ACTA UNIVERSITATIS LODZIENSIS

FOLIA OECONOMICA 54, 1986

Czesław Domański*, Andrzej S. Tomaszewicz*

A METHOD FOR COMPUTING THE POWER OF THE TEST BASED ON THE NUMBER OF RUNS IN THE CASE OF THE SECOND ORDER AUTOCORRELATION PROCESS

1. Introduction

The paper presents a generalization of the results obtained in Domański, Tomaszewicz (1980) in the case of the second order autocorrelation. We shall consider the sequence of random variables

$$X_1, X_2, \ldots, X_n, \quad (n > 2), \tag{1}$$

about which the following assumptions are made.

1. Each of the variables has a two-point zero-one distribution

$$P(X_t = 0) + P(X_t = 1) = 1, \text{ for } t = 1, 2, \ldots, n.$$

2. The variables X_t are linked into a second order Markov chain

$$P(X_t = x_t \mid X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}, \ldots, X_1 = x_1) \tag{2}$$
$$= P(X_t = x_t \mid X_{t-1} = x_{t-1}, X_{t-2} = x_{t-2}),$$

for $t = 1, 2, \ldots, n$ and arbitrary $x_t, x_{t-1}, \ldots, x_1 \in \{0, 1\}$.

3. The chain X_t is stationary (in the narrower sense)

$$P(X_t = x_t, X_{t+1} = x_{t+1}, \ldots, X_n = x_n)$$
$$= P(X_1 = x_t, X_2 = x_t, \ldots, X_{n-t+1} = x_n).$$

*Lecturers, Institute of Econometrics and Statistics, University of Łódź.

for $t = 1, 2, \ldots, n$ and arbitrary $x_{t+1}, x_{t+2}, \ldots, x_n \in \{0, 1\}$.

We aim at constructing an algorithm for determining the distribution of the number of runs in (1), i.e. the random variable

$$R_n = 1 + \operatorname{card}\left\{t : 2 \leqslant t \leqslant n,\ X_{t-1} \neq X_t\right\}. \tag{3}$$

We shall begin our considerations from the specification in paragraphs 2-5, of the basic properties of joint distributions of subsequent variables of the chain (1); because of the assumption 2 we shall confine ourselves to the distributions of three variables at most. Paragraphs 6 and 7 are dealing with the distribution of the number of runs R_n and test power based on this statistic.

2. A Univariate Distribution

Assume the notation

$$P(X_t = 1) = p, \quad P(X_t = 0) = q = 1 - p.$$

The formulae:

$$EX_t = p, \quad EX_t^2 = p, \quad D^2X_t = pq,$$

are generally known.

3. A Bivariate Distribution

This distribution is two-parametric. Four probabilities

$$p_{hj} = P(X_{t-1} = h,\ X_t = j), \quad h, j = 0, 1, \tag{4}$$

are dependent on

$$p_{00} + p_{01} + p_{10} + p_{11} = 1 \tag{5}$$

and on the stationarity condition

$$p_{01} + p_{11} = p$$

and

$$p_{10} + p_{11} = p. \tag{6}$$

Hence

$$p_{01} = p_{10}. \tag{7}$$

If in a special case X_{t-1} and X_t are independent, then $p_{11} = p^2$; assume generally

$$p_{11} = p^2 + \varkappa. \tag{8}$$

Hence, both from (6) and (7)

$$p_{01} = p_{10} = p - p_{11} = pq - \varkappa \tag{9}$$

and

$$p_{00} = q - p_{01} = q^2 + \varkappa. \tag{10}$$

Of course,

$$EX_{t-1}X_t = p_{11} = p^2 + \varkappa,$$

thus

$$\mathrm{cov}(X_{t-1}, X_t) = EX_{t-1}X_t - EX_{t-1}EX_t = p^2 + \varkappa - p^2 = \varkappa.$$

Therefore, the correlation coefficient between X_{t-1} and X_t is described by the formula

$$\varrho = \frac{\mathrm{cov}(X_{t-1}, X_t)}{DX_{t-1}\ DX_t} = \frac{\varkappa}{pq}$$

and thus

$$\varkappa = pq\varrho.$$

We shall write the bivariate distribution X_t, X_{t-1} in the table.

