International Working Group of Price Indices
Stockholm, November 15-17, 1995
Jörgen Dalén

A statistical interpretation of CPI comparability

Objectives

1. The pursuit of comparability for statistical measures between Member States is a major objective of the harmonisation process in the European Union. In the case of Consumer Price Indices, the comparability concept is particularly crucial and it also seems to be the first set of statistics for which the striving for comparability has reached an operational phase, involving detailed methods and procedures.

2. From the Maastricht Treaty it can be conjectured that politicians and economists will use the Harmonised Indices of Consumer Prices (HICP) in, e.g., the following way: Look at 12-month changes in all 15 Member States' HICPs. For all countries averages of such numbers are computed over a run of, say, 6-12 months. These averages are compared and the three ("best") countries with lowest such averages are identified. Other countries' averages are compared with these best countries' averages (an average of the three) and based on this comparison it is decided whether a country's inflation rate is more than 1.5% larger than that of the three best. For this comparison to be reasonable, it is necessary that the HICPs are comparable.

3. In this paper, an attempt shall be made to formulate a statistical model (framework) for making quantitative estimates/judgements relating to the comparability concept. The use of such a model will mainly be related to future estimates of differences in inflation figures. We will also analyse the risk for a so called *Maastricht error*, whereby we mean being on the wrong side of the 1.5% demarcation line when estimating the differences in inflation rates.

International Working Group of Price Indices
Stockholm, November 15-17, 1995
Jörgen Dalén

A model

4. From a statistician's point of view comparability must be discussed in terms of errors. We will formulate a model in which there are G *sources of non-comparability* between countries which could be looked at as *errors* when we consider differences between HICPs as our target variables. These errors sum up to *total non-comparability* or *total error*. In the sequel, we will talk about errors for short. We will look at all errors as having both a systematic and a random componenent. The interpretation of this assumption runs like this: Over many years a certain error in a single country has an expected value, β, and a random variation, σ, around this expected value. In the process of harmonisation, judgements will be made on *future* values of β and σ, based on theoretical knowledge and empirical analyses on *past* data.

5. When $\beta \neq 0$ the error generates a bias. Note that the distinction between bias and random error in this model has to do with the error distribution over time and is not the same as that of, e.g., survey sampling. Note also that for a single country bias must be determined in relation to a fixed reference point. But when comparing several countries, it does not matter how this reference point is set.

6. We now formulate the following model for the inflation rate R_k of K countries, 1...,k,...K and for G errors, 1...,g,...G.

$\hat{R}_k = R_k + B_k$, where $\hat{R}_k$ is the estimated inflation rate,

$B_k = \sum_g B_{kg}$ and $B_{kg} = \beta_{kg} + b_{kg}$, β_{kg} being a constant bias for error source g and b_{kg} a random part of this error, distributed as $N(0, \sigma_{kg})$.

We will also use the symbols $\beta_k = \sum_g \beta_{kg}$, $b_k = \sum_g b_{kg}$ and $\sigma_k^2 = \sum_k \sigma_{kg}^2$.

The interpretation of this model is that we have a true inflation rate for every country, defined e.g. by a set of Commission rules. For a number (G) of reasons, the estimated value deviates from this true value. Each of these errors

International Working Group of Price Indices
Stockholm, November 15-17, 1995
Jörgen Dalén

generates a deviation with two components: a constant component β_{kg} and a random component b_{kg}. We will include all kinds of errors in this model from basic index construction to sampling. (In the case of sampling all β_{kg} will be close to zero and the b_{kg} (σ_{kg}) are the important part.) We will, in the sequel, assume that all random errors are independent both in the country and in the error dimensions. This is not automatically true but it could serve as a norm for our organisation of the error structure.

7. With this model random errors will cover errors that are random in a narrow sense (such as sampling errors) as well as errors of a more systematic nature where the random component expresses a time variation around an average level. The latter type of error has earlier been labelled short-term bias. We motivate this choice of model with its use in connection with uncertain statements about non-comparability in the future.

8. In the harmonisation perspective the difference in inflation rate, D and its estimate, $\hat{D}$, between two countries, say 1 and 2, is central. We have

$$\hat{D}_{12} = (\hat{R}_1 - \hat{R}_2) = D_{12} + (\beta_1 - \beta_2) + (b_1 - b_2)$$

That is, this estimate is disturbed by the *differences* in the constant biases as well as the random errors between pairs of countries.

9. We define total error of a two-country comparison as mean squared error:

$$\text{mse}(\hat{D}_{12}) = E(\hat{D}_{12} - D_{12})^2 = (\beta_1 - \beta_2)^2 + \sigma_1^2 + \sigma_2^2.$$

A 95% confidence interval for D_{12} will then be

$$\hat{D}_{12} + (\beta_1 - \beta_2) - 1.96\sqrt{\sigma_1^2 + \sigma_2^2}, \hat{D}_{12} + (\beta_1 - \beta_2) + 1.96\sqrt{\sigma_1^2 + \sigma_2^2}$$

Application of the model

10. In this Section we will look at some examples of errors and give some preliminary judgements of their sizes in order to demonstrate, how the model

could be used. Remember that errors here are defined as sources of non-comparability.

