Optimal allocation of prices in practice - using the Neyman formula as tool

Alexandra Beisteiner, Statistics Austria

Paper presented at the Joint UNECE/ILO Meeting on Consumer Price Indices
8 – 9 May 2008, Palais des Nations, Geneva
Pre-conference version, March 2008

1. Introduction

One of the most important and cost intensive items in a consumer price index is the collection of price information on a monthly basis. It is important to use available resources in the best and cost efficient way to get the best result under given circumstances. In case of the CPI it is to optimize the precision of the CPI or the inflation rate when stratified samples are used. It is obvious that for areas with a high relative expenditure weight more prices should be collected than for areas with a low expenditure weight. In other words this means that resources should be put in those areas which have a big influence on the inflation rate. A second issue that can be taken into account is the diversity of price dispersion within one good or service. When the deviation of prices or of the price development is rather high than more prices should be collected than for homogeneous goods and services. In such a situation it is not necessary to reinvent the wheel but to use some standard procedure based on mathematical sampling theory called the Neyman formula. The Neyman formula answers the question of optimal allocation of observations and to maximize precision of the estimates at the same time given stratified sample with a fixed budget.

2. The Austrian price survey

The Austrian CPI survey is based on a national regulation that states that the prices have to be collected by the National Statistical Institute that is Statistics Austria, on the one hand and by twenty city governments on the other hand. The monetary compensation for the city governments is fixed in the national regulation and a value guarantee rule with 3% per annum is given. In the year 2007 the compensation for the city governments was 280.094,20 €, on average 40.182 prices per month are collected. A simple division result in an amount of 0.58 € per regional collected price. This amount does not include plausibility checks and the computation of quality adjustments as these are done by Statistics Austria. In addition Statistics Austria is collecting about 3.800 but the transferred budget from the government also includes the validation, the revision of the basket of goods and services, methodological improvements and the computation of the index. This is the reason why the computation of a comparable value for the central collected prices is not straightforward although not impossible.

The regional price collection is done with the help of price collection sheets which are standardized for all regions. The item descriptions which are printed on the collection sheets are predetermined by Statistics Austria and kept constant over an index period or at least over one year. Each good and service is assigned to a branches of trade. This classification is only a national auxiliary classification that combines those goods and services which are typically bought in one shop. E.g. the most food

items can be and are typically bought in a supermarket therefore all food items are in one branches. All garments for women are typically sold in one shop therefore all women garments are summarized in one branches. Once a shop is selected all items on the branches list have to be collected. This means that no matter how important a good is in terms of weight in the basket of goods the same number of prices is collected when goods are assigned to the same branches.

The central price collection is done by the staff of Statistics Austria by email, fax, telephone inquiries, and personal data collection in the field and on the internet. The main reasons for the central price collection are the level of difficulty and the complexity of varieties connected with the price observation and quality adjustment of a good or service on the one hand and the mode of price setting on a central level on the other hand, that is e.g. that prices for cigarettes are the same for the whole country. In comparison with the regional price collection the central price collection is more flexible as information on changes in the item descriptions are distributed expeditiously. As the central staff is more involved in the validation of the data and in methodological considerations there is a deeper understanding of CPI methodology and needs.

3. Initial Sampling – Starting a new survey

Let's assume a situation where a certain country is starting with a new CPI survey. The first thing that needs to be known are the relative expenditure shares for a breakdown using a classification which is in the case of CPIs commonly the COICOP[1] classification or even in more detail. Table 1 gives a small extract of a basket of goods for milk products. In the first column the items are named together with a national coding, in the second column the relative expenditure shares of each item are given. Before starting the survey the number of needed price observations is an important question. Option I shows a situation where the distribution of price observation is not corresponding to the respective weights. The number of price observations is evenly distributed over the items. In Option II the relative expenditure weights are fully taken into account. As convenience food has only a low expenditure weight only a few prices are selected for this item. In contrary for the item whole milk the most price observations are assigned as this is the item with the biggest weight. In both options, I and II, the total number of price observations sum up to 100, which gives in the case of Option II exactly the values of the expenditure weights. If for example more money is available the total number of prices could be increased to a total of 250 price observations in Option III. Now the item values are again computed with the help of the relative expenditure weights: for each cell the total number of price observations, that is 250, is multiplied by the relative expenditure weight standardized to a sum of 1. For convenience food 250 * 0.02 = 5 price observations have to be collected. For whole milk the result is 250 * 0.68 = 170 price observations.

