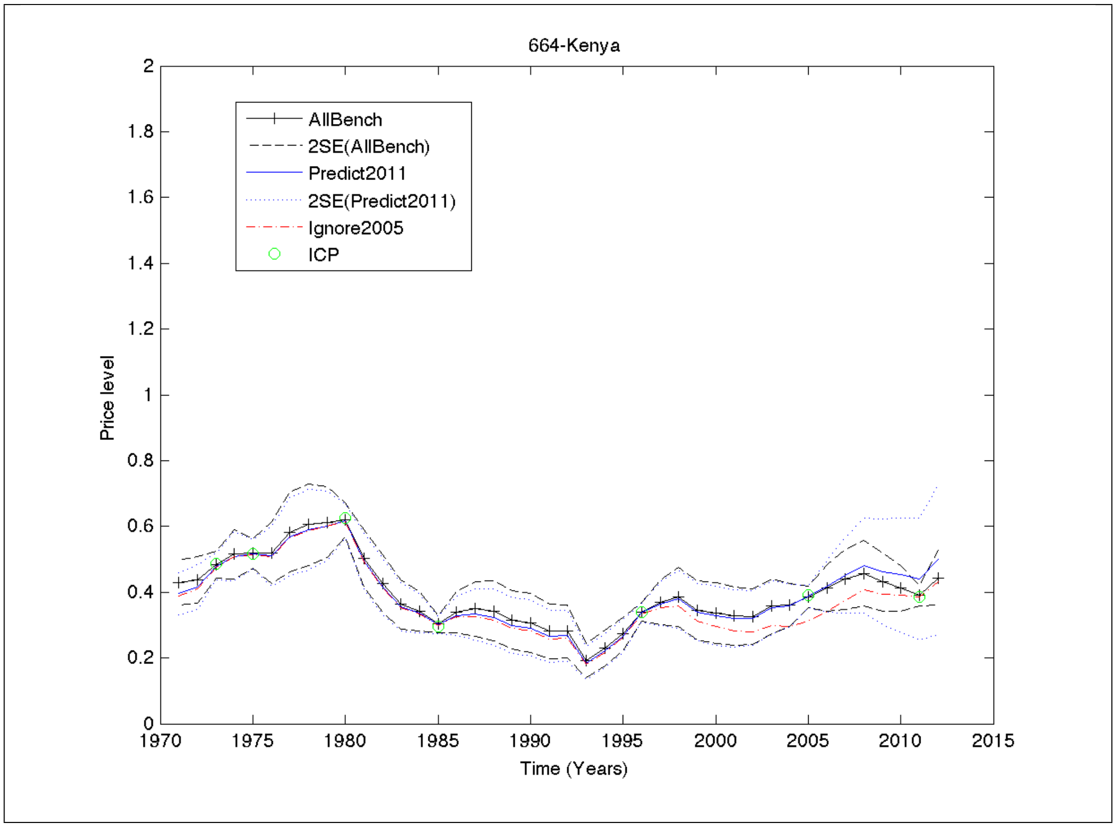


Figure 18: UQICD - GDP Level - Latin American Countries



