Original Research Article

Likelihood Ratio Search Procedure for Optimum Number of States in a Hidden Markov Manpower Model

Abstract

This paper aims at obtaining the optimum number of states for a hidden Markov manpower model, which, hitherto, has been chosen arbitrarily. A search procedure that attains this optimum number after a few steps across a series of *N* hidden Markov manpower models is proposed. The likelihood ratio statistic is employed to conduct pairwise model comparison tests on the *N* hidden Markov manpower models ordered according to their level of parsimony. The illustration shows the usefulness of the procedure in choosing the right number of states for a hidden Markov manpower model to avoid the wrong model specification.

Keywords: *Statistical manpower planning, heterogeneity, hidden Markov model, number of states, likelihood ratio test*

1. **Introduction**

Hidden Markov model (HMM) is an important stochastic model widely applied in different fields of research to represent systems whose internal dynamics are not completely observable, but possesses Markov property [1,2]. Such systems are in most cases viewed as having two parts: one observable part and one unobserved part. By the instrument of HMM the unobserved part can be studied through the observable part [3,4].

HMM has been applied in the area of statistical manpower planning. See, for example, [2,4,5]. A manpower system is a system of personnel groups on the basis of existing ranks or carders of personnel in the system, where members of a group have the potential of moving to other groups according to some probability distribution [6-8]. Ugwuowo and McClean[9] emphasize that apart from the observable classes of a manpower system due to observable heterogeneity in personnel transitions there are hidden classes due to the problem of unobserved heterogeneity. Observable classes of a manpower system are determined by observing the manpower system structure and data; see, for example, [10,11]. The hidden classes cannot be chosen in the same way because the data from where they evolve are not observable. The practice has been for researchers to decide the number of hidden states subjectively. This is not good enough because it may lead to assigning wrong number of hidden states to a given manpower system; it may also lead to a given manpower system being assigned different number of hidden states by different researchers or manpower planners. This paper aims to address these problems.

A manpower system can be represented by a number of functional models, see, for example, [12-15], of which Markov model is one. In a classical Markov manpower model (CMMM) only the observable classes are considered, where personnel inter-class transition probabilities are realized according to a Markov chain [16]. In a hidden Markov manpower model (HMMM) both the observable and the hidden classes are considered, where each observable class is assumed to contain equal number of hidden personnel classes. The division of the hidden personnel classes within each observable class is assumed to be based on different sources of hidden heterogeneity influencing members' transitions to other observable classes. Guerry[4] assumes that these sources of hidden heterogeneity lead to two groups of personnel: those that move more rapidly, called ‘movers’, characterized by higher probabilities of such movement, and those that move more slowly or do not move at all, called ‘stayers’, characterized by smaller values of probabilities of such movement. This assumption gives the number of hidden states to be two. Udom and Ebedoro[2] increased the number of hidden states in a manpower system to three to represent the class of ‘movers’, ‘mediocres’ and ‘stayers’. Ossai et al.[5], in an attempt to check whether the property of parsimony or homogeneity is more important in a manpower system, increased the number of hidden classes to five, representing the class of ‘high movers’, ‘movers’, ‘above mediocres’, ‘mediocres’ and ‘stayers’, with the values of probabilities of moving to other observable classes decreasing in this order. The authors’ choices of the number of hidden classes in these cases are not guided by any rule nor are there any check whether they are the best choices to make. Again, all the authors, with the exception of [5], compared only their new HMMM with the CMMM and concluded on the superiority of their HMMM based on this. However, Udom and Ebedoro[2] emphasize the need to search for the optimum number of hidden states to be included in a HMMM. In the current paper, a search procedure that can lead to the realization of the optimum number of states in a HMMM is formulated.

