ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 216, 2008

*Grzegorz Kończak**

ON TESTING THE HYPOTHESIS OF STABILITY OF THE RATIO OF TWO RANDOM VARIABLES

ABSTRACT. The problem of testing the hypothesis of the stability of expected value of the ratio of random variables is considered in the article. The ratio of random variables X and Y under assumptions of the autoregression models is analyzed. A test which makes it possible to verify the mentioned hypothesis is proposed. The problem being considered in the article appears in the statistical quality control when the proportions of dimensions of the product or the proportion of components in mixtures are required to be stable.

Key words: ratio, testing stability, autoregression.

1. INTRODUCTION

Problems connected with the estimation of the parameters of the ratio of random variables can be found in various practical statistical analyses. The problem of testing the hypothesis of the stability of expected value of the ratio of random variables is considered in the article. It is assumed that measurements of random variables X and Y in n periods are made. The ratio of two random variables X and Y is analyzed in the paper. The stability of the expected value of the ratio in periods $t = 1, 2, \ldots, n$ is assumed and the hypothesis of the stability of the ratio of X and Y for period $t = n+1$ is verified. A test which makes it possible to verify the mentioned hypothesis is proposed and its properties are presented in the paper.

The problem being considered in the article appears in the statistical quality control when the proportions of dimensions of the product or the proportion of components in mixtures are to be stable. The expected values of the random variables can change but the ratio of these variables should be stable over time. The advantages of the use of the proposed test instead of simultaneous testing the expected values of random variables X and Y are presented.

*Ph.D., Department of Statistics, Katowice University of Economics, koncz@ae.katowice.pl.

II. BASIC DEFINITIONS

Let $\{Y_t, t = 1,...,n\}$ and $\{X_t, t = 1,...,n\}$ be two normally distributed time series with constant means μ_Y and μ_X respectively. Let us assume that the random variables Y_t and X_t are independent and that their variances are stable over time and are equal to σ_Y^2 and σ_X^2. These assumptions can be written as follows:

$$\begin{gathered} E(Y_1) = E(Y_2) = ... = E(Y_n) = \mu_Y , \\ E(X_1) = E(X_2) = ... = E(X_n) = \mu_X , \\ D^2(Y_1) = D^2(Y_2) = ...D^2(Y_n) = \sigma_Y^2 , \\ D^2(X_t) = D^2(X_2) = ... = D^2(X_n) = \sigma_X^2 , \\ Cov(X_i, Y_j) = 0 \quad \text{for } i, j = 1, 2, \ldots, n, \end{gathered} \tag{1}$$

Under these assumptions the ratio of random variables Y_t and X_t

$$Z = \frac{Y_t}{X_t} \quad \text{for } t = n+1$$

will be considered.

III. RATIO OF RANDOM VARIABLES Y_t AND X_t

Let $x_1, x_2, ..., x_{n+1}$ and $y_1, y_2, ..., y_{n+1}$ be two random samples. Let us assume that the expected value of the ratio Z is stable for periods $t = 1, 2, \ldots, n$. It can be written as follows:

$$E\left(\frac{Y_t}{X_t}\right) = r = const . \tag{2}$$

There are various methods of determining the expected value of the ratio of two random variables. If $f_Y(y)$ and $f_X(x)$ are probability densities of random variables Y and X then the probability density of the random variable Z (M. G. Kendall, 1945) can be written as follows

$$f(z)=\int_{-\infty}^{\infty}|x|f_X(x)f_Y(zx)dx. \quad (3)$$

The probability density (3) can be used for calculating expected value of the ratio (2).

The ratio of normal variables was considered by R.C. Geary (1930), G. Marsaglia (1965), D. V. Hinkley (1969), A. Cedilnik et al (2004, 2006). In the case of the ratio of two normal random variables with zero expected values the ratio has Cauchy distribution and has no expected value, variance and no moments of higher orders.

