# [02.01]

# Report of the 2011 ICP Preliminary Results


7TH Technical Advisory Group Meeting

September 17-18, 2012

Washington, DC

## 1. Themes Covered

- Versions of CPD
- A data validation algorithm for CPD prices
- Results/Problem Areas in ICP 2011

## 2. Versions of CPD

Assume there are three countries and four products

Country 1 prices products 1 and 4. Product 1 is important.

Country 2 prices products 1, 2 and 3. Product 2 is important.

Country 2 prices products 2, 3 and 4. Products 2 and 4 are important.

- *Standard CPD*

The $(X'X)\hat{\beta} = X'y$ system takes the following form:

$$\left(\begin{array}{ccc|cccc} 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 3 & 0 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 2 \end{array}\right) \left(\begin{array}{c} \pi_1 \\ \pi_2 \\ \pi_3 \\ \hline \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{array}\right) = \left(\begin{array}{c} \Sigma_{n=1}^4 \ln p_{n1} \\ \Sigma_{n=1}^4 \ln p_{n2} \\ \Sigma_{n=1}^4 \ln p_{n3} \\ \hline \Sigma_{k=1}^3 \ln p_{1k} \\ \Sigma_{k=1}^3 \ln p_{2k} \\ \Sigma_{k=1}^3 \ln p_{3k} \\ \Sigma_{k=1}^3 \ln p_{4k} \end{array}\right)$$

- *Importance-Weighted CPD*

$$\left(\begin{array}{ccc|cccc} 2 & 0 & 0 & 2 & 0 & 0 & 1 \\ 0 & 4 & 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 5 & 0 & 2 & 1 & 2 \\ \hline 2 & 1 & 0 & 3 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 & 4 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 2 & 0 \\ 1 & 0 & 2 & 0 & 0 & 0 & 3 \end{array}\right)\left(\begin{array}{c} \pi_1 \\ \pi_2 \\ \pi_3 \\ \hline \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{array}\right) = \left(\begin{array}{c} \Sigma_{n=1}^{4} w_{n1} \ln p_{n1} \\ \Sigma_{n=1}^{4} w_{n2} \ln p_{n2} \\ \Sigma_{n=1}^{4} w_{n3} \ln p_{n3} \\ \hline \Sigma_{k=1}^{3} w_{1k} \ln p_{1k} \\ \Sigma_{k=1}^{3} w_{2k} \ln p_{2k} \\ \Sigma_{k=1}^{3} w_{3k} \ln p_{3k} \\ \Sigma_{k=1}^{3} w_{4k} \ln p_{4k} \end{array}\right),$$

where $w_{nk} = 0$ if product $n$ is missing in country $k$, it equals 1 if it is present but not important, and it equals 2 if it is present and important.

- *Availability-Weighted CPD*

Given that the price quotes from country 1 are scarcer, it could be argued that each of its prices should be given more weight. Below we demonstrate one way of doing this.

$$\left(\begin{array}{ccc|cccc} 1 & 0 & 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1 & 0 & 1/3 & 1/3 & 1/3 \\ \hline 1/2 & 1/3 & 0 & 5/6 & 0 & 0 & 0 \\ 0 & 1/3 & 1/3 & 0 & 2/3 & 0 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 2/3 & 0 \\ 1/2 & 0 & 1/3 & 0 & 0 & 0 & 5/6 \end{array}\right)\left(\begin{array}{c} \pi_1 \\ \pi_2 \\ \pi_3 \\ \hline \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4 \end{array}\right)=\left(\begin{array}{c} \Sigma_{n=1}^4 w_{n1} \ln p_{n1} \\ \Sigma_{n=1}^4 w_{n2} \ln p_{n2} \\ \Sigma_{n=1}^4 w_{n3} \ln p_{n3} \\ \hline \Sigma_{k=1}^3 w_{1k} \ln p_{1k} \\ \Sigma_{k=1}^3 w_{2k} \ln p_{2k} \\ \Sigma_{k=1}^3 w_{3k} \ln p_{3k} \\ \Sigma_{k=1}^3 w_{4k} \ln p_{4k} \end{array}\right)$$