The problem addressed in this paper is, therefore, how to choose the optimum number of hidden states for a HMMM. To tackle this problem, a search procedure for locating this optimum number is proposed for any manpower system data where HMMM is applicable. The proposed search procedure utilizes the likelihood ratio statistic in pairwise comparison test to locate the HMMM that has the optimum number of hidden states among a number of HMMMs ordered according to a desired property. The likelihood ratio test is chosen because it is considered as the only statistical test which can be used to directly compare the goodness-of-fit of two models [2,17].

2. **Specification of HMMM**

The formulation of HMM for manpower systems, HMMM in the current paper, has been done in a number of works in this area; see [2,4,5]. Here we specify and highlight important components of HMMM relevant to the current aim.

For a HMMM, there are k observable classes $C_1, \ldots, C_k$ which form the states of a stochastic process $\{Y_t\}$. Each class C_i $(i = 1, \ldots, k)$ of $\{Y_t\}$ is further subdivided into $k^{'}$ hidden classes, $H_1^i, \ldots, H_{k^{'}}^i$. The $k^{'}$ hidden classes in each observable class again form the states of an underlying Markov chain $\{X_t^i\}$; what $k^{'}$ hidden states stand for depend on the type of HMMM, that is the value of $k^{'}$. Ossai et al.[5] introduced the model type HMM$k^{'}$ to represent a HMM for a manpower system with $k^{'}$ hidden states in each observable class, where the states are described according to the personnel’s ability to move to other observable classes. In the current work we represent a HMMM with $k^{'}$ hidden states per observable class by HMMM$k^{'}$.

For the process $\{Y_t\}$, the following transition probabilities are defined,

$$p_{ij} = \mathrm{P}(Y_{t+1} = C_j | Y_t = C_i) \quad \text{for } i = 1, \ldots, k;\ j = 1, \ldots, k+1 \qquad (2.1)$$

(2.1) defines the transition probability that the next state of the process $\{Y_t\}$ is C_j given that its current state is C_i . C_{k+1} is the state of having left the manpower system; that is,

C_{k+1} is the wastage state. For the hidden Markov chain $\{X_t^i\}$, the following transition probabilities are also defined,

$$\eta_{lm}^i = \mathrm{P}(X_{t+1}^i = H_m^i \mid X_t^i = H_l^i) \quad ; \quad l, m = 1, \ldots, k' \quad . \quad (2.2)$$

The HMMM is established upon the distributional dependence of the two processes $\{Y_t\}$ and $\{X_t^i\}$ such that the transition probability of a personnel moving from any hidden class H_l^i $(l = 1, \ldots, k')$ of $\{X_t^i\}$ within the observable class C_i to another observable class C_j $(j = 1, \ldots, k+1)$ of $\{Y_t\}$ is given by

$$p_{lj}^i = \mathrm{P}(Y_{t+1} = C_j \mid Y_t = C_i, X_t^i = H_l^i) = \mathrm{P}(Y_{t+1} = C_j \mid X_t^i = H_l^i) \quad (2.3)$$

Let $X_t^i = H_l^i$ and $n_{ij}(t)$ be the observed number of personnel who move from C_i to each of C_j $(j = 1, \ldots, k+1)$ with probability p_{lj}^i within the time period t to $t+1$. Since $\sum_{j=1}^{k+1} p_{lj}^i = 1$ for the realizations $n_{i1}(t), \ldots, n_{ik}(t), n_{i,k+1}(t)$ then the random vector defined by $M_t^i = (n_{i1}(t), \ldots, n_{ik}(t), n_{i,k+1}(t))$ has a multinomial distribution. That is, if $Q_{l,v_i(t)}^i = \mathrm{P}(M_t^i = v_i(t) \mid X_t^i = H_l^i)$, where $v_i(t)$ is any realization of the observation vector $(n_{i1}(t), \ldots, n_{ik}(t), n_{i,k+1}(t))$ at t and $O_i(t) = \sum_{j=1}^{k+1} n_{ij}(t)$ then

$$Q_{l,v_i(t)}^i = \binom{O_i(t)}{n_{i1}(t), \ldots, n_{i,k+1)}(t)} \prod_{j=1}^{k+1} (p_{lj}^i)^{n_{ij}(t)}$$

$$l = 1, \ldots, k'; t = 1, \ldots, T \quad (2.4)$$

The specified parameters of HMMMk' need to be estimated. For HMMM1 the maximum likelihood method is applied to obtain the estimator of p_{ij} as

