



Research Report  
Statistical Research Unit  
Department of Economics  
University of Gothenburg  
Sweden

---

**Tests of Markov Order and Homogeneity in a  
Markov Chain Model for Analyzing  
Rehabilitation**

**Robert Jonsson**

**Research Report  
2011:7  
ISSN 0349-8034**

Mailing address:  
Statistical Research Unit  
P.O. Box 640  
SE 405 30 Göteborg  
Sweden

Phone  
Nat: 031-786 00 00  
Int: +46 31 786 00 00

Home Page:  
<http://www.statistics.gu.se/>

**Research Report**  
**2011:7**  
**ISSN 0349-8034**

Mailing address:  
Statistical Research Unit  
P.O. Box 640  
SE 405 30 Göteborg  
Sweden

Phone  
Nat: 031-786 00 00  
Int: +46 31 786 00 00

Home Page:  
<http://www.statistics.gu.se/>

# Tests of Markov Order and Homogeneity in a Markov Chain

## Model for Analyzing Rehabilitation

ROBERT JONSSON

[Robert.Jonsson@economics.gu.se](mailto:Robert.Jonsson@economics.gu.se)

*Department of Economics, Goteborg University, Box 640, 405 30 Goteborg, Sweden*

**Abstract.** A three-state non-homogeneous Markov chain (MC) of order  $m \geq 0$ , denoted  $M(m)$ , was previously introduced by the author. The model was used to analyze work resumption among sick-listed patients. It was demonstrated that wrong assumptions about the Markov order  $m$  and about homogeneity can seriously invalidate predictions of future health states. In this paper focus is on tests (estimation) of  $m$  and of homogeneity. When testing for Markov order it is suggested to test  $M(m)$  against  $M(m+1)$  with  $m$  sequentially chosen as 0, 1, 2, ..., until the null hypothesis can't be rejected. Two test statistics are used, one based on the Maximum Likelihood ratio (MLR) and one based on a chi-square criterion. Also more formal test strategies based on Akaike's and Baye's information criteria are considered. Tests of homogeneity are based on MLR statistics. The performance of the tests is evaluated in simulation studies. The tests are applied to rehabilitation data where it is concluded that the rehabilitation process develops according to a non-homogeneous Markov chain of order 2, possibly changing to a homogeneous chain of order 1 towards the end of the period.

**Keywords:** Likelihood ratio, Test power, Bias of tests

## 1. Introduction

In an earlier paper the author introduced a three-state Markov chain (MC) model for analyzing the progress of patient's health [9]. The states were denoted 2, 1 and 0 and represented the development from 'Acute diseased' to 'Improved' and further to 'Healthy'. Another example is progression through the stages: HIV infection, AIDS and finally death. The MC model is allowed to be non-homogeneous in discrete time and also to have any Markov order  $m$ , where  $m$  refers to the number of time points back in history that has to be considered when assigning a transition probability one step ahead. In [9] it was demonstrated that assuming homogeneity, when the MC actually is non-homogeneous, can seriously invalidate calculations of predictive probabilities two or more steps ahead. It was also demonstrated that prediction probabilities are heavily dependent on the choice of  $m$ .

Although correct specifications of both heterogeneity and of Markov order are of crucial importance for the analysis, few attempts have been made to compare and evaluate various tests that can be used in order to find a proper structure of the MC. In these respects one may notice a difference between various disciplines. Econometricians have since long ago used non-homogeneous MC models, or models with time varying transition probabilities in their terminology (see e.g. Diebold et. al. [ 3] and Van den Berg [14]). However, the latter seem to only cover the case  $m = 1$ . Among statisticians oriented toward natural sciences focus has been on testing for a proper value of  $m$  in homogeneous MCs ([7], [2], [10], [11]), with a few exceptions (see e.g. [8]).

In this paper various tests of Markov order and of homogeneity are considered in the three-state MC and the performances of the tests are studied in Monte Carlo simulations. The methods are then applied to data of work resumption after rehabilitation. The paper ends with a concluding discussion. Throughout the paper the notation  $M(m)$  will be used for a MC of order  $m$ .

## 2. Assumptions, notations and some basic results

Let  $X_t^{(m)}$ ,  $t = 1, 2, \dots, T$  denote the health state at time  $t$  for a MC of order  $m$ , with the possible outcomes 0 (Healthy), 1 (Improved) and 2 (Acute diseased). The probabilities of these

outcomes are  $P(X_t^{(m)} = j)$ ,  $j = 0, 1, 2$ . At  $t=1$  only the states 1 and 2 are available and for these initial states the notations  $\pi_2 = P(X_1^{(m)} = 2)$  and  $1 - \pi_2 = P(X_1^{(m)} = 1)$  are used. From one state at time  $t$  to the following state at time  $t+1$  only the following transitions are possible:  $2_t \rightarrow (2_{t+1} \text{ or } 1_{t+1})$ ,  $1_t \rightarrow (1_{t+1} \text{ or } 0_{t+1})$  and  $0_t \rightarrow 0_{t+1}$ . Here the notation  $j_t$  has been used to denote that state  $j$  is occupied at time  $t$ . Transitions to state 2 can thus only take place from state 2. Therefore, omitting the index  $m$  for simplicity,

$$P(X_{t+1} = 2 | X_{t-s} = 2 \dots X_t = 2) = P(X_{t+1} = 2 | X_t = 2) = \beta_t \quad (1)$$

$$P(X_{t+1} = 1 | X_{t-s} = 2 \dots X_t = 2) = P(X_{t+1} = 1 | X_t = 2) = 1 - \beta_t$$

When the last preceding state at time  $t$  is 1, the earlier states can be either 1 or 2. Such transitions are denoted by  $\alpha_t$ -probabilities as in Table 1. The latter has to be denoted in such a way that they reflect the preceding states. The following notations will be used:

$$\begin{aligned} P(X_{t+1} = 1 | X_{t-s+1} = 1 \dots X_t = 1) &= \alpha_t(1^s) \\ P(X_{t+1} = 1 | X_{t-s} = 2, X_{t-s+1} = 1 \dots X_t = 1) &= \alpha_t(2, 1^s) \end{aligned} \quad (2)$$

**Table 1** Schematic illustration of probabilities for transitions from the states at time  $t$  to the states at time  $t+1$ .

		State at $t+1$		
		0	1	2
State at $t$	0	1	0	0
	1	$1 - \alpha_t$	$\alpha_t$	0
	2	0	$1 - \beta_t$	$\beta_t$

In general, let  $(i_s \dots j_t)$  denote the event that a sequence of states is occupied, from state  $i$  at time  $s$  to state  $j$  at time  $t$ . In analogy with the notations for the transition probabilities in (1) and (2) which were designated by Greek symbols, the following notations are used for the transition frequencies:

$$\begin{aligned}
& B_t = \text{Number of transitions from } (2_t) \text{ to } (2_{t+1}). \\
& A_t(1^s) = \text{"-"} \quad (1_{t-s+1} \dots 1_t) \text{ to } (1_{t+1}) \\
& A_t(2,1^s) = \text{"-"} \quad (2,1_{t-s+1} \dots 1_t) \text{ to } (1_{t+1})
\end{aligned} \tag{3}$$

The state frequencies or risk masses, i.e. the number of persons being in a state just before transitions occur, are denoted in the following way:

$$\begin{aligned}
& N_t(2) = \text{Number of persons in state } 2_t \\
& N_t(1^s) = \text{"-"} \quad (1_{t-s+1} \dots 1_t) \\
& N_t(2,1^s) = \text{"-"} \quad (2,1_{t-s+1} \dots 1_t)
\end{aligned} \tag{4}$$

The quantities in (3) and (4) are related. If no subjects in the sample disappear between the transitions, then e.g.  $B_t = N_{t+1}(2)$ . Since withdrawals may occur in practice both notations in (3) and (4) will be used. The total fixed sample size is denoted by  $n$ . An illustration of these frequencies is shown in Table 2 for a Markov chain of order  $m=2$ , or  $M(2)$ . Here one may notice that  $A_t(1^2) + A_t(2,1) = A_t(1)$  and  $N_t(1^2) + N_t(2,1) = N_t(1)$ .

**Table 2** Transition- and state frequencies in the three-state model for progress of health when the Markov order is  $m=2$ .