If $X \sim N(0,1)$ and $Y \sim N(0,1)$, then the ratio (2) has Cauchy distribution ($Z \sim C(0,1)$). In the case when $X \sim N(0,\sigma_X)$ and $Y \sim N(0,\sigma_Y)$ the ratio has Cauchy distribution $Z \sim C\left(0,\frac{\sigma_Y}{\sigma_X}\right)$. The ratio of two arbitrary normal random variables $X \sim N(\mu_Y,\sigma_X)$ and $Y \sim N(\mu_Y,\sigma_Y)$ for $\mu_Y \neq 0$ and $\mu_X \neq 0$ is discussed in G. Marsaglia (1965) and leads to Cauchy-like distribution, but if $\frac{\mu_X}{\sigma_X} >> 0$ then it leads to normal distribution (D.V. Hinkley, 1969).

The Taylor series expansion can be used for estimating parameters (if they exist) of the ratio (2). The expected value and variance of the ratio can be written as follows:

$$E\left(\frac{Y}{X}\right)\approx\frac{\mu_Y}{\mu_X}+\frac{\mu_Y}{\mu_X^3}D^2(X)-\frac{1}{\mu_X^2}\sigma_{XY}, \quad (4)$$

$$D^2\left(\frac{Y}{X}\right)\approx\frac{\mu_Y^2}{\mu_X^2}-\frac{\mu_Y}{\mu_X}+\left(3\frac{\mu_Y^2}{\mu_X^4}-\frac{\mu_Y}{\mu_X^3}\right)\sigma_X^2+\frac{1}{\mu_X^2}\sigma_Y^2. \quad (5)$$

For estimating the parameters of complex estimators the resampling methods are often used. The resampling methods used most often for estimating the parameters of function of random variables are jackknife and bootstrap.

IV. THE CASE OF AUTOCORRELATED DATA

The standard assumptions in various statistical controls analysis are that the data generated by the process being in control are normally and independently distributed. This assumption is made in Shewhart's control chart model which can be written as follows

$$X_t = \mu + \varepsilon_t \quad t = 1, 2, \ldots,$$

where ε_t ($t = 1, 2, \ldots$) are normally and independently distributed with mean 0 and standard deviation σ. The assumption of independent observations is fulfilled in many real processes. In real processes the data are often correlated over time (D.C. Montgomery, C. M. Mastrangelo, 2000). Numerous industrial processes produce data that change over time (R. L. Mason, J. C. Young, 2000).

Let us assume the first-order autoregressive model AR(1) for time series Y_t and X_t. We can write then (G. E. P. Box, G. M. Jenkins, 1983)

$$\widetilde{X}_t = \varphi_X \widetilde{X}_{t-1} + \varepsilon_{Xt} \tag{6}$$

and

$$\widetilde{Y}_t = \varphi_Y \widetilde{Y}_{t-1} + \varepsilon_{Yt}, \tag{7}$$

where $\widetilde{X}_t = X_t - \mu_X$, $\widetilde{Y}_t = Y_t - \mu_Y$ and $\varepsilon_{Xt}, \varepsilon_{Yt}$ are normally and independently distributed with mean 0 and standard deviations σ_{Xt}, σ_{Yt} for $t = 1, 2, 3, \ldots, n$. These processes are stationary when $-1 < \varphi_X < 1$ and $-1 < \varphi_Y < 1$. The autoregressive models (6) and (7) can be written as follows:

$$X_t = \varphi_X X_{t-1} + \mu_X (1 - \varphi_X) + \varepsilon_{Xt} \tag{8}$$

and

$$Y_t = \varphi_Y Y_{t-1} + \mu_Y (1 - \varphi_Y) + \varepsilon_{Yt}. \tag{9}$$

V. TEST OF STABILITY OF THE RATIO OF RANDOM VARIABLES

Let $x_1, x_2, ..., x_{n+1}$ and $y_1, y_2, ..., y_{n+1}$ be the observed time series (t = 1, 2, ..., n). Under the assumption (2) the hypothesis of the stability of the expected value of the ratio will be tested. It can be written as follows:

$$H_0 : E\left(\frac{Y_{n+1}}{X_{n+1}}\right) = E\left(\frac{\sum_{t=1}^{n} Y_t}{\sum_{t=1}^{n} X_t}\right) \tag{10}$$

against the alternative

$$H_1 : E\left(\frac{Y_{n+1}}{X_{n+1}}\right) \neq E\left(\frac{\sum_{t=1}^{n} Y_t}{\sum_{t=1}^{n} X_t}\right). \tag{11}$$