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} n_{ij}(t)}{\sum_{t=1}^{T} \sum_{j=1}^{k+1} n_{ij}(t)} \quad (2.5)$$

For HMMMk', $k' > 2$, the Expectation-Maximization (EM) re-estimation algorithm is applied. Only the main results of the EM algorithm for the estimation of the parameters of HMMMk' are specified in the current paper. For more details see [2,4,5].

For $t = 1, \ldots, T$, the joint likelihood of the manpower flows from C_i $(i = 1, \ldots, k)$ being a specified sequence of observations $M_1^i = v_i(1), \ldots, M_T^i = v_i(T)$ is given by

$$L_T^i = \mathrm{P}(M_1^i = v_i(1), \ldots, M_T^i = v_i(T) \mid \pi^i, \eta^i, P^i) \quad (2.6)$$

π^i is some initial distribution vector of π_l^i, $l = 1, \ldots, k'$; π_l^i is the initial probability distribution of $\beta_l^i(t) = \mathrm{P}(X_t^i = H_l^i \mid M_1^i = v_i(1), \ldots, M_T^i = v_i(T))$; η^i is the transition probability matrix of $\{X_t^i\}$, that is the $k' \times k'$ matrix of η_{lm}^i; and P^i is the matrix of the vectors p_{ij}. (2.6) gives

$$L_T^i = \sum_{l=1}^{k'} [\mathrm{P}(X_1^i = H_l^i \mid \pi^i) \prod_{t=2}^{T} \mathrm{P}(X_t^i = H_m^i \mid X_{t-1}^i = H_l^i, \eta^i) \prod_{t=1}^{T} \mathrm{P}(M_t^i = v_i(t) \mid X_t^i = H_l^i, P^i)] \quad (2.7)$$

The EM algorithm formulation for HMMMk' is concluded by the maximization of the expected log likelihood given from (2.7) by

$$\text{ELog}L_T^i = \sum_{l=1}^{k'} \beta_l^i(1)\log\pi_l^i + \sum_{t=2}^{T}\sum_{l=1}^{k'}\sum_{m=1}^{k'} \gamma_{lm}^i(t)\log\eta_{lm}^i + \sum_{t=1}^{T}\sum_{l=1}^{k'} \beta_l^i(t)\log Q_{l,v_i(t)}^i \quad (2.8)$$

where $\gamma_l^i(t) = P(X_t^i = H_l^i, X_{t+1}^i = H_m^i \mid M_1^i = v_i(1), \ldots, M_T^i = v_i(T))$. The maximization of (2.8) with respect to the stochastic constraints on the parameters π_l^i, η_{lm}^i and $Q_{l,v_i(t)}^i$, through the method of the Lagrange multipliers results to the following formulas.

$$\pi_l^i = \beta_l^i(1), l = 1, \ldots, k' \quad (2.9)$$

$$\eta_{lm}^i = \frac{\sum_{t=2}^{T} \gamma_{lm}^i(t)}{\sum_{m=1}^{k'} \sum_{t=2}^{T} \gamma_{lm}^i(t)}, \quad l, m = 1, \ldots, k' \quad (2.10)$$

and

$$p_{lj}^i = \frac{\sum_{t=2}^{T} \beta_l^i(t)\, n_{ij}(t)}{\sum_{t=1}^{T} \beta_l^i(t)\, O_i(t)}, \quad l = 1, \ldots, k', \quad j = 1, \ldots, k+1 \quad (2.11)$$

Given the observed data of a manpower system, $v_i(1), \ldots, v_i(T)$, the algorithm is implemented through the formulas in (2.4), (2.9), (2.10) and (2.11) by first choosing initial values for π_l^i, η_{lm}^i and p_{ljr}^i and then using these values to obtain starting values for $Q_{l,v_i(t)}^i$. This iterative process continues until the convergence of the estimates of the parameters. This is utilized in Section 4 to obtain the estimates of the transition probabilities in the manpower data used for illustration.

