A C T A U N I V E R S I T A T I S L O D Z I E N S I S

FOLIA OECONOMICA 68, 1987

Czesław Domański*, Andrzej Stefan Tomaszewicz*

ANALYSIS OF THE POWER OF THE INDEPENDENCE TEST BASED ON THE NUMBER OF RUNS IN THE CASE OF THE SECOND-ORDER MARKOV CHAIN

1. Introduction

Most of the publications concerning the application of run tests to verify random variables independence hypothesis base on the theory of the first-order Markov chains. In this paper we present an attempt to estimate the power of run tests under the assumption that the random variables generate a second-order Markov chain. It concerns the situation when the information about the value of the random variable at the previous period not fully determines its distribution at the current period.

Let

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

be a sequence of random variables such that:

1^o each of the variables has a zero-one distribution

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

2^o the variables X_t are said to form a second-order Markov chain, i.e.

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

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

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

3^o the chain X_t is stationary

$$(4) \quad P(X_t = x_t, X_{t+1} = x_{t+1}, \ldots, X_n = x_n) =$$
$$= P(X_1 = x_t, X_2 = x_t, \ldots, X_{n-t-1} = x_n).$$

Let us consider a random variable which is a number of runs in the sequence (1)

$$(5) \quad R_n = 1 + \mathrm{card}\{t : 2 \leqslant t \leqslant n, X_{t-1} \neq X_t\}.$$

In accordance with [2] we shall assume the following symbols for stationary probabilities:

$$(6) \quad p_{ghj} = P(X_{t-2} = g, X_{t-1} = h, X_t = j),$$

$$(7) \quad p_{hj} = P(X_{t-1} = h, X_t = j)$$

$$(8) \quad p_j = P(X_t = j),$$

for $g, h, j = 0, 1$ and $t = 1, 2, \ldots, n$ for $t \geqslant 3$ in (6) and $t \geqslant 2$ in (7). Transition probabilities:

$$(9) \quad p_{j|gh} = P(X_t = j | X_{t-2} = g, X_{t-1} = h)$$

for $g, h, j = 0, 1$; $t = 3, 4, \ldots, n$; of course

$$(10) \quad p_{ghj} = p_{gh} \cdot p_{j|gh}.$$

The coefficient of the first-order autocorrelation is

$$(11) \quad \varrho = 1 - \frac{p_{1|00}}{p_{1|00} + p_{0|10}} - \frac{p_{0|11}}{p_{1|01} + p_{0|11}}.$$

The second-order autocorrelation coefficient for linear regression of the first kind is

$$(12)\quad \varrho_2 = 1 - \frac{(p_{0|10} + p_{1|01})(p_{1|00} + p_{0|11})}{p_{1|00} + p_{0|10}}.$$

2. The number of runs distribution

We shall present the distribution of the number of runs R_n (cf. (5)) in the sequence (1). Let $Q_{hj}(n,r)$ be the probability of the sequence (1) to contain r runs and its two last elements to be h and j. Hence

$$(13)\quad Q_{hj}(n, r) = P(R_n = r, X_{n-1} = h, X_n = j),\quad h = 0,1$$

and

$$(14)\quad Q(n, r) = \sum_h \sum_j Q_{hj}(n, r).$$

For n = 2, of course,

$$(15)\quad \begin{aligned} Q_{00}(2, 1) &= p_{00}, \\ Q_{01}(2, 2) &= p_{01}, \\ Q_{10}(2, 2) &= p_{10}, \\ Q_{11}(2, 1) &= p_{11}. \end{aligned}$$

Hence

$$(16)\quad \begin{aligned} Q(2, 1) &= p_{00} + p_{11}, \\ Q(2, 2) &= p_{01} + p_{10}. \end{aligned}$$

In [2] we have given a general formula

$$(17)\quad Q_{hj}(n, r) = \sum_{g=0}^{1} Q_{gh}(n - 1, r - \delta_{hj})(1 - j + (2j - 1)w_{gh}),$$

in which

$$1 - j + (2j - 1)w_{gh} = P(X_n = j | X_{n-2} = g, X_{n-1} = h) =$$

$$= \begin{cases} u_{gh} & \text{for } j = 0 \\ w_{gh} & \text{for } j = 1. \end{cases}$$

In particular

$$(18) \quad \begin{aligned} Q_{00}(n, r) &= Q_{00}(n-1, r)u_{00} + Q_{10}(n-1, r)u_{10}, \\ Q_{01}(n, r) &= Q_{00}(n-1, r-1)w_{00} + Q_{10}(n-1, r-1)w_{10}, \\ Q_{10}(n, r) &= Q_{01}(n-1, r-1)u_{01} + Q_{11}(n-1, r-1)u_{11}, \\ Q_{11}(n, r) &= Q_{01}(n-1, r)w_{01} + Q_{11}(n-1, r)w_{11}. \end{aligned}$$

The formula (18) with the initial conditions (15) has been the basis for the numerical determination of the probability of the number of runs R_n in the second-order Markov process.