		State at $t+1$			Total
		0	1	2	
State at $(t-1,t)$	(0,0)	$N_t(0)$	0	0	$N_t(0)$
	(1,1)	$N_t(1^2) - A_t(1^2)$	$A_t(1^2)$	0	$N_t(1^2)$
	(2,1)	$N_t(2,1) - A_t(2,1)$	$A_t(2,1)$	0	$N_t(2,1)$
	(2,2)	0	$N_t(2) - B_t$	$B_t$	$N_t(2)$
					$n$

The random marginal quantities to the right in Table 2 will in the sequel be termed working sample sizes and these will be used for estimating the  $\alpha$  and  $\beta$  parameters.

The likelihood  $L^{(m)}$  of all observations in the three-state model of Markov order  $m$  is rather complicated and an expression is therefore omitted here and is given in the Appendix A1. (See also [9] for some special cases.) The ML estimators of the parameters are easily found from the likelihood,

$$\hat{\beta}_t = \frac{B_t}{N_t(2)}, \quad \hat{\alpha}_t(1^s) = \frac{A_t(1^s)}{N_t(1^s)}, \quad \hat{\alpha}_t(2,1^s) = \frac{A_t(2,1^s)}{N_t(2,1^s)} \quad (5)$$

In the homogeneous case,

$$\hat{\beta} = \frac{\sum_{t=1}^{T-1} B_t}{\sum_{t=1}^{T-1} N_t(2)}, \quad \hat{\alpha}(1^s) = \frac{\sum_{t=1}^{T-1} A_t(1^s)}{\sum_{t=1}^{T-1} N_t(1^s)}, \quad \hat{\alpha}(2,1^s) = \frac{\sum_{t=1}^{T-1} A_t(2,1^s)}{\sum_{t=1}^{T-1} N_t(2,1^s)} \quad (6)$$

The above relations hold only for the case  $m \geq 1$ . When  $m = 0$  (independent transitions)  $P(1_{t+1}|1_t) = P(1_{t+1}|2_t)$ , from which it follows that  $\alpha_t(1) + \beta_t = 1$ . By putting this relation into the Likelihood one easily gets the following ML estimate of  $\alpha_t(1)$  (under independency denoted  $\alpha_t$ )

$$\hat{\alpha}_t = \frac{A_t(1) + N_t(2) - B_t}{N_t(1) + N_t(2)} \quad (7)$$

In the homogeneous case the index  $t$  is dropped from the parameter and all terms in the numerator and denominator are preceded by a summation symbol, just as in (6).

Some properties of the estimators in (5)-(7) are discussed in [9]. It can be shown that they are unbiased, but the variances of the estimators do not behave as in the case where the sample sizes in the denominators are fixed quantities.

### 3. Tests of Markov order

In this section, first some test statistics based on the Maximum Likelihood Ratio (MLR)  $\hat{\Lambda}^{(m,m^*)} = \hat{L}^{(m)} / \hat{L}^{(m^*)}$  are presented for testing  $M(m)$  against  $M(m^*)$ ,  $m < m^*$ . Focus will be on the case  $m^* = m + 1$ , with  $m$  sequentially chosen as  $0, 1, 2, \dots$ . The process is repeated until  $H_0$  can't be rejected. This strategy is different from the one recommended by Hoel [7], where it is

instead advocated to start with a large value of  $m$  and test successively for lower order. Hoel's approach may be risky because tests of larger values of  $m$  are based on more estimated parameters of lower precision due to smaller working sample sizes. This in turn decreases the power of the tests and opens for the possibility that a large value of  $m$  is falsely accepted. Various test strategies are discussed in Section 3.3 below.

The case  $m = 0$  (independency) is considered first, followed by more general cases. It will be assumed that the sample sizes are so large that  $-2 \ln \hat{\Lambda}$  has a chi-square distribution under  $H_0$ , with degrees of freedom (df) equal to the difference between the number of estimated parameters under  $H_0$  and  $H_1$  (cf. Ch. 22.7 in Stuart *et al.* [13]). Also an alternative chi-square procedure for testing  $M(m)$  against  $M(m+1)$  is presented. Finally some criteria for estimating a proper Markov order are discussed. Throughout this section focus will be on heterogeneous MCs. Expressions for the homogenous case are obtained by simply dropping the indices of the estimated parameters and by performing a summation over all frequencies.

### 3.1 Maximum Likelihood Ratio tests of $M(m)$ against $M(m+1)$

Consider first test of  $M(0)$  against  $M(1)$ . From the Likelihood given in Appendix A1 it is easily shown that the MLR statistic based on transitions from time  $i$  to time  $i+1$ ,  $i = 1 \dots T-1$  can be written

$$\hat{\Lambda}_i^{(0,1)} = \left( \frac{1 - \hat{\alpha}_i}{\hat{\beta}_i} \right)^{B_i} \left( \frac{\hat{\alpha}_i}{\hat{\alpha}_i(1)} \right)^{A_i(1)} \left( \frac{\hat{\alpha}_i}{1 - \hat{\beta}_i} \right)^{N_i(2) - B_i} \left( \frac{1 - \hat{\alpha}_i}{1 - \hat{\alpha}_i(1)} \right)^{N_i(1) - A_i(1)} \quad (8)$$

Here the parameter estimates are given in (5)-(7). Under  $H_0$ ,  $-2 \ln \hat{\Lambda}_i^{(0,1)}$  has an asymptotic  $\chi^2(1)$ -distribution and  $H_0$  is rejected at the 5 % level if  $-2 \ln \hat{\Lambda}_i^{(0,1)}$  exceeds the 95 % percentile of the  $\chi^{(2)}(1)$ -distribution. When the same hypothesis is tested for a sequence of transitions from time  $s \geq 1$  to time  $t \leq T - 1$  the MLR-statistic is

$$\hat{\Lambda}_{s,t}^{(0,1)} = \prod_{i=s}^t \hat{\Lambda}_i^{(0,1)} \quad (9)$$

Now  $H_0$  is rejected at the 5 % level if  $-2 \ln \hat{\Lambda}_{s,t}^{(0,1)}$  exceeds the 95 % percentile of the  $\chi^2(t - s + 1)$ -distribution.



In the homogeneous case the statistic corresponding to (9) is

$$\hat{\Lambda}_{s,t}^{(0,1)} = \left( \frac{\hat{\alpha}}{\hat{\alpha}(1)} \right)^{\sum_i A_i(1)} \left( \frac{1-\hat{\alpha}}{\hat{\beta}} \right)^{\sum_i B_i} \left( \frac{\hat{\alpha}}{1-\hat{\beta}} \right)^{\sum_i N_i(2) - \sum_i B_i} \left( \frac{1-\hat{\alpha}}{1-\hat{\alpha}(1)} \right)^{\sum_i N_i(1) - \sum_i B_i} \quad (10)$$

Under  $H_0$   $-2 \ln \hat{\Lambda}_{s,t}^{(0,1)}$  now has a  $\chi^2(1)$ -distribution.

The MLR test statistic based on transitions from time  $i$  to time  $i+1$ ,  $i = m+1 \dots T-1$ , for testing  $M(m)$  against  $M(m+1)$ ,  $m \geq 1$ , in the non-homogeneous case can be expressed as

$$\hat{\Lambda}_i^{(m,m+1)} = \left( \frac{\hat{\alpha}_i(1^m)}{\hat{\alpha}_i(1^{m+1})} \right)^{A_i(1^{m+1})} \left( \frac{1-\hat{\alpha}_i(1^m)}{1-\hat{\alpha}_i(1^{m+1})} \right)^{N_i(1^{m+1})-A_i(1^{m+1})} \left( \frac{\hat{\alpha}_i(1^m)}{\hat{\alpha}_i(2,1^m)} \right)^{A_i(2,1^m)} \left( \frac{1-\hat{\alpha}_i(1^m)}{1-\hat{\alpha}_i(2,1^m)} \right)^{N_i(2,1^m)-A_i(2,1^m)} \quad (11)$$

The expression in (11) is obtained after simple but somewhat tedious calculations where the relations  $A_i(1^{m+1}) + A_i(2,1^m) = A_i(1^m)$  and  $N_i(1^{m+1}) + N_i(2,1^m) = N_i(1^m)$  have been used.