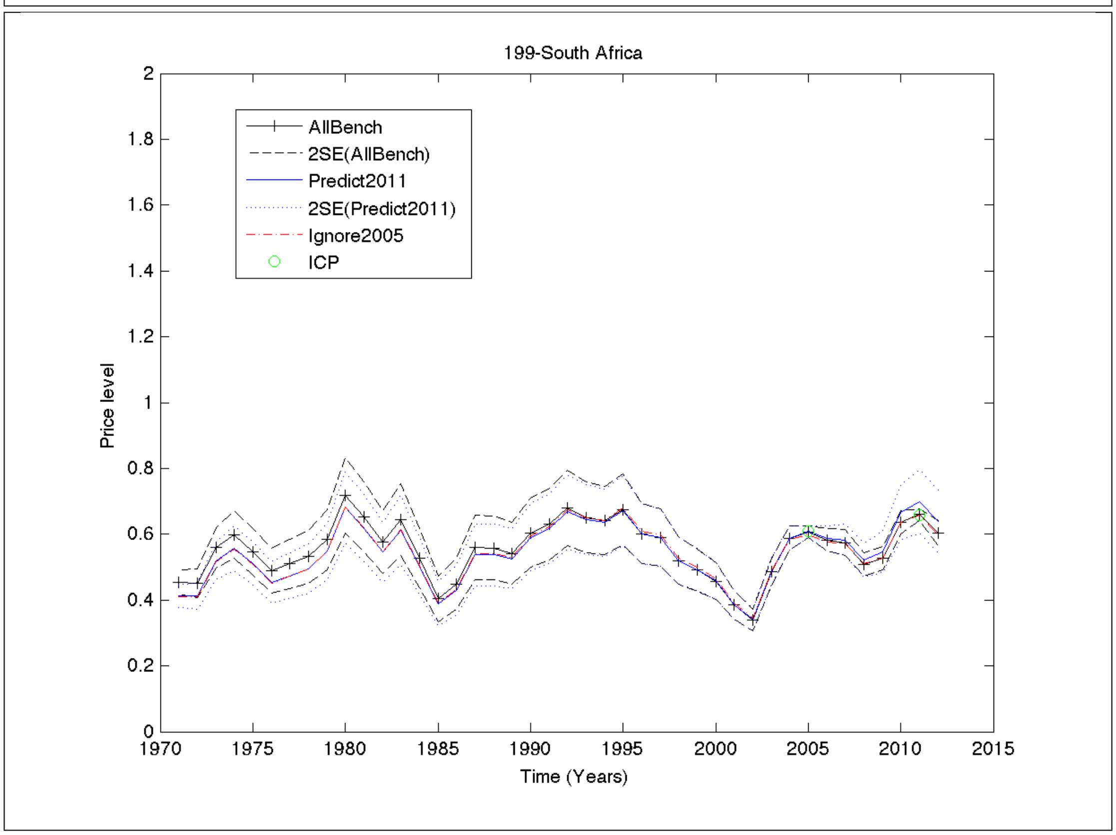
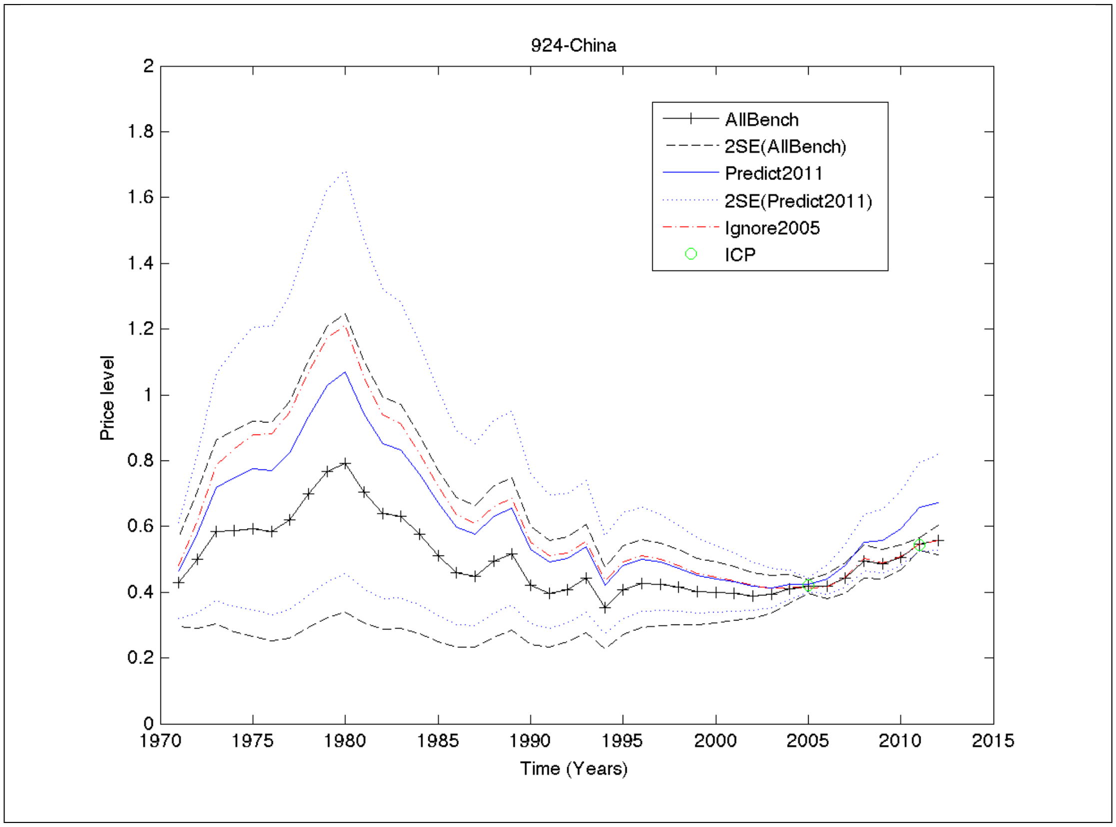


