Markov Chain Model and Stationary Test: A Case Study on Malaysia Social Security (SOCSO)

(Model Rantaian Markov dan Ujian Kepegunan: Suatu Kajian Kes terhadap Keselamatan Sosial Malaysia (PERKESO))

SHAMSHIMAH SAMSUDDIN* & NORISZURA ISMAIL

ABSTRACT

Recently, the Markov chain model, which is a model that depends on the probability of transition, has been widely used in areas related to health problems. This article aims to build the yearly transition model for the health state of workers who contribute under the Employment Injury Scheme (EIS) SOCSO in Malaysia using the Markov chain model. In addition, the stationary test is carried out to confirm whether the model can be used for the projection of transition probabilities of the contributors' health levels.

Keyworkds: Markov chain; stationary; transition probabilities

ABSTRAK

Sejak kebelakangan ini, model rantaian Markov, iaitu suatu model yang bergantung kepada kebarangkalian peralihan, telah banyak digunakan dalam bidang berkaitan masalah kesihatan. Kertas ini bertujuan untuk membina model peralihan tahunan bagi tahap kesihatan pekerja yang mencarum di bawah Skim Bencana Pekerjaan (SBP) PERKESO di Malaysia menggunakan model rantaian Markov. Selain itu, ujian kepegunan dijalankan untuk mengesahkan sama ada model yang diperoleh boleh digunakan untuk pengunjuran kebarangkalian peralihan bagi tahap kesihatan pencarum.

Kata kunci: Kebarangkalian peralihan; kepegunan; rantaian Markov

INTRODUCTION

A finite Markov chain is a discrete time parameter stochastic process in which the future state of the system is dependent only on the present state and is independent of the previous states. The behaviour of a Markov chain model largely depends on the transition probabilities which can represented by the probability matrix. Recently, the Markov chain model has been applied in many areas related to health problems. Examples of applications in the field of disability can be found in Hariyanto et al. (2014) who used Markov model to estimate the transition probabilities amongst level of disability based on the Australian Survey of Disability, Aging and Carer, and in Blanco-Encomienda (2013) who used Markov model for projecting health care expenditures.

The stationary transition probability matrix can be estimated using the maximum likelihood estimation. Examples of past studies that use maximum likelihood estimate of stationary transition probability matrix can be found in Aktaş (2015) who estimate the health states using the Danish longitudinal data for every third year on a three category state for years 1962-1965, and Jain (1986) who estimate the state of health of chronic bronchial asthma among patients according to seasonal patterns. According to Anderson and Goodman (1957), the maximum likelihood estimates and their asymptotic distribution were obtained when they repeated observations of the chain.

The stationarity of the transition probabilities can be tested using the stationary test of χ^2. Further discussion on the Markov chain's stationary test can can be found in Anderson and Goodman (1957), Billingsley (1961) and Tainiter (1963).

This paper aims to model the health state at yearly intervals for workers who contribute under the Employment Injury Scheme (EIS) SOCSO in Malaysia using Markov chain model and maximum likelihood estimation, and to test the stationarity of the transition matrix using the χ^2 test. In this study, we assume that the transition matrix is a square matrix which is constructed from different time points.

DATA AND METHODS

SAMPLE DATA

The sample data is obtained from SOCSO, which is an organization in Malaysia that acts as a safety net when an employee becomes unemployed, temporary or permanently disabled, or dead. In this study, the transition model is used to estimate the movement of a worker from one health state to another. Four states are considered: state 1 for work/active (A), state 2 for temporary disability (TD), state 3 for permanent disability (PD) and state 4 for death (D). These states represent the levels of disabilities which are based on the decision of a panel of doctors appointed

by SOCSO (Malaysia 1969). A more detailed discussion on the issues of disability among workers who contribute under SOCSO in Malaysia can be found in Samsuddin and Ismail (2016, 2015).

Table 1 shows the transition count matrices for each pair of years in 2009-2013 for a male worker aged 34 and a female worker aged 35. It should be noted that the male and female mean ages for the population data in years 2009-2013 are 34 and 35, respectively. Specifically, Table 1 provide the matrices of transition count of a male aged 34 (mean age for male worker) and a female aged 35 (mean age for female worker) for four pairs of years (2009-2010, 2010-2011, 2011-2012 and 2012-2013).

