

# **Weighing things up: sources of data for CPI weights**

**Patrick Kelly and Lee Everts**

Statistics South Africa

**Paper presented at the the UNECE/ILO meeting of CPI experts**

**30 May – 1 June 2012**

## **ABSTRACT**

A survey of household expenditure (HES) typically provides the source of the weights for a Consumer Price Index. However, the HES may have a number of limitations including high levels of aggregation, coding problems, under reporting and irregular data quality. These problems may be experienced more in certain product groups than others. The normative literature suggests the use of a limited set of alternative data sources, primarily from the National Accounts, to compensate for problems in the HES. A more extensive incorporation of alternative data sources may be helpful to remedy the range of problems.

This paper examines the South African experience of changing the method of conducting the HES and challenges thrown up by this. Alternative source data including industry surveys, scanner data from retail firms and administrative data sources provide supplementary data from which to compile the final weights set.

## **A. Introduction**

The weighting structure of a CPI is critical as it determines the impact that price changes of different products and services will have on the overall rate of inflation. Most countries rely on a household expenditure survey as the primary source of weights, but these may have inherent weaknesses. In 2009 South Africa introduced new weights based on a revamped expenditure survey. Despite the improvements in the survey method, it proved necessary to adjust the results to arrive at final weights for the CPI.

This paper highlights important advice offered by the ILO manual and Practical guide to Producing CPIs. The South African experience of household expenditure surveys is discussed with a focus on reasons for possible under-reporting of expenditure. Practical examples are provided to illustrate how to adjust the survey data using auxiliary sources.

## **B. Why adjustments are needed**

The Consumer Price Index Manual: Theory and Practice and the Practical guide to producing consumer price indices both provide an overview on the use of additional data sources. Little guidance is available to CPI practitioners on the relative usefulness of different sources or the mechanics of actually making adjustments.

In most countries the Household Budget Survey (HBS) is the primary data source for CPI weights. However, the as the CPI Manual states, “Under reporting is probably the most serious and common problem affecting the HBS”. There are typically a number of areas in which underreporting occurs. These include:

- Frequently purchased items such as food and non alcoholic beverages. This may be because the respondent cannot recall the details of all purchases, or experiences ‘respondent fatigue’, from having to record extensive details over the survey period (usually 2 or 4 weeks);
- Alcoholic beverages and tobacco. Typically respondents may be embarrassed about the level of expenditures on such ‘socially unacceptable’ items and therefore under report them.

Other reporting problems may be an over reporting of expenditure on luxury items or telescoping, where large items, such as motor vehicles, may be reported as having been purchased in the survey period, when in fact they were purchased in a previous period.

This under-reporting necessitates that “results from the HES should be compared and combined with statistics from other sources when constructing CPI weights” (ILO manual).

Under reporting is not the only reason why weights from a HES may need to be adjusted. There are methodological factors that also need to be taken into account.

If the scope of the CPI is domestic, it should include expenditure by foreigners in the country. This would mainly be tourist expenditure, which normally would require adjustment to values for hotels and restaurants.

Owner occupied housing is another area in which it is not usually possible to simply extract a value from an HES<sup>1</sup>. The selection of a weight may require a series of computations, including the use of external data, to derive a weight that is in line with the method used to price housing.

A third methodological reason for adjusting the weights concerns used vehicles. The HES will record total sales of used vehicles. However, the CPI weight should reflect only net sales to the household sector. This requires that only sales to households from businesses

---

<sup>1</sup> Owner occupied housing adjustments are not discussed in this paper.

(e.g. ex fleet stock) and the dealer margin on household to household sales, should be included in the value for used vehicles.

### **C. Sources of alternative and complementary data**

The National Accounts household consumption expenditure account is proposed as a useful alternative when an HBS is not conducted frequently. However, the possibility is that because the National Accounts also draws heavily on the HBS, the estimates may also be dated. Furthermore, the National Accounts data is only available at a national level, whereas the CPI may compile elementary aggregates at geographically dispersed levels. National Accounts estimates are the result of balancing different data sources, and making a number of assumptions, that may not be consistent with the objectives of a CPI.

A useful complementary source may be data obtained from retail firms. This could either be in the form of scanner data or a detailed survey of retailers. Scanner data is mainly available from large supermarket chains, typically in more developed countries. This has been found to be a reliable data source in Iceland for example (Gudnason, 1998). An alternative to scanner data itself, may be aggregated product sales data from supermarket chains. Sales data may be available from the accounting systems of the firms and could reflect quantities and/or the value of sales per product.

Should retail firms not be willing to divulge firm or product specific information, it might be possible to rely on information collected through sample surveys of the retail sector. While these surveys have the advantage of capturing the whole range of different retail stores, they may be limited in terms of geographic level and more importantly, product detail. Some retail chains provide information to market research companies who then aggregate the information. A major drawback of this data is that it is not classified according to COICOP which makes comparisons with the HES difficult.

