Estimating the First Semester Data of the ICP- Africa Reference Year 2005 using a Model Based Method

Abdoulaye Adam[1] and Codjo N. Noumon[2]

Summary:
In most African countries participating to the 2005 International Comparison Program (ICP) round, the collection of goods and services consumed by households started in the second semester of the reference year 2005. As national annual averages are required to compute Purchasing Power Parities (PPPs), the missing data need to be dealt with. One way to address the issue is to estimate the missing prices of the months when the data collection has not started. The paper presents and tests a method of estimating the prices from January 2005 to the month preceding ICP price data collection. The method is model based and uses all the available information in years 2005 and 2006 to estimate the 2005 first semester prices.

The model is tested using the 2004 and 2005 Consumer Price Index (CPI) data of Togo, and 2005 and 2006 CPI data of Senegal, and then fit to two sets of ICP data with different data quality. To assess the quality of the fit, the correlation coefficient between observed and predicted values and the ratios of the predicted values to the observed values were used. Highly significant correlations coefficients were obtained and the predicted values are very closed to the observed values. But as expected, the model is sensitive to the quality of the input data. If predicted values look odd, chances are high that there is something wrong with the input data. The sensitivity of the model to the quality of the input data can therefore be used to fix problems in the data prior to computing the PPPs.

Key Words: *Average Prices, Correlation, International Comparison Program, ICP-Africa, Model, Purchasing Power Parities, Variation.*

Résumé:
Dans la majorité des pays africains participant au Programme de Comparaison Internationale (PCI), la collecte des prix des biens et services de la consommation des ménages a démarré au second semestre de l'année de référence 2005.

[1]Principal Statistician, African Development Bank Group, BP 323-1002 Belvédère,Tunis,Tunisia, email :a.adam@afdb.org
[2]PhD student in economics, Department of Economics, University of Montreal C.P. 6128, succursale Centre-ville QC H3C 3J7 Canada email : c.noumon@gmail.com

Comme les moyennes annuelles nationales sont requises pour le calcul des parités de pouvoir d'achat (PPA), il est donc nécessaire de résoudre le problème des données du premier semestre 2005 qui n'ont pas été collectées. Une possibilité est d'estimer ces données. L'article présente et teste une méthode d'estimation des données de janvier 2005 au mois précédent la collecte des données des prix. La méthode est basée sur un modèle et utilise toute l'information disponible en 2005 et 2006 pour faire une estimation des prix du premier semestre 2005.

Le modèle est testé en utilisant les données de 2004 et 2005 de l'indice des prix à la consommation (IPC) du Togo, les données de 2005-2006 de l'IPC du Sénégal et deux séries de données du PCI de différente qualité. La corrélation entre les valeurs observées et les valeurs prédites et les rapports entre ces deux ensembles de données sont utilisées pour tester la qualité du modèle. Les coefficients de corrélations sont très significatifs et les valeurs prédites par le modèle sont très proches des valeurs observées. Cependant comme il fallait s'y attendre, le modèle est sensible à la qualité des données d'entrée. La sensibilité du modèle aux données d'entrée peut être utilisée pour améliorer les données avant le calcul des PPA.

Mots clés: *Prix Moyens, Corrélation, Programme de Comparaison Internationale, PCI-Afrique, Modèle, Parités de pouvoir d'achat, Variation*

1. Context

The International Comparison Program (ICP) aims at generating purchasing power parity (PPP) to facilitate cross-country comparisons of GDP and its sub-aggregates in real terms without price and exchange rate distortions. To that end, national annual average prices of goods and services are required so as to match the corresponding national accounts data. As the reference year is 2005, the PPPs estimation rests on the collection of prices representative of the whole transactions carried out in 2005 within 160 countries for roughly 1000 goods and services intended for final use.

The ICP multilateral data requirements are as follows: (i) the items should be comparable and representative, represent all the components of the aggregate being compared and match on the features that affect prices; (ii) national annual average prices are used as input for PPP computation. The last requirement suggests a national coverage and a data collection during the entire period of the reference year. If these conditions are not met then some adjustments are needed to come up with estimates of annual national average prices.