11. The first error that we will look at is *basic index construction*. By this term we mean the macro index formula, length of the index link and timing of chaining. (If two countries both use 5-year Laspeyres links, there may still be a source of uncomparability if the base years differ). In this case we may assume that random errors are negligible but we have a constant as well as a time-varying bias. American studies (and Swedish long- vs short-term index) point at a possible long-term bias of around 0.15 between a Laspeyres and a superlative index. This figure may serve as an upper bound to the constant bias between a pair of HICPs. The variation around this figure is such that we may have a σ of about 0.1 for a single country. Little is known as to the effects of different lengths of Laspeyres' links.

12. The second example is the *elementary aggregate formula*. Here we know that the average of relatives (R) has a constant (mean) bias compared with other formulae but also a time variation around this. The estimates of difference between the R and other formulae for a certain country could, for example, be β=0.5 with a time variation of σ=0.1. When comparing other formulae, e.g. the ratio of averages (A) and the geometric mean (G), we do not find any evidence of constant bias but important "varying bias" which may be treated in our model as a σ of about 0.15.

13. Thirdly, we have *sampling of outlets and items* for which there is no evidence of any constant bias (although technically this would be possible due to non-probability sampling, undercoverage etc.). Estimates of sampling errors in different countries range from 0.1 to 0.2 for a one-year change expressed as standard deviations, σ.

International Working Group of Price Indices
Stockholm, November 15-17, 1995
Jörgen Dalén

14. Now we focus on a particular pair of Member States 1 and 2. Let's say we have a picture like this:

	Constant bias $(\beta_{1g}-\beta_{2g})$	Random error $(\sigma^2_{1g}+\sigma^2_{2g})$
Basic index construction (fixed base vs chain with long-term index)	0.15	$2*0.1^2$
Elementary aggregate formula (R vs A)	0.5	$2*0.1^2$
Sampling	0.0	$0.1^2+0.2^2$
SUM	0.65	0.09

In this case we would obtain an mse=$0.65^2+0.09=0.5125$. A 95% confidence interval would be $[0.65-1.96\sqrt{0.09},\ 0.65+1.96\sqrt{0.09}]$ or from 0.06 to 1.14. This is just an example of how an analysis according to the model presented here would work, for two specified countries with known procedures. Within this framework one could, e.g., analyse the effects of introducing harmonised methods and procedures, making more efficient samples etc by looking at this scheme before and after the changes under consideration.

15. At this point we must also note that a target variable of the Maastricht Treaty is the difference between one country's inflation rate and the average of the three best countries. An average of several countries' inflation rates has, in general, a smaller error than that of a single country. This fact carries over to the difference between one country's inflation rate and that average.

Errors in a Maastricht-type estimate

16. In this Section we analyse a very special problem. Given that the true difference between the inflation rate of a certain country and the average of the three best countries, D, is smaller than 1.5%, what is the risk (probability)

for a *Maastricht error* defined as the estimated difference being greater than 1.5%? Terminology and notation will change slightly in this Section. Instead of bias differences we will talk about biases for short, denoted β. And by random error, denoted σ, we will mean the aggregate random error for a comparison.

17. The above question is posed for different combinations of bias and random error. (The analysis of the dual problem with D>1.5% and the risk of obtaining an estimate smaller than 1.5% is completely symmetric.) In particular, under what circumstances is it correct to say that bias is a more serious and detrimental type of error than random error? We immediately note that a negative bias always decreases the risk for an estimated difference greater than 1.5%, since the expected value of the estimated difference would then be even farther away from 1.5%. In the sequel we therefore only analyse the situation with a bias greater than 0.

18. We will assume that the random errors are normally distributed. This means that we have a random variable $\hat{D}$ distributed $N(D+\beta,\sigma)$. In Diagram 1 we give a 3-dimensional illustration of this probability as a function of β and σ. In paragraphs 19-21 below, we will set up three different analysis frameworks.

19. In framework 1 we postulate a constant mean square error $C=\beta^2+\sigma^2$. Holding the total error C constant, when does $P(\hat{D}>1.5)$ decrease with decreasing β or, equivalently, increase with increasing β? We will consider this question - which is a natural interpretation of the basic question of whether it is better to decrease β than σ - for various fixed values of D<1.5. We have

$$P(\hat{D}>1.5)=1-\Phi\left(\frac{1.5-D-\beta}{\sqrt{C-\beta^2}}\right),$$

where Φ is the standard normal distribution function.