Figure 19: UQICD - GDP Level - African Countries



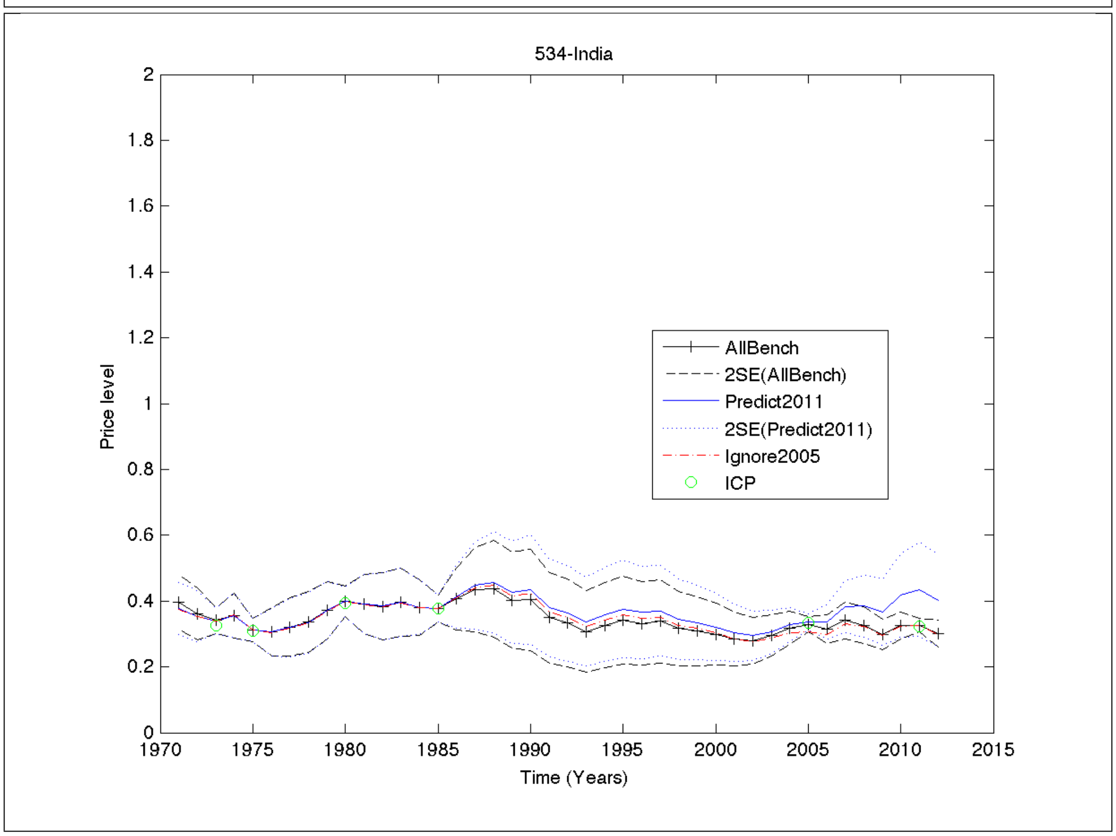
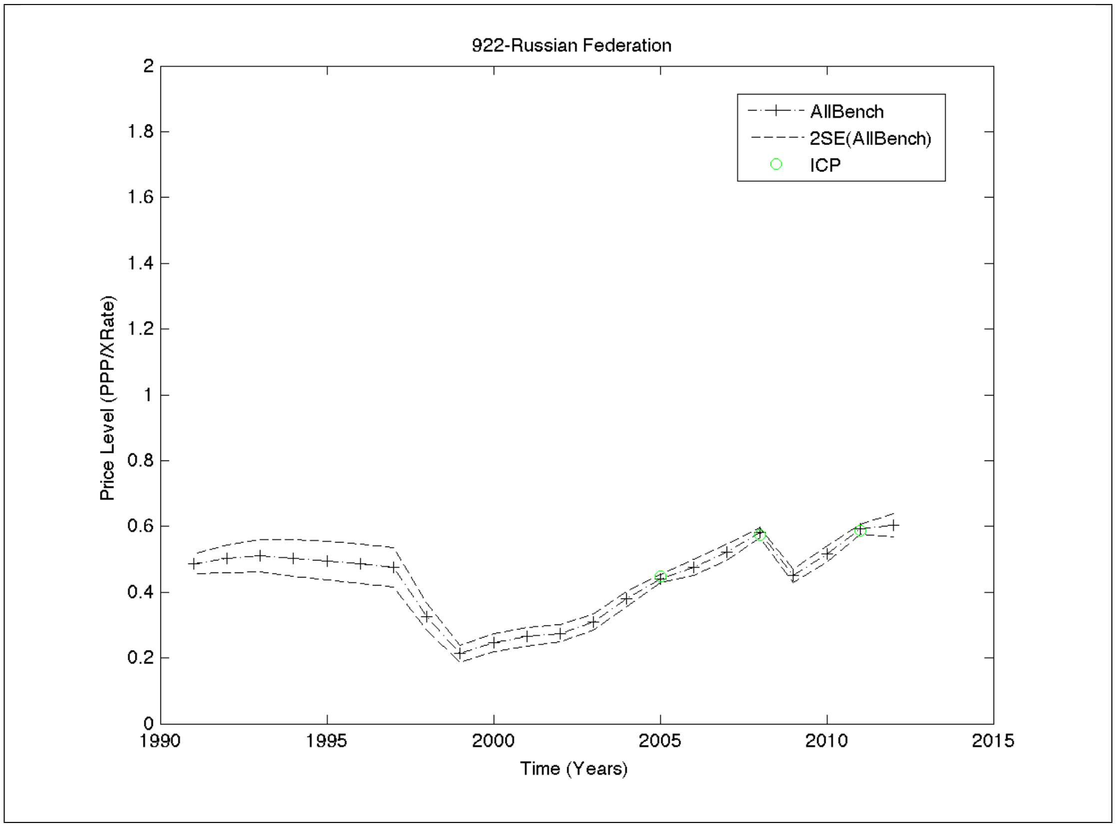