In most African countries participating in the 2005 ICP round, the collection of prices of goods and services consumed by households started in the third quarter of 2005. As national annual averages are required to compute PPPs, an obvious question is how to deal with the missing data. The question can be addressed in two different ways: (i) no action is taken and the available average prices are used to compute PPPs as if they were annual averages, (ii) estimate the missing prices of the months when the data collection has not started. The first solution would be acceptable if the inflation rate is very small or even equal to zero as it is implicitly assuming the estimation of the first semester price values by the rest of the 2005 data. Also the estimators are not known and we therefore cannot assess their properties. As shown in appendix A, the assumption of zero to low inflation rates does not stand in many countries participating in the ICP-Africa.

Deciding on the second alternative, the estimation of the missing prices can then be done in several ways. Possible procedures are: (i) a backward projection, which consists of estimating the ICP prices from the data collected within the framework of the Consumer Price Index (CPI) compilation, or from information collected from centralized sources (administrative prices, car dealers, telecommunication, water and electricity services, etc.); (ii) a back extrapolation using the CPI indices and the collected prices, and/or (iii) use of econometric methods. The implementation of the first method requires matching the ICP and the CPI baskets of goods and services and determining the products common to both of them. The number of common products is at most 30% of the number of products of the ICP list in most countries. Supplementary surveys are then necessary to assess price changes between the two semesters of 2005, and this is the major constraint to using this method. The second method rests on the quality of the CPI indices and the back extrapolated data. The feasibility of the third method depends on the availability and quality of the information needed to build a model in view of predicting the prices of the months prior to the ICP data collection in 2005.

The ICP-Africa data collection, which started in the second semester of 2005 in most African countries, continued in 2006. An econometric approach using the 2005 second semester data and the 2006 data to carry out the estimation of the price data (prior to the ICP data collection in 2005) is considered. However the two data sets are disjoint in time, especially in the period (first semester of 2005), for which data need to be estimated. Products common to both ICP and CPI have been determined in some countries and their 2005 price recorded. These prices can be used to link

the two data sets. When the price information on products common to both ICP and CPI is not available, the 2005 CPI monthly indices can be used to back extrapolate the ICP data of the first month of collection so as to link the 2005 and 2006 data sets. The back extrapolation need to be done for all months of data collection of the 2005 first semester. To that end, detailed CPI indices are needed and the more detail (products, classes and groups of products) the better.

2. A Model Based Method

The method is model based and uses all the available price data of the reference year 2005, the 2006 price data and the 2005 price data of products common to both ICP and CPI, or the back extrapolated data using the CPI indices and the first month of data collection. A simple model is used to explain the total average price variation: the total average price variation is due to three factors, product, month and year. It is assumed that the average month effect is the same for a group of homogeneous products. The same assumption is made for the factor year. The 113 basic headings are grouped into 7 categories of 'similar' products/services as indicated in the table below.

Table 1: Groups of Basic Headings

Group	**Basic Heading**
Food and Beverage	1-43
Clothing and Footwear	44-54
Housing, Water and Electricity	55-60
Furniture and Housing Equipment	61-79, 108-113
Transport and Communication	80-93
Recreation	94-104
Restaurants and Hotels	105-107

The model is applied to each group separately and the end results are merged to get the estimated prices for each participating country.

2.1 Back Extrapolation and Model Specification

The CPI index is a measure of change in the prices of many commodities from the reference period to another period. For most practical purposes, it can be regarded as a 'weighted mean' of the change in the relative prices of commodities under consideration in the two periods. For a single product, it is just the price relative of the prices in the two periods:

$$I_t = \frac{P_t}{P_0} \qquad (1)$$

where P_t and P_0 are the prices in period t and in the reference period 0 respectively. Then it can be established that

$$P_{t-i} = \frac{I_{t-i}}{I_t} P_t \qquad (2)$$

When the prices of products common to ICP and CPI are not available, the CPI monthly indices can be used as indicated in the above formula to determine the ICP prices in months prior to the ICP data collection. In practice, equation (2) was used with the index at the available level (product, subclass, class or group) as the detail of that information varies from country to country.

As indicated the assumption that the total average price variation is the sum of a variation due to products, a variation due to months and a variation due to years is used to specify a model. In view of doing the estimation, the effects of the different factors Product, Month, and Year need to be estimated. Least squares with dummy variables can be used. However prices tend to vary by proportion. This implies that standard deviations are proportional to means, and thus variances are proportional to the squares of means. As a consequence, the observed price averages do not have the same variance and thus we cannot use the Ordinary Least Squares (OLS) to do the estimation. We can either look for a transformation to stabilize the variance and then use OLS as heteroscedasticity is one of the requirements of OLS.