- *Availability-Importance-Weighted CPD*

$$
\left(\begin{array}{ccc|cccc}
1 & 0 & 0 & 2/3 & 0 & 0 & 1/3 \\
0 & 1 & 0 & 1/4 & 2/4 & 1/4 & 0 \\
0 & 0 & 1 & 0 & 2/5 & 1/5 & 2/5 \\
\hline
2/3 & 1/4 & 0 & 2/3+1/4 & 0 & 0 & 0 \\
0 & 2/4 & 2/5 & 0 & 2/4+2/5 & 0 & 0 \\
0 & 1/4 & 1/5 & 0 & 0 & 1/4+1/5 & 0 \\
1/3 & 0 & 2/5 & 0 & 0 & 0 & 1/3+2/5
\end{array}\right)
\left(\begin{array}{c}
\pi_1 \\ \pi_2 \\ \pi_3 \\ \hline \eta_1 \\ \eta_2 \\ \eta_3 \\ \eta_4
\end{array}\right)
$$

$$
= \left(\begin{array}{c}
\Sigma_{n=1}^{4} w_{n1} \ln p_{n1} \\
\Sigma_{n=1}^{4} w_{n2} \ln p_{n2} \\
\Sigma_{n=1}^{4} w_{n3} \ln p_{n3} \\
\hline
\Sigma_{k=1}^{3} w_{1k} \ln p_{1k} \\
\Sigma_{k=1}^{3} w_{2k} \ln p_{2k} \\
\Sigma_{k=1}^{3} w_{3k} \ln p_{3k} \\
\Sigma_{k=1}^{3} w_{4k} \ln p_{4k}
\end{array}\right)
$$

The weights $w_{nj}$ used in $X'y$ now vary for each country-product over the range 0, 1/5, 1/4, 1/3, 2/5, 1/2, and 2/3.

## 3. An Algorithm for Validating the CPD Price Indexes

The algorithm selects each country in a region in turn as the base $b$ and then proceeds as follows.

Let $P^i_{bk}$ denote the CPD price index for basic heading $i$ in country $k$ with country $b$ as the base.

For country $k$ find the maximum ratio of its CPD price indexes as follows:

$$MaxCPDRatio_k = \max_{i,j=1,\ldots,I}\left(\frac{P^i_{bk}}{P^j_{bk}}\right).$$

If $MaxCPDRatio_k \leq Z$, then move on to the next country.

If $MaxCPDRatio_k > Z$, then we conclude that there may be a problem with the price data of one of the following four country-basic headings:

(i) Heading $max$ in the base country $b$.

(ii) Heading $max$ in country $k$.

(iii) Heading $min$ in the base country $b$.

(iv) Heading $min$ in country $k$.

The remainder of the algorithm is focused on establishing which of these four country-basic headings needs checking.

Calculate the geometric mean of the basic heading price indexes for country $k$, excluding the *max* and *min* headings, as follows:

$$GM_{bk} = \prod_{i \neq max\ or\ min}^{I} (P_{bk}^{i})^{1/(I-2)},$$

$$\text{Now let} \quad A = \frac{P_{bk}^{max}}{GM_{bk}}, \quad B = \frac{GM_{bk}}{P_{bk}^{min}}.$$

If $A > B$ we conclude that it is one of the *max* headings that needs checking [i.e., (i) or (ii)].

If $A < B$ we conclude that it is one of the *min* headings that needs checking [i.e., (iii) or (iv)].

In what follows we will assume that $A > B$. The algorithm proceeds in an analogous way if $A < B$.

Calculate the geometric mean of the country price indexes for the basic heading $max$, both including and excluding country $k$ as follows:

$$GM_{max} = \prod_{j=1}^{K} (P_{bj}^{i})^{1/K}, \quad GM_{max}/k = \prod_{j \neq k}^{K} (P_{bj}^{i})^{1/(K-1)}.$$

Now calculate

$$C = \frac{GM_{max}/k}{GM_{max}}.$$

If $C < 1/2$ we conclude that it is country-basic heading (iv) that needs checking. If $C > 1/2$ we conclude that it is country-basic heading (iii) that needs checking.

The selected country-basic heading is now deleted, and the algorithm is rerun.

This process continues until the inequality $MaxCPDRatio_k \leq Z$ is satisfied.

The algorithm then moves on to the next country in the region.

Once all countries have been checked, the algorithm updates the base country and repeats the process.

The algorithm terminates once all countries have been used as the base country.

## 4. Results for ICP 2012

- *Validation of CPD Price Indexes*

Our results for the ICP 2011 data, with $Z = 40$ are shown in Table 1.

No outliers are detected in the Asia and CIS regions.

481 are detected in Africa

5 in Eurostat-OECD

82 in Latin America

46 in Western Asia.