Table 1

Joint distribution of two zero-one variables

X_t X_{t-1}	0	1	Sum
0	$q(q + p\varrho)$	$pq(1 - \varrho)$	q
1	$pq(1 - \varrho)$	$p(p + q\varrho)$	p
Sum	q	p	1

We shall assume the following notation for conditional probabilities

$$w_h = P(X_t = 1 \mid X_{t-1} = h),\ u_h = 1 - w_n = P(X_t = 0 \mid X_t = h), \quad h = 0,1. \tag{11}$$

We have

$$w_o = \frac{p_{01}}{q} = p(1 - \varrho), \qquad u_o = q + p\varrho, \tag{12}$$

$$w_1 = \frac{p_{11}}{\varrho} = p + q\varrho, \qquad u_1 = q(1 - \varrho).$$

Hence $w_o + u_1 = 1 - \varrho$, thus

$$\varrho = 1 - w_o - u_1 \tag{13}$$

and

$$p = \frac{w_o}{w_o + u_1}, \quad q = \frac{u_1}{w_o + u_1}. \tag{14}$$

If variables (1) are linked into the first order Markov chain, i.e.

$$P(X_t = x_t \mid X_{t-1} = x_{t-1}, \ldots, X_1 = x_1) = P(X_t = x_t \mid X_{t-1} = x_{t-1}) \tag{15}$$

for $t = 2, \ldots, n$ and arbitrary $x_{t-1}, x_{t-2}, \ldots, x_1$, then the probabilities (11) form its transition matrix

$$M = \begin{bmatrix} 1 - w_o & w_o \\ u_1 & 1 - u_1 \end{bmatrix}.$$

The regression function of the first type

$$E(X_t | X_{t-1} = h) = P(X_t = 1 | X_{t-1} = h) = w_h$$

is linear, since (cf. (12))

$$w_h = p(1 - \varrho) + h\varrho.$$

4. A Three-Dimensional Distribution. A General Case

Eight probabilities

$$p_{hjk} = P(X_{t-2} = h, X_{t-1} = j, X_t = k), \; h,j,k = 0, 1$$

of the joint distribution (X_{t-2}, X_{t-1}, X_t) are linked by the dependence

$$\sum_{h=0}^{1} \sum_{j=0}^{1} \sum_{k=C}^{1} p_{hjk} = 1$$

and by two stationarity conditions of univariate marginal distributions

$$P(X_{t-2} = 1) = P(X_{t-1} = 1) = P(X_t = 1)$$

and by one stationarity condition of bivariate marginal distributions

$$P(X_{t-2} = 1, X_{t-1} = 1) = P(X_{t-1} = 1, X_t = 1).$$

The distribution(X_{t-2}, X_{t-1}, X_1) is therefore four-parametric.

Assume that transition probabilities, i.e. conditional probabilities

$$w_{hj} = P(X_t = 1 \mid X_{t-2} = h, X_{t-1} = j) \tag{16}$$

are given.

Denoting, similarly as previously (cf. (11))

$$u_{hj} = 1 - w_{hj},\ h,\ j = 0,\ 1. \tag{17}$$

we express other characteristics of the distribution (X_{t-2}, X_{t-1}, X_t) as a function of four parameters

$$w_{00},\ w_{01},\ w_{10},\ w_{11}.$$

Besides, we shall use the notation from (4), (8)-(10) for bivariate distributions (X_{t-2}, X_{t-1}) and (X_{t-1}, X_t) and

$$\begin{aligned} v_0 &= w_{00} + u_{10}, \\ v_1 &= w_{01} + u_{11}, \\ v &= w_{00}v_1 + u_{11}v_0, \\ d_1 &= w_{00}u_{11}, \\ d_2 &= w_{01}u_{10}, \\ d &= d_1 - d_2 \end{aligned} \tag{18}$$