Let x be the average price and *m* be its expected value. Let g be the trans formation such that $\sigma^2_{g(x)} = Constante = C^2$. Then assuming that g is a derivable function, we have

$$g(x)-g(m)\approx g'_m(x-m) \quad \text{then}$$

$$g(x)-g(m)\approx \underset{2}{g'_m}(x-m)$$

$$E\left(g(x)-g(m)\right)^2\approx g^{\wr}(x-m)^2 \qquad (3)$$

$$\sigma^2_{g(x)}=\left(\frac{dg}{dm}\right)^2\sigma^2_x=C^2$$

If the variance is proportional to the square of the mean

$$(\sigma^2_x \approx m^2), \text{ then } \left(\frac{dg}{dm}\right)^2=\frac{C^2}{m^2} \text{ or}$$

$$g(m) = \int\frac{Cdm}{m} = CLog(m) + Cte \qquad (4)$$

Equation (4) indicates that the logarithmic transformation is to be used. We do not know the values of the means but we can use the observed price means as a first approximation.

An alternative to using the logarithmic transformation is to use a weighted least squares using the reciprocal of the squares of price as weights.

In the model specification, the logarithm of the average prices (or the average prices) are regressed against three sets of dummy variables: one set contains a dummy for each product, the second set a dummy for each month and the third set a dummy for each year. The dependent variable is the natural log of the monthly average prices (or the monthly average prices) and the independent variables are the three sets of dummy variables. If average prices are used as dependent variable then weighted regression should be used with $1/(Average\ Price)^2$ as weights.

As a consequence of having a dummy variable for each product, the model is fit without an intercept and with one less dummy than the number of months and years for respectively the second and third sets of dummy variables. The model can be specified as:

$$Ln(P_{ijk}) = \alpha_1 X_{i1} + \alpha_2 X_{i2} + \cdots + \alpha_N X_{iN} + \beta_1 Y_{1j} + \beta_2 Y_{2j} + \cdots + \beta_{11} Y_{11\,j} + \gamma_k Z_k + \varepsilon_{ijk} \quad (5)$$

or

$$P_{ijk} = \alpha_1 X_{i1} + \alpha_2 X_{i2} + \cdots + \alpha_N X_{iN} + \beta_1 Y_{1j} + \beta_2 Y_{2j} + \cdots + \beta_{11} Y_{11\,j} + \gamma_k Z_k + \varepsilon_{ijk} \quad (6)$$

With $1/P_{ijk}^2$ as weights
where i = 1, 2,3, ……..12 (Monthly data collection)
j = 1, 2, …, N N is the number of items observed in the country out of the 853 items of the ICP, k = 1, 2 (Year).

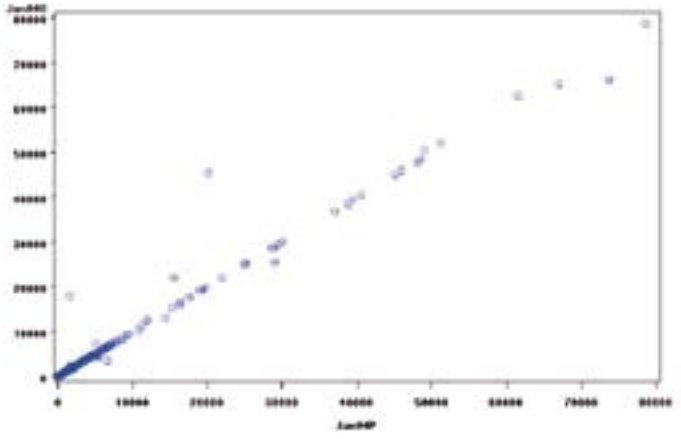
P_{ijk} is the average price of the j^{th} item in month i during year k. $Ln(P_{ijk})$ is the natural logarithm of P_{ijk} ε_{ijk} is a random error, X_{ij} and Y_{ij} are two sets of dummy variables. Each of the N items observed in the country out of the 853 products of the ICP list is represented by an X dummy variable, each of the 12 months except one (which can be chosen to be any one among the 12 months) is represented by a Y dummy variable, and one of the years (2005 or 2006) is represented by a Z dummy variable. The values of Xs, Ys and Z are zero or one.