3. Likelihood Ratio Search Procedure for Optimum k'

The likelihood ratio test is a statistical test widely applied in different areas of research; see, for example, [18,19]. It compares the performance of two models on the basis of the ratio of their likelihoods. For two hidden Markov manpower models HMMM*l* and HMMM*m*, the likelihood ratio statistic, L_r, is given as follows [2,4,5].

$$L_r = -2\log\left(\frac{L_{\text{HMMM}l}}{L_{\text{HMMM}m}}\right) \sim \chi_\alpha^2(\nu) \quad (3.1)$$

L_r has the chi square distribution with ν degrees of freedom. The hypothesis being tested is that the two models, HMMM*l* and HMMM*m*, fit the data equally well. This hypothesis of equality of fit is rejected if $L_r > \chi_\alpha^2(\nu)$; the rejection implies that the second model, HMMM*m*, fits the data better. This is applied in the following procedure for obtaining the optimum value of k'.

The procedure:

Step 1: Set $k' = 1, \ldots, N$; $N = 5$may be adequate for Markov manpower models; see [5]
.

Step 2: Assume HMMMk' exists for all k' ($k' = 1,\dots,N$) and estimate the transition probabilities for each HMMMk', using the EM algorithm and formulas in Section 2, to confirm the existence of the models.

Step 3: Obtain the log likelihood for each HMMMk'.

Step 4: Arrange HMMMk's in descending order or decreasing level of parsimony (see [5]) based on the value of k'; following this order, carry out pairwise L_r comparison test of the form: HMMM1 versus HMMM2, HMMM2 versus HMMM3, HMMM3 versus HMMM4 and HMMM4 Versus HMMM5, and so on.

Step 5: Mark the pair where equality of model performance first occurs; choose the k' of the first model in this pair as the optimum number of hidden states for the manpower system considered. Note that the search is stopped at the first occurrence of model equality since this is ordered comparison, ordered according to a decreasing level of a desired property.

Step 6: If HMMM1 versus HMMM2 happen to be the pair where equality of model performance first occurs then $k' = 1$is the optimum, which implies that there is no need to use HMMM for the system in the first place; that is, there exists no significant hidden heterogeneity in the manpower system considered.

Two values need to be computed for each paired comparison in the procedure for optimum value of number of hidden states of a HMMM. These are the values of L_r and ν. The computation of L_r, in equation (3.1), is straightforward after obtaining the likelihoods of the models. The computation ν is undertaken in the following subsection.

3.1 **Computation of the Degrees of Freedom for L_r Tests**

The value of the degrees of freedom, ν, for the L_r tests is given by

$$\nu = (\text{no. of free parameters of HMMM}m - \text{no. of free parameters of HMMM}l).$$

We obtain general formulas for ν in three model comparison cases. These cases cover all cases that may be of interest in the current likelihood ratio search procedure for optimum k'. For any k and k' the following results are obtained.

Case 1: HMMM1 versus HMMMk'

Under this case, the general formula for ν can be obtained as

$$\nu = (k' - 1)k^2 + (k'^2 - 1)k$$

(3.2)

To show how the result in (3.2) is obtained: each of the k observable classes of the HMMM1 contributes $(k + 1 - 1) = k$ free parameters. This gives k^2 free parameters for the k observable classes in HMMM1. In the HMMMk', each C_i ($i = 1,\dots,k$) has k' unobserved classes from where transitions can originate and $k + 1$ destination classes to where transitions can be made. This gives $(k + 1 - 1)k' = kk'$ free parameters for probabilities of transiting to the observable classes. Also, there are $(k' - 1)k'$ and $(k' - 1)$ free parameters from transitions within the unobserved classes and the initial states respectively. Hence, ν becomes

$$\nu = k(kk' + (k' - 1)k' + (k' - 1)) - k^2 \quad (3.3)$$

Equation (3.3) results to the given general formula for ν as

$$\nu = (k' - 1)k^2 + (k'^2 - 1)k.$$

Case 2: HMMM$(k' - 1)$ versus HMMMk'