Figure 20: UQICD - GDP Level - Asian Countries



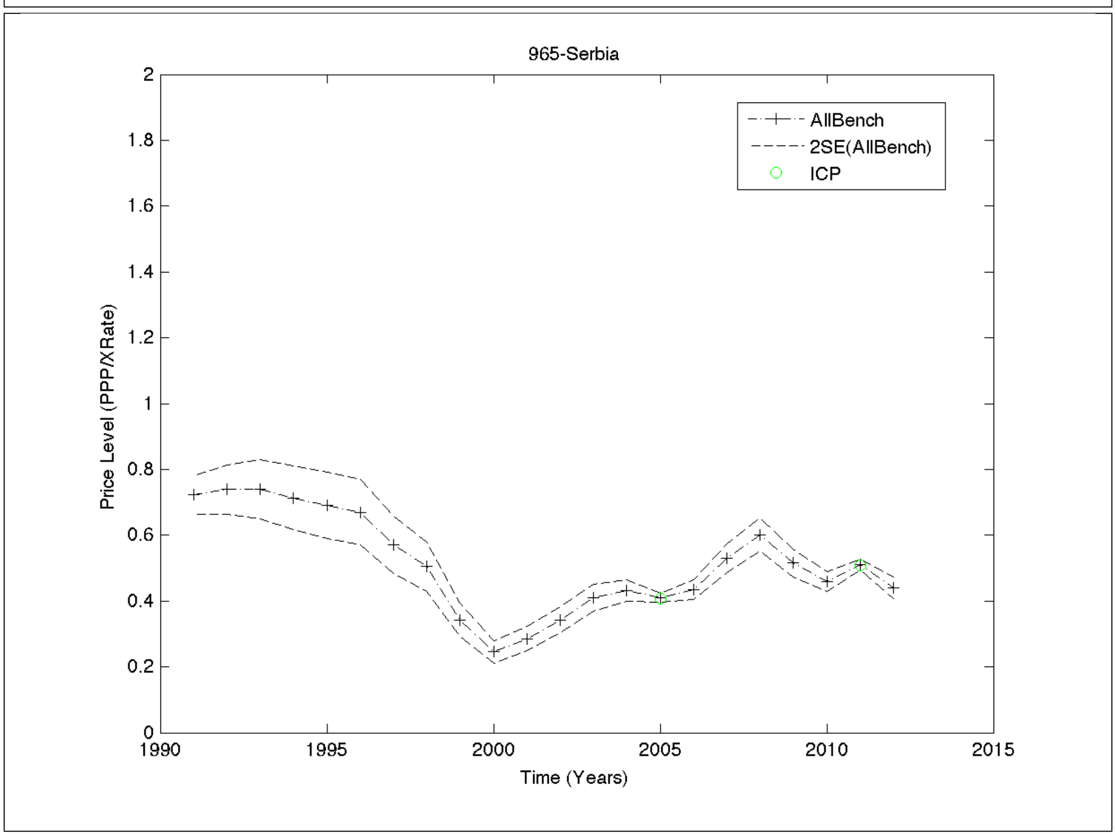


Figure 21: UQICD - GDP Level - Newer Formed Countries

A The RRD smoothing weights

In this appendix we present the form of the weights in equation (15), where we express the smoothed estimates of log PPPs as $\widetilde{p_t} = \sum_{j=1}^{T} \omega_{jT} y_j$. We use wide "$\widehat{\ }$" and "$\widetilde{\ }$" to denote the the Kalman filter and Kalman smoother estimates of log PPP. The smoothed estimates are those based on the full sample; e.g. the weight given to a particular ICP benchmark, uses all the information in the sample ($t = 1, \ldots T$ and $i = 1, \ldots N$), and the weights are a decreasing function away from the particular benchmark.

The presentation here is brief and the interested reader should go to Koopman and Harvey (2003) for a full treatment.