Under  $H_0$ ,  $-2 \ln \hat{\Lambda}_i^{(m,m+1)}$  has an asymptotic  $\chi^2(1)$ -distribution. When the hypothesis is tested for a sequence of transitions from time  $s \geq m+1$  to  $t \leq T-1$  the MLR statistic is analogous to that in (9) and has the same asymptotic distribution. The test statistic in the homogeneous case is obtained in the same way as in (10).

A general expression of  $\hat{\Lambda}_i^{(m,m^*)}$  for testing  $M(m)$  against  $M(m^*)$ ,  $m < m^*$ , is given in the Appendix A2.

### 3.2 An alternative test of $M(m)$ against $M(m+1)$

A test of Markov order 0 against 1 can be based on the difference between  $\hat{\alpha}_t(1^1)$  in (5) and  $\hat{\alpha}_t$  in (7). For  $m \geq 1$  a test of Markov order  $m$  against  $m+1$  may be based on the differences be  $\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(1^m)$  or  $\hat{\alpha}_t(2,1^m) - \hat{\alpha}_t(1^m)$ . In Appendix A3 it is shown that both differences can be expressed in terms of  $\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(2,1^m)$ . Under  $H_0: M(m)$ ,  $m \geq 1$ , the latter quantity has mean 0. Since  $(A_t(1^{m+1}) | N_t(1^{m+1}))$  and  $(A_t(2,1^m) | N_t(2,1^m))$  are independent it follows that a roughly unbiased estimator of  $V(\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(2,1^m))$  under  $H_0$  is given by (see Lemma (c) in Section 3.3 in [8])

$$\hat{V}_0(\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(2, 1^m)) = \hat{\alpha}_t(1^m) [1 - \hat{\alpha}_t(1^m)] (N_t^{-1}(1^{m+1}) + N_t^{-1}(2, 1^m)),$$

$$\text{where } \hat{\alpha}_t(1^m) = \frac{A_t(1^{m+1}) + A_t(2, 1^m)}{N_t(1^{m+1}) + N_t(2, 1^m)} \quad (12)$$

It follows that a test of Markov order  $m$  against  $m+1$ ,  $m \geq 1$ , can be based on the statistic

$$X_t^2 = \frac{[\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(2, 1^m)]^2}{\hat{V}_0(\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(2, 1^m))} \quad , \quad (13a)$$

where the variance in the denominator is given in (12). Under  $H_0$ ,  $X_t^2$  is asymptotically distributed as a chi-square variable with  $df=1$ .

By similar arguments it follows that a test of  $M(0)$  against  $M(1)$  can be based on

$$X_t^2 = \frac{[\hat{\alpha}_t(1) - \hat{\alpha}_t]^2}{\hat{\alpha}_t(1 - \hat{\alpha}_t) \cdot \frac{N_t(2)}{N_t(1)(N_t(1) + N_t(2))}} \quad , \quad (13b)$$

where  $\hat{\alpha}_t$  is given in (7), and  $X_t^2$  has the same asymptotic distribution as in (13a).

### 3.3 Test strategies

The various tests in the preceding sections give no clear-cut answer to the problem of what Markov order to choose in a given situation. One way to deal with the problem is to perform all possible tests of  $M(m)$  against  $M(m^*)$ , for  $0 \leq m < m^*$ , where  $m^*$  is some more or less arbitrarily chosen upper limit for the Markov order. This strategy was used by Gabriel and Neumann [4] when analyzing the frequently quoted Tel Aviv precipitation data, and they drew the conclusion that  $m=1$ . No study seems to have been published, where the benefit of such a multiple hypothesis strategy is evaluated. When the number of tests involved is large it is clear that there is a risk of obtaining biased estimates of  $m$ .

For a large value of the Markov order the number of parameters to estimate increases rapidly. So, in small samples it is especially important that  $m$  is small. In the latter situation one may use the strategy mentioned in the beginning of Section 3 to sequentially testing for higher and higher Markov orders. More formal strategies are Akaike's information criterion (AIC) (Akaike [1]) and Baye's information criterion (BIC) [12]. A convenient formulation of

AIC and BIC is the following [10]. Consider the sequence of MLR statistics,  $\hat{\Lambda}^{(m,m^*)}$ , for testing  $M(m)$  against  $M(m^*)$   $m = 0, 1, \dots, m^* - 1$ . Let  $df(m, m^*)$  be the df of the test and put

$$\begin{aligned} AIC(m) &= -2 \ln \hat{\Lambda}^{(m,m^*)} - 2df(m, m^*) \\ BIC(m) &= -2 \ln \hat{\Lambda}^{(m,m^*)} - \ln(\text{sample size}) \cdot df(m, m^*) \end{aligned} \quad (14)$$

The estimate of  $m$  based on AIC,  $\hat{m}_{AIC}$ , is chosen such that  $AIC(\hat{m}_{AIC}) = \min AIC(m)$ , while the estimate based on BIC,  $\hat{m}_{BIC}$ , is similarly obtained from  $BIC(\hat{m}_{BIC}) = \min BIC(m)$ . The only difference between  $AIC(m)$  and  $BIC(m)$  is that different penalty functions are used, and it is seen that  $AIC(m) > BIC(m)$  for sample sizes larger than 8. In the case when transitions are possible to all states, ‘sample size’ in (14) is simply synonymous with  $n$ , the total sample size. In the present model it will be clear from the simulation study in Section X that ‘sample size’ should be put equal to the working sample size (cf. Section 2).

The asymptotic properties of AIC- and BIC criteria are well known when applied to Markov chains. The AIC criterion has been proved to be inconsistent as the sample size tends to infinity and overestimates the true Markov order, while BIC is consistent [10]. In finite samples the picture is somewhat unclear. For the Tel Aviv precipitation data mentioned above, Gates and Tong [5] obtained  $\hat{m}_{AIC} = 2$  while Katz [9] got  $\hat{m}_{BIC} = 1$ , both using  $m^* = 3$ .

In the last cited study simulations of a two-state Markov chain were performed with  $m=1$ . The AIC- and BIC criteria were then applied simultaneously in order to estimate  $m$  from samples of various sizes. It was found that both criteria gave estimates of  $m$  that were heavily dependent on the sample size. In small samples both criteria suggested the estimates  $\hat{m} = 0$  with highest probability. As sample sizes increased both criteria tended to favor the estimate  $\hat{m} = 1$ . But this process went slower for the BIC criterion than for the AIC criterion. In fact, sample sizes of 1000 and more were needed in order for the BIC criterion to give the estimate  $\hat{m} = 1$  the highest chance, while the AIC criterion only required sample sizes of 300 and more. Thus, although BIC seems to work better than AIC in very large samples, it is not clear that this is the case in small sample cases.

#### 4. Tests of homogeneity

When facing the problem whether a MC is homogeneous or not, it is a good advice to first plot estimates of the various transition probabilities against time, as in Figure 2 in [9]. Such a plot may be accompanied by a table showing the results from tests of homogeneity in an

interval from time  $u$  to time  $v, 1 \leq u < v \leq T - 1$ . An example of the latter is given in Table 11 of Section 6. Let any subset of  $[1, \dots, T - 1]$  be denoted by  $S$ . Test of homogeneity can be performed for the parameters  $\beta_t, \alpha_t(1^s)$  and  $\alpha_t(2, 1^s)$ . Below the MLR statistics for these tests are given.

$$H_0 : \beta_t = \beta \text{ against } H_1 : \beta_t \neq \beta \text{ for } t \in S$$

By inserting the estimators in (5) and (6) into the likelihood given in Appendix X one gets the MLR-statistic

$$\hat{\Lambda} = \frac{\hat{\beta}_t^{\sum_t B_t} (1 - \hat{\beta}_t)^{\sum_t N_t(2) - \sum_t B_t}}{\prod_t \hat{\beta}_t^{B_t} (1 - \hat{\beta}_t)^{N_t(2) - B_t}}, t \in S \quad (15)$$

Under  $H_0$   $-2 \ln \hat{\Lambda}$  has an asymptotic chi-square distribution. The degrees of freedom (df) is the difference between the number of estimated parameters in the numerator and denominator. E.g. if  $S = [1, \dots, T - 1]$  then df is  $(T-1)-1 = T-2$ .

$$H_0 : \alpha_t(1^s) = \alpha(1^s) \text{ against } H_1 : \alpha_t(1^s) \neq \alpha(1^s) \text{ for } t \in S$$

The MLR statistic now becomes

$$\hat{\Lambda} = \frac{\hat{\alpha}(1^s)^{\sum_t A_t(1^s)} [1 - \hat{\alpha}(1^s)]^{\sum_t N_t(1^s) - \sum_t A_t(1^s)}}{\prod_t \hat{\alpha}_t(1^s)^{A_t(1^s)} [1 - \hat{\alpha}_t(1^s)]^{N_t(1^s) - A_t(1^s)}}, t \in S \quad (16)$$

In this case the largest subset  $S$  is  $[s, \dots, T - 1]$  in which case df in the asymptotic chi-square distribution is  $T - s - 1$ . Test of homogeneity of  $\alpha_t(2, 1^s)$  is performed in a similar way.