2.2 Model Fitting and Testing

For purpose of testing the model, the 2004 and 2005 CPI data of Togo and the 2005 and the 2006 CPI data of Senegal are used. The various inflation rates are low (1.7%) for Senegal and medium (6.8% for Togo). The ICP-Africa situation is simulated by retaining the second semester of 2004, the first semester of 2005 CPI data of Togo and the second semester of 2005 and the first semester of 2006 of Senegal. The average prices of January, February, March, April and May were deleted for year 2004 for Togo and 2005 for Senegal so as to simulate the ICP-Africa situation. Their indices and the June data and index are then used to back extrapolate and fill the gaps left by the deleted prices.

For both countries, the model was then fit to the data sets with a dummy variable for each product, a dummy variable for each month except December and a dummy variable for the year 2005. The purpose is to estimate the missing observations in the first months of 2004 and 2005 for Togo and Senegal respectively. To that end, predicted values from the model are used. The model fits the data very well in both cases. The plot below presents a graph of predicted values against the observed values for the month of January 2004 for Togo.

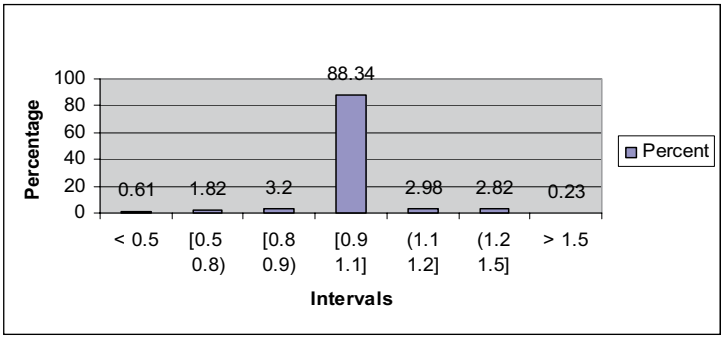
Figure 1: Togo's 2004-2005 CPI data - Comparison of January 2004 observed and predicted prices (that are less than or equal to 100,000 CFA)



The correlation coefficient between the observed and predicted prices in the Togo data set is 0.98 (n= 356, P < 0.01). The 45 degree line indicates that the observed and the predicted values are very close for most products.

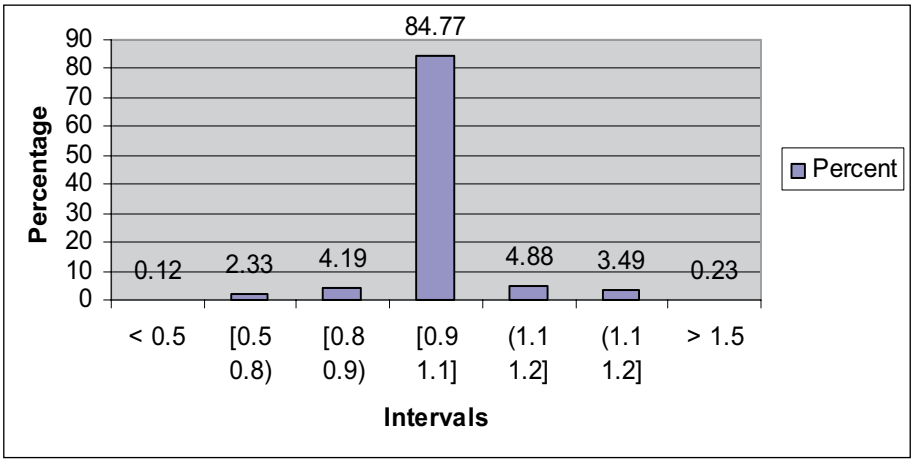
To assess the quality of the fit, the ratios of the predicted values to the observed values and the correlation between the two sets of values were used. A distribution of these rations indicates that 88.34% of the predicted values are within ±0.1 of the observed values and 94.52% are within ± 0.2. The correlation between the predicted and the observed values is 0.999 (n = 1810 P < 0.0001).

Figure 2: Togo's 2004-2005 CPI data – Distribution of ratios of predicted to observed prices



In Senegal, the observed and predicted price values have also a correlation coefficient of 0.999 (n=860, P < 0.0001), 84.77% of the ratios of the predicted to the observations values are within ±0.1 and 93.84% are within ± 0.2.