Under this case, the general formula for ν can be obtained as

$$\nu = k^2 + (2k' - 1)k \quad (3.4)$$

In this case equation (3.4) is obtained as follows:

For HMMM$(k' - 1)$

No. of free parameters from each C_i $= (k + 1 - 1)(k' - 1) = k(k' - 1)$

No. of free parameters from hidden classes in C_i $= (k' - 2)k'$

No. of free parameters from the initial states for C_i $= (k' - 2)$

Combining these results and those for HMMMk' in Case1, ν becomes

$$\nu = k(kk' + (k' - 1)k' + (k' - 1) - k(k' - 1) - (k' - 2)k' - (k' - 2)) \quad (3.5)$$

The simplification of (3.5) results to the general formula for ν given in (3.4).

Case 3: HMMM$(k' - 2)$ versus HMMMk'

Under this case, the general formula for ν can be obtained as

$$\nu = 2k^2 + 4(k' - 1)k \quad (3.6)$$

In this case equation (3.6) is similarly obtained as follows:

For HMMM$(k' - 2)$

No. of free parameters from each C_i $= (k + 1 - 1)(k' - 2) = k(k' - 2)$

No. of free parameters from hidden classes in C_i $= (k' - 3)k'$

No. of free parameters from the initial states for C_i $= (k' - 3)$

Combining these results and those for HMMMk' in Case1, ν becomes

$$\nu = k(kk' + (k' - 1)k' + (k' - 1) - k(k' - 2) - (k' - 3)k' - (k' - 3)) \quad (3.7)$$

Simplifying (3.7) gives the general formula for ν in (3.6).

The general formulas for ν in the cases considered above are utilized to obtain some specific formulas that may be needed in executing the likelihood ratio search procedure for optimum k'. These specific formulas are given in Table 1.

Table 1: Formulas for the degrees of freedom (ν) for some specific L_r tests

Models Compared	Formula for ν
HMMM1 Vs HMMM2	$k^2 + 3k$
HMMM1 Vs HMMM3	$2k^2 + 8k$
HMMM1 Vs HMMM4	$3k^2 + 15k$
HMMM1 Vs HMMM5	$4k^2 + 24k$
HMMM2 Vs	$k^2 + 5k$

HMMM3	
HMMM3 Vs HMMM4	$k^2 + 7k$
HMMM4 Vs HMMM5	$k^2 + 9k$

4. Numerical Illustration

The developments in this paper are illustrated using the manpower data presented by Ossai et al. (2022). The manpower data are shown in Table 2 and contain senior academic personnel inter-class flow numbers for a university system for 8 periods of time, $t = 1, \ldots, 8$. In Table 2, C_1, C_2, C_3 and C_4 represent the class of senior lecturers, readers, professors, and leavers respectfully. The values in the table represent the number of personnel involved in moving from the given row class to the given column class for the given period of time. For example, the first value, 807, is the number of senior lecturers who remained as senior lecturers after time $t = 1$. The last value in the same row, 25, is the number of senior lecturers who left the system after time $t = 1$, and so on. For the current illustration, let there be four investigators or researchers R1, R2, R3 and R4 interested in modelling the manpower system using hidden Markov model approach. Assume R1 chooses two hidden classes for the system (i.e. HMMM2), R2 chooses three hidden classes for the system (i.e. HMMM3), R3 chooses four hidden classes for the system (i.e. HMMM4) and R4 chooses five hidden classes for the system (i.e. HMMM5). The problem is that of knowing the optimum number of hidden states for the system.