## 5. Comparative simulation studies

The performance of the various tests that were presented in Sections 3-4 will be studied by simulations. First, the conditions under which the simulations were carried out are given. Then results about the power of some tests of Markov order are presented, followed by a section where the AIC and BIC procedures in Section 3.3 are compared. Finally, the power functions of the proposed tests of homogeneity are studied.

### 5.1 Design of the simulation studies

Tests of Markov order were applied to the following cases: M(0) against M(1), M(0) against M(2) and M(1) against M(2). The tests were studied by simulating realizations of  $X_t$  ( Cf. Section 2.) for  $t = 1, \dots, 4$ . Frequencies of the original state  $X_1$  were obtained by assigning the states 2 and 1 to each subject in a sample of size  $n$  with probabilities  $\pi_2$  and  $1 - \pi_2$ , respectively.  $n$  was chosen as 100, 200, 500 and 1000 whereas  $\pi_2$  was 0.5 and 0.9. The probabilities  $\beta_t$  of remaining in state 2 were  $\beta_1 = 0.60$ ,  $\beta_2 = 0.65$  and  $\beta_3 = 0.70$ . When testing M(0) against M(1) the probability that a subject remains in state 1 was chosen as  $\alpha_t(1) = 1 - \beta_t + \delta^{(1)}$ . Here  $\delta^{(1)} = 0$  corresponds to M(0) or independent transitions whereas  $\delta^{(1)} \neq 0$  gives rise to various M(1)-models. For tests of M(0) against M(2) and M(1) against M(2) the  $\alpha_t$ -probabilities were  $\alpha_t(1) = 1 - \beta_t + \delta^{(1)}$  and for  $t = 2, 3$   $\alpha_t(1^2) = 1 - \beta_t + \delta^{(1^2)}$ ,  $\alpha_t(2,1) = 1 - \beta_t + \delta^{(2,1)}$ . Here  $\delta^{(1^2)} = \delta^{(2,1)} = 0$  gives M(0) - models,  $\delta^{(1^2)} = \delta^{(2,1)} \neq 0$  gives M(1)-models and  $\delta^{(1^2)} \neq \delta^{(2,1)}$  gives M(2)-models. The values of the  $\delta$ -parameters that were chosen in the simulations are given in Tables 3-5 below.

When comparing the AIC and BIC criteria for estimating the Markov order  $m$  the estimates  $\hat{m}_{AIC}$  and  $\hat{m}_{BIC}$  were computed from the same simulated data sets. The actual Markov orders were chosen as  $m = 0, 1, 2$ , and for each value of  $m$  the relative frequencies of  $\hat{m}_{AIC}$  and  $\hat{m}_{BIC}$  were observed. The  $\beta_t$  and  $\alpha_t$  parameters were chosen as in the last paragraph above with  $\delta^{(1)} = 0$  in the M(0) chain and  $\delta^{(1)} = 0.15$  in the M(1) chain. In the M(2) chain the values of the  $\alpha_t(1^2)$  and  $\alpha_t(2,1)$  parameters for  $t = 2, 3$  were chosen such that the transition probabilities

$P(X_{t+1}^{(2)} = 1 | X_t^{(2)} = 1)$  were roughly the same as  $\alpha_t(1)$  in the M(1) model. (See Appendix A1 for details.)

For the estimation of  $m$  based on BIC the ‘*sample size*’ in (14) was put equal to the total working sample size  $N_t(1) + N_t(2)$ . In this way the performance of the method based on BIC was much improved compared to when the fixed total sample size  $n$  was used.

To study tests of homogeneity only M(1)-models were used with  $\pi_2 = 0.9$ . Realizations of  $X_t$  were obtained at  $t = 1, \dots, 7$ , and the  $\beta_t$  and  $\alpha_t$  -parameters were chosen as either constant (0.60) or increasing (0.60, ..., (0.05) ..., 0.90). Thus there were four different cases in total. Power functions of the tests were calculated for gradually increasing time sets, from  $S = [1, 2]$  to  $S = [1, \dots, 7]$ . The chosen values of the parameters were of about the same magnitude as those that were observed in empirical data (cf. Section 6.)

A stumbling block when simulating MCs is the possibility of getting zero frequencies due to extinction of certain states. In the present three-state model where diseased subjects successively become more healthy there is clearly a risk for zeros in the numerators in (5), or even in the denominators which in turn leads to missing values. In the tests for Markov order it was found that the proportion of missing estimates never exceeded 2 %, the largest proportion being obtained with  $n = 100$ . However, in the tests for homogeneity the proportion of missing estimates could be nearly 30 % obtained for  $S = [1, \dots, 7]$  and  $n = 100$ .

Throughout Section 5 each reported figure is based on 10,000 simulations, disregarding the missing estimates just mentioned. About 99 % of all values of the test powers are found within reported value  $\pm 0.005$  for powers close to 0.05 and 0.95, and within reported value  $\pm 0.012$  for powers near 0.50.

## 5.2 Power of some tests of Markov order

Table 3 summarizes the results for test of M(0) against M(1). From top to bottom there are three chi-square statistics and three MLR statistics based on transitions at the single times 1, 2 and 3. These are followed by two MLR statistics based on transitions at times (1, 2) and (2, 3). Finally, there is one MLR statistic based on transitions at the times (1, 2, 3).

**Table 3.** Simulated powers of various statistics for testing M(0) against M(1) at the 5 % significance level.  $\delta^{(1)}$  is the difference between the parameters under M(0) and M(1). Two values of the null power that are suspiciously large are shown within parentheses.

Statistic	$\delta^{(1)}$	$\pi_2 = 0.5$				$\pi_2 = 0.9$			
		$n =$				$n =$			
		100	200	500	1000	100	200	500	1000
$X_1^2$	0	.052	.051	.047	.048	.041	.046	.052	.051
	0.10	.172	.296	.617	.887	.096	.141	.282	.484
	0.15	.338	.574	.927	.998	.150	.255	.539	.819
$X_2^2$	0	.051	.052	.050	.048	.049	.050	.049	.052
	0.10	.135	.231	.484	.777	.175	.280	.603	.871
	0.15	.254	.453	.825	.985	.311	.550	.912	.997
$X_3^2$	0	.047	(.058)	.051	.050	.050	.050	.052	.048
	0.10	.105	.178	.367	.629	.145	.243	.506	.811
	0.15	.189	.333	.688	.931	.265	.468	.851	.989
$-2\ln \Lambda_1^{(0,1)}$	0	.052	.051	.048	.048	.045	.049	.053	.052
	0.10	.172	.296	.617	.887	.098	.138	.277	.478
	0.15	.338	.574	.927	.998	.150	.249	.533	.815
$-2\ln \Lambda_2^{(0,1)}$	0	.053	.053	.050	.048	.051	.051	.050	.051
	0.10	.140	.235	.487	.778	.175	.280	.602	.870
	0.15	.260	.457	.826	.985	.311	.550	.915	.997
$-2\ln \Lambda_3^{(0,1)}$	0	.051	(.060)	.052	.050	.052	.051	.052	.049
	0.10	.116	.185	.375	.634	.149	.246	.507	.811
	0.15	.201	.341	.693	.932	.270	.472	.852	.990
$-2\ln \Lambda_{12}^{(0,1)}$	0	.053	.050	.054	.051	.048	.050	.050	.052
	0.10	.210	.369	.758	.968	.172	.287	.639	.911
	0.15	.419	.711	.986	1.000	.322	.584	.946	.999
$-2\ln \Lambda_{23}^{(0,1)}$	0	.052	.054	.051	.052	.053	.051	.049	.051
	0.10	.164	.285	.610	.898	.208	.376	.759	.970
	0.15	.317	.572	.930	1.000	.411	.713	.985	1.000
$-2\ln \Lambda_{13}^{(0,1)}$	0	.053	.056	.050	.052	.051	.052	.053	.052
	0.10	.225	.409	.816	.985	.205	.383	.787	.979
	0.15	.463	.767	.993	1.000	.421	.736	.991	1.000

When  $\delta^{(1)} = 0$  the null power should be 0.050. However, in Table 3 there are two values of the null power that are too large in order to be accepted and these are given within parentheses. Notice the close agreement between the powers of the chi-square statistic and the

MLR statistic., although the latter seems to have slightly larger power. As expected, the non-null power increased with increasing  $n$  for given  $\delta^{(1)}$  and  $\pi_2$ .

