REVUE FRANÇAISE D'AUTOMATIQUE, D'INFORMATIQUE ET DE RECHERCHE OPÉRATIONNELLE. RECHERCHE OPÉRATIONNELLE

J. C. HAYYA
E. M. SANIGA
R. A. SCHAUL

Uncertainty in price indices

Revue française d'automatique, d'informatique et de recherche opérationnelle. Recherche opérationnelle, tome 16, nº 1 (1982), p. 33-43.

<http://www.numdam.org/item?id=RO_1982__16_1_33_0>

R.A.I.R.O. Recherche opérationnelle/Operations Research
(vol. 16, n° 1, février 1982, p. 33 à 43).

UNCERTAINTY IN PRICE INDICES (*)

by J. C. HAYYA (1), E. M. SANIGA (2), R. A. SCHAUL (3)

Résumé. — *Dans cet article, nous présentons l'argument que les indices de coût devraient être considérés comme variables aléatoires, si quelques-uns de leurs composants sont aléatoires.*

Si le numérateur et le dénominateur d'un indice sont fondés sur des échantillons assez larges, on peut utiliser le théorème central limite, et supposer que l'indice est le quotient de deux variables aléatoires normales.

On peut en déduire la construction d'un intervalle de confiance pour l'index; cette construction ne nécessite pas la connaissance de la distribution exacte de l'indice lui-même.

Puisque intervalles de confiance et tests d'hypothèse sont équivalents, il en résulte que ces derniers peuvent être accomplis par l'usage des premiers. Ainsi, par l'usage d'intervalles de confiance, nous présentons un test de l'hypothèse de stabilité temporale d'un indice.

Nous concluons en présentant le cas d'un indice du coût des habitations nouvelles pour une famille.

Keywords: Probability Intervals, Price Indices

Abstract. — *In this paper, we argue that price indices should be treated as random variables if some of the components used in their construction are random. If the numerator and the denominator of an index are based on large enough sample sizes, one may be justified in invoking the central limit theorem and assuming that the index is a ratio of two normally distributed variables. The construction of a probability interval for the index follows; this construction does not depend on a knowledge of the exact distribution of the index itself. Since probability intervals and hypothesis tests are mathematically equivalent, it follows that the latter can be accomplished by means of the former. Hence, using probability intervals, we suggest a test of the hypothesis of temporal stability.*

We conclude with a case study on the construction of a price index for new one-family houses sold.

Keywords: Probability Intervals, Price Indices

1. INTRODUCTION

Unless price indices are based on a census, they are ratios of possibly related weighted sums of sample observations. Some, though not all, of these observations may be random variables. If these random observations are drawn from populations with finite variance and if the sample is large enough, then, by the central limit theorem, the numerator and denominator of the index can be

(*) Received in november 1981.
(1) Pennsylvania State University.
(2) University of Delaware.
(3) Bureau of the Census.

expected to approximately follow a normal distribution. Furthermore, if some mild restrictions as set forth by Geary [1], and Hayya, Armstrong and Gressis [3] are met, inference about the index in the form of probability intervals and hypothesis tests can be made. And inference techniques for price indices are necessary since price indices are very much used by policy makers to guide their actions.

If a price index is constructed such that some of its components are random, then the price index is a point estimate; it would not incorporate the uncertainty inherent in the sampling process as does a probability interval. A rise or fall in the value of the point estimate may not reflect an actual change. Yet, in the belief that a price index is nonrandom, policy makers behave as if any change in the value of the index is significant. The purpose of this paper is to provide analysts with techniques for including the inherent uncertainty in price indices in their decisions.

Our first concern is the development of probability bands for price indices. Hypothesis tests, which are mathematically equivalent, are briefly discussed. Our methods are based on a weighted aggregate index, I_k, where:

$$I_k = \sum_{i=1}^{n} p_{ki} q_{ji} \Big/ \sum_{i=1}^{n} p_{0i} q_{ji}, \tag{1}$$

with p_{ki} the price of the ith commodity in period k, q_{ji} the quantity or weight of commodity i in period j, and p_{0i} the price of commodity i in the base period [4]. The choice of j specifies the index; e. g., if $j=0$, I_k is the Laspeyres index; if $j=k$, I_k is the Paasche index. While our procedures are developed for weighted indices, they would nevertheless apply to any index so long as the assumptions we set forth are met.