Table 2 shows the transition count matrices of a male worker aged 34, M^{male}, and a female worker aged 35, M^{female}, for the combined years of 2009-2013. The transition count matrices are obtained using,

$$M = \sum_{t=1}^{4} M_t = \left[n_a^{ij} \right]$$

where n_a^{ij} is the transition count from state i to state j for a worker aged a; t is the pair of years; and $\sum_{t=1}^{4} n_a^{ij}(t) = n_a^{ij}$.

MARKOV CHAIN MODEL

A Markov chain is a stochastic process that can be parameterized by empirically estimating the transition probabilities between discrete states (Taylor & Karlin 1998). In this study, the transition probabilities are assumed to depend only on age, gender, year and disability category, without considering the duration and past disability experience of how or when an employee arrived in his/her category. These conditions were also stated in Jung (2006), who carried out a study on disability states where

TABLE 1. Transition count matrices for each pair of years in 2009-2013

	Male, Age=34					Female, Age=35			
	M_1=2009-2010					M_1=2009 -2010			
	1	2	3	4		1	2	3	4
1	3234	785	376	16	1	837	169	45	1
2	869	56	14	0	2	158	5	0	0
3	328	11	6	0	3	48	3	0	0
4	0	0	0	0	4	0	0	0	0
	M_2=2010-2011					M_2=2010-2011			
1	3235	831	372	30	1	782	184	75	2
2	806	36	14	1	2	191	6	2	0
3	371	6	6	0	3	72	0	0	0
4	0	0	0	26	4	0	0	0	7
	M_3=2011-2012					M_3=2011-2012			
1	3400	823	418	28	1	758	222	74	4
2	871	39	9	2	2	201	8	2	0
3	385	9	10	1	3	77	1	1	0
4	0	0	0	28	4	0	0	0	1
	M_4=2012-2013					M_4=2012-2013			
1	3217	853	366	13	1	760	229	68	4
2	770	32	8	0	2	230	2	0	0
3	397	8	7	0	3	65	0	0	0
4	0	0	0	14	4	0	0	0	1

TABLE 2. Transition count matrices for all years in 2009-2013

$$M^{male} = \begin{bmatrix} 13086 & 3292 & 1532 & 87 \\ 3316 & 163 & 45 & 3 \\ 1481 & 34 & 29 & 1 \\ 0 & 0 & 0 & 68 \end{bmatrix}, M^{female} = \begin{bmatrix} 3137 & 804 & 262 & 11 \\ 780 & 21 & 4 & 0 \\ 262 & 34 & 1 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$$

the transition probabilities have the combined effects of age, gender, income, and education.

Assuming a homogeneous Markov process, the matrix of transition rates is denoted by P, and is expressed as follows:

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \left| \begin{array}{cccc} P_a^{11} & P_a^{12} & P_a^{13} & P_a^{14} \\ P_a^{21} & P_a^{22} & P_a^{23} & P_a^{24} \\ P_a^{31} & P_a^{32} & P_a^{33} & P_a^{34} \\ P_a^{41} & P_a^{42} & P_a^{43} & P_a^{44} \end{array} \right| \end{array}$$

where the probability of X_{n+1} at age a and being in state j; given that X_n is in state i; is called the one-step transition probability and is denoted by $_{n,n+1}P_a^{ij}$. That is, $_{n,n+1}P_a^{ij} = \Pr\{X_{n+1} = j | X_n = i\}$. Since the transition probabilities are assumed to have annual probabilities, the quantities P_a^{ij} satisfy the condition $P_a^{ij} \geq 0$ and $\sum_{j=0}^{\infty} P_a^{ij} = 1$ for $i, j \in \Omega$.

The maximum likelihood estimates for P_a^{ij} are,

$$\hat{p}_a^{ij} = \frac{n_a^{ij}}{n_a^i} = \frac{n_a^{ij}(t)}{\sum_{t=1}^{T} n_a^i}$$

where $n_a^{ij}(t)$ is the number of observed transitions at age a from state i to state j at time t; $n_a^{ij}(t)$ is the total row of each time point t; and $n_a^i(t) = \sum_j n_{ij}(t)$.

STATIONARY TEST

A Markov process is said to be stationary if $P_a^{ij}(t)$ is independent of time t. Therefore, the hypothesis for testing the stationarity of transition probabilities are,

$$H_0 : P_a^{ij}(t) = P_a^{ij} \quad t = 1, 2, \ldots, T$$

$$H_1 : P_a^{ij}(t) \neq P_a^{ij} \quad t = 1, 2, \ldots, T$$

where the likelihood ratio criterion is $\lambda = \prod_t \prod_{ij} \left[\frac{p_a^{ij}}{\hat{p}_a^{ij}(t)} \right]^{n_a^{ij}(t)}$.