Following the steps of the proposed search procedure we assume the existence of all the HMMMk' for all k' ($k' = 1, \ldots, 5$). This is confirmed by estimating the parameters of the models through the EM formulas in Section 2, and then observing the estimated transition probabilities to ensure the hidden classes are actually represented by their different transition probabilities. This has been done for the data of Table 2 by Ossai et al (2022), and they show that all the five HMMM exist for the manpower data. For example, the transition probability matrices for CMMM (i.e. HMMM1) and HMMM5, as obtained by Ossai et al (2022), are shown below as P_1 and P_5 respectively.Next, the log likelihood for each HMMMk' is obtained. Next, HMMMk' is arranged in increasing order of the value of k'; following this, pairwise L_r comparison tests are carry out. The results of these steps are shown in Table 3. Table 3 also includes the results for the comparison of HMMM1 and all the other four HMMMs for further emphasis on the choice of the four researchers, as contained in the discussion section.

Table 2: A university senior academic manpower flow data

	C_1	C_2	C_3	C_4
C_1: $t=1$	807	50	40	25
$t=2$	801	102	20	34
$t=3$	788	81	18	20
$t=4$	794	32	34	41
$t=5$	820	72	27	37
$t=6$	815	42	36	52
$t=7$	826	61	30	45
$t=8$	840	55	45	30
C_2: $t=1$	0	182	31	10
$t=2$	0	201	30	10
$t=3$	0	194	28	12
$t=4$	0	198	30	12
$t=5$	0	211	35	21
$t=6$	0	190	21	15
$t=7$	0	185	42	26
$t=8$	0	205	37	28
C_3: $t=1$	0	0	560	56
$t=2$	0	0	534	50
$t=3$	0	0	521	54
$t=4$	0	0	550	46
$t=5$	0	0	578	32
$t=6$	0	0	570	64
$t=7$	0	0	540	93
$t=8$	0	0	548	71

$$P_1 = \begin{array}{c} \\ C_1 \\ C_2 \\ C_3 \end{array} \begin{array}{c} \begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \end{array} \\ \begin{pmatrix} 0.863 & 0.066 & 0.033 & 0.038 \\ 0 & 0.801 & 0.130 & 0.069 \\ 0 & 0 & 0.904 & 0.096 \end{pmatrix} \end{array};$$

$$P_5 = \begin{array}{c} C_1 \\ \\ C_2 \\ \\ C_3 \end{array} \left(\begin{array}{cccc} C_1 & C_2 & C_3 & C_4 \\ 0.859 & 0.063 & 0.031 & 0.047 \\ 0.858 & 0.075 & 0.028 & 0.039 \\ 0.871 & 0.055 & 0.045 & 0.029 \\ 0.852 & 0.098 & 0.020 & 0.029 \\ 0.872 & 0.040 & 0.038 & 0.050 \\ \hdashline 0 & 0.746 & 0.151 & 0.103 \\ 0 & 0.841 & 0.093 & 0.066 \\ 0 & 0.828 & 0.123 & 0.049 \\ 0 & 0.790 & 0.131 & 0.079 \\ 0 & 0.824 & 0.131 & 0.045 \\ \hdashline 0 & 0 & 0.853 & 0.147 \\ 0 & 0 & 0.892 & 0.108 \\ 0 & 0 & 0.910 & 0.090 \\ 0 & 0 & 0.948 & 0.052 \\ 0 & 0 & 0.923 & 0.077 \end{array} \right)$$