The equations of the model in state-space form are repeated for ease of presentation. They are the measurement equations in a benchmark year

$$y_t = \begin{bmatrix} 0 \\ \hat{p}_t \\ \tilde{p}_t \end{bmatrix} = Z_t p_t + \zeta_t \tag{17}$$

where, $\hat{p}_t$ is a prediction from the price level regression, $\tilde{p}_t$ is a benchmark ICP observation, and

$$y_t = \begin{bmatrix} 0 \\ \hat{p}_t \end{bmatrix} = Z_t p_t + \zeta_t \tag{18}$$

in non-benchmark years.

The vector y_t is $N \times 1$ and in non-benchmark years, the elements are zero for the reference country and an $N-1$ vector of regression predictions

The transition equations

$$p_t = p_{t-1} + c_t + \eta_t \tag{19}$$

where $\zeta_t \sim N(0, H_t)$ and $\eta_t \sim N(0, Q_t)$, and the two where the variance covariance H_t takes the form

$$H_t^{NB} = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 S_{np} \Omega S_{np}^{'} \end{bmatrix} \tag{20}$$

in non-benchmark years, and

$$H_t^{B} = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 S_{np} \Omega S_{np}^{'} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\xi^2 S_p V_t S_p^{'} \end{bmatrix} \tag{21}$$

in benchmark years.

It will be convenient to re-write the equations as follows

$$y_t = Z_t p_t^* + \zeta_t \tag{22}$$

$$p_t^* = p_{t-1}^* + \eta_t \tag{23}$$

where $p_t^* = p_t - \sum_{j=1}^{t} c_j$. This transformation incorporates the movement in the deflators as observed in the data (see definition of c_t below equation (13)).

The estimated log of PPP at time t can then be written as the following weighted sum

$$\widetilde{p_t^*} = \sum_{j=1}^{T} w_j(\widetilde{p_{t|T}^*}) y_j \tag{24}$$

where the weights sum to one and the expression (24) indicates that *the adjustment provided by the weights* is from information about the movement of PPPs between benchmarks *after the deflator movement has been incorporated*. This information include national accounts data, ICP benchmarks and the influence of movements in the price level of other trading partners which is brought into the weights through the cross-sectional correlation information gather by the price level regression used by UQICD.

To provide an expression for the weight, $w_j(\widetilde{p_{t|T}^*})$, we need to show the form of the Kalman filter to be able to provide the required definitions of the components of the weights.

The Kalman filter Equations and the Smoothed Weights Given the initial state vector p_0^* and Ψ_0, the Kalman filter provides an optimal linear one-step-ahead estimates of the underlying states ($p_{t+1|t}^*$) their corresponding PMSE ($\Psi_{t+1|t}$), where:

$$p_{t+1|t}^* = p_{t|t-1}^* + \Psi_{t|t-1} Z_t' F_t^{-1} \nu_t \tag{25}$$

$$\Psi_{t+1|t} = \Psi_{t|t-1} + Q_t - \Psi_{t|t-1} Z_t' F_t^{-1} Z_t \Psi_{t|t-1} \tag{26}$$

where,

$p_{t+1|t}^*$ is the Kalman filter prediction of the log of PPP for $t+1$given information up to and including time period t, which has mean squared prediction error given by $\Psi_{t+1|t}$

$\nu_t = y_t - Z_t p_{t|t-1}^*$, measures the difference between the predicted log PPP at time t and the observed, and $\nu_t \sim N(0, F_t)$

$F_t = Z_t \left[\Psi_0 + \sum_{j=1}^{t-1} Q_j - \sum_{j=1}^{t-1} \Psi_{j|j-1} Z_j' (Z_j \Psi_{j|j-1} Z_j' + H_j^k)^{-1} Z_j \Psi_{j|j-1} \right] Z_t' + H_t^k$, where $k = B, NB$

Here we note that F_t is a combined measure of the current and past uncertainty in the measurement of ICP and regression predictions (from the H_t^k) matrices and the deflator movements, from the Q_t matrices. Note that if t is *the first benchmark* in the sample then $F_t = \Psi_0 + \sum_{j=1}^{t-1} Q_j - \sum_{j=1}^{t-1} \Psi_{j|j-1} (\Psi_{j|j-1} + H_j^{NB})^{-1} \Psi_{j|j-1} + H_t^B$.

The smoothed weights are given by

$$w_j(\widetilde{p_{t|T}^*}) = \begin{cases} (I - \Psi_{t|t-1} N_{t-1}) w_j(\widehat{p_{t|t-1}^*}) & j = t-1, t-2, \ldots, 1 \\ \Psi_{t|t-1} D_t & j = t \\ \Psi_{t|t-1} B_{j,t-1}' D_j & j = t+1, t+2, \ldots, T \end{cases} \tag{27}$$

with definitions,

$D_j = Z_j F_j^{-1} - L_j^{-1} N_j \Psi_{j|j-1} Z_j' F_j^{-1}$; $j = t, t+1, \ldots, T$.