In part 2 of the paper, we develop the inference techniques for price indices and present some numerical examples. An application involving the price index for new one-family houses sold is discussed in 3. A brief summary is given in 4.

2. CONFIDENCE BANDS AND HYPOTHESIS TESTS FOR PRICE INDICES

Let $T=(w, y)$ have the bivariate normal distribution with parameters μ_w, μ_y, σ_w^2, σ_y^2, and $\rho_{wy}=\rho$, where $\rho=\sigma_{wy}/\sigma_w \sigma_y$.

We assume $\mu_y - c\,\sigma_y > 0$ for some c large enough so that $P(y \leq 0)$ is negligible, say $c=3$ or 4. Note that if $c=3$ then $P\{y<0]=.0044$ and if $c=4$ then $P[y<0]=.0001$. We also have $P(y=0)=0$.

Now:

$$P\{a<I<b\}=P\{a<w/y<b\}=P\{ay<w<by,\ y>0\}$$
$$+P\{by<w<ay,\ y<0\}.$$

Then:

$$|P\{ay<w<by\}-P\{a<I<b\}|=|P\{ay<w<by\}-P\{ay<w<by,\ y>0\}|$$
$$+P\{by<w<ay,\ y<0\}\leqq 1-P\{y>0\}+P\{y<0\}=2\ P\{y<0\}.$$

Since $2\ P\{y<0\}$ is negligible, we have:

$$P\{a<I<b\}\simeq P\{ay<w<by\}.$$

So we seek $(a,\ b)$ such that:

$$P\{ay<w<by\}=1-\alpha. \tag{2}$$

Also, since

$$w-xy\sim N(\mu_w-x\,\mu_y,\ \sigma_w^2-2\ x\,\sigma_{wy}+x^2\,\sigma_y^2),$$

for any real number x, it follows that:

$$P(w-xy<0)=P(Z<-(\mu_w-x\ \mu_y)/(\sigma_w^2-2\ x\,\sigma_{wy}+x^2\,\sigma_y^2)^{1/2}), \tag{3}$$

where $Z\sim N(0,\ 1)$.

To find a and b such that (2) holds, we solve:

$$P(w-by<0)\geqq 1-\alpha/2, \tag{4}$$

and

$$P(w-ay<0)\leqq \alpha/2. \tag{5}$$

The inequalities (4) and (5) are used if an exact $1-\alpha$ internal cannot be developed; their direction ensures that $(a,\ b)$ will be *at least* a $(1-\alpha)$ probability interval.

Let Z_α be the α th percentile of a $N(0,\ 1)$ distribution. From (3) and (4) we seek $x=b$ such that:

$$-(\mu_w-x\,\mu_y)/(\sigma_w^2-2\ x\,\sigma_{wy}+x^2\,\sigma_y^2)^{1/2}\geqq Z_{1-\alpha/2}, \tag{6}$$

i. e..

$$x\,\mu_y-\mu_w\geqq Z_{1-\alpha/2}\,(\sigma_w^2-2\ x\,\sigma_{wy}+x^2\,\sigma_y^2)^{1/2}. \tag{7}$$

Since the r. h. s. of (7) is positive, the l. h. s. must also be positive. Hence, under the condition $x \geqq \mu_w/\mu_y$, we can square both sides of (7) yielding:

$$(x\,\mu_y-\mu_w)^2 \geqq Z^2_{1-\alpha/2}\,(\sigma^2_w-2\,\sigma_{wy}+x^2\,\sigma^2_y). \tag{8}$$

In a similar manner, using (3) and (5) we seek $x=a$ such that (8) holds with $x \leqq \mu_w/\mu_y$. After some algebra, (8) becomes:

$$f(x)=(\mu^2_y-Z^2_{\alpha/2}\,\sigma^2_y)\,x^2-2\,(\mu_w\,\mu_y-\sigma_w\,\sigma_y\,\rho\,Z^2_{\alpha/2})\,x+(\mu^2_w-Z^2_{\alpha/2}\,\sigma^2_w) \geqq 0. \tag{9}$$