From the obvious dependences

$$p_{hj1} = p_{hj}w_{hj}, \quad h,\ j = 0,\ 1,$$

the system of equations

$$\begin{aligned} p_{00}w_{00} + p_{10}w_{10} &= p_{01}, \\ p_{01}w_{01} + p_{11}w_{11} &= p_{11}, \end{aligned}$$

follows.

Taking into account formulae (8)-(10) and solving this system vs. p and we obtain

$$p = \frac{w_{00}v_1}{v}, \quad q = \frac{u_{11}v_0}{v} \tag{19}$$

and

$$\varrho = \frac{-d}{v_0 v_1}. \tag{20}$$

In some cases it is more convenient to use another form of this formula

$$\varrho = 1 - \frac{w_{00}}{v_0} - \frac{u_{11}}{v_1} \tag{21}$$

and hence both from (8), (9) and (10)

$$p_{00} = q\,(q + p\varrho) = \frac{u_{10}u_{11}}{v}, \tag{22}$$

$$p_{01} = p_{10} = pq(1 - \varrho) = \frac{w_{00}u_{11}}{v},$$

$$p_{11} = p(p + q\varrho) = \frac{w_{00}w_{01}}{v}.$$

Let ϱ_2 denote a second order autocorrelation coefficient of the process (1), i.e. the correlation coefficient between X_t and X_{t-2}.

Similarly as (9) we have

$$P(X_{t-2} = 0, X_t = 1) = pq(1 - \varrho_2).$$

On the other hand

$$P(X_{t-2} = 0, X_t = 1) = p_{001} + p_{010} = p_{00}w_{00} + p_{01}w_{01}.$$

Thus, after taking into account (20) and (19)

$$\frac{u_{10}u_{11}}{v}\,w_{00} + \frac{w_{00}u_{11}}{v}\,w_{01} = \frac{w_{00}v_1\ u_{11}\ v_0}{v^2}\,(1 - \varrho_2)$$

and hence

$$\varrho_2 = 1 - \frac{(u_{10} + w_{01})\ v}{v_0 v_1}\ . \tag{23}$$

5. A Three-Dimensional Distribution with Linear Regression

The first-type regression in the distribution (X_{t-2}, X_{t-1}, X_t) is not usually linear due to (16).

A conditional expected value can be written in the form

$$E(X_t | X_{t-2} = h,\ X_{t-1} = j) = w_{hj}$$

and thus

$$\begin{aligned} E(X_t \mid X_{t-2} &= h,\ X_{t-1} = j) \\ &= w_{00} + h(w_{10} - w_{00}) + j(w_{01} - w_{00}) \\ &+ hj\ (w_{11} - w_{10} - w_{01} + w_{00}). \end{aligned}$$

Therefore, we obtain the linearity condition

$$w_{11} - w_{10} - w_{01} + w_{00} = 0,$$

which, taking into account (18) can be written in the form

$$w_{00} + u_{10} = w_{01} + u_{11}, \tag{24}$$

or (cf. (18))

$$v_0 = v_1. \tag{25}$$

When condition (25) is satisfied, we have from (18) and (17)

$$v = (w_{00} + u_{11})v_0. \tag{26}$$

Formulae (22) are therefore simplified to

$$p = \frac{w_{00}}{w_{00} + u_{11}},\quad q = \frac{u_{11}}{w_{00} + u_{11}}. \tag{27}$$

Similarly, instead of (21) we can write

$$\varrho = 1 - \frac{w_{00} + u_{11}}{v_0}$$

or

$$\varrho = \frac{u_{10} - u_{11}}{w_{00} + u_{10}}. \qquad (28)$$

From (23) and (26) we have

$$\varrho_2 = 1 - \frac{(u_{10} + w_{01})(w_{00} + u_{11})}{w_{00} + u_{10}}. \qquad (29)$$