Figure 3: Senegal's 2005-2006 CPI data: Distribution of ratios of predicted to observed prices

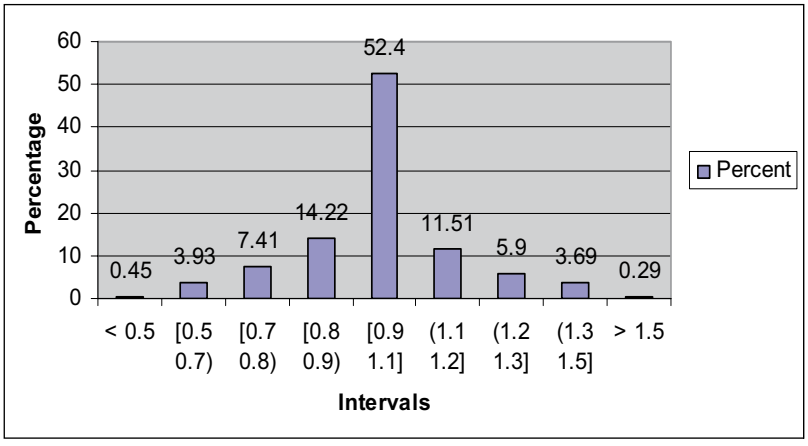


Thus, in both cases, the model capability to predict the missing price information is very good.

2.3 Applying the Model to ICP Data

In view of further testing the model, an ICP-Africa data set from one of the participating country is used. The country is referred to as Country1. Its data set consists of partially validated monthly average data from June 2005 to September 2006. The observations of June, July, August and September 2005 were deleted and their CPI indices were used to back extrapolate the October 2005 data for each of the deleted month. The model was fit to the data set consisting of observed values from October 2005 to September 2006 and back extrapolated values for the month of June, July, August and September so as to simulate the ICP situation on hands. The ratios of the predicted to the observed values were computed and the distribution is presented in the plot below.

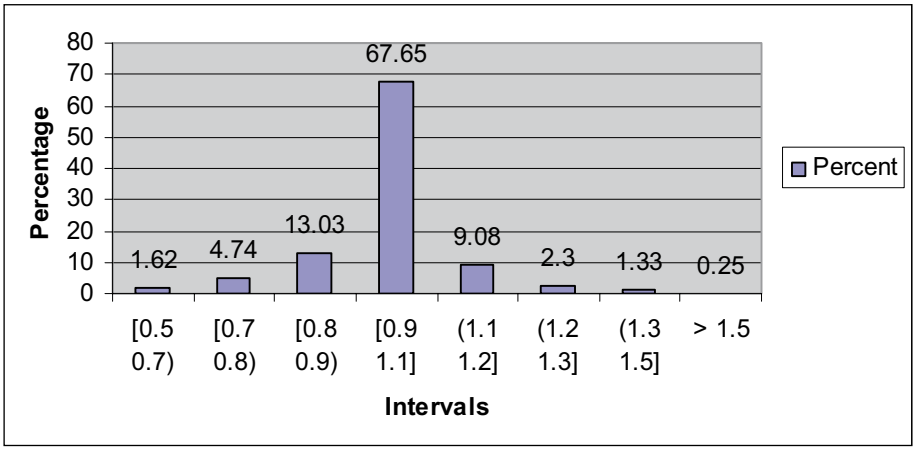
Figure 4: Distribution of ratios of predicted to observed prices of a country's ICP data at an early stage of data validation



The correlation between the observed and the predicted values is 0.99 (n=2441, P < 0.0001), 52% of the predicted values are within ± 0.1 of the observed values, 78.13% are within ± 0.2, and 91.44% are within ± 0.3.

A second county referred to as Country 2 with more elaborated validated data is considered. The ratios of the predicted to the observed values of average prices of the months of June, July, August and September 2005 were computed. The correlation between the observed and the predicted values is 0.999 (n=2785, P < 0.0001). The distribution of the ratios indicates that 67.65% of the ratios are in the interval [0.9, 1.1], 89.76% are in the interval [0.8, 1.2] and 96.80% are in the interval [0.7 1.3]. Comparisons between observed and predicted values of some average prices of the months of June, July, August and September 2005 are presented in appendix B. Points on and near the 45 degree line indicate that the observed and the predicted values are very close for most products.