We require $f(x)=0$ to have two real roots, a and b, such that $a \leqq \mu_w/\mu_y \leqq b$. Two real roots $a<b$ exist and equations (4) and (5) hold if: (*See* the Appendix for details):

(*a*) $\mu_y > c\,\sigma_y$ for some constant c

that determines the precision sought (10)

(*b*) $Z_{1-\alpha/2} < \mu_y/\sigma_y$; (11)

(*c*) also if $\rho=+1$, $\mu_y/\sigma_y \neq \mu_w/\sigma_w$. (12)

that is, if $w=ay+b$ for some $a>0$, we must have $b \neq 0$, otherwise $I=a$.

Note that condition (11) implies that the probability of y assuming negative values is negligible. (Note also that this condition is met if y can assume only positive values.) This is in accord with. (1) Geary [1], who suggested $\mu_y/\sigma_y>3$ or in terms of the coefficient of variation, $\sigma_y/\mu_y<0.33$; and (2) Hayya, Armstrong, and Gressis [3] who found via simulation that for approximations at the 0.05 significance level, $\sigma_y/\mu_y<0.39$. Finally, we mention that the random variable w need not be restricted to any positive values; it only must not tend toward a constant. Hayya, *et al.* [3] suggest $\mu_w/\sigma_w<200$ for approximations at the 0.05 level of significance.

In summary, probability bands for price indices can be constructed by finding the roots of the quadratic (9) if the numerator and denominator are approximately normal and the conditions (10), (11), and (12) are met.

Now consider the null hypothesis $H_0 : E(I_t)=E(I_{t+1})$; this states that an index I has the same expectation in period t as in the subsequent period $t+1$, or, in other words, exhibits temporal stability. Assume the numerators and deonominators of I_t and I_{t+1} are normal under the central limit theorem. Further assume the conditions (10), (11), and (12) hold. A *conservative* test of the hypothesis at the α significance level can be achieved by constructing separate $(1-\alpha)^{1/2}$ probability intervals for I_t and I_{t+1}. If the two intervals intersect we would not reject the hypothesis of temporal stability at the α level of significance. We can similarly test the hypothesis $E(I_t)=E(I_{t+k})$, for any k greater than one; we can also test the hypothesis $E(I_t)=E(J_{t+k})$ for any other index J.

A numerical example

Consider a Laspeyres index with:

$$w=\sum_{i=1}^{n} p_{ki}\, q_{0\,i}, \tag{13}$$

and:

$$y=\sum_{i=1}^{n} p_{0\,i}\, q_{0\,i}. \tag{14}$$

If the sample size n is large, w and y can be assumed to be normal by the central limit theorem. With small n and q normal, the the individual product terms in (13) and (14) can be assumed to be normal if [3]:

$$(\mu_p/\sigma_p)\,(\mu_q/\sigma_q)\geqq 100, \tag{15}$$

and p and q are statistically independent. With p and q not statistically independent, the same result holds if the r. h. s. of (8) is at least 250 [3].

As an example, suppose we wish to construct a 95 percent probability interval for $I=w/y$ in period t with $w\sim N\,(800{,}10^4)$, $y\sim N(500{,}10^4)$, and $\rho=0$. Here:

$$Z_{.975}\leqq \mu_y/\sigma_y=5,$$

and the conditions (10), (11), and (12) hold. Hence, the quadratic (9) has a real solution. Substituting in (9) gives $211{,}584\ x^2-800{,}000\ x+601{,}584=0$, and solving for x yields (1.03, 2.75) as the 95 percent probability interval. Further, suppose we construct another probability interval for I in a subsequent period $t+1$, with (1.80, 4.60) as the result. We would not reject the hypothesis of temporal stability at the $\alpha=0.0975$ level of significance since the intervals intersect. [Note that $(1-\alpha)^{1/2}=.95\Rightarrow\alpha=0.0975$.]