One can note that that the powers behaved differently for  $\pi_2 = 0.5$  and  $\pi_2 = 0.9$ . Consider e.g. the power of  $-2 \ln \hat{\Lambda}_t^{(0,1)}$ ,  $t = 1, 2, 3$ , for a specific value of  $\delta^{(1)} > 0$ . When  $\pi_2 = 0.5$  the power steadily decreased with increasing  $t$ , but when  $\pi_2 = 0.9$  the power was largest for  $t = 2$ . There is a simple explanation for this, namely that it can be shown that in the former case the expected working sample size,  $E(N_t(1))$ , is strictly decreasing with  $t$ , whereas in the latter case it has a maximum at  $t = 2$ . A further conclusion from Table 3 is that the power was increased if transitions from several time points were used.

The behavior of four power functions for testing  $M(0)$  against  $M(2)$  is illustrated in Table 4. Two powers are based on transitions from the separate time points 2 and 3, one power is based on transitions at (2,3) and one power from transitions at (1,2,3). The null power was close to 0.05 except when  $n = 100$ , in which case the null power sometimes was far below 0.05. This unwanted performance of the MLR test, with a local bias of the test, is well known (see e.g. [6]). But it is interesting that this pattern occurred more frequently with test of  $M(0)$  against  $M(2)$  than with test of  $M(0)$  against  $M(1)$ . This raises the question whether tests of  $M(m)$  against  $M(m^*)$  in general tend to be more biased as  $m^*$  deviates from  $m$ .

Table 5 shows some values of the power function when testing  $M(1)$  against  $M(2)$ . The case  $\delta^{(1)} \neq 0, \delta^{(1^2)} = \delta^{(2,1)} = 0$  corresponds to the situation under  $M(1)$ . Also here there was a close agreement between the powers based on the statistics  $X_t^{(1,2)}$  and  $-2 \ln \hat{\Lambda}_t^{(1,2)}$ ,  $t = 2, 3$ . As in Tables 3 and 4 the power was increased if data from more time points were used. (There was one exception from this when  $\pi_2 = 0.9$  and  $n = 100$ .) Comparing the non-null powers in Table 5 with those in Table 4 where test of the more deviating hypothesis  $M(0)$  against  $M(2)$  was considered, it is striking that the powers in Table 4 always dominated the powers in Table 5.



**Table 4.** Simulated powers of some statistics for testing  $M(0)$  against  $M(2)$  at the 5 % significance level. The  $\delta$ -parameters express the differences between the parameters under  $H_1$  and  $H_0$  (see Section 4.1). Figures within parentheses show null powers that deviate too much from the scheduled value 0.050.

Statistic	$\delta^{(1)} \delta^{(1^2)} \delta^{(2,1)}$	$\pi_2 = 0.5$				$\pi_2 = 0.9$			
		$n =$				$n =$			
		100	200	500	1000	100	200	500	1000
$-2 \ln \Lambda_2^{(0,2)}$	0 0 0	.052	.053	.051	.048	(.030)	.045	.050	.051
	.10 .15 .05	.161	.286	.614	.905	.073	.143	.299	.559
$-2 \ln \Lambda_3^{(0,2)}$	0 0 0	(.042)	.053	.051	.050	.050	.053	.052	(.055)
	.10 .15 .05	.127	.220	.508	.814	.141	.255	.550	.856
$-2 \ln \Lambda_{23}^{(0,2)}$	0 0 0	.047	.053	.050	.049	(.034)	.050	.052	.053
	.10 .15 .05	.185	.365	.779	.980	.130	.278	.618	.921
$-2 \ln \Lambda_{13}^{(0,2)}$	0 0 0	.050	.053	.049	.050	(.032)	.047	.051	.051
	.10 .15 .05	.240	.474	.895	.997	.151	.303	.678	.949

**Table 5.** Simulated powers of some statistics for testing  $M(1)$  against  $M(2)$  at the 5 % significance level. The  $\delta$ -parameters and the figures within parentheses are interpreted as in Table 4.

Statistic	$\delta^{(1)} \delta^{(1^2)} \delta^{(2,1)}$	$\pi_2 = 0.5$				$\pi_2 = 0.9$			
		$n =$				$n =$			
		100	200	500	1000	100	200	500	1000
$X_2^2$	.10 .10 .10	.053	.049	.054	.051	(.021)	(.043)	.051	.050
	.10 .15 .05	.109	.159	.324	.562	.044	.094	.162	.278
	.10 .25 .05	.260	.470	.858	.988	.082	.215	.475	.764
$X_3^2$	.10 .10 .10	.046	.052	.051	.049	.052	.048	.051	.052
	.10 .15 .05	.067	.116	.227	.389	.090	.138	.275	.488
	.10 .25 .05	.174	.326	.682	.928	.230	.401	.764	.968
$-2 \ln \Lambda_2^{(1,2)}$	.10 .10 .10	(.055)	.050	.054	.051	(.024)	.051	.054	.051
	.10 .15 .05	.113	.161	.326	.563	.044	.095	.159	.272
	.10 .25 .05	.265	.473	.858	.988	.084	.215	.472	.760
$-2 \ln \Lambda_3^{(1,2)}$	.10 .10 .10	.050	.054	.052	.050	.056	.049	.051	.053
	.10 .15 .05	.075	.125	.228	.394	.094	.139	.275	.488
	.10 .25 .05	.185	.333	.685	.929	.238	.405	.765	.968
$-2 \ln \Lambda_{23}^{(1,2)}$	.10 .10 .10	.054	.054	.052	.052	(.038)	.049	.053	.052
	.10 .15 .05	.110	.183	.392	.677	.075	.147	.298	.541
	.10 .25 .05	.314	.575	.947	.990	.214	.445	.842	.990

## 5.5 Comparison between AIC and BIC

Table 6 summarizes the performance of the AIC and BIC procedures for estimating the Markov order  $m$ . For  $m = 0$  it is evident that BIC performed better than AIC. The latter was not improved as  $n$  increased, thus showing the same inconsistency that was noticed in [10]. It is also peculiar that AIC placed Markov order 2 in front of Markov order 1. For  $m = 1$  both procedures gave poor estimates unless  $n$  was 500 or larger. In this case, as well as for  $m = 2$

**Table 6.** Comparison between the AIC and BIC procedures for estimating the Markov order  $m$ . The  $m$ -values to the left are those that were used in the simulations, while the six columns to the right show the relative frequencies (%) of the estimates of  $m$ .