3. The randomized run test

Let us assume that we are verifying the hypothesis of the independence of the sequence of random variables (1) chained in a second-order Markov chain. By way of simplification we shall assume that the condition of the linearity of regression is satisfied (considerations for the general case are not substantially different)

$$(19) \quad w_{11} - w_{10} - w_{01} + w_{00} = 0.$$

Using the test based on the R_n statistic (cf. (5)) we are to verify the hypothesis of the independence of X_t

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

According to the alternative hypothesis ($H_1 : \varrho < 0, H_1' : \varrho > 0, H_1'' : \varrho = 0$) we shall assume respectively a right-sided, left-sided, or a two-sided critical region for the run test.

In comparative studies of independence tests we can consider randomized tests, which ensure identical probabilities of making a type I error (cf. [1]).

For a left-sided region we have the critical value

$$(20) \qquad r_\alpha = \max\left\{r : P_o(R_n \leqslant r) \leqslant \alpha\right\}$$

and

$$(21) \qquad p_\alpha = \frac{\alpha - P_o(R_n \leqslant r_\alpha)}{P_o(R_n = r_\alpha + 1)}$$

where P_o denotes probability calculated for H_o assumed to be true.

The test is as follows:

1° H_o is rejected when $R_n < r_\alpha + 1$,

2° H_o is accepted when $R_n > r_\alpha + 1$,

3° H_o is rejected with the probability p_α, when $R_n = r_\alpha + 1$.

The power of the randomized run test is

$$(22) \qquad 1 - \beta = p_1(R_n \leqslant r) + p_\alpha p_1(R_n = r_\alpha + 1).$$

All the probabilities, the power of run test included can be calculated from the recursive formulae (17).

4. Scope of the study

In the three-parameter stochastic process the power of a run test is a function of five parameters: n, p, ϱ, ϱ_2, α, where n is the sample size, α is the significance level, and $p = p_1$, ϱ and ϱ_2 have been defined by (8), (11), and (12), respectively.

We have replaced the parameter ϱ_2 by the difference

$$(23) \qquad \delta = \varrho_2 - \varrho^2$$

measuring the "distance" of the given process from the first-order process, because we have $\varrho_2 = \varrho^2$ (cf. [2]) in the second-order Markov chain.

So the power of a randomized run test can be expressed as the function

$$(24) \qquad P(R_n \leqslant r_\alpha) + p_\alpha P(R_n = r_\alpha) = \varphi(n, p, \varrho, \delta, \alpha).$$

The quantiles r_{α} and p_{α} correspond to the significance level according to the formulae (20) and (21).

It seems impossible to find the analytical form of the function φ (24). That is why we have attempted to formulate some conclusions concerning the power of test obtained from numerical experiments. The values of the power function (24) have been determined for

$n = 5, 10, 20, 30, 50, 100, 150, 200,$

$p = 0.5, 0.6, 0.7, 0.75, 0.8, 0.9, 0.95,$

$\varrho = 0, 0.5, 0.9,$

$\delta = 0, 0.1, 0.2,$

$\alpha = 0.02, 0.05, 0.10.$

Because of the cost of computations, we have limited our considerations to positive autocorrelation, i.e. $\varrho > 0$, and accordingly, we consider only a left-sided run test (a small number of runs is the evidence of positive autocorrelation). The case of negative autocorrelation is presented only for a symmetrical distribution $p = 0,5$, calculations being made also for $\varrho = -0.9$ and -0.5.

Remembering the fact that not for all triples (p, ϱ, δ) there are stationary distributions

$$(25) \quad (X_{t-2}, X_{t-1}, X_t)$$

we have derived respective inequalities in the subsequent part of this paper.