The hypothesis of interest is that the random variable represented by the T rows have the same distribution, so that the data are homogeneous.

The χ^2-test of homogeneity can be used for the hypothesis:

$$\chi_i^2 = \sum_{ij} \frac{n_a^i(t-1)\left[\hat{p}_a^{ij}(t) - \hat{p}_a^{ij}\right]^2}{\hat{p}_a^{ij}},$$

and if the null hypothesis is true, the χ_i^2 has the usual limiting distribution with $m(m-1)(T-1)$ degrees of freedom, where m is the number of states and T is the number of pairs of years.

RESULTS AND DISCUSSION

Table 3 shows the transition probabilities for each pair of years in 2009-2013, respectively, for male employee aged 34 and female employee aged 35. The transition probabilities are obtained using the values in Tables 1-2.

Table 4 shows the maximum likelihood estimates of transition probabilities for all years in 2009-2013. The transition matrices are arranged according to gender and level of disability starting at any state, with the exception of an absorbing state (death). The estimates of transition probabilities for other ages can be obtained using the same approach. The results in Tables 3-4 indicate that different age and gender have different transition probabilities.

After obtaining the maximum likelihood estimates, the test of stationary is performed using the χ^2-test. The χ^2 is compared with $P(X > \chi^2)$ which is less than or equal to the significant level, $\alpha = 0.05$. The results of the χ^2-test are shown in Table 5, which shows that the null hypothesis, H_0, is not rejected, implying that the estimated transition probabilities, P_a^{ij}, for a male employee aged 34 and a female employee aged 35 are constant and do not depend on time t. These stationary results are consistent with other past studies which used the Markov chain model, such as the ones carried out in (Jain 1986).

The stationary condition of the transition probabilities also indicate that the Markov model can be further used for predicting the future health state of workers in Malaysia who contribute under SOCSO. As an example, let p_t denote the row vector of a worker's health state in period t. The estimate of health state in the next period can be obtained using $p_{i+1} = p_t \times P$.

CONCLUSION

This study has estimated the transition probabilities of health states using Markov chain model and maximum likelihood method, and carry out the stationary test using the χ^2-test. The estimated transition probabilities are based on the yearly data of health levels of the contributors under the Employment Injury Scheme (EIS) SOCSO in Malaysia in years 2009-2013. In this study, the transition probabilities of a male employee aged 34 and a female employee aged 35 are estimated. These ages are selected because they respectively represent the mean age of male and female workers in years 2009-2013.

The stationary tests performed on the transition probabilities of a male worker aged 34 and a female worker aged 35 show that the estimates are stationary, and the Markov model can be used for further study in predicting the future health state of workers in Malaysia who contribute under the SOCSO. The predictions can also be further applied in other areas related to health status, such as hospital administrators, social security administrator

TABLE 3. Transition probabilities $\hat{p}_a^{ij}(t)$ for male aged 34 and female aged 35

		Male					Female		
		M_1=2009-2010					$M_{1=}$2009 -2010		
	1	2	3	4		1	2	3	4
1	0.7332	0.1780	0.0852	0.0036	1	0.7956	0.1606	0.0428	0.0010
2	0.9255	0.0596	0.0149	0.0000	2	0.9693	0.0307	0.0000	0.0000
3	0.9507	0.0319	0.0174	0.0000	3	0.9412	0.0588	0.0000	0.0000
4	0.0000	0.0000	0.0000	0.0000	4	0.0000	0.0000	0.0000	1.0000
		M_2=2010-2011					M_2=2010-2011		
1	0.7240	0.1860	0.0833	0.0067	1	0.7498	0.1764	0.0719	0.0019
2	0.9405	0.0420	0.0163	0.0012	2	0.9598	0.0302	0.0101	0.0000
3	0.9687	0.0157	0.0157	0.0000	3	1.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	1.0000	4	0.0000	0.0000	0.0000	1.0000
		M_3=2011-2012					M_3=2011-2012		
1	0.7282	0.1763	0.0895	0.0060	1	0.7164	0.2098	0.0699	0.0038
2	0.9478	0.0424	0.0098	0.0022	2	0.9526	0.0379	0.0095	0.0000
3	0.9506	0.0222	0.0247	0.0025	3	0.9747	0.0127	0.0127	0.0000
4	0.000	0.0000	0.0000	1.0000	4	0.0000	0.0000	0.0000	1.0000
		M_4=2012-2013					M_4=2012-2013		
1	0.7231	0.1917	0.0823	0.0029	1	0.7163	0.2158	0.0641	0.0038
2	0.9506	0.0395	0.0099	0.0000	2	0.9914	0.0086	0.0000	0.0000
3	0.9636	0.0194	0.0170	0.0000	3	1.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	1.0000	4	0.0000	0.0000	0.0000	1.0000