Drawbacks of these data sources mainly concern possible differences in the scope of the survey and firm data. These would not capture which sales are to private households (CPI scope) and which may be to businesses, government or other out-of- scope populations.

A third complementary source may be official administrative data. Government agencies may monitor the sales of certain products or levy specific taxes on products. Again, it would be necessary to make a certain assumptions to extract the value of sales to households specifically. Furthermore, geographic or product breakdowns may not be readily available from these sources.

### **D. The South African Income and Expenditure Survey and its limitations**

The weights of the South African CPI are largely based on a household survey of consumer expenditure known as the Income and Expenditure Survey (IES). The Survey is conducted once every 5 years. The current weights draw on the 2005/6 survey and the results of the 2010/11 survey will be incorporated into the weights in 2013. In 2008/09, Stats SA ran a Living Conditions Survey (LCS). This survey is aimed at measuring poverty levels, and includes a full household expenditure module.

The scope of the IES and LCS is identical to the CPI with two exceptions. Firstly, tourist expenditure (expenditure by tourists not resident in the reference country) is not included. Secondly, institutional households such as hospital and schools are not surveyed. While the exclusion of tourist expenditure is expected to create some bias in specific COICOP groups (e.g. hotels and restaurants), institutional households are not a significant feature of the South African social landscape and their exclusion is therefore not seen as posing a significant risk of bias.

The IES and LCS run for a full 12 month period so as to capture seasonal expenditure. There are two collection instruments. Firstly, each household completes an expenditure

diary. This diary records all expenditures within the survey period such as food, electricity, clothing, personal care items etc. The second instrument is a recall questionnaire. This questionnaire captures basic demographic and housing variables, income information and expenditure on durables. The IES 05/06 marked a departure from previous IES surveys which had been run according to a traditional model of a snapshot survey. In the past the IES had conducted the entire survey within one month, using only a recall questionnaire.

The change in method yielded a surprise in the form of a markedly lower level for food expenditure. This difference was not only noticeable when compared to previous IES results but also when compared to the National Accounts estimate for household consumption on food - although this admittedly partly relied on the earlier IES. Alcoholic beverages and tobacco was the second category that appeared to under report when compared to the national accounts (but not previous IESs). Both of these are in line with the international experience as discussed above. However, no noticeable underreporting on luxury or 'big ticket' items was observed.

A further identified problem with both the IES and LCS is an inability to capture the finer details of expenditure within product groups. For example, the level of expenditure on men's clothing may be realistic, but the breakdown between shirts, trousers etc is not. This problem is most marked in food where there are a high number of products purchased by consumers and it is necessary to ensure granular detail in the weights and basket.

In order to address this problem, data is sourced from retail firms. This is alike to scanner data, but without reference to brand, size and the like. It has been possible to source information on the sales of the top 1000 products sold across a range of retail chains.

## E. Practical example of adjusting weights

South Africa first used external data sources to make adjustments to weights in 2008. Preparations are currently underway for the release of the new set of weights aimed for implementation in 2013. In order to show clearly how weights might be adjusted, an adjustment exercise was conducted using the South African Living Conditions Survey (LCS).

**Table 1: Summary of reasons for adjustments and sources of data**

Group	Reason	External data sources used
Food and non-alcoholic beverages (COICOP 01)	Under-reporting in the LCS	Retail trade survey, Large sample survey (LSS)
Alcoholic beverages and tobacco (COICOP 02)	Under-reporting in the LCS	Liquor Boards, Large sample survey of Retail Trade (LSS)
Used vehicles (COICOP 071)	Use of the net weight	National Accounts Private Consumption Expenditure (PCE)
Car hire (COICOP 07)	Adjusting for tourism expenditure	South African Vehicle Rentals and Leasing Association (SAVRALA)
Package holidays (COICOP 096)	Adjusting for tourism expenditure	Tourism satellite account (TSA)
Restaurants and hotels (COICOP 11)	Adjusting for tourism expenditure, as well as under-reporting	Accommodation; Food and beverages; and Domestic tourism surveys

### 1. Food and non-alcoholic beverages

For the 2008 reweighting, the food and non-alcoholic beverages category was adjusted for under-reporting using Large Sample Survey of Retail (2005) results. Expenditure on Food and non-alcoholic beverages was increased from R100,971 billion to R143,402 billion

In 2008/09, the published results of the LCS contained an adjustment for Food and non-alcoholic beverages from R125,070 billion to R174,098 billion.