On the base of the observed values the parameters of the regression models (8) and (9) can be estimated and the theoretical values $\hat{x}_1, \hat{x}_2, ..., \hat{x}_{n+1}$ and $\hat{y}_1, \hat{y}_2, ..., \hat{y}_{n+1}$ of time series can be calculated. The residuals $e_{y,t} = y_t - \hat{y}_t$ and $e_{x,t} = x_t - \hat{x}_t$ (t = 1, 2, ..., n+1) can also be then calculated. Under the H_0 hypothesis the residuals have normal distributions with means 0 and variances σ_Y^2 and σ_X^2 respectively.

From the above we have that $\widetilde{e}_{y,t} = \frac{e_{y,t}}{\sigma_Y}$ and $\widetilde{e}_{X,t} = \frac{e_{X,t}}{\sigma_X}$ have normal distributions with mean 0 and standard deviations 1. To verify the hypothesis (10) the following statistic will be used:

$$T = \frac{U}{V}. \tag{12}$$

where the statistics U and V are given by:

$$U = \frac{1}{2}\left(\widetilde{e}_{Y,n+1} - \widetilde{e}_{X,n+1}\right)^2, \tag{13}$$

$$V = \sum_{t=1}^{n} \widetilde{e}_{Y,t}^2 + \sum_{t=1}^{n} \widetilde{e}_{X,t}^2. \tag{14}$$

Under the H_0 hypothesis the statistic U has chi square distribution with 1 degree of freedom and the statistic V has chi square distribution with $2n$ degrees of freedom. The ratio statistic T given by (12) has F distribution with 1 and $2n$ degrees of freedom.

VI. MONTE CARLO STUDY

The approximate values of the probabilities of rejection the hypothesis (10) for various discordance levels in models (8) and (9) are obtained in Monte Carlo analysis. In the simulation process observations $x_1, x_2, ..., x_{10}$ and $y_1, y_2, ..., y_{10}$ were generated using the (8) and (9) models. The $11th$ observations for both time series were generated using (15) and (16).

$$Y_t = \varphi_Y Y_{t-1} + \mu_Y(1 - \varphi_Y) + k_Y \sigma_Y + \varepsilon_{Yt} \tag{15}$$

and

$$X_t = \varphi_X X_{t-1} + \mu_X(1 - \varphi_X) + k_X \sigma_X + \varepsilon_{Xt} \tag{16}$$

where $k_Y, k_X = 0, 1, 2, 3$.

Then the statistic (12) was calculated and the values were compared to critical value $F_{0,05;\ 1,\ 20}$ which is equal to 4.35. The procedure was repeated 1000 times for each combination k_Y and k_X (k_Y, k_X = 0, 1, 2, 3) with $\mu_Y = 11, \mu_X = 10, \sigma_Y = \sigma_X = 1$. The results of Monte Carlo study are presented in Table 1 and on the Figure 1.

Table 1

The estimated values of probabilities of rejection the null hypothesis

k_X	k_Y			
	0	1	2	3
0	0.050*	0.067	0.176	0.371
1	0.093	0.035	0.071	0.191
2	0.166	0.063	0.037	0.048
3	0.304	0.094	0.110	0.080

*) the significance level (not simulated)
Source: Monte Carlo study.