3. CONFIDENCE BANDS FOR THE BUREAU OF THE CENSUS' PRICE INDEX FOR NEW ONE-FAMILY HOUSES SOLD

The price index for new one-family houses sold measures changes in selling price. These houses must be similar to those sold in 1967 with respect to eight characteristics [5], p. 2: (i) floor area, (ii) number of stories, (iii) number of bathrooms, (iv) air conditioning, (v) type of parking facility, (vi) type of foundation, (vii) geographic division within region, and (viii) location within a metropolitan area. (Recently, the Census Bureau has made some small changes in this procedure. Also, the new base is 1972.) Using a monthly sample of about 1,000 observations, the Census Bureau collects information (via the Housing Sales Survey) about these attributes and the transaction prices [4]. The index is calculated as follows. First, estimate the coefficients B_c in the following regression;

$$Y = X_c\, B_c + \varepsilon,$$

where c stands for the current period, Y is the transaction price of new one-family houses, and X is the design matrix of the attributes. These attributes are represented by 27 independent variables, several of which are dummy 0-1 variables. Second, calculate the means $\overline{Y}_0$ and $\overline{X}_{0i}$, $i=1, 2, \ldots, 27$. The price index is then:

$$I_c = \sum_{i=1}^{27} \hat{B}_{ci}\, \overline{X}_{0i} / \overline{Y}_0, \tag{16}$$

where $\overline{X}_{00} = 1$ (constant term).

If ε is normal then $\hat{B}_{ci}$ is normal. Also, $\overline{X}_{0i}$ and $\overline{Y}_0$ are approximately normal by the central limit theorem. The product $\hat{B}_{ci}\, \overline{X}_{0i}$ is approximately normal for each i if [2]:

$$n^{1/2} \left(\frac{B_{ci}}{\sigma_{B_{ci}}} \right) \left(\frac{E(X_i)}{\sigma_{X_i}} \right) \geq 100.$$

This restriction is easily met due to the large sample size. Further, the numerator in (16) would tend toward normality since it is a sum of products. Consequently, the price index may then be assumed to be a ratio of two normally distributed variables.

Calculation of the probability interval for the index requires an estimate of μ_w and σ_w^2. We assume the $\hat{B}_i$ are statistically independent of the X_{0i}. Then:

$$\mu_w = E\left(\sum_{i=1}^{p} B_i X_{0i} \right) = \sum_{i=1}^{p} \mu_{B_i} \mu_{X_{0i}}. \tag{17}$$

Consequently·

$$\hat{\mu}_w = \sum_{i=1}^{p} \hat{B}_i \overline{X}_{0i}. \tag{18}$$

Also:

$$\operatorname{Var}(W)=\sum_{i=1}^{p}\left(\sigma_{B_i}^2\,\sigma_{X_{0i}}^2+\sigma_{B_i}^2\,\mu_{X_{0i}}^2+\sigma_{X_{0i}}^2\,\mu_{B_i}^2\right)+2\sum_{i<j}\left(E(B_i B_j)\,E(X_i X_j)-\mu_{X_i}\mu_{X_j}\mu_{B_i}\mu_{B_j}\right). \quad (19)$$

For convenience, and also in practice, the sum of the covariance terms in (19) may be assumed negligible relative to Var (W). Thus, our estimate $\hat{\sigma}_w^2$ can be written:

$$\hat{\sigma}_w^2=\sum_{i=1}^{p}\left(\hat{\sigma}_{B_i}^2\,\hat{\sigma}_{X_{0i}}^2+\hat{\sigma}_{B_i}^2\,\overline{X}_{0i}^2+\hat{\sigma}_{X_{0i}}^2\,\hat{B}_i^2\right). \quad (20)$$

We present in Table I the probability intervals for the price index for new one-family houses sold for the third quarter of 1976. The intervals are calculated for $\rho=-1.0\ (0.2)\ 1.0$, where ρ denotes the unknown correlation between the numerator and denominator of the index.

TABLE I

The probability interval for the price index for new one-family houses sold, third quarter, 1976 [base - 100 (1963)]

	Correlation, ρ										
Band. . .	−1.0	−0.8	−0.6	−0.4	−0.2	0.0	0.2	0.4	0.6	0.8	1.0
a.	1.914	1.915	1.917	1.918	1.920	1.922	1.923	1.926	1.928	1.931	1.937
b.	1.967	1.965	1.964	1.962	1.961	1.959	1.957	1.955	1.952	1.949	1.943
%.	2.7	2.6	2.4	2.3	2.1	1.9	1.8	1.5	1.2	0.9	0.3

Estimated Index Value: $I=w/y=1.940$. Confidence bounds: $a<I<b$.