		$\hat{m}_{AIC} =$			$\hat{m}_{BIC} =$		
$\pi_2 = 0.5$	$n$	0	1	2	0	1	2
$m = 0$	100	81.6	7.8	10.6	97.4	2.4	0.2
	200	78.1	7.5	14.4	97.4	2.4	0.2
	500	78.9	7.5	13.6	98.7	1.2	0.1
	1000	78.6	7.7	13.7	99.2	0.8	0.0
$m = 1$	100	56.1	33.2	10.7	75.8	21.5	2.7
	200	40.3	47.1	12.6	68.2	29.6	2.2
	500	13.1	71.8	15.1	42.5	55.7	1.8
	1000	1.6	82.6	15.8	14.2	84.4	1.4
$m = 2$	100	13.2	4.0	82.8	34.0	5.0	61.0
	200	0.9	0.4	98.7	8.0	0.8	91.2
	500	0.0	0.0	100.0	0.0	0.0	100.0
	1000	0.0	0.0	100.0	0.0	0.0	100.0
$\pi_2 = 0.9$	$n$	0	1	2	0	1	2
$m = 0$	100	77.6	8.2	14.2	96.6	2.8	0.6
	200	78.7	7.1	14.2	97.4	2.3	0.3
	500	78.2	7.9	13.9	98.6	1.2	0.2
	1000	78.4	7.9	13.7	99.3	0.7	0.0
$m = 1$	100	47.2	41.2	11.7	70.3	26.9	2.8
	200	28.4	58.5	13.1	56.5	41.0	2.5
	500	5.2	79.1	15.8	24.0	74.0	2.0
	1000	0.2	84.0	15.8	3.6	95.1	1.3
$m = 2$	100	24.2	14.2	61.6	49.6	14.4	36.0
	200	5.0	5.4	89.6	24.0	10.3	65.7
	500	0.0	0.1	99.9	1.0	1.2	97.8
	1000	0.0	0.0	100.0	0.0	0.0	100.0

AIC performed better.

As was mentioned in Section 3.3, the BIC criterion has been found to be superior to AIC in large samples for homogeneous MCs of order less than three. In the present study, where the parameter values varied in time, a somewhat different conclusion is reached.

### 5.6 Power of some tests of homogeneity

Tests of homogeneity for the case  $m = 1$  can be based on the statistics in Section 4. Tests can be performed for one of the two type of parameters, or for both. The four parameter settings are described in Table 7, where the notation  $PS_t$  is used for a set of parameter values that gradually increase from 0.60 to  $0.60+0.05(t-1)$ ,  $t=2,\dots,7$ .

**Table 7.** Summary of some desired properties of the power when performing various tests of homogeneity in the parameter settings (A)-(D).

		Homogeneity of		
		$\beta_t$	$\alpha_t(1)$	$\beta_t, \alpha_t(1)$
Parameter Setting	(A) $\beta_t = \alpha_t(1) = 0.60$	.05	.05	.05
	(B) $\beta_t \in PS_t, \alpha_t(1) = 0.60$	>.05	.05	>.05
	(C) $\beta_t = 0.60, \alpha_t(1) \in PS_t$	.05	>.05	>.05
	(D) $\beta_t \in PS_t, \alpha_t(1) \in PS_t$	>.05	>.05	>.05

The results from the simulations are given in Table 8. It is seen that the null-power behaved as desired, possibly with a few exceptions when  $n \leq 200$  or  $t \geq 6$ . In the latter case the random sample sizes  $N_t(1)$  and  $N_t(2)$  were too small for the chi-square approximation to hold, and sometimes even zero. On the whole the non-null power increased with  $n$  and also with  $t$ . The largest power was obtained when homogeneity of both  $\beta_t$  and  $\alpha_t(1)$  was tested and there was a true shift in both parameters.

**Table 8.** Values of the power when testing for homogeneity in the four parameter settings (A)-(D) given in Table 7.

		Homogeneity of $\beta_t$				Homogeneity of $\alpha_t(1)$				Homogeneity of $\beta_t, \alpha_t(1)$			
		$n =$				$n =$				$n =$			
	$t$	100	200	500	1000	100	200	500	1000	100	200	500	1000
A	2	.050	.049	.050	.052	.047	.056	.047	.053	.045	.053	.049	.052
	3	.052	.051	.052	.050	.046	.052	.048	.051	.052	.049	.052	.052
	4	.055	.051	.052	.052	.048	.050	.047	.049	.051	.048	.050	.053
	5	.050	.050	.053	.054	.048	.054	.047	.052	.050	.050	.051	.053
	6	.042	.051	.052	.054	.049	.052	.048	.053	.043	.054	.050	.055
	7	.028	.047	.053	.054	.049	.053	.046	.053	.028	.034	.054	.035
	B	2	.093	.130	.268	.472	.045	.056	.051	.051	.073	.111	.205
3		.139	.252	.578	.885	.046	.052	.050	.050	.109	.187	.463	.798
4		.225	.436	.861	.995	.048	.051	.051	.049	.162	.329	.763	.982
5		.325	.649	.977	1.000	.048	.053	.050	.050	.231	.519	.943	1.000
6		.429	.835	.999	1.000	.047	.054	.049	.051	.309	.714	.995	1.000
7		.531	.945	1.000	1.000	.051	.054	.053	.051	.247	.649	.991	1.000
C		2	.055	.054	.051	.051	.060	.073	.102	.151	.061	.067	.085
	3	.055	.056	.049	.051	.090	.128	.281	.493	.081	.105	.209	.388
	4	.055	.052	.051	.053	.155	.264	.629	.916	.124	.201	.506	.842
	5	.054	.052	.052	.050	.275	.521	.934	1.000	.207	.401	.866	.997
	6	.042	.053	.050	.051	.470	.810	.999	1.000	.344	.691	.993	1.000
	7	.030	.046	.049	.050	.729	.977	1.000	1.000	.277	.623	.988	1.000
	D	2	.089	.128	.259	.476	.062	.073	.111	.153	.084	.119	.247
3		.138	.252	.568	.879	.087	.133	.269	.493	.142	.264	.619	.924
4		.220	.435	.855	.993	.148	.260	.610	.905	.254	.512	.936	1.000
5		.324	.641	.979	1.000	.256	.495	.917	.998	.425	.801	.998	1.000
6		.427	.831	.998	1.000	.438	.787	.997	1.000	.638	.967	1.000	1.000
7		.538	.946	1.000	1.000	.666	.963	1.000	1.000	.571	.952	1.000	1.000

## 6. An application: Test of Markov order and homogeneity in rehabilitation data

### 6.1 Markov order

In a Swedish study 2440 sick-listed women and 1801 sick-listed men were followed after start of a rehabilitation period. At the beginning of each quarter  $t = 1, \dots, 8$  each person was classified into one of the states 0 (Healthy), 1 (Improved) and 2 (Acute diseased) and initially

all persons were in state 1 or 2. The classification into the states was made by social insurance authorities. Results from the study, including further details about data, estimated transition probabilities and prediction probabilities, have been reported earlier [9]. In this paper focus will be on tests.

A first step in the analysis was to get knowledge about the dependency structure. Tables 9a and 9b show values of the chi-square statistics  $\chi_t^2$  and the MLR statistic  $-2 \ln \hat{\Lambda}_t^{(m,m+1)}$  for testing  $M(m)$  against  $M(m+1)$ ,  $m = 0,1,2$ . Both statistics have an asymptotic chi-square distribution with 1 degree of freedom under the null hypothesis  $M(m)$ . Since the 95 % percentile in the latter distribution is 3.84 the Markov orders 0 and 1 were rejected in most cases, but not Markov order 2. Two exceptions from this were obtained for the MLR test at quarter 6 for women and quarter 7 for men. But here one has to take into consideration that the working sample sizes were very small, e.g.  $N_t(2,1)$  was far below 20 for both sexes. When testing  $M(1)$  against  $M(2)$  for a sequence of transitions from quarter 2 to quarter 6 the MLR statistic for women was 44, to be compared with 11.07 which is the 95 % percentile of the chi-square distribution on 5 (  $6-2+1$ ) degrees of freedom. So Markov order 1 is strongly rejected. On the other hand, when  $M(2)$  was tested against  $M(3)$  the MLR statistic was 3.7 which is far below the rejection limit 9.45 (df =  $6-3+1$ ) so there was no reason to reject a Markov order of 2. Similar results were obtained for men.

The use of AIC and BIC criteria for estimating Markov order gave a somewhat unclear picture, as is seen in Table 10. In most cases a Markov order of 2 is suggested. But there are exceptions at the quarters 6 and 7 where a Markov order of 1 is suggested in a similar way as was seen in Table 9b. The total AICs computed from all quarters were for women 9629 ( $m = 0$ ), 18 ( $m = 1$ ), -6.3 ( $m = 2$ ) and for men 5855 ( $m = 0$ ), 58 ( $m = 1$ ), -4.8 ( $m = 2$ ), thus suggesting that  $m = 2$ .