TABLE 4. ML estimates of transition probabilities P_a^{ij} for male aged 34 and female aged 35

For male   For female

$$\hat{p}_{34}^{ij} = \begin{bmatrix} 0.7272 & 0.1829 & 0.0851 & 0.0048 \\ 0.9403 & 0.0462 & 0.0127 & 0.0008 \\ 0.9587 & 0.0220 & 0.0187 & 0.0006 \\ 0.0000 & 0.0000 & 0.0000 & 1.0000 \end{bmatrix}, \hat{p}_{35}^{ij} = \begin{bmatrix} 0.74442 & 0.19079 & 0.06218 & 0.00026 \\ 0.96891 & 0.02610 & 0.00498 & 0.00001 \\ 0.98116 & 0.01502 & 0.00378 & 0.00004 \\ 0.00000 & 0.00000 & 0.00000 & 1.00000 \end{bmatrix}$$

TABLE 5. Results of χ^2–test for male and female

		Male		
Past state	χ^2	$P(X > \chi^2)$	Degrees of freedom	Decision of H_0
1	15.1657	1.00	36	Not Reject
2	11.2515	1.00	36	Not Reject
3	6.2971	1.00	36	Not Reject
4	0.0000	1.00	36	Not Reject
Total	83.8534	0.96	144	Not Reject

		Female		
Past state	χ^2	$P(X > \chi^2)$	Degrees of freedom	Decision of H_0
1	29.0689	0.79	36	Not Reject
2	8.2394	1.00	36	Not Reject
3	11.1873	1.00	36	Not Reject
4	0.0000	1.00	36	Not Reject
Total	48.4956	1.00	144	Not Reject

and policy makers. Specifically, the projection of health care expenditures related to disability and death can be carried out by multiplying the estimated probabilities (temporary disability, permanent disability, and death) with the estimated costs. Further studies regarding the projection of health care expenditures are currently being conducted.

REFERENCES

Aktaş, S. 2015. Testing the transition probabilities in square contingency tables. *Journal of Scientific Research & Reports* 8(4): 1-6.

Anderson, T.W. & Goodman, L.A. 1957. Statistical inference about Markov chains. *The Annals of Mathematical Statistics* 28(1): 89-110.

Billingsley, P. 1961. Statistical methods in Markov chains. *Ann. Math. Statist.* 31(1)*:* 12-40.

Blanco-Encomienda, F.J. 2013. A multi-state Markov model for projecting health care spending. *European Scientific Journal* 9(21): 745-752.

Hariyanto, E.A., Dickson, D. & Pitt, D.G.W. 2014. Estimation of disability transition probabilities in Australia I: Preliminary. *Annals of Actuarial Science* 8(01): 131-155.

Jain, S. 1986. Markov chain model and its application. *Computers and Biomedical Research* 19(4): 374-378.

Jung, J. 2006. *Estimating Markov Transition Probabilities Between Health States in the HRS Dataset*. Indiana University-Bloomington.

Malaysia. 1969. Akta Keselamatan Sosial Pekerja 1969.

Samsuddin, S. & Ismail, N. 2015. Isu hilang upaya dikalangan pencarum PERKESO di Malaysia. *Malaysia Labour Review* 11(2): 85-94.

Samsuddin, S. & Ismail, N. 2016. Multi-state markov model for disability: A case of Malaysia social security (SOCSO). *Innovations Through Mathematical and Statistical Research: Proceedings of the 2nd International Conference on Mathematical Sciences and Statistics (ICMSS2016)*.

Tainiter, M. 1963. Estimation, hypothesis testing and parameter correlation for Markov Chains. *IEEE Transactions on Reliability* 12(4): 26-35.

Taylor, H.M. & Karlin, S. 1998. *An Introduction to Stochastic Modeling*. 3rd edition. San Diego: Academic Press.

Shamshimah binti Samsuddin*
Pusat Pengajian Sains Aktuari
Fakulti Sains dan Teknologi
Universiti Teknologi MARA
40450 Shah Alam, Selangor Darul Ehsan
Malaysia

Noriszura binti Ismail
Pusat Pengajian Matematik
Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor Darul Ehsan
Malaysia

*Corresponding author; email: shamshimah@tmsk.uitm.edu.my