5. Constraints on the parameters of the second-order Markov chain

There is no loss of generality in assuming that $p \geq 0.5$. Due to the traditional notation $q = 1 - p$ we assume

$$(26) \quad p \geq q.$$

We also include the inequalities

$$(27) \quad |\varrho| \leq 1 \quad \text{and} \quad |\varrho_2| \leq 1.$$

In some of the subsequent expressions we can find zero in the denominator (in extreme cases: $p = 1$, $|\varrho| = 0$, or $|\varrho_2| = 1$). Since it does not have any influence on final results, we have omitted detailed comments on this problem.

5.1. Constraints on ϱ with given p

Transition probabilities (28)

$$p_{j|h} = \frac{p_{jh}}{p_h}$$

fulfil the following inequalities

$$(28) \qquad 0 \leqslant p_{j|h} \leqslant 1 \qquad \text{for } h, j = 0, 1.$$

Since

$$p_{0|h} + p_{1|h} = 1 \qquad \text{for } h = 0,1,$$

(28) will hold if and only if

$$0 \leqslant p_{1|0} \leqslant 1 \quad \text{and} \quad 0 \leqslant p_{0|1} \leqslant 1.$$

In light of [1] (p. 4, the formulae (14)), we have

$$(29) \qquad p_{1|0} = p(1 - \varrho) \quad \text{and} \quad p_{0|1} = q(1 - \varrho).$$

Consequently, by virtue of (26) and (27)

$$0 \leqslant p_{0|1} \leqslant p_{1|0}.$$

So, for (28) to hold it is necessary and sufficient that

$$p_{1|0} \leqslant 1.$$

In light of (29) and after solving the last inequality with regard to ϱ we obtain the condition for which we have searched:

$$(30) \qquad \varrho \geqslant -q/p.$$

5.2. Constraints on ϱ_2 with p and ϱ given

Like in the former case, conditional probabilities are the constraint here:

(31) $0 \leq p_{j|gh} \leq 1$ for g, h, j = 0, 1.

Since (cf. [1], p. 8, formulae (31))

$$p_{1|00} = p(1-\varrho)y = yp_{1|0},$$

$$p_{1|01} = (p + q\varrho)y = yp_{1|1},$$

$$p_{0|10} = (p\varrho + q)y = yp_{0|1},$$

$$p_{1|11} = (1-\varrho)y = yp_{0|1},$$

where

$$y = \frac{1-\varrho_2}{1-\varrho^2},$$

the inequalities (31) are equivalent to the following

$$p_{j|h} \leq \frac{1}{y} \qquad \text{for h, j = 0, 1}$$

and these in turn to the conditions

$$p(1-\varrho) \leq \frac{1}{y}, \qquad p + q\varrho \leq \frac{1}{y}$$

(because $p_{0|1} \leq p_{1|0}$ and $p_{0|0} \leq p_{1|1}$), which can be written as

$$y \leq \frac{1}{m},$$

where

$$m = \max\{p(1-\varrho),\ p + q\varrho\}.$$

For $0 \leq p(1-\varrho) \leq p + q\varrho$, hence

(32) $$\varrho_2 \geq 1 - \frac{1-\varrho^2}{p + q\varrho} \quad \text{for } \varrho \geq 0,$$

otherwise $m = p(1 - \varrho)$, so

$$(33) \quad \varrho_2 \geq 1 - \frac{1 + \varrho}{p} \quad \text{for } \varrho \leq 0.$$

6. Results and conclusions

In accordance with the information given in § 4, the study of the power of test based on the number of runs in the case of the second-order Markov chain has been conducted in dependence on five parameters (n, p, ϱ, δ, α). We have picked up several values of these parameters and we have found the power of run test in different sections. Some of them are presented in Tab. 1-6. On the basis of the results shown in the tables we can formulate the following conclusions:

1) for $\varrho = 0$ there is a difference in power of the dependence of δ; with the increase of p the power decreases for small n, and it is stable for large n (cf. Tab. 1);

2) when $\varrho = 0,5$ the difference of power depending on δ is small; the power of the test decreases (cf. Tab. 2);

3) when ϱ increases the differences of power caused by δ decrease; for $n \geq 50$ and $\varrho > 0,5$ the power of the test is 1, regardless of δ (cf. Tab. 3);

4) for $p = 0,5$ if $\varrho = -0,5$ the power of the test increases together with the increase of n (cf. Tab. 4);

5) the power of the test for $\varrho = 0$ does not depend on n regardles of p (cf. Tabl. 5 and Tab. 6);

6) for small $n(n \leq 50)$ the power of the test decreases together with the increase of p.