**Table 9a.** Chi-square statistics for tests of Markov order.

		Quarter						
Sex	Markov order	1	2	3	4	5	6	7
Women	$M(0)$ vs $M(1)$	1134	1406	1581	1559	1612	1609	1497
	$M(1)$ vs $M(2)$	-	11.4	8.6	14.9	21.3	4.3	8.1
	$M(2)$ vs $M(3)$	-	-	0.19	0.95	0.99	0.17	0.39
Men	$M(0)$ vs $M(1)$	707	816	928	938	955	995	956
	$M(1)$ vs $M(2)$	-	21.7	27.5	12.6	29.2	55.0	4.6
	$M(2)$ vs $M(3)$	-	-	0.20	0.01	1.93	0.30	0.29

**Table 9b.** MLR statistics for tests of Markov order.

		Quarter						
Sex	Markov order	1	2	3	4	5	6	7
Women	$M(0)$ vs $M(1)$	1236	1609	1890	1885	2003	2018	1825
	$M(1)$ vs $M(2)$	-	10.1	6.7	9.4	11.5	2.6	4.0
	$M(2)$ vs $M(3)$	-	-	0.02	0.84	1.9	0.15	0.76
Men	$M(0)$ vs $M(1)$	766	920	1082	1134	1167	1241	1184
	$M(1)$ vs $M(2)$	-	19.1	20.9	9.7	16.2	23.3	2.8
	$M(2)$ vs $M(3)$	-	-	2.5	0.01	1.6	0.58	0.56

## 6.2 Homogeneity

In a non-homogeneous MC of order 2 there are three transition probabilities to estimate at each  $t \geq 2$ ,  $\beta_t$ ,  $\alpha_t(1^2)$  and  $\alpha_t(2,1)$ . Plots of the latter against  $t$  were presented in Figure 2 in [9], where it is apparent that the transition probabilities are non-stationary at earlier quarters. Table 11 summarizes the results from tests of homogeneity by means of the MLR statistics suggested in Section 4. Test of stationarity of  $\alpha_t(2,1)$  was not performed due to very small working sample sizes. The results in Table 11 agree very well with the pattern that was noticed in the plots, i.e. a non-homogeneous pattern at early quarters that later turns into a

**Table 10.** Estimation of Markov order by means of the AIC and BIC criteria, (a) for women and (b) for men. The smallest values of the AIC- and BIC statistics at each quarter are marked by an asterisk.

(a)

		Quarter				
Criterion	Markov order	3	4	5	6	7
AIC	0	1890	1889	2011	2015	1824
	1	2.7	6.3	7.5	-3.24*	-1.2
	2	-2.0*	-1.2*	-0.08*	-1.9	-1.3*
BIC	0	1875	1874	1996	2000	1809
	1	-7.5*	-3.8	-0.53	-11.2*	-9.1*
	2	-7.1	-6.2*	-5.1*	-6.8	-6.2

(b)

		Quarter				
Criterion	Markov order	3	4	5	6	7
AIC	0	1099	1138	1178	1259	1181
	1	19.4	5.7	11.8	17.9	-2.7*
	2	0.52*	-1.9*	-0.42*	-1.4*	-1.4
BIC	0	1085	1124	1165	1246	1168
	1	10.1	-3.4	4.9	11.1	-9.4*
	2	-4.1*	-6.6*	-4.9*	-5.8*	-5.8

homogeneous pattern. In Table 11 one may notice that it is easier to reject homogeneity during longer periods. E.g. in Table 11 (a) for women, homogeneity was not rejected at transitions from 3 to 4, from 4 to 5 and from 5 to 6, but homogeneity was strongly rejected if transitions from 2 to 6 were considered.

**Table 11.** MLR statistics for testing homogeneity from quarter  $u$  to quarter  $v$  among men and women, (a) for  $\beta_t$  and (b) for  $\alpha_t(1^2)$ . Values within parentheses indicate non-significance at the 5 % level.

(a)

		Women						Men					
		$v$						$v$					
		2	3	4	5	6	7	2	3	4	5	6	7
$u$	1	53	121	171	231	279	283	15	36	83	115	140	145
	2		11	18	35	47	50		(3.5)	24	36	45	45
	3			(.32)	5.7	9.3	20			9.6	13	15	17
	4				(3.1)	(4.7)	18				(.01)	(.01)	(5.1)
	5					(.03)	18					(.01)	(4.4)
	6						13						(3.4)

(b)

		Women					Men				
		$v$					$v$				
		3	4	5	6	7	3	4	5	6	7
$u$	2	9.8	18	37	57	78	10	12	24	63	99
	3		(.70)	8.5	17	28		(.35)	(5.2)	31	53
	4			4.1	8.7	14			4.9	27	44
	5				(.66)	(2.3)				8.9	15
	6					(.48)					(.59)

## 7. Conclusions

In [9] it was demonstrated that a correct specification of Markov order and homogeneity is highly important. It is therefore interesting to study to which degree it is possible to make correct specifications based on tests. In Section 6 one statistic based on the chi-square criterion and one based on the MLR criterion was evaluated when testing for Markov order  $m$ . Also the use of AIC and BIC was studied for obtaining an estimate of  $m$ . From Tables 3 and 5 it is concluded that a sample size ( $n$ ) of at least 500, or even 1000, is needed in order to reach an acceptable level of the power. The procedures based on AIC and BIC also require large sample sizes in order to be reliable. However, even with  $n = 1000$  AIC works less well when



estimating Markov orders less than 2. The difference in power between the two statistics for test of Markov order was small, possibly with a slight advantage for the MLR statistic.

When testing for homogeneity the largest power was obtained when (i) a long period of transitions was used and (ii) homogeneity of several different parameters were considered simultaneously, and there actually was a shift in all of them. In the application (Section 6) it was found that homogeneity was rejected when the test was based on a long period of transitions, but not when the test was based on a single transition mostly towards the end of the studied period. The latter might be explained by the fact that there actually was a stabilization of the parameters towards the end, but this could not be detected by a 'long-period test'.

In conclusion, it is recommended to use all procedures that were considered in Sections 3 and 4 in the same way as was done in Section 6. One should also be aware of the fact that failure of rejecting a hypothesis may be due to a small working sample size, even if the fixed total sample size ( $n$ ) is large.

### **Acknowledgements**

The research was supported by the National Social Insurance Board in Sweden (RFV), Dnr 3124/99-UFU.

### **References**

1. Akaike, H., "A new look at the statistical model identification", *IEEE Transactions on Automatic Control* 19, 716-723, 1974.
2. Avery, P.J. and Henderson, D.A., "Fitting Markov chain models to discrete state series such as DNA sequences", *Applied Statistics* 48, 53-61, 1999.
3. Diebold, F.X., Lee, J-H. and Weinbach, G.C., "Regime switching with time-varying probabilities", in *Nonstationary Time Series Analysis and Cointegration* (Hargreaves, C. ed.), 283-302, Oxford University Press, Oxford, 1994.
4. Gabriel, K. and Neumann, J., "A Markov chain model for daily rainfall occurrence at Tel Aviv", *Quarterly Journal of the Royal Meteorological Society* 88, 90-95, 1962.

5. Gates, P. and Tong, H., "On Markov chain modeling to some weather data", *Journal of Applied Meteorology* 15, 1145-1151, 1976.
6. Hayakawa, T., "The likelihood ratio criterion for composite hypothesis under a local alternative", *Biometrika* 62, 451-460, 1975.
7. Hoel, P.G., "A test for Markoff chains", *Biometrika* 41, 430-433, 1954.
8. Hughes, J.P., Guttorp, P. and Charles, S.P., "A non-homogeneous hidden Markov model for precipitation occurrence", *Applied Statistics* 48, 15-30, 1999.
9. Jonsson, R., "A Markov chain model for analyzing the progression of patient's health state", *Research Report 2011:6*, Statistical Research Unit, Göteborg University, Sweden, 2011.
10. Katz, R.W., "On some criteria for estimating order of a Markov chain", *Technometrics* 23, 243-249, 1981.
11. Lowry, W.P. and Guthrie, D., "Markov chains of order greater than one", *Monthly Weather Review* 96, 798-801, 1968.
12. Schwartz, G., "Estimating the dimension of a model", *Annals of Statistics* 6, 461-464, 1999.
13. Stuart, A., Ord, J.K. and Arnold, S., *Kendalls Advanced Theory of Statistics Vol. 2A*, Arnold, London, 1999.
14. Van den Berg, G., in *Specification, Identification and Multiple Duration* (Heckman, J. and Leamer, E. eds.), *Handbook of Econometrics*, Vol. V, North Holland, Amsterdam, 2001.