5. **Results and Discussion**

The estimates of the transition probabilities of the manpower ersonnels making various movements across the classes of the manpower system of Table 2, for the different HMMMs, are contained in their transition probability matrices. All the transition probability matrices for the five HMMMs can be seen in Ossai et al. The matrices P_1 and P_5 included in the current paper are for HMMM1 and HMMM5 respectively, and are obtained using the EM algorithm and the specified formulas in section 2. P_1 and P_5 show enough evidence that hidden classes exist in the manpower data up to $k' = 5$ since probabilities of the classes of high movers, movers, above mediocres, mediocres and stayers can be distinctly identified. For example, in P_1 the probability of a senior lecturer moving to the rank of a professor is $(P_1)_{13} = 0.033$; but in P_5 this same probability has five values, corresponding to the five hidden classes in decreasing order of magnitude. These five values are: $(P_5)_{33} = 0.045$, $(P_5)_{53} = 0.038$, $(P_5)_{13} = 0.031$, $(P_5)_{23} = 0.028$, $(P_5)_{43} = 0.020$.

Each of the independent researchers needs only P_1 and one other relevant transition probability matrix to verify the existence of his model; R1 needs P_1 and P_2, R2 needs P_1 and P_3, R3 needs P_1 and P_4 and R4 needs P_1 and P_5. And from the results obtained by the application of the likelihood ratio search procedure of sections 3 and 3.1, shown in Table 3, each of the four researchers would conclude that his chosen HMMM is adequate (superior to CMMM). This is because for R1: $L_r = 79.0935 > \chi^2_{0.05}(18) = 28.869$; for R2: $L_r = 100.2824 > \chi^2_{0.05}(42) = 55.758$; for R3: $L_r = 116.9538 > \chi^2_{0.05}(72) = 90.531$; and for R4: $L_r = 137.7263 > \chi^2_{0.05}(108) = 113.145$. However, by the use of the proposed procedure, only R1 chose the optimum number of hidden states for the manpower data, which is $k' = 2$. This is because by the likelihood ratio search procedure the pair where equality of model performance first occurred is HMMM2 Versus HMMM3 (Table 3), giving the optimum k' to be equal to 2 as HMMM2 is the first model in this pair. In other words, even though R1, R2 and R3 can conclude independently that their models are adequate for the data, that of R1 is the best specified model.

Table 3: Results of pairwise model comparison tests

Models Compared	Log likelihood	L_r	ν	Chi-square Value at $\propto = 0.05$	Decision by the four researchers	Decision Based on the search procedure
HMMM1 Vs HMMM2	-207.16303 (-167.61626)	79.0935	18	28.869	R1's model is superior	HMMM2 better
HMMM1 Vs HMMM3	-207.16303 (-157.02183)	100.2824	42	55.758	R2's model is superior	Not needed
HMMM1 Vs HMMM4	-207.16303 (-148.68615)	116.9538	72	90.531	R3's model is superior	Not needed
HMMM1 Vs HMMM5	-207.16303 (-138.29987)	137.7263	108	113.145	R4's model is superior	Not needed
HMMM2 Vs HMMM3	-167.61626 (-157.02183)	21.1889	24	36.415	Not applicable	Perform equally
HMMM3 Vs HMMM4	-157.02183 (-148.68615)	16.6714	30	43.773	Not applicable	Perform equally
HMMM4 Vs HMMM5	-148.68615 (-138.29987)	20.7726	36	55.758	Not applicable	Perform equally

(Note: The second model has the value in bracket in the second column)

6. Conclusion

In this paper, a procedure for choosing the optimum number of hidden states in a hidden Markov manpower model has been proposed and formulated. The proposed likelihood ratio search procedure has shown to be useful in preventing the choice of models which can be judged through a statistical test to be significantly adequate for a given manpower data, but which may not be the best in terms of parsimony in parameter inclusion. The procedure, therefore, leads to the selection of the best hidden Markov manpower model for a given manpower data.

For further research, the sensitivity of the proposed search procedure can be considered. The desired property upon which the manpower models are ordered and compared as possible candidates for yielding the optimum number of hidden states can be varied.

References

[1] Rabiner LR, Juang B H. An introduction to hidden Markov models. IEEE ASSP MAGAZINE. 1986;3 (1):4.