$B_{t,t-1} = I$; $B_{t,j} = L_{t-1} L_{t-2} \ldots L_{j+1}$, $j = 1, \ldots, t-2$

$B_{t,j-1} = B_{t,j} L_j$, $j = t-1, t-2, \ldots, 1$

$L_t = I - \Psi_{t|t-1} Z_t' F_t^{-1} Z_t$

$N_{t-1} = Z_t' F_t^{-1} Z_t + L_t' N_t L_t$; for $t = T-2, \ldots, 1$

$N_{T-1} = Z_T' F_T^{-1} Z_T$

$N_T = 0$

and where $w_j(\widehat{p_{t|t-1}^*})$ are the Kalman filtering weights,

$$w_t(\widehat{p_{t|t}^*}) = \Psi_{t|t-1} Z_t' F_t^{-1}$$

$$w_j(\widehat{p^*_{t|t}}) = (I - \Psi_{t|t-1} Z_t' F_t^{-1} Z_t) w_j(\widehat{p^*_{t|t-1}}); \; j = t-1, t-2, \ldots, 1$$

$$w_j(\widehat{p^*_{t|t}}) = 0; \; j > t+1, \ldots, T$$

The filter weights are highest at t and are a decreasing function of past information.

Similarly to the filter weights, at time t the weight, in (27), is highest on the current time period information; however, the information on either side of the time period enters the weight (i.e. $j \lesseqgtr t$). For example, if the sample goes up to 2012, for t =2011, the highest weight will be given to the 2011 information with a discounted memory back to the start of the sample; although in practice most of the non-zero weights will be from the immediate past years. Some weight, although likely to be small, will be from the 2012 information.

B Bootstrap PMSE of the states

The Appendix briefly details the Bootstrap procedures used to improve standard errors computation for the RRD method. Consider the state space model in the RRD method (please refer back to Section 2) with the observation equations (in (11)) and the transition equations (13). The observation equations can be rewritten in the form:

$$y_t - B_t X_t \theta = Z_t p_t + \zeta_t \tag{28}$$

Let $y_{t,new} = y_t - B_t X_t \theta$ an $N \times 1$ vector. The observation equations accommodating both benchmark and non-benchmark observations can be re-written in the form:

$$y_{t,new} = Z_t p_t + \zeta_t \tag{29}$$

where Z_t is a $N \times N$ matrix mapping the observations to the states and $\zeta_t \sim N(0, H_t)$, where the variance covariance H_t takes the form

$$H_t = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 S_{np} \Omega S_{np}' \end{bmatrix} \tag{30}$$

in non-benchmark years, and

$$H_t = \begin{bmatrix} 0 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_u^2 S_{np} \Omega S_{np}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\xi^2 S_p V_t S_p' \end{bmatrix} \tag{31}$$

in benchmark years.

From Equation 9, $\Omega = (I - \phi W_t)^{-1}[(I - \phi W_t)^{-1}]'$ with W_t is the known spatial weight matrix and ϕ is the spatial correlation coefficient $\phi < 1$, S_{np} and S_p are selection matrices.

The transition equations disturbance is given by $\eta_t \sim N(0, Q_t)$ and $Q_t = \sigma_\eta^2 V_t$ (see definition of V_t below equation (6)).

The disturbances ζ_t and η_t are uncorrelated with each other in all time periods. Given the initial state vector p_0 and Ψ_0, the Kalman filter provides optimal unbiased linear one-step-ahead estimates of the underlying states ($p_{t|t-1}$) their corresponding PMSE ($\Psi_{t|t-1}$), where:

$$p_{t|t-1} = p_{t-1|t-2} + c_t + K_t F_{t-1}^{-1} U_{t-1} \tag{32}$$

$$\Psi_{t|t-1} = \Psi_{t-1|t-2} - K_t F_{t-1}^{-1} K_t' + Q_t \tag{33}$$

with $K_t = \Psi_{t-1|t-2} Z'_{t-1}$ as the filter gain, $U_t = y_{t,new} - Z_t p_{t|t-1}$ is the one-step-ahead vector of innovations and their covariance $F_t = Z_t \Psi_{t|t-1} Z'_t + H_t$, given the vector of hyperparameters $\delta = (\sigma_u^2, \sigma_\eta^2, \sigma_\xi^2, \phi)$

The state space model described by equations 28 to 13 can be written in the Innovation Form (IF) so that they will only depend on a unique disturbance vector U_t instead of two as below:

$$p_{t|t-1} = p_{t-1|t-2} + c_t + K_t F_{t-1}^{-1} U_{t-1} \tag{34}$$

$$y_{t,new} = Z_t p_{t|t-1} + U_t \tag{35}$$

The bootstrap procedure intended to be used is the nonparametric method proposed by Rodriguez and Ruiz (2012). It is based on resampling from the residuals of the estimated model and does not assume any particular error distribution following these steps:

Step 1: The standardized estimated innovations, $\hat{U}_t^s = \hat{F}_t^{-1/2} \hat{U}_t$ are resampled to obtain $\{U_1^*, U_2^*, \dots U_t^*\}$, and then the bootstrap replicates, $\{y_{1,new}^*, y_{2,new}^*, \dots y_{T,new}^*\}$ are obtained from the IF as the following:

$$y_{t,new}^* = Z_t p_{t|t-1}^* + U_t^* \tag{36}$$

$$p_{t+1|t}^* = p_{t|t-1}^* + c_t + K_t F_{t-1}^{-1} U_{t-1}^* \tag{37}$$

Step 2: The hyperparameters are estimated Maximum Likelihood, obtaining $\hat{\delta}^*$.

Step 3: The Kalman filter is run once with the original observations $\{y_{1,new}, y_{2,new}, \dots y_{T,new}\}$ and $\hat{\delta}^*$ to obtain $\theta(\hat{\delta}^*)$, $\hat{p}_{t|t-1}(\hat{\delta}^*)$ and $\hat{\Psi}_{t|t-1}(\hat{\delta}^*)$ for $t = 1, ..., T$.

Step 4: Repeat step 1 to 3 B times to obtain B bootstrap replicates, the bootstrap PMSE is obtained as follows

$$\widehat{PMSE}_{t|t-1} = \frac{1}{B} \sum_{j=1}^{B} \hat{\Psi}_{t|t-1}(\hat{\delta}^{*(j)}) + \frac{1}{B} \sum_{j=1}^{B} [\hat{p}_{t|t-1}(\hat{\delta}^{*(j)}) - \bar{p}_{t|t-1}^*][\hat{p}_{t|t-1}(\hat{\delta}^{*(j)}) - \bar{p}_{t|t-1}^*]' \tag{38}$$

where $\bar{p}_{t|t-1}^* = \frac{1}{B} \sum_{j=1}^{B} \hat{p}_{t|t-1}(\hat{\delta}^{*(j)})$

There are three advantages in this method compared to other bootstrap methods. Firstly, the biases of the PMSE produced are smaller than those of the asymptotic procedures of Hamilton (1986) and the bootstrap PMSE procedure of Pfeffermann and Tiller (2005). Secondly, the Kalman filter is run with the bootstrap estimates of the parameters and the original time series so the PMSE in step 4 is conditional on the information contained in the original series, which make full use of the data. Moreover, the procedure is computationally simpler since it avoids running the filter two more times as in the procedure proposed by Pfeffermann and Tiller (2005).

C Components - Initial regression results

These regressions are obtained using an unbalanced panel of the data from ICP participating countries and benchmark years.

Table 1: Private Consumption Regression

Variables	Estimate	Standard Error	t-statistics
D_asean	-0.1572	0.0657	-2.3913
D_mercsr	-0.0818	0.0733	-1.1169
D_nafta	-0.1079	0.0708	-1.5233
D_island	0.0586	0.0464	1.2641
D_landlock	-0.0112	0.0398	-0.2821
D_comst1	-0.3753	0.2076	-1.8080
D_comst2	-0.1543	0.1723	-0.8956
D_wcfa	0.1515	0.0822	1.8433
D_Eurocorp	0.1228	0.0426	2.8822
D_Europeg	0.1067	0.0836	1.2763
D_spacific	-0.0132	0.0866	-0.1521
D_usd	0.0365	0.0435	0.8393
Agric	-0.0112	0.0019	-5.7502
Labpop	-0.0094	0.0041	-2.2844
Life	-0.0022	0.0033	-0.6556
Nontrade	-0.0005	0.0027	-0.1723
Black_I	0.0042	0.0070	0.5906
Phones	0.0017	0.0002	10.9197
Secendaschl	0.0000	0.0001	0.3539
Trade	-0.0003	0.0003	-1.0364
Service	-0.0025	0.0020	-1.2162
Internet	0.0010	0.0016	0.6155
Capital control	0.0109	0.0062	1.7513
Mobile	-0.0003	0.0011	-0.2889
	Number of observations	491	
	R2	0.7407	
	Adjusted R2	0.7214	