The LCS report explained that “The value was derived from using total sales data that Stats SA collects every month. Total sales on food for the period September 2008 to August 2009 (the LCS year) were used. While total sales from specialised food stores were not a problem to derive, it was difficult to isolate food sales from general dealers as their total sales included non-food items as well. The proportion of food and non-alcoholic beverages was obtained from retail trade data collected by Stats SA every five years. The proportion of food and non-alcoholic beverages sales to total sales in general dealers amounted to 67%. This was then applied to total sales in general dealers. Using this information, the total sales in 2009 amounted to R183 billion.”

## 2. Alcoholic beverages and tobacco

Alcoholic beverages and tobacco are massively underreported in the household survey, in line with international experience. In South Africa and fixed value excise tax is imposed on the sale of each bottle of alcohol and each pack of cigarettes.

The total alcoholic beverages sold in South Africa for the period 2009 was R57,3 billion (ASA, 2009). To derive the weight for alcoholic beverages for personal consumption, sales through restaurants and hotels and to businesses (intermediate consumption) must be excluded. The following process was followed:

**Table 2: Alcoholic beverages**

Item	Source	Value
Total sales	ASA 2009	R57,300,000,000
Less: Sales to Restaurants and hotels	Food and beverage Survey 2009	R 3,051,602,145
Total sales ex sales to restaurants and hotels		R54,248,397,855
Less: Business expenditure on alcoholic beverages	Food and beverage survey (Split between private and business consumption of all food and beverages; respectively 78% and 22%)	R54,248,397,855 x 0.78
Alcoholic beverages		R42,313,750,327

Total expenditure on tobacco was derived by using the Large Sample Survey of the retail sector for the proportional split between alcoholic beverages and tobacco, and in addition the value for alcoholic beverages was used as a baseline to derive tobacco. The following process was followed:

**Table 3: Tobacco**

Item	Source	Value
Alcoholic beverages	Table 2	R42,313,750,327
LSS Proportions for alcoholic beverages and tobacco	Large Sample Survey (LSS); Alcoholic beverages 62% and Tobacco 38%	
Calculate total alcoholic beverages and tobacco		$R42,313,750,327 \times 100 / 62$
Alcoholic beverages and tobacco		R68,247,984,398
Less; Alcoholic beverages		R42,313,750,327
Tobacco		R25,934,234,071

**3. Used vehicles**

Used vehicles are the only class of second hand goods included in the South African CPI. Previously, the gross sales value for used vehicles was used as the weight in the South African CPI. However, because households' sales of durables constitute negative expenditures, the weights for second-hand goods should be based on households' net expenditures: i.e. total purchases less sales. It is therefore necessary to calculate a net weight based on available data sources. For the analysis, the net weight was derived by using the average national accounts proportions for new and used vehicles of 2008 and 2009. The net weight was calculated as follows:

**Table 4: Used vehicles**

Item	Source	Value
Used vehicles	LCS 2008/09	R23,216,384,868
Proportions of purchases of vehicles	National accounts average proportion 2008 and 2009; New vehicles 80.6% and Used vehicles 19.4%	
New vehicles	LCS 2008/09	R35,415,467,550
Calculate total purchase of vehicles		$R35,415,467,550 \times 100 / 80.6$
Total purchase of vehicles		R43,939,786,042
Less: New vehicles		R35,415,467,550
Used vehicles		R 8,524,318,492

#### 4. Car hire

Because the HES does not capture the expenditure of foreign tourists inside South Africa, it is necessary to make an adjustment to these figures. The calculation was done as follows:

**Table 5: Car hire**

Item	Source	Value
Car hire	LCS 2008/09	R 2,188,518,308
Proportions of car hire domestic and international	SAVRALA <sup>2</sup> ; Domestic 82% and International 18%	
Calculation of total car hire		R 2,188,518,308 x 100 / 82
Car hire		R 2,668,924,766

#### 5. Package holidays

Adjustment of package holidays for tourism expenditure. The calculation was done as follow:

**Table 6: Package holidays**

Item	Source	Value
Package holidays	LCS 2008/09	R 1,376,023,689
Package holidays	Tourism Satellite Account 2008 and 2009	R 4,663,500,000
Proportion of personal and business expenditure	Tourism Satellite Account 2010 and 2009; Personal 56% and Business 44%	
Package holidays (TSA) less business expenditure		R 4,663,500,000 x 0.56
Package holidays		R 2,611,560,000

#### 6. Restaurants and Hotels

Restaurants and hotels (excluding school and university accommodation) are adjusted for under-reporting as well as tourism expenditure. The CPI 2008 reweighting only adjusted for purchases in restaurants, but not hotel accommodation. The IES 2005/06 for restaurants was adjusted to R15,007 billion from R9,633 billion.