# Table 1. Outlier Country-Basic Heading Prices Detected by Our Algorithm

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Africa** | | | | | | | | | |
| 1101112 | GNQ | NGA | | | | | | | |
| 1101113 | DZA | KEN | NGA | | | | | | |
| 1101115 | SLE | TUN | | | | | | | |
| 1101121 | EGY_A | NGA | TZA | | | | | | |
| 1101123 | EGY_A | ETH | GAB | KEN | MUS | | | | |
| 1101125 | AGO | BWA | COM | CPV | GIN | KEN | | | |
| 1101131 | CIV | KEN | | | | | | | |
| 1101132 | CPV | ETH | | | | | | | |
| 1101142 | GNQ | KEN | | | | | | | |
| 1101143 | AGO | BFA | ETH | KEN | | | | | |
| 1101151 | KEN | | | | | | | | |
| 1101161 | BFA | CIV | GMB | | | | | | |
| 1101162 | CIV | | | | | | | | |
| 1101171 | EGY_A | ETH | KEN | NGA | UGA | | | | |
| 1101172 | CPV | ETH | GMB | LBR | MRT | NAM | NGA | SDN_A | SEN |
| 1101173 | NGA | | | | | | | | |
| 1101181 | LBR | STP | SYC | | | | | | |
| 1101182 | ETH | | | | | | | | |
| 1102111 | CAF | GNQ | SEN | ZAF | | | | | |

Applying similar reasoning to the between-region links, two region-basic heading price indexes stand out as big outliers (on the high side). These are:

Africa: 1104511 (Electricity)

CIS: 1104511 (Electricity)

- *Anomalies in the Aggregate Level Results*

(i) Individual Countries

**Africa**:

Ethiopia (too low)

Liberia (too high)

# Table 2. Household Consumption Value Shares (Standard CPD Method)

| | CAR Shares | CAR Rescaled USA = 1 | CAR Per Capita USA=1 | Unfixed GEKS Shares | Within Region GEKS Shares |
|---|---|---|---|---|---|
| AGO | 0.0005921 | 0.0024320 | 0.0386257 | 0.0006126 | 0.0221191 |
| BDI | 0.0000653 | 0.0002681 | 0.0097411 | 0.0000670 | 0.0024382 |
| BEN | 0.0001972 | 0.0008101 | 0.0277401 | 0.0001935 | 0.0073684 |
| BFA | 0.0002115 | 0.0008686 | 0.0159511 | 0.0002155 | 0.0079003 |
| BWA | 0.0001516 | 0.0006225 | 0.0955204 | 0.0001586 | 0.0056621 |
| CAF | 0.0000842 | 0.0003459 | 0.0240219 | 0.0000820 | 0.0031461 |
| CIV | 0.0003840 | 0.0015774 | 0.0243895 | 0.0004125 | 0.0143472 |
| CMR | 0.0005880 | 0.0024153 | 0.0375719 | 0.0005844 | 0.0219674 |
| COD | 0.0003034 | 0.0012461 | 0.0057303 | 0.0003073 | 0.0113334 |
| COG | 0.0001000 | 0.0004108 | 0.0309202 | 0.0001038 | 0.0037363 |
| COM | 0.0000031 | 0.0000128 | 0.0052940 | 0.0000021 | 0.0001165 |
| CPV | 0.0000355 | 0.0001460 | 0.0908592 | 0.0000365 | 0.0013276 |
| DJI | 0.0000179 | 0.0000736 | 0.0253382 | 0.0000194 | 0.0006698 |
| DZA | 0.0043021 | 0.0176717 | 0.1530386 | 0.0046140 | 0.1607279 |
| EGY_A | 0.0020331 | 0.0083513 | 0.0315278 | 0.0020504 | 0.0759569 |
| ETH | 0.0000729 | 0.0002994 | 0.0011009 | 0.0000721 | 0.0027228 |
| GAB | 0.0000646 | 0.0002653 | 0.0538818 | 0.0000391 | 0.0024131 |
| GHA | 0.0005556 | 0.0022820 | 0.0284816 | 0.0003539 | 0.0207557 |
| GIN | 0.0001565 | 0.0006427 | 0.0195929 | 0.0001578 | 0.0058459 |

**Latin America**:

Panama (too high)

Uruguay (too high)

Cost Rica (too high)

Cuba (too high)

Dominican Republic (too high)

**Western Asia**:

UAE (too high)

Kuwait (too low)

Saudi Arabia (too low)

Qatar (too low)

Liberia has the second highest per capita income in the Africa region (after the Seychelles). In contrast, in ICP 2005 Liberia was one of the poorest countries in Africa.

Ethiopia in our results is 10 times poorer than any other country in Africa. By comparison, in ICP 2005, while still poor, it was not the poorest.

In Latin America, we find that Uruguay, Costa Rica, Cuba and Dominican Republic all have per capita incomes more than 50 percent higher than Brazil's.

Panama's per capita income is more than 4 times that of Brazil.

In Western Asia, the UAE comes out more than 7 times richer per capita than any other country.