These conclusions give an idea of the power of run test in the case of the second-order Markov chain especially of the dependence of the results on the parameter δ. However, in order to obtain more precise results,further and more detailed research should be conducted.

Table 1

The power of the run test (per mille) for $\rho = 0$ and $\alpha = 0.05$

p	n	δ -0.2	-0.1	0.0	0.1	0.2
0.50	5	26	36	50	67	86
0.60		31	40	50	62	77
0.70		38	44	50	57	65
0.75		41	45	50	55	61
0.80		43	46	50	54	58
0.90		•	48	50	52	53
0.95		•	•	50	51	52
0.50	50	23	35	50	68	89
0.60		23	35	50	68	89
0.70		24	36	50	67	88
0.75		25	36	50	67	88
0.80		25	36	50	67	89
0.90		•	37	50	67	88
0.95		•	37	50	64	83
0.50	200	22	35	50	68	89
0.60		22	35	50	68	89
0.70		23	35	50	68	89
0.75		23	35	50	68	89
0.80		24	35	50	68	89
0.90		•	36	50	68	89
0.95		•	•	50	67	89

Table 8

The power of the run test (per mille) for $\varrho = 0.05$ and $\alpha = 0.05$

p	n	δ -0.2	-0.1	0.0	0.1	0.2
0.50	5	191	221	253	288	327
0.60		.	.	.	219	245
0.70		.	112	123	134	147
0.75		87	93	100	108	116
0.80		75	.	84	89	94
0.90		.	.	63	64	66
0.95		.	.	56	56	57
0.50	50	997	991	980	963	941
0.60		.	.	.	945	919
0.70		.	914	884	854	824
0.75		855	824	796	771	747
0.80		709	.	680	667	656
0.90		.	.	356	390	425
0.95		.	.	179	210	248
0.50	200	1000	1000	1000	1000	1000
0.60		.	.	.	1000	1000
0.70		.	1000	1000	1000	1000
0.75		1000	1000	1000	1000	997
0.80		1000	.	1000	993	984
0.90		.	.	877	848	819
0.95		.	.	564	569	575

T a b l e 3

The power of the run test for (ρ, δ) $p = 0.5$ and $\alpha = 0.05, 0.10$

ρ	n	δ = -0.2		-0.1		0.0		0.1		0.2	
		α 0.05	α 0.10	α 0.05	α 0.10	α 0.05	α 0.10	α 0.05	α 0.10	α 0.05	α 0.10
-0.9	5	•	•	•	•	0	0	7	10	•	•
-0.5		0	1	1	5	3	11	9	23	18	40
0.0		36	62	36	79	50	100	67	124	86	151
0.5		191	319	221	348	253	380	288	415	327	464
0.9		•	•	•	•	652	840	607	797	•	•
-0.9	20	•	•	•	•	0	•	8	15	•	•
-0.5		0	0	1	1	1	2	4	11	14	32
0.0		15	38	29	67	50	100	83	148	125	208
0.5		332	431	404	495	480	620	559	289	636	752
0.9		•	•	•	•	980	994	990	995	•	•
-0.9	50	•	•	•	•	0	0	10	18	•	•
-0.5		0	0	0	1	1	1	2	6	12	25
0.0		11	28	25	55	50	100	92	166	157	299
0.5		453	610	559	705	665	792	763	864	846	980
0.9		•	•	•	•	1000	1000	1000	1000	•	•

Table 4

The power of the run test (per mille) for p = 0.5 and ρ = -0.5

n	α	δ -0.2	-0.1	0.0	0.1	0.2
5	0.10	319	348	380	415	455
10		618	617	615	614	613
15		803	777	754	733	715
20		902	871	842	814	787
30		979	960	938	112	883
50		1000	997	992	982	967
100		1000	1000	1000	1000	999
150		1000	1000	1000	1000	1000
200		1000	1000	1000	1000	1000
5	0.05	191	221	253	288	327
10		390	410	430	451	472
15		602	600	598	596	595
20		765	743	723	705	689
30		933	904	875	846	817
50		997	991	980	963	941
100		1000	1000	1000	999	997
150		1000	1000	1000	1000	1000
200		1000	1000	1000	1000	1000
5	0.02	77	88	101	115	131
10		232	267	302	338	374
15		388	408	427	446	464
20		557	559	560	562	563
30		818	789	764	741	719
50		985	968	947	922	893
100		1000	1000	999	997	992
150		1000	1000	1000	1000	1000
200		1000	1000	1000	1000	1000