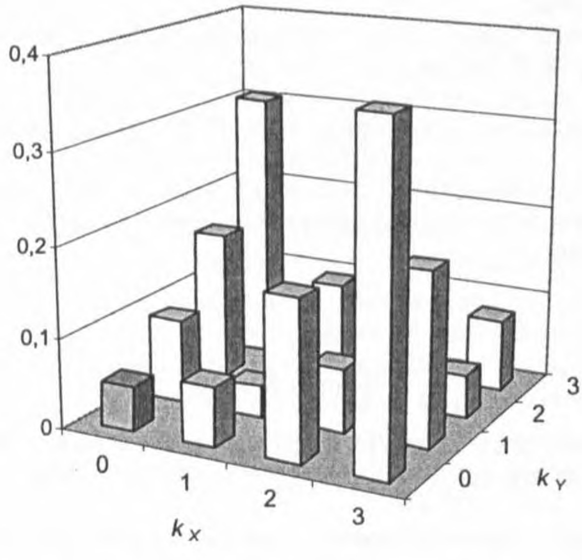

Fig. 1 The estimated values of probabilities of rejection the null hypothesis

The probabilities of rejection the hypothesis (10) are greater for $k_Y = 3$, $k_X = 0$ or $k_Y = 0$, $k_X = 3$ (there are the greatest changes in the ratio). If $k_Y = k_X$ then the probability of the rejection the null hypothesis is close to $\alpha = 0.05$ (changes in the ratio are small).

The proposed test can be used for testing the stability of expected value of the ratio of random variables for which autoregression models are assumed. The probability of rejection the hypothesis of stability increases when the expected value of one of the analyzed variables change and is almost close to significant level α for consistence proportional changes of the expected values of both of the random variables.

REFERENCES

Box G. E. P., Jenkins G. M. (1983) *Analiza szeregów czasowych. Prognozowanie i sterowanie*, Państwowe Wydawnictwo Naukowe, Warszawa.

Cedilnik A., Kosmelj K., Blejec A. (2004) *The Distribution of the Ratio of Jointly Normal Variables*, Metodoloski zvezki, vol. 1, no. 1, s. 99–108.

Cedilnik A., Kosmelj K., Blejec A. (2006) *Ratio of Two Random Variables: A Note on the Existence of its Moments*, Metodoloski zvezki, vol. 3, no. 1, s. 1–7.

Geary R. C. (1930) *The frequency distribution of the quotient of two normal variates*, Journal of the Royal Statistical Society 93, 442–6, 1930

Hinkley D. V. (1969) *On the ratio of two correlated normal random variables*, Biometrika 56, s. 635-639.

Kendall M. G. (1945) *The Advanced Theory of Statistics*, Charles Griffin & Company Limited, Londyn.

Marsaglia G. (1965) *Ratios of normal variables and ratios of sums of uniforms variables*. JASA, 60, s. 163–204.

Mason R. L., Young J.C. (2000), *Autocorrelation in Multivariate Processes*, w: Statistical process monitoring and optimization, s. 387–394, Marcel Dekker Inc, New York – Basel.

Montgomery D. C., Mastrangelo C. M. (2000) *Process monitoring with autocorrelated data*, in: Statistical process monitoring and optimization s. 139–160, Marcel Dekker, New York-Basel.

Grzegorz Kończak

O TESTOWANIU HIPOTEZY O STABILNOŚCI WARTOŚCI OCZEKIWANEJ ILORAZU DWÓCH ZMIENNYCH LOSOWYCH

W artykule zaprezentowano propozycję testu pozwalającego na weryfikację hipotezy o stabilności wartości oczekiwanej ilorazu zmiennych losowych. W rozważaniach przyjęto, że w kolejnych *n* okresach czasowych dokonywane są pomiary wartości zmiennych losowych Y i X. Przedmiotem analizy jest iloraz zmiennych losowych $Z = Y/X$ przy założeniu modelu autoregresji. Problem przedstawiony w artykule jest spotykany w zagadnieniach statystycznej kontroli jakości, gdy niezbędne jest zachowanie odpowiedniej proporcji np. dla wymiarów produktu czy proporcji składników w mieszaninach.