The last row in table I shows the width of the interval as a percentage of the point estimate of the index; in this application, the interval is relatively narrow which is undoubtedly due to the large sample sizes used. Table I also shows that the width of the interval is a decreasing function of ρ and that the estimated value of the index may not be at the midpoint of the interval. These results are true in general; proofs are available from the authors upon request.

4. SUMMARY AND CONCLUSIONS

Since index numbers are based on sampling, it would be useful to policy makers if they are treated as sample statistics and subjected to the inference procedures of hypothesis testing and the construction of probability intervals. A change in the value of an index may not be statistically significant, and using inference techniques the policy maker need take no action.

In this paper, we present a method for the construction of probability intervals and for the testing of hypotheses concerning the expectation of an index number. For the latter, we suggest a conservative method for testing the temporal stability of an index; some extensions are presented. A case study utilizing data from the Bureau of the Census is included as an example.

The authors are grateful to a referee for several helpful suggestions.

APPENDIX

ANALYSIS OF THE QUADRATIC [equation (9)].

Consider equation (9):

$$f(x)=(\mu_y^2-Z_{\alpha/2}^2\,\sigma_y^2)\,x^2-2(\mu_w\,\mu_y-\sigma_w\,\sigma_y\,\rho_{wy}\,Z_{\alpha/2}^2)\,x+(\mu_w^2-Z_{\alpha/2}^2\,\sigma_\omega^2)\geqq 0. \quad (A.1)$$

We seek two real numbers a and b such that

$$f(a)=f(b)=0, \quad (A.2)$$

with:

$$a\leqq\frac{\mu_w}{\mu_y}\leqq b. \quad (A.3)$$

A necessary condition for $f(x)=0$ to have two roots is that it be a proper quadratic, i. e.:

$$\mu_y^2-Z_{\alpha/2}^2\,\sigma_y^2\neq 0,$$

or:

$$Z_{\alpha/2}\neq\frac{\mu_y}{\sigma_y}. \quad (A.4)$$

We also require the reduced discriminant, $\Delta'=(B/2)^2-AC$, of $f(x)$ to be positive. From (A.1):

$$\Delta'=(\mu_w\,\mu_y-\sigma_w\,\sigma_y\,\rho\,Z_{\alpha/2}^2)^2-(\mu_y^2-Z_{\alpha/2}^2\,\sigma_y^2)\,(\mu_w^2-Z_{\alpha/2}^2\,\sigma_w^2)=Z_{\alpha/2}^2\,g(Z_{\alpha/2},\,\rho), \quad (A.5)$$

where:

$$g(z, \rho) = -\sigma_w^2 \sigma_y^2 (1-\rho^2) Z^2 + \mu_y^2 \sigma_w^2 + \mu_w^2 \sigma_y^2 - 2 \mu_y \mu_w \sigma_y \sigma_w \rho. \quad (A.6)$$

if $|\rho| = 1$, then:

$$g(z, 1) = (\mu_y \sigma_w \pm \mu_w \sigma_y)^2 > 0, \quad (A.7)$$

if (12) holds, i.e.:
if:

$$\rho = +1 \text{ then } \frac{\mu_y}{\sigma_y} \neq \frac{\mu_w}{\sigma_w}. \quad (A.8)$$

If $|\rho| < 1$, we require $g(z, \rho) > 0$, *i.e.*:

$$z^2 < \frac{\mu_y^2 \sigma_w^2 + \mu_w^2 \sigma_y^2 - 2 \mu_y \mu_w \sigma_y \sigma_w \rho}{\sigma_w^2 \sigma_y^2 (1-\rho^2)} = \frac{\mu_y^2}{\sigma_y^2} + \frac{(\mu_w \sigma_y - \mu_y \sigma_w \rho)^2}{\sigma_w^2 \sigma_y^2 (1-\rho^2)}. \quad (A.9)$$

We require the two roots of $f(x)$, *say* a and b, to be such that:

$$a \leqq \frac{\mu_w}{\mu_y} \leqq b, \quad (A.11)$$

with at least one strict inequality.