Figure 5: Distribution of ratios of predicted to observed prices of a country's ICP data at an advanced stage of data validation



The average price predictions presented in figures 4 and 5 clearly indicate that the quality of the predictions is heavily dependent on the quality of the input data. This point cannot be overemphasized. It is therefore important to fully and thoroughly validate the data if this method is to be used.

3. Conclusion

The estimation of the 2005 first semester data using CPI data and data collected during the second semester of 2005 is feasible for countries with low inflation rate. The inflation rate in 2005 is low (under 5%) in some countries participating in this ICP round. Those countries would qualify for such procedure. However, the number of common products to both ICP and CPI in most of those countries is small and supplementary surveys would be necessary to assess price changes between the two semesters of 2005. This is a major constraint to using this method.

Considering the available average prices only in the computation of PPPs as if they were annual averages is also sensible for countries with very low inflation rate. But it is implicitly assuming the estimation of the first semester price values by the rest of the 2005 data. The estimators are not known and we therefore cannot assess their properties.

The proposed model based approach is built on the basis of the available information: the 2005 and 2006 ICP price data and some of the CPI 2005

data (products common to both ICP and CPI). When the price information on products common to both ICP and CPI is not available, the 2005 CPI monthly indices can be used to back extrapolate the ICP data of the first month of collection so as to link the 2005 and 2006 data sets. Testing the model using CPI and ICP data sets provide some useful results but indicate that the model is sensitive to the quality of the input data. This sensitivity can be used to fix problems in the input data. If estimation looks odd, chances are high that there is something odd with the input data.

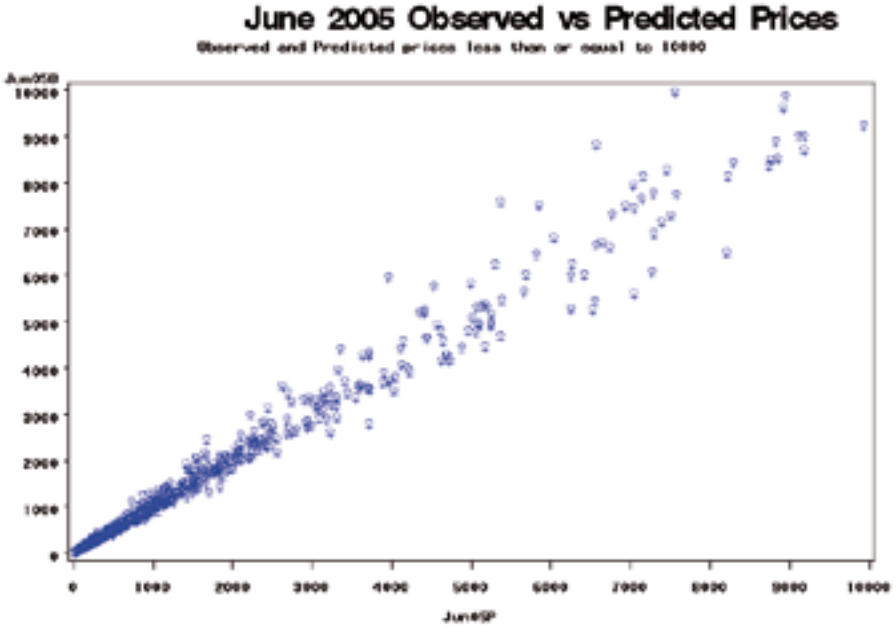
The quality of the estimations in all methods depends heavily on the quality of the available ICP and CPI data. All methods require a review of the CPI list to determine products which are common to ICP and CPI. This review provides some insights into the integration of ICP and CPI in terms of the lists of products.