This principle can be applied for each sample within the CPI. When for example the sample within whole milk needs to be optimized than the total sample size is 170 in Variant III. Now the regions or supermarket chains become the relevant variables for stratification and the respective market shares

[1] COICOP = Classification of Individual Consumption by Purpose.

substitute the expenditure weights. In general this kind of sampling procedure is not a speciality for CPI samples but a common approach for all kind of samples.

Table 1: Relative expenditure weights and three variants of price observation distribution for milk products

Item	expenditure weights	Option I	Option II	Option III
47 convenience food	2%	25	2	5
48 whole milk	68%	25	68	170
49 milk cocktail	16%	25	16	40
50 sour cream	14%	25	14	35
	100%	100	100	250

4. The Neyman Formula

In section 3 one part of the Neyman formula was already used to allocate price observations to items when a sample is newly build. In a dynamic setting where each month prices are observed and the composition of the sample has a history it can be of interest to reconsider the distribution of price observations over the whole basket of goods and services to optimise the precision of the inflation estimate under given resources. A new element that is used in the Neyman formula is the variance or standard deviation within an item of the basket of goods.

$$n_h = n * \left[(N_h * \sigma_h) / \sqrt{c_h} \right] / \left[\sum (N_i \sigma_i) / \sqrt{c_i} \right] \qquad \text{(Formulae 1)}$$

Table 2: Variable definition of formulae 1 and meaning in a CPI setting

	meaning in formulae 1	meaning in CPI setting
n_h	sample size for stratum h	number of prices observed for weighted item in the basket of goods
n	total sample size	number of prices observed for the whole CPI
N_h	population size for stratum h	item weight (relative expenditure shares)
σ_h	standard deviation of stratum h	standard deviation of price relatives for weighted item in the basket of goods
c_h	direct cost to sample an individual element from stratum h	cost to sample an individual price for a specific item in the basket of goods

Table 2 fives details about what is needed to compute the estimated number of price observations. All needed data but the standard deviations are easily to achieve. The latter needs more considerations. The first option that could be used is the standard deviation of the index numbers at a given point in time. In this case the individual index numbers of each price series, that is the current price divided by the price of the base period, are the statistics in focus. A second option is the variance of the inflation rate that is the twelve month rate of change as this is the statistic in focus for monetary policy. In this case the current price has to be related to the price twelve month ago.

There is no clear answer given to this question here as no evaluation of these two options was made. A more pragmatic view is taken which is based on the availability of the data. The first option the (quality adjusted) price relatives compared to the base period are available in each month therefore the usage of this statistics is easily implemented in any further analysis. The computation of the twelve

month rate of change for the individual price series is not in the standard procedure and has to be computed by increasing efforts without knowing if this procedure would improve the resulting estimates for optimal allocation.

5. A numerical example using Austrian CPI data

This section gives practical guidance on how to use the Neyman formulae in practice. Here the assumption is made that the needed statistics, number of price observations and price relatives, are easily available. This must not be the case in all NSIs. In Table 3 Formula 1 is used for the above presented example based on Austrian CPI data for milk products. The expenditure weights are the same as in Table 1.

Step 1: For each item in the basket of goods compute the standard deviation of the individual index numbers

To compute the standard deviations the individual price data are needed. There are two options that can be used. The first ones are the price relatives in a given month that is the current price divided by the price of the base period. As a second option the twelve month rate of change could be used that is the price of the current month divided by the price of the same month of the previous year (times hundred minus hundred). The standard deviations are than computed, based on the price relatives or the rates of change within an item in the basket of goods. If the price trend is homogeneous the variance is low, if the price trend is heterogeneous the variance is high. In principle more prices should be collected for heterogeneous items than for homogeneous.

Step 2: For each item in the basket of goods compute the absolute number of prices for a given mode of price collection (central v. regional)

The absolute number of prices by item in the basket of goods needs to be summarized by the original data base from the CPI survey. If there are two (or more) modes of price collection, the number of prices should be cross classified by the mode of collection. In Table 3 only the data from the regional collection are presented. To integrate the central price collection an own column should be added.