Looking at this as a function of β and differentiating, we learn that this probability increases with β when (1.5-D)<C/β but decreases when (1.5-D)>C/β. Since C/β>√C this means that $P(\hat{D}>1.5)$ always increases with β when 1.5-D<√C ⇔D>1.5-√C. For D<1.5-√C, $P(\hat{D}>1.5)$ increases with β to a maximum point which occurs when β=C/(1.5-D), after which it decreases. That is, for values of D close to 1.5 an increase of β always gives a greater risk for a Maastricht error but for D far from 1.5% the issue is less clear-cut. In another dimension, the larger the total error is, the more important is it to decrease bias. In Diagram 2 we present the probability $P(\hat{D}>1.5)$ for four cases, all with C=1: D=0.2, 0.5, 0.8 and 1.2. We see that for D=0.5 and larger the probabilities increase with β all the way but for D=0.2 the probability reaches a maximum (at β=1/1.3=0.77) after which it decreases.

20. In framework 2, starting with arbitary β and σ, we compare $P(\hat{D}>1.5)$ when decreasing bias with a certain amount ε with the same probability when decreasing random error with the same amount. We thus take the difference between two probabilities:

$$P\left[\hat{D}>1.5|\hat{D}\propto N(D+\beta,\sigma-\varepsilon)\right]-P\left[\hat{D}>1.5|\hat{D}\propto N(D+\beta-\varepsilon,\sigma)\right]=$$
$$\Phi\left(\frac{1.5-D-\beta+\varepsilon}{\sigma}\right)-\Phi\left(\frac{1.5-D-\beta}{\sigma-\varepsilon}\right),\text{ where }0<\varepsilon<\min(\beta,\sigma).$$

We ask whether this difference is positive, since this means that a bias decrease results in a greater decrease of the risk for a Maastricht error than the same decrease of the random error. Since Φ is a monotonically increasing function the question of the difference between the terms being larger than zero is equivalent with the difference between the expressions inside the parentheses being larger than 0. This leads us to solve the inequality

$$\frac{1.5-D-\beta+\varepsilon}{\sigma}-\frac{1.5-D-\beta}{\sigma-\varepsilon}>0\Leftrightarrow 1.5-D<\beta+\sigma-\varepsilon$$

The interpretation of this solution is that a bias decrease reduces the risk more when D is closer to 1.5 than the sum of the two types of errors. In the opposite situation, however, a random error reduction would be better. We illustrate these results in Diagram 3 for four values of D. (By net probability gain in Diagrams 3-4 we mean the above difference between two probabilities.)

21. In framework 3, we make the same analysis as in framework 2 but with a multiplicative factor k instead of an additive one. We thus take the difference between the following two probabilities:

$$P\left[\hat{D} > 1.5 | \hat{D} \propto N(D+\beta, k\sigma)\right] - P\left[\hat{D} > 1.5 | \hat{D} \propto N(D+k\beta, \sigma)\right] =$$

$$\Phi\left(\frac{1.5 - D - k\beta}{\sigma}\right) - \Phi\left(\frac{1.5 - D - \beta}{k\sigma}\right), \text{ where } 0 < k < 1.$$

We again ask whether this difference is positive leading us to solve the inequality

$$\frac{1.5 - D - k\beta}{\sigma} - \frac{1.5 - D - \beta}{k\sigma} > 0 \Leftrightarrow 1.5 - D < \beta(k+1) \Leftarrow 1.5 - D < \beta$$

The interpretation of this solution is that a bias decrease always reduces the risk more when D is closer to 1.5 than the amount of the bias. However, when 1.5-D>2β it would be better to decrease random error. The interval $\beta \leq 1.5 - D \leq 2\beta$ represents a gray zone, where the effect depends on the value of k. We illustrate these results in Diagram 4 for four values of D.

22. The above analysis leads to the following conclusions:

- When D, the true difference, is close to 1.5%, or when errors are large, a bias reduction decreases the risk for a Maastricht error more than an equally large reduction in random error. At the same time, this is the situation when an error matters most, since the risk is then great in the first place.
- When the true difference is far from 1.5% or when the errors are small, the conclusion is less clear-cut and a reduction of random error *may* have a larger effect on the risk for a Maastricht error.

- For negative biases the risk for a Maastricht error actually decreases the larger the negative bias is!

Conclusions

23. In this paper we have presented a model intended to serve as a conceptual framework when discussing and quantifying the comparability between HICPs for guiding the process of setting up rules in different areas. The figures given as examples could be a result of an expert judgement process guided by theoretical knowledge and empirical analyses. A scheme such as the one presented in paragraph 14 could be set up for any pair of countries and any set of sources of non-comparability. The effect of a particular rule could also be analysed by comparing such schemes before and after the implementation of that rule.

An analysis of the risk for an estimate at the wrong side of the 1.5% demarcation line shows that, when the true difference is close to this value, a bias reduction is more important than a reduction in random error.

Acknowledgement: The research behind this report was supported by EUROSTAT, the Statistical Office of the European Communities. Jesper Jacobsen has contributed significantly to the analysis in paragraphs 16-21.