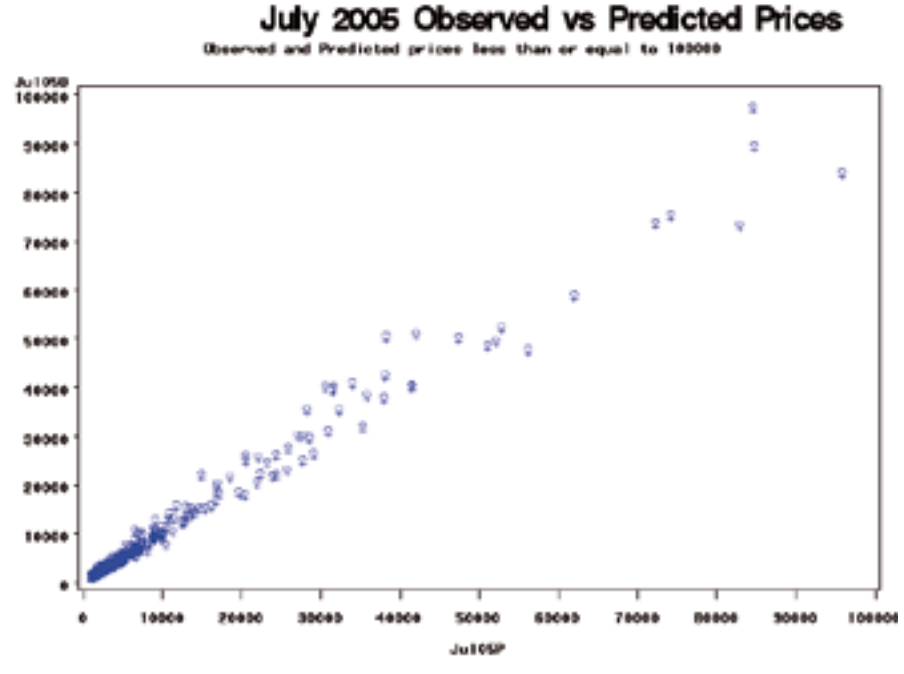
Appendix A: Consumer Price – Inflation (%)[1]

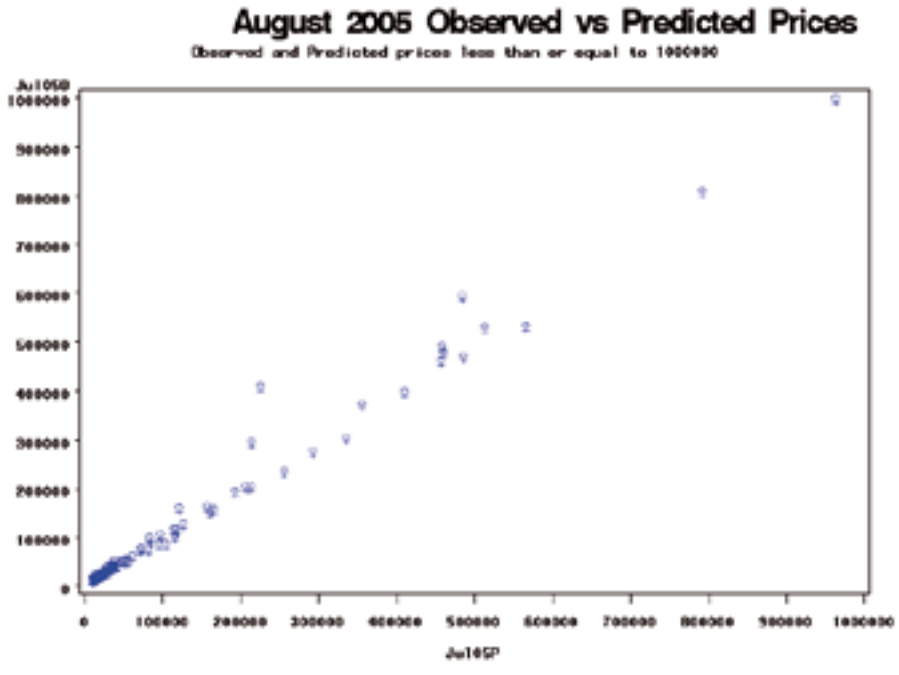
COUNTRY	INFLATION	COUNTRY	INFLATION
ALGERIA	3.9	LIBYA	...
ANGOLA	28.3	MADAGASCAR	25.2
BENIN	2.6	MALAWI	14.8
BOTSWANA	6.9	MALI	3.4
BURKINA FASO	3.4	MAURITANIA	16.5
BURUNDI	19.2	MAURITIUS	16.5
CAMEROON	2.0	MOROCCO	1.2
CAPE VERDE	-0.6	MOZAMBIQUE	5.7
CENT. AFR. REP.	3.8	NAMIBIA	1.8
CHAD	7.2	NIGER	6.7
COMOROS	3..5	NIGERIA	....
CONGO	2.0	RWANDA	12.1
CONGO (DRC)	15.5	SAO T. PRINC.	...
COTE D'IVOIRE	4.9	SENEGAL	0.5
DJIBOUTI	3.5	SEYCHELLES	....
EGYPT	6.7	SIERRA LEONE	13.9
EQUAT. GUINEA	7.6	SOMALIA	....
ERITREA	...	SOUTH AFRICA	3.0
ETHIOPIA	...	SUDAN	...
GABON	-0.3	SWAZILAND	3.6
GAMBIA	-5.3	TANZANIA	...
GHANA	13.0	TOGO	6.6
GUINEA	26.8	TUNISIA	1.4
GUINEA BISSAU	3.9	UGANDA	10.5
KENYA	17.9	ZAMBIA	18.4
LESOTHO	4.1	ZIMBABWE	128.0
LIBERIA	....		
AFRICA			**7.9**