Step 3: Compute the product of the expenditure weights and the standard deviation of the item (Step 1)

The next step is to connect the relative weights of the items with the respective standard deviations. For convenience food the product is 2% * 0.2101 = 0.4202, for whole milk it is 68% * 0.1269. Although the standard deviation of convenience food is almost twice as high as the one of whole milk the resulting product for whole milk is twenty times higher as the expenditure weight is much higher.

Step 4: Compute the sum of the products of Step 3

Having computed the products of weights * standard deviations the total sum of this products needs to be computed.

Step 5: Calculate the percentage of the product (Step 3) of an individual item in relation to the sum of these products (Step 4)

The next step is to standardize the products of weights * standard deviations to 100%. For each item the product is divided by the total sum of the products and multiplied by 100. This step gives the optimal allocation in relative values.

Step 6: Multiply the computed percentage by the total sample size of the whole sample

To get the absolute optimal allocation values the relative values from step 5 are multiplied by the total sample size. The current total sample size is 469. 2.76% of 469 are 13 price observation which would be optimal for convenience food or 26.61% * 169 = 266 price observations for whole milk. By taking the current total number of prices the number of price observations is held constant. Under the assumption that each local collected price costs the same the budget for the survey is fixed. The total number could be increased ore decreased if different resources or a different level of accuracy is desired.

Step 7: Compute the difference between the estimated number of prices (Step 6) and the actual number of observed prices (Step 2)

In the last step the differences of the estimated prices and the current number of prices are computed. If the value is negative (estimated minus current) the current number of prices is too high and could be lowered. If the difference is positive more price observations could be allocated to this item.

Table 3: Example for the application of the Neyman formula for milk products from Table 2

Item	expendi-ture weights	Step 1: Std.dev. of price relatives	Step 2: current number of prices	Step 3: weight * Std.dev.	Step 5: % of Step 3	Step 6: estimated number of prices	Step 7: difference
47 convenience food	2%	0.2101	90	0.4202	2.76	13	-77
48 whole milk	68%	0.1269	185	8.6292	56.61	266	+81
49 milk cocktail	16%	0.2481	97	3.9696	26.04	122	+25
50 sour cream	14%	0.1588	97	2.2232	14.59	68	-29
	100%		**469**	**15.2422 (Step 4)**		469	

Remark: Std.dev. = standard deviation

6. Special cases

The example in Table 3 is an easy one as for each item in the basket of goods the number of prices is greater than one and a corresponding standard deviation is available from the general CPI database.

But there are some product groups which need a special consideration. The items list described here does not mean that in each NSI this items need a special consideration but the general procedures lying behind these items need a special treatment and that these special treatments need special considerations.

The first product group are new cars. In the Austrian HICP / CPI the certain models are own weighted elementary aggregates in the basket of goods. A certain model of a brand has the same level as whole milk as a whole. The basket of goods and services for new cars looks like the one for milk products and an example is given in Table 4. As the publication of the corresponding values is constrained by confidentiality issues, values that are based on less than four enterprises are not allowed to be published, the example contains fictious values. The sub-basket for new cars is treated like the sub-basket for milk products. The computation of standard deviation form the general CPI database which is based on the national coding, that is e.g. the code 48 for whole milk doesn't allow to compute the standard deviation for new cars in a straightforward way as the national code is set at the individual brand and model level. Here a special by analysis has to be made and all codes which belong to the new cars basket have to be recoded to one value to estimate the standard deviation. As corresponding values for Table 3 the number of price observations for new cars, that is 4 in Table 4 and the standard deviation 6.4550 has to be taken over to Table 3. Then the further steps have to be computed as described in Section 5. Similar treated areas in the Austrian HICP are newspapers, used cars

Table 4: Basket of goods and services for new cars, index numbers and weights

Item	expenditure weights	Number of price observations	Index numbers in month t
586 brand A model 1	14%	1	150.0
857 brand A model 2	28%	1	135.0
861 brand B model 3	40%	1	145.0
862 brand B model 4	18%	1	140.0
Sum	100%	**4**	
Mean			142.5
Standard deviation			**6.4550**