The distribution (X_{t-2}, X_{t-1}, X_t) with the linearity regression condition (24) is three-parametric. All its characteristics can be described by three conditional probabilities (cf. formulae (27)-(29)) or by parameters p, ϱ, ϱ_2. The easy transformations lead to the formulae

$$\begin{aligned} w_{00} &= \frac{p(1 - \varrho_2)}{1 + \varrho}, \\ w_{01} &= \frac{(p + q\varrho)(1 - \varrho_2)}{1 - \varrho^2}, \qquad (30) \\ u_{10} &= \frac{(p\varrho + q)(1 - \varrho_2)}{1 - \varrho^2}, \\ u_{11} &= \frac{q(1 - \varrho_2)}{1 + \varrho}. \end{aligned}$$

In a special case the variables (1) are linked into the first order Markov chain, i.e. the condition stronger than (2) holds:

$$P(X_t = x_t \mid X_{t-1} = x_{t-1}, \ldots, X_1 = x_1)$$
$$= P(X_t = x_t \mid X_{t-1} = x_{t-1}).$$

Then the equalities of conditional probabilities must be fulfilled

$$w_{00} = w_{10},\ w_{11} = w_{01},$$

i.e.

$u_{10} = 1 - w_{00}, w_{01} = 1 - u_{11}.$

Hence from (28) and (29)

$$\varrho = 1 - w_{00} - u_{11}$$

and

$$\varrho_2 = (1 - w_{00} - u_{11})^2 = \varrho^2 .$$

6. The Distribution of the Number of Runs

Now we shall consider the distribution of the number of runs R_n in the sequence (1). We shall define by $Q_{hj}(n, r)$ the probability that the sequence (1) contains r runs and its two last elements are h and j

$$Q_{hj}(n, r) = P(R_n = r, X_{n-1} = h, X_n = j), h = 0, 1,$$

and

$$Q(n, r) = \sum_h \sum_j Q_{hj}(n, r). \tag{31}$$

For n = 2, of course

$$Q_{00}(2, 1) = p_{00}, \tag{32}$$
$$Q_{01}(2, 2) = p_{01},$$
$$Q_{10}(2, 2) = p_{10},$$
$$Q_{11}(2, 1) = p_{11}.$$

Hence

$$Q(2, 1) = p_{00} + p_{11} ,$$
$$Q(2, 2) = p_{01} + p_{10}.$$

For n > 2, because of the assumption (2) and the dependence

$$R_{n-1} = \begin{cases} R_n & \text{for } h = j \\ R_{n-1} & \text{for } h \neq j \end{cases}$$

we have

$$
\begin{aligned}
Q_{hj}(n, r) &= P(R_n = r, X_{n-1} = h, X_n = j) \\
&= \sum_{g=0}^{1} P(R_n = r, X_{n-2} = g, X_{n-1} = h, X_n = j) \\
&= \sum_{g=0}^{1} P(R_{n-1} = r - \delta_{hj}, X_{n-1} = g, X_{n-1} = h) \cdot \\
&\quad P(X_n = j \mid X_{n-2} = g, X_{n-1} = h)
\end{aligned}
$$

Therefore, we obtain the general formula

$$Q_{hj}(n, r) = \sum_{9=0}^{1} Q_{gn}(n - 1, r - \delta_{hj}) \left(1 - j + (2_j - 1) w_{gh}\right) \quad (33)$$

in which

$$
\begin{aligned}
1 - j + (2j - 1)w_{gh} &= P(X_n = j \mid X_{n-2} = g, X_{n-1} = h) = \\
&= \begin{cases} u_{gh} & \text{for } j = 0, \\ w_{gh} & \text{for } j = 1. \end{cases}
\end{aligned}
$$