[2] Udom AU, Ebedoro UG. On multinomial hidden Markov model for hierarchical manpower systems. Communications in Statistics – Theory and Methods. 2021;50: 1370-1386. Available: https://doi.org/10.1080/03610926.2019.1650185.

[3] Rabiner LR. A tutorial on hidden Markov model and selected applications in speech recognition. Proceedings of the IEEE. 1989;77(2):257–86. Available: https://doi.org/10.1109/5.18626

[4] Guerry MA. Hidden heterogeneity in manpower systems: A Markov-switching model approach. European Journal of Operational Research. 2011;210(1):106-113. Available: https://doi.org/10.1016/j.ejor.2010.10.039

[5] Ossai EO, Ezra PN, Ohanuba FO, Eze MN. Homogeneity versus parsimony in Markov manpower models: A hidden markov chain approach. Asian Journal of Probability and Statistics. 2022;20(4):82-93. Available:https://doi.org/10.9734/ajpas/2022/v20i4441

[6] Seal HL. The mathematics of a population composed of *k* stationary strata each recruited from the stratum below and supported at the lower level by a uniform number of annual entrants. Biometrika. 1945;33:226-230.

[7] Bartholomew DJ. Stochastic models for social processes. 3rd ed. Chichester: Wiley; 1982.

[8] Bartholomew DJ, Forbes AF, McClean SI. Statistical techniques for manpower planning. 2nd ed. Chichester: Wiley; 1991.

[9] Ugwuowo FI, McClean S I. Modelling heterogeneity in a manpower system: A review. Applied Stochastic Models in Business and Industry. 2000;16(2):99–110. Available: https://doi.org/10.1002/1526-4025(200004/06)16:2<99::AID-SMB385>3.0.CO;2-3

[10] De Feyter T. Modelling heterogeneity in manpower planning: Dividing the personnel system into more homogeneous subgroups. Appl Stoch Models Bus Ind. 2006;22(4):321–334.

[11] Carette P, Guerry M-A. Markov models for duration-dependent transitions: Selecting the states using duration values or duration intervals? *Statistical Methods and Applications.* 2022. Available: https://doi.org/10.1007/s10260-022-00637-2.

[12] McClean S, Papadopoulou AA, Tsaklidis G. Discrete time reward models for homogeneous semi-Markov systems. Communications in Statistics – Theory and Methods. 2004;33 (3):623-638. Available:https://doi.org/10.1081/STA-120028688

[13] Guerry M-A, De Feyter T. An extended and tractable approach on the convergence problem of the mixed push-pull manpower model. Applied Mathematics and Computation. 2011;217:9062-9071.

[14] Nwaigwe CC. A renewal model with duration in a grade for manpower planning. Journal of Interdisciplinary Mathematics. 2011;14(4):463-469. Available: https://doi.org/10.1080/09720502.2011.10700764

[15] Verbeken B, Guerry M-A. Discrete Time hybrid semi-Markov models in manpower planning. Mathematics. 2021; 9(14):1681. Available:https://doi.org/10.3390/math9141681.

[16] Guerry M-A, De Feyter T. Markovian approaches in modelingworkforce systems. Journal of Current Issues in Finance, Business and Economics. 2009;2(4):1-20.

[17] Hélie S. An introduction to model selection: Tools and algorithms. Tutorial in Quantitative Methods for Psychology. 2006;2:1–10. Available:https://doi.org/10.20982/tqmp.02.1.p001.

[18] Perneger TV. How to use likelihood ratios to interpret evidence from randomized trials. Journal of Clinical Epidemiology. 2021;136:235-242. Available:https://doi.org/10.1016/j.jclinepi.2021.04.010.

[19] De Rochemonteix M, Napolioni V, Sanyal N, Belloy ME, Caporaso NE, Landi MT, et al. A likelihood ratio test for gene-environment interaction based on the trend effect of genotype under an additive risk model using the gene-environment independence assumption. American Journal of Epidemiology. 2021;190 (1):129–141. Available:https://doi.org/10.1093/aje/kwaa132.