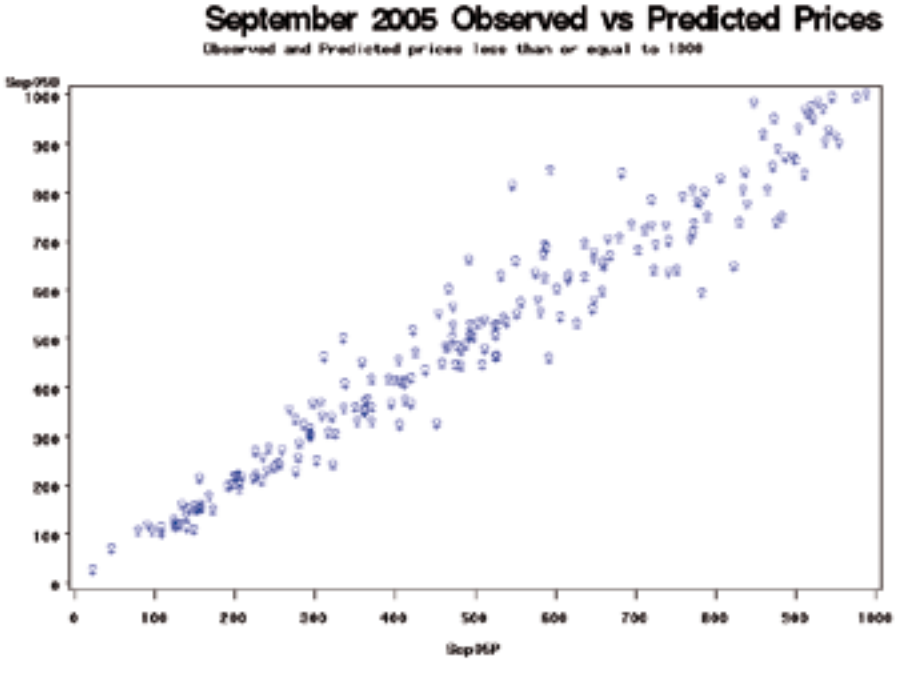
[1] Source : Country Report and ADB Statistics Department Estimates

Appendix B Comparison of observed and predicted prices of a country ICP data at an advanced stage of DATA Validation






August 2005 Observed vs Predicted Prices
Observed and Predicted prices less than or equal to 1000000
Jul05P


September 2005 Observed vs Predicted Prices
Observed and Predicted prices less than or equal to 1000
Sep05P

References

Gregory C. Chow (1983). Econometrics Mc Graw-Hill Book Company.

Irving B. Kravis, Zoltan Kenessey, Alan Heston and Robert Summers (1975). A system of international comparisons of gross product and purchasing power, The Johns Hopkins University Press. Baltimore and London.

Kirk M. Wolter, (1985). Introduction to Variance Estimation. Springer-Verlag New York Inc.

Michel Mouyelo-Katoula and Kantilal Munnsad. (1993). Comparison of price levels and economic aggregates - The results of 22 African countries.

Rudolf J. Freund, Ramon C. Littell, (1986). SAS System for Regression, SAS Series in Statistical Applications, SAS Institute Inc.

Snedecor and Cochran (1980) Statistical Methods. The Iowa State University Press.

The World Bank, (2004) ICP 2003-2006 Handbook, available at http://www.worldbank.org/data:icp