In particular

$$
\begin{aligned}
Q_{00}(n, r) &= Q_{00}(n - 1, r)u_{00} + Q_{10}(n - 1, r)\, u_{10}, \\
Q_{01}(n, r) &= Q_{00}(n - 1, r - 1)w_{00} + Q_{10}(n - 1, r - 1)w_{10}, \quad (34) \\
Q_{10}(n, r) &= Q_{01}(n - 1, r - 1)u_{01} + Q_{11}(n - 1, r - 1)u_{11}, \\
Q_{11}(n, r) &= Q_{01}(n - 1, r)w_{01} + Q_{11}(n - 1, r)w_{11}.
\end{aligned}
$$

Formula (33) with initial conditions (32) can be a basis for numerical determination of probability distribution of the number of runs R_n in the second order Markov process.

7. The Power of Randomized Run Tests

Let us assume that we verify the hypothesis of independence of the sequence of random variables (1) linked into the second order Markov chain. We assume for simplicity that the linearity regression condition (24) is satisfied[1], i.e. the joint distribution (X_{t-2}, X_{t-1}, X_t) is three-parametric. Using the test based on R_n statistic we can verify the independence hypothesis X_t:

$$H_0 : \varrho = \varrho_2 = 0.$$

Depending on the alternative hypothesis ($H_1 : \varrho < 0, H_1 : \varrho > 0, H_1 : \varrho \neq 0$) we assume respectively right-hand-side, left-hand-side or two-sided critical region for the R_n statistic.

For the purposes of a comparative analysis of the power of run tests it is wortwhile to consider the randomized tests which guarantee the same probabilities of the first type error (cf. Domański, Tomaszewicz 1980).

For the left-hand-side region we have the critical value

$$r_\alpha = \max\left\{ r : P_0(R \leqslant r) \leqslant \alpha \right\} \tag{35}$$

and

$$p_\alpha = \frac{\alpha - P_0(R \leqslant r_\alpha)}{P_0(R = r_\alpha + 1)}, \tag{36}$$

where P_0 denotes probability calculated under the assumption that H_0 is true.

The test works as follows:

H_0 is rejected if $R \leqslant r_\alpha$,

H_0 is accepted if $R > r_\alpha + 1$,

H_0 is rejected with probability p_α if $R = r_\alpha + 1$.

The power of the randomized test of runs is expressed by the formula

$$1 - \beta = P_1(R \leqslant r_\alpha) + p_\alpha\ P_1(R = r_\alpha + 1). \tag{37}$$

[1] The considerations for the general case do not differ significantly.

All probabilities, and therefore the power of the run test can be calculated using recursive formulae (33).

References

Domański Cz., Tomaszewicz A. S., 1980, Variants of Tests Based on the Length of Runs. Paper presented at the Conference on Problems of Building and Estimation of Large Econometric Models, Polanica.

Czesław Domański, Andrzej S. Tomaszewicz

METODA OBLICZANIA MOCY TESTU OPARTEGO NA LICZBIE SERII W PRZYPADKU ŁAŃCUCHA MARKOWA DRUGIEGO RZĘDU

Artykuł dotyczy podstawowych własności testów opartych na liczbie serii weryfikujących hipotezę o niezależności ciągu zmiennych losowych

$X_1, X_2, \ldots, X_n$,

przy założeniu, że powiązane są one w łańcuch Markowa drugiego rzędu, tj.

$$P(X_t|X_1 = x_1, \ldots, X_{t-1} = x_{t-1}) = P(X_1|X_{t-2} = X_{t-2}, X_{t-1} = x_{t-1}), \text{ dla } t = 3, 4, \ldots, n.$$

Autorzy ograniczyli swe rozważania do przypadku, gdy zmienne X_i mają rozkład dwupunktowy, a ponadto trójwymiarowy rozkład X_{t-2}, X_{t-1}, X_t jest stacjonarny dla $t = 3, 4, \ldots, n$.