The per capita incomes in Kuwait, Saudi Arabia and Qatar all seem too low.

(ii) Between-Region Links

In comparison with the Eurostat-OECD region, per capita incomes in Western Asia, Africa and CIS seem too low.

Example: Based on CIS comparison, Russia's per capita income is 0.101 (USA = 1).
Based on Eurostat-OECD comparison, Russia's per capita income is 0.283 (USA = 1).

## (iii) Standard Versus Weighted CPD

Can use the following metric to see which countries' results change the most when importance weights are used:

$$D_k^{AB} = \frac{|s_k^A - s_k^B|}{(s_k^A + s_k^B)/2}.$$

The biggest change is observed for Ghana, followed by Zambia and Bahrain.

Ghana identified more products as important than any other country in Africa barring Nigeria.

Neither Zambia nor Bahrain were unusual in terms of the number of products they identified as important.

An overall measure of the change in the results from one method to the next is provided by the following metric:

$$M_{AB} = \frac{1}{K} \sum_{k=1}^{K} D_k^{AB}.$$

A comparison of the standard CPD and importance-weighted CPD methods yields a $M_{AB}$ coefficient of 0.0178.

# Table 3. Household Consumption Value Shares (Importance-Weighted CPD Method)

| | CAR Shares | CAR Rescaled USA = 1 | CAR Per Capita USA=1 | Unfixed GEKS Shares | Within Region GEKS Shares | $D_k^{AB}$ |
|---|---|---|---|---|---|---|
| AGO | 0.0005937 | 0.0024342 | 0.0386608 | 0.0006154 | 0.0220111 | 0.0027805 |
| BDI | 0.0000668 | 0.0002737 | 0.0099460 | 0.0000681 | 0.0024751 | 0.0226854 |
| BEN | 0.0001997 | 0.0008189 | 0.0280403 | 0.0001962 | 0.0074050 | 0.0126342 |
| BFA | 0.0002080 | 0.0008527 | 0.0156582 | 0.0002042 | 0.0077104 | 0.0166593 |
| BWA | 0.0001510 | 0.0006192 | 0.0950064 | 0.0001560 | 0.0055990 | 0.0035218 |
| CAF | 0.0000846 | 0.0003468 | 0.0240819 | 0.0000806 | 0.0031357 | 0.0043670 |
| CIV | 0.0003831 | 0.0015705 | 0.0242825 | 0.0004115 | 0.0142016 | 0.0025215 |
| CMR | 0.0005890 | 0.0024151 | 0.0375688 | 0.0005808 | 0.0218385 | 0.0017890 |
| COD | 0.0003112 | 0.0012760 | 0.0058679 | 0.0003144 | 0.0115384 | 0.0256037 |
| COG | 0.0000997 | 0.0004086 | 0.0307539 | 0.0001034 | 0.0036947 | 0.0035207 |
| COM | 0.0000031 | 0.0000127 | 0.0052451 | 0.0000021 | 0.0001148 | 0.0074213 |
| CPV | 0.0000357 | 0.0001462 | 0.0910149 | 0.0000365 | 0.0013222 | 0.0035855 |
| DJI | 0.0000178 | 0.0000730 | 0.0251228 | 0.0000192 | 0.0006602 | 0.0066631 |
| DZA | 0.0044357 | 0.0181861 | 0.1574932 | 0.0047459 | 0.1644493 | 0.0305631 |
| EGY_A | 0.0019965 | 0.0081857 | 0.0309024 | 0.0020224 | 0.0740196 | 0.0181597 |
| ETH | 0.0000794 | 0.0003256 | 0.0011975 | 0.0000767 | 0.0029446 | 0.0859313 |
| GAB | 0.0000645 | 0.0002644 | 0.0536992 | 0.0000397 | 0.0023910 | 0.0015225 |
| GHA | 0.0004759 | 0.0019510 | 0.0243504 | 0.0002845 | 0.0176424 | 0.1545296 |

## (iv) Inequality as a Measure of Noise?

More noise in a method should lead to a higher level to measured inequality.

Hence I calculated Theil's inequality measure for both methods.

$$T = \frac{1}{K} \sum_{k=1}^{K} \left[ \frac{y_k}{\bar{y}} \ln \left( \frac{y_k}{\bar{y}} \right) \right]$$

$T_{CPD} = 0.52536$

$T_{WCPD} = 0.52496$

Is this evidence that WCPD is doing a better job?

## 5. Conclusion

- Is availability weighting worth considering?

- Our data validation algorithm might be worth trying.

- Some data problems have been identified that need checking.

- Results do not seem that sensitive to the choice between CPD and WCPD (when important products are given a weight of 2).

# Thank you