A second area which differs in the treatment compared to the majority of price observations are telecommunication services. The prices for the tariff packages and the different providers are entered into a distinct database as the computational procedures in the general CPI database can not cope with the needs for telecommunication services. In the general CPI database only the price trend for all packages and providers is entered which does not allow computing number of observed prices and standard deviations. This does not mean of course that standard deviations and number of observed prices are not available but it does mean that the values that are computed are not the true ones and better estimates are available. In this case the number of prices and the standard deviations has to be computed from the telecommunication services database and has to be entered manually into the database for the Neyman computation. This procedure is also needed for air fares.

The last area which has to be excluded from this kind of calculation is rents. Rents are observed in an own survey, the micro census survey. Form there only the general price trend is computed and entered into the CPI database. In principle it would be possible to integrate the rents into the estimation but as the sample for rents is not under direct control of the CPI section this area is left out of the estimation.

7. Interpreting the results

In Austria this kind of analysis are done every five years well in advance of a general revision of the basket of goods. The aim is not to follow the suggested changes in a strict re-allocation of the price observations but to use the results to detect areas where more observations are needed and lower the resources in those areas where the Neyman allocation suggests decreasing the number of price observations. The aim is to have more objective criteria for decisions about changes in the price observation systems than to base the allocation of resources on rather subjective criteria. During the last decade no major revision of the price survey was undertaken which means the mode of price collection stayed rather constant and only minor changes – de- or increasing the number of observations within a given mode – were implemented.

The second issue is that the items in the basket of goods are assigned to national branches of trade and hence are observed in the same frequency as soon as they belong to one of those branches. If the result suggests that the number of prices for whole branches of trade should be increased then the number of outlets for a given branches is increased. It is more difficult to increase or decrease the number of price observations for a certain item within a branches while the other items should stay constant. It gets even trickier when there is a certain pattern of increases and decreases within a certain branches. Table 5 gives an example for woman garments. In the present situation there are fewer observations for shirts and woman suit with skirt than for woman suit with trousers and skirts. To increase the number of observations for shirts and woman suit with skirt the item description for these two items could be duplicated on the price observation survey. If only one shirt is observed in one shop then the number of observed prices for shirts is twice as high after this procedure. This procedure is equivalent to the price observation survey for whole milk which is given in Annex 1. For all milk products only one item description is on the survey but for whole milk there are two item descriptions. The duplication of item descriptions can be done in a strict sense, the same item description is duplicated or a second different item description can be added to ensure a good representation of the reality. It is more difficult to decrease the number of price observations when only one item description is on the price observation sheet. A solution here could be to have two price observation branches for the same branches of trade. The first one includes all items of the branches of trades the second one only those for which a higher number of price observations is desired. This kind of approach would give maximum flexibility in the question of needed number of price observations in the Austrian CPI setting although it is not implemented at present.

Table 5: Neyman allocation for basket of goods for woman garments

item	Observed number of prices	Estimated number of prices
304 woman shirt	121	+545
304 woman suit with skirt	121	+545
301 woman suit with trousers	230	-146
301 woman skirt	236	-141

Finally it has to be said that implementing changes into the running CPI survey have to be done with great care as the computation of the inflation rate has to be guaranteed at any point in time. Although it is important to have statistics that provide objective information for decisions on the structure of the sample national circumstances and a framework in which changes can only be introduced at a slow pace have to be taken into account. In Austria the resulting values of the Neyman formula are not taken as absolute values but as hints for smooth changes within the given framework.

Annex 1

Branche: 2 (food)

Geschäft: …..

47 0	**Convenience food, refrigerated,** (e.g. Chef Menü, Gourmet Menü, etc.)

48 0	**Whole milk,** fat content 3,5 % bis 3,6 %, 1 l,

48 1	**Ultraheated milk** from the refrigaretor 1 L, extended live milk

49 0	**milk cocktail,** 330 - 750 ml with fruits a: cocktail with yoghurt, b: cocktail with milk, c: cocktail with whey, d: cocktail with butter milk

50 0	**sour cream, fat content** 15 %, 200 - 300 ml