Table 5

The power of the run test (per mille) for p = 0.5 and ϱ = 0.0

n	α	δ -0.2	-0.1	0.0	0.1	0.2
5	0.10	68	83	100	120	144
10		60	79	100	123	148
15		60	79	100	123	147
20		60	79	100	123	147
30		60	79	100	123	147
50		59	79	100	123	147
100		59	78	100	123	147
150		58	78	100	123	147
200		58	78	100	123	148
5	0.05	26	37	50	67	86
10		26	37	50	66	86
15		24	36	50	67	87
20		24	35	50	67	88
30		23	35	50	68	88
50		23	35	50	68	89
100		22	35	50	68	90
150		22	35	50	68	90
200		22	35	50	68	90
5	0.02	10	15	20	27	35
10		7	12	20	31	46
15		7	12	20	31	45
20		7	12	20	31	45
30		7	12	20	31	46
50		6	12	20	31	47
100		6	12	20	31	47
150		6	12	20	31	47
200		6	12	20	32	47

Table 6

The power of the run test (per mille) for (n, δ) and (p, ρ) α = 0.05

n	(p, ρ)	δ -0.2	-0.1	0.0	0.1	0.2
5	(0, 5, 0.5)	191	221	253	288	327
10		390	410	430	451	472
15		602	600	598	596	595
20		765	743	723	705	689
30		933	904	975	946	917
50		997	991	980	963	941
100		1000	1000	1000	1000	997
150		1000	1000	1000	1000	1000
200		1000	1000	1000	1000	1000
5	(0, 75, 0.0)	41	45	50	55	61
10		29	38	50	65	84
15		29	38	50	65	84
20		26	37	50	67	87
30		26	37	50	66	86
50		25	36	50	68	88
100		24	36	50	68	89
150		23	35	50	68	89
200		23	35	50	68	89
5	(0, 75, 0, 50)	87	93	100	108	116
10		148	173	203	239	281
15		248	277	308	342	379
20		371	379	422	445	467
30		561	565	569	571	574
50		855	824	796	771	747
100		998	990	978	960	937
150		1000	1000	998	993	985
200		1000	1000	1000	1000	997

Bibliography

[1] D o m a ń s k i Cz., T o m a s z e w i c z A. S. (1980):Variants of Tests Based on the Length of Runs, Conference Problems of Building and Estimation of Large Econometric Models, Polanica.

[2] D o m a ń s k i Cz., T o m a s z e w i c z A. S. (1982): A Method for Computing the Power of the Test Based on the Number of Runs in the Case of Second Order Autocorrelation Process,The 9th Conference on Problems of Building and Estimation of Large Econometric Models, Łódź.

Czesław Domański, Andrzej Stefan Tomaszewicz

ANALIZA MOCY TESTÓW NIEZALEŻNOŚCI OPARTYCH NA LICZBIE SERII W PRZYPADKU ŁAŃCUCHA MARKOWA DRUGIEGO RZĘDU

Artykuł przedstawia numeryczną analizę mocy testu losowości próby losowej (niezależności elementów próby), opartego na liczbie serii, w przypadku łańcucha Markowa drugiego rzędu. W przypadku procesu stochastycznego o trzech parametrach moc testu serii jest funkcją pięciu niezależnych zmiennych (n, p, ϱ, ϱ_2, α), gdzie n jest wielkością próby, p - stacjonarnym prawdopodobieństwem, ϱ - współczynnikiem autokorelacji pierwszego rzędu, ϱ_2 - współczynnikiem autokorelacji drugiego rzędu w przypadku regresji liniowej pierwszego rodzaju, α - poziomem istotności. Na podstawie naszych badań można sformułować, między innymi, następujące wnioski:

1) dla $\varrho = -0,5$, $p = 0,5$ moc testu rośnie wraz ze wzrostem n;

2) moc testu dla $\varrho = 0$ nie zależy od n, bez względu na p;

3) dla małych n ($n \leq 50$) moc testu zmniejsza się ze wzrostem p;

4) gdy ϱ rośnie różnica w mocy spowodowana przez $\delta = \varrho_2 - \varrho^2$ (miara odległości między danym procesem i procesem pierwszego rzędu) maleje.

Dla $n \geq 50$ i $\varrho > 0,5$ moc testu wynosi 1, bez względu na δ.