<sup>2</sup> Southern African Vehicle Rental and leasing Association

For the analysis of the LCS 2008/09, both groups were adjusted. The calculations were done as follows:

**Table 7: Restaurants**

Item	Source	Value
Restaurants	LCS 2008/09	R14,130,581,427
Restaurants	Food and beverages survey 2008 and 2009	R 37,985,100,000
Proportion of personal and business expenditure	Food and beverages industry report 2009; Personal 78% and Business 22%	
Restaurants (Food and Beverages Survey) less business expenditure		R 37,985,100,000 x 0.78
Restaurants		R29,682,380,000

**Table 8: Hotels**

Item	Source	Value
Hotels	LCS 2008/09	R 5,907,560,017
Hotels	Accommodation survey (2008 and 2009)	R12,674,800,000
Proportion of personal and business expenditure	Domestic Tourism Survey 2010 (DTS); Personal 87% and Business 23%	
Hotels (Accommodation survey) less business expenditure		R12,674,800,000 x 0.87
Hotels		R 11,027,076,000

## 7. Overall impact of the changes

The overall nominal impact of changes results in a 7.4% increase in expenditure from R902 billion to R969 billion.

**Table 9: Total changes**

COICOP	COICOP description	Values	
		LCS (2008/09)	CPI (2008/09)
01	Food and non-alcoholic beverages	175,098	175,098
02	Alcoholic beverages and tobacco	8,800	68,248
03	Clothing and footwear	43,715	43,715

04	Housing, water, electricity, gas and other fuels	225,536	225,536
05	Furnishings, household equipment and routine maintenance of the dwelling	48,573	48,573
06	Health	11,960	11,960
07	Transport	138,152	123,958
08	Communication	30,558	30,558
09	Recreation and culture	38,622	39,857
10	Education	25,197	25,197
11	Restaurants and hotels	21,356	42,028
12	Miscellaneous goods and services n.e.c	134,833	134,833
TOTAL		902,402	969,562
		<b>Proportions</b>	
<b>COICOP</b>	<b>COICOP description</b>	<b>LCS (2008/09)</b>	<b>CPI (2008/09)</b>
01	Food and non-alcoholic beverages	19.40	18.06
02	Alcoholic beverages and tobacco	0.98	7.04
03	Clothing and footwear	4.84	4.51
04	Housing, water, electricity, gas and other fuels	24.99	23.26
05	Furnishings, household equipment and routine maintenance of the dwelling	5.38	5.01
06	Health	1.33	1.23
07	Transport	15.31	12.78
08	Communication	3.39	3.15
09	Recreation and culture	4.28	4.11
10	Education	2.79	2.60
11	Restaurants and hotels	2.37	4.33
12	Miscellaneous goods and services n.e.c	14.94	13.91
TOTAL		100.00	100.00

## 8. Example of lower-level adjustment

As discussed in section B, the HES might not provide the detailed proportions between different products within a group. The example that will be used to adjust indicator product proportions is the group for Grain and Cereals (COICOP 0111), specifically bread. The adjustment is done in the following manner:

**Table 10: Adjustment of the proportion of bread**

Item	Source	Value
Bread (Value):	LCS 2008/09	
White bread		R 8,029,093,068
Brown bread		R12,272,666,023
Total		R20,301,759,091
Bread (Proportion):	Derived from line above	
White bread		39.5%
Brown bread		60.5%
Bread (Proportions):	Retailers sales - Chain stores	
White bread		52.6%
Brown bread		47.4%
Retailer sales value proportions to LCS total value for bread		
White bread		$R20,301,759,091 \times 52.6\%$
Brown bread		$R20,301,759,091 \times 47.4\%$
Value (weight) for bread:		
White bread		R10,676,940,494
Brown bread		R 9,624,818,597
Total		R20,301,759,091

## F. Concluding thoughts

Addressing deficiencies in the HES can have a significant impact on the values of expenditure used to weight the CPI. These adjustments are necessary to properly reflect the actual budget shares of different expenditure groups. However, price statisticians need to weigh up the relative merits of alternative data sources against the risks inherent in combining data from different sources and the assumptions necessary to ensure consistency with the scope of the CPI. The results of adjustments may influence the rates of consumer inflation for several years.

## REFERENCES

ASA, 2009, The South African Liquor industry: Our contribution. Accessed at <http://www.ara.co.za/uploads/DTI-convention-brochure%20final.pdf>

Gudnason, R, 1998, Comparison of different sources for weights of selected groups in the Icelandic CPI: Consumers' bar code cash receipts vs scanner data from supermarkets. International Working Group on Price Statistics.

International Labour Organisation, 2010 Consumer Price Index: Theory and Practice, ILO, Switzerland

Statistics South Africa, Living Conditions Survey Report. Accessed at [www.statssa.gov.za/publications/statsdownload.asp?PPN=P0310&SCH=5017](http://www.statssa.gov.za/publications/statsdownload.asp?PPN=P0310&SCH=5017)

United Nations, 2009, Practical guide to producing Consumer Price Indices, United Nations, Switzerland