Next, from (A.1):

$$f\left(\frac{\mu_w}{\mu_y}\right) = -Z_{\alpha/2}^2 \left\{ \sigma_y^2 \frac{\mu_w^2}{\mu_y^2} - 2 \sigma_w \frac{\mu_w}{\mu_y} \sigma_y \rho + \sigma_w^2 \right\}. \quad (A.12)$$

The quantity in the brackets in (A.12) is non-negative since its lower bound is:

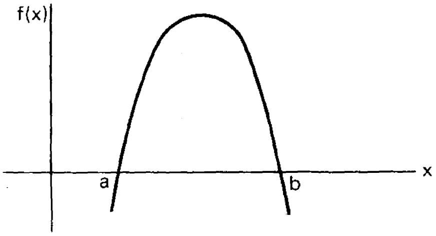
$$\sigma_y^2 \frac{\mu_w^2}{\mu_y^2} - 2 \sigma_w \frac{\mu_w}{\mu_y} \sigma_y + \sigma_w^2 = \left(\sigma_y \frac{\mu_w}{\mu_y} - \sigma_w \right)^2 \geqq 0,$$

which is attained for $\rho = +1$.

Then:

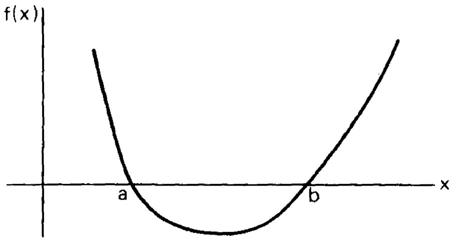
$$f\left(\frac{\mu_w}{\mu_y}\right) \leqq -Z_{\alpha/2}^2 \left(\sigma_y \frac{\mu_w}{\mu_y} - \sigma_w \right)^2 < 0,$$

since (A.8) holds. Next, if the coefficient of x^2 in (A.1) is negative, we have the following:



Since $f(\mu_w/\mu_y) < 0$, $\mu_w/\mu_y \in [a, b]$ which contradicts the required conditions (A.3).

If the coefficient of x^2 in (A.1) is positive, we have the following:



And since $f(\mu_w/\mu_y) < 0$, we have $\mu_w/\mu_y \in]a, b[$, i.e., conditions (A.3) are satisfied.

We thus have the condition:

$$\mu_y^2 - Z_{\alpha/2}^2 \sigma_y^2 > 0, \quad \text{i.e.,} \qquad Z_{1-\alpha/2} < \frac{\mu_y}{\sigma_y}. \tag{A.12}$$

Since (A.12) must hold, then (A.9) holds *a fortiori*. Thus, the required conditions are:

$$\left.\begin{aligned} &\frac{\mu_y}{\sigma_y} > c, \qquad c \geqq 3 \text{ or } 4 \\ &Z_1 - \frac{\alpha}{2} < \frac{\mu_y}{\sigma_y} \end{aligned}\right\} \tag{A.13}$$

Also if:

$$\rho = +1 \text{ then } \frac{\mu_y}{\sigma_y} \neq \frac{\mu_w}{\sigma_w}. \tag{A.14}$$

REFERENCES

1. R. C. GEARY, *The Frequency Distribution of the Quotient of Two Normal Variables*, Royal Statistics Society Journal, Vol. 93, 1930, pp. 442-446.
2. J. C. HAYYA and W. L. FERRARA, *On Normal Approximations of the Frequency Functions of Standard Forms Where the Main Variables are Normally Distributed*, Management Science, Vol. 19, No. 2, October 1972, pp. 173-186.
3. J. HAYYA, D. ARMSTRONG and N. GRESSIS, *A Note on the Ratio of Two Normally Distributed Variables*, Management Science, Vol. 21, No. 11, July 1975, pp. 1338-1341.
4. U.S. Department of Commerce, Bureau of the Census, *Value of New Construction Put in Place*, 1947 *to* 1974, C30-74S, issued December 1975.
5. U.S. Department of Commerce, Bureau of the Census, *Construction Reports: Price Index of New One-Family Houses Sold*, Second Quarter 1976, C27-76-Q2, issued September 1976.