Table 2: Government Expenditure regression

Variables	Estimate	Standard Error	t-statistics
D_anz	-0.5738	0.3151	-1.8212
D_asean	-0.2923	0.0956	-3.0581
d_ca	0.3903	0.2210	1.7655
D_mercsr	-0.0848	0.1064	-0.7974
D_nafta	0.0645	0.1030	0.6256
D_island	-0.0698	0.0677	-1.0306
D_landlock	0.1629	0.0584	2.7907
D_comst1	-0.7576	0.3011	-2.5161
D_comst2	0.3547	0.2511	1.4126
D_wcfa	0.3538	0.1193	2.9656
D_safricac	0.2119	0.1860	1.1391
D_eurocorp	0.2261	0.0618	3.6554
D_europeg	0.0235	0.1214	0.1938
D_spacific	0.6153	0.2966	2.0742
D_usd	-0.1515	0.0652	-2.3252
Agric	-0.0168	0.0029	-5.8226
Labpop	-0.0264	0.0060	-4.4300
Life	0.0061	0.0049	1.2228
Nontrade	-0.0016	0.0040	-0.4060
D_black1	-0.0019	0.1039	-0.0185
D_black2	-0.0704	0.1550	-0.4540
Black_I	0.0031	0.0141	0.2179
Phones	0.0024	0.0002	10.6390
Secendaschl	0.0002	0.0001	1.6233
Trade	-0.0016	0.0005	-3.4317
Service	-0.0009	0.0030	-0.2859
Internet	0.0002	0.0023	0.0918
Capital control	0.0277	0.0092	3.0231
Mobile	0.0024	0.0016	1.4662
	Number of observations	491	
	R2	0.8135	
	Adjusted R2	0.7979	

Table 3: Gross Investment regression

Variables	Estimate	Standard Error	t-statistics
D_euro	0.1525	0.0421	3.6268
D_mercsr	0.0137	0.0831	0.1650
D_landlock	0.0056	0.0458	0.1217
D_comst1	-0.5908	0.2365	-2.4976
D_comst2	0.3835	0.1969	1.9472
D_safricac	-0.1378	0.1453	-0.9487
D_spacific	0.0783	0.0873	0.8962
D_usd	-0.1013	0.0485	-2.0902
Agric	-0.0100	0.0022	-4.4831
Labpop	-0.0215	0.0047	-4.6014
Life	-0.0153	0.0038	-4.0112
Nontrade	0.0038	0.0030	1.2557
D_black1	0.0929	0.0657	1.4138
D_black2	-0.0790	0.0935	-0.8446
Phones	0.0017	0.0002	9.4700
Secendaschl	0.0002	0.0001	2.0560
Trade	-0.0009	0.0004	-2.5972
Service	0.0023	0.0023	1.0064
Internet	-0.0005	0.0014	-0.3832
Capital control	0.0137	0.0069	1.9790
	Number of observations	491	
	R2	0.6344	
	Adjusted R2	0.6106	

D Variables Definitions

Table 4: Definition of variables

Variables	Definition
Agric	Agriculture, value added (% of GDP)
Black_I	0 if black market premium in exchange for 5 year periods< 20%;2 if >100%
Capital control	International capital market controls
D_anz	Dummy for Australia-New Zealand ANZD agreement.
D_asean	Dummy for ASEAN countries
D_black1	Black market premium in exchange for 5 year periods > 20%= 1
D_black2	Black market premium in exchange for 5 year periods > 100%=1
D_ca	Dummy for CACM (Central Amerian Common Market) countries
D_comst1	Dummy variable for in transition from communist regime
D_comst2	Dummy variable for countries either current or former communist rule
D_euro	Dummy for countries which have used the euro since 1999
D_eurocorp	Dummy variable for countries which have used the euro since 1999 and have a cooparative exchange rate arrangement gion before 1999
D_europeg	Dummy variable for countries with currencies (CFA_franc) pegged to European Euro
D_island	Dummy variable for islands
D_landlock	Dummy variables for landlock countries
D_mercsr	Dummy for MERCOSUR (an economic and political agreement among some south American countries) countries.
D_nafta	Dummy for NAFTA (North American Free trade Agreement) countries
D_safricac	Dummy variable for countries with currency union with or fix to South African rand
D_spacific	Dummy for South Pacific Trade and Economic Co-Operation Agreement
D_usd	Dummy variable for countries with currencies either pegged to the US$ for substantial amounts of time or use US$ as the legal tender - during the post-Bretton Woods era (1973 onwards)
D_wcfa	Dummy variable for countries with common west african CFA franc currency
Internet	Internet users (per 100 people)
Labpop	Labor force as percentage of total population. For developing countries the labor force is simply defined as the "economically active" population, which is itself based on age groups
Life	Life expectancy at birth, total (years)
Mobile	Mobile cellular subscriptions per 100 people
Nontrade	Non-tradable sector value added (% of GDP) - definition 2: sum of Construction,Wholesale, retail trade, restaurants and hotels, Transport, storage and communication and "Other Activities"
Phones	Telephone mainlines (per 1,000 people)
Secendaschl	School enrollment, secondary per ’000 gross enrolment
Service	Service, value added (% of GDP).
Trade	Trade (% of GDP)