## Appendix

### A1. The Likelihood of the three-state Markov chain of order $m$

The likelihood  $L^{(m)}$  of all observations in the three-state model of Markov order  $m$  can be expressed as

$$L^{(m)} = C \cdot \prod_{t=1}^{T-1} \beta_t^{B_t} (1 - \beta_t)^{N_t(2) - B_t} \cdot F^{(m)} \cdot G^{(m)}$$

Here  $C$  is a constant that does not depend on the  $\alpha$ - and  $\beta$ -parameters, and

$$F^{(m)} = \begin{cases} \prod_{t=1}^{T-1} [\alpha_t(1)]^{A_t(1)} [1 - \alpha_t(1)]^{N_t(1) - A_t(1)}, & \text{if } m = 1 \\ \prod_{t=1}^{m-1} [\alpha_t(1^t)]^{A_t(1^t)} [1 - \alpha_t(1^t)]^{N_t(1^t) - A_t(1^t)} \prod_{t=m}^{T-1} [\alpha_t(1^m)]^{A_t(1^m)} [1 - \alpha_t(1^m)]^{N_t(1^m) - A_t(1^m)}, & \text{if } m \geq 2 \end{cases}$$

$$G^{(m)} = \begin{cases} 1, & \text{if } m = 1 \\ \prod_{i=2}^m \prod_{t=i}^{T-1} [\alpha_t(2, 1^{i-1})]^{A_t(2, 1^{i-1})} [1 - \alpha_t(2, 1^{i-1})]^{N_t(2, 1^{i-1}) - A_t(2, 1^{i-1})}, & \text{if } m \geq 2 \end{cases}$$

A proof of the above expression is outlined in the Appendix A1 in [9].

### A2. The MLR statistic for testing $M(m)$ against $M(m^*)$ , $0 < m < m^*$

By the same arguments that were used in Section 3.1 it follows that the MLR statistic is

$$\hat{\Lambda}_i^{(m, m^*)} = \left( \frac{\hat{\alpha}_i(1^m)}{\hat{\alpha}_i(1^{m^*})} \right)^{A_i(1^{m^*})} \left( \frac{1 - \hat{\alpha}_i(1^m)}{1 - \hat{\alpha}_i(1^{m^*})} \right)^{N_i(1^{m^*}) - A_i(1^{m^*})} \prod_{s=m}^{m^*-1} \left( \frac{\hat{\alpha}_i(1^m)}{\hat{\alpha}_i(2, 1^s)} \right)^{A_i(2, 1^s)} \left( \frac{1 - \hat{\alpha}_i(1^m)}{1 - \hat{\alpha}_i(2, 1^s)} \right)^{N_i(2, 1^s) - A_i(2, 1^s)}$$

Under the hypothesis of  $M(m)$ ,  $-2 \ln \hat{\Lambda}_i^{(m, m^*)}$  has a chi-square distribution with  $\text{df} = m^* - m$ .

When the same hypothesis is tested for a sequence of transitions from  $s \geq m + 1$  to  $t \leq T - 1$ ,

the MLR statistic is  $\hat{\Lambda}_{s,t}^{(m, m^*)} = \prod_{i=s}^t \hat{\Lambda}_i^{(m, m^*)}$  where  $-2 \ln \hat{\Lambda}_{s,t}^{(m, m^*)}$  is chi-square distributed with  $\text{df} =$

$(m^* - m)(t - s + 1)$ .

### A3. Proof of the relations leading to the test statistic in (13a)

In Section 3.2 it was stated that two differences between estimators can be expressed in terms of one common difference. The ML estimator of  $\alpha_t(1^m)$  is

$$\frac{A_t(1^m)}{N_t(1^m)} = \frac{A_t(1^{m+1}) + A_t(2,1^m)}{N_t(1^{m+1}) + N_t(2,1^m)} = \frac{\hat{\alpha}_t(1^{m+1}) \cdot N_t(1^{m+1}) + \hat{\alpha}_t(2,1^m) \cdot N_t(2,1^m)}{N_t(1^{m+1}) + N_t(2,1^m)}.$$

From this it follows that

$$\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(1^m) = [\hat{\alpha}_t(1^{m+1}) - \hat{\alpha}_t(2,1^m)] \cdot \frac{N_t(2,1^m)}{N_t(1^m)}$$

Similarly one obtains

$$\hat{\alpha}_t(2,1^m) - \hat{\alpha}_t(1^m) = [\hat{\alpha}_t(2,1^m) - \hat{\alpha}_t(1^{m+1})] \cdot \frac{N_t(1^{m+1})}{N_t(1^m)}.$$

## Research Report

2007:1	Andersson, E.:	Effect of dependency in systems for multivariate surveillance.
2007:2	Frisén, M.:	Optimal Sequential Surveillance for Finance, Public Health and other areas.
2007:3	Bock, D.:	Consequences of using the probability of a false alarm as the false alarm measure.
2007:4	Frisén, M.:	Principles for Multivariate Surveillance.
2007:5	Andersson, E., Bock, D. & Frisé, M.:	Modeling influenza incidence for the purpose of on-line monitoring.
2007:6	Bock, D., Andersson, E. & Frisé, M.:	Statistical Surveillance of Epidemics: Peak Detection of Influenza in Sweden.
2007:7	Andersson, E., Kuhlmann, S., Linde, A & Frisé, M.:	Predictions by early indicators of the progress of the influenza in Sweden.
2007:8	Bock, D., Andersson, E. & Frisé, M.:	Similarities and differences between statistical surveillance and certain decision rules in finance.
2007:9	Bock, D.:	Evaluations of likelihood based surveillance of volatility.
2007:10	Bock, D. & Pettersson, K.:	Explorative analysis of spatial aspects on the Swedish influenza data.
2007:11	Frisén, M. & Andersson, E.:	On-line detection of outbreaks.
2007:12	Frisén, M., Andersson, E. & Schiöler, L.:	A non-parametric system for on-line outbreak detection of epidemics.
2007:13	Frisén, M., Andersson, E. & Pettersson, K.:	Estimation of outbreak regression.
2007:14	Pettersson, K.:	Unimodal regression in the two-parameter exponential family with constant dispersion parameter.
2007:15		

2008:1	Frisén, M.	Introduction to financial surveillance.
2008:2	Jonsson, R.	When does Heckman's two-step procedure for censored data work and when does it not?
2008:3	Andersson, E.	Hotelling's T2 Method in Multivariate On-Line Surveillance. On the Delay of an Alarm.
2008:4	Schiöler, L. & Frisé, M.	On statistical surveillance of the performance of fund managers.
2008:5	Schiöler, L.	Explorative analysis of spatial patterns of influenza incidences in Sweden 1999 – 2008.
2008:6	Schiöler, L.	Aspects of Surveillance of Outbreaks.
2008:7	Andersson, E & Frisé, M.	Statistiska varningssystem för hälsorisker
2009:1	Frisé, M., Andersson, E. & Schiöler, L.	Evaluation of Multivariate Surveillance
2009:2	Frisé, M., Andersson, E. & Schiöler, L.	Sufficient Reduction in Multivariate Surveillance
2010:1	Schiöler, L.	Modelling the spatial patterns of influenza incidence in Sweden
2010:2	Schiöler, L. & Frisé, M.	Multivariate outbreak detection
2010:3	Jonsson, R.	Relative Efficiency of a Quantile Method for Estimating Parameters in Censored Two-Parameter Weibull Distributions
2010:4	Jonsson, R.	A CUSUM procedure for detection of outbreaks in Poisson distributed medical health events
2011:1	Jonsson, R.	Simple conservative confidence intervals for comparing matched proportions
2011:2	Frisé, M.	On multivariate control charts
2011:3	Frisé, M.	Methods and evaluations for surveillance in industry, business, finance, and public health
2011:4	Knoth, S & Frisé, M.	Minimax Optimality of CUSUM for an Autoregressive Model
2011:5	Marianne Frisé	Inference principles for multivariate surveillance
2011:6	Robert Jonsson	A Markov Chain Model for Analysing the Progression of Patient's Health States