Monika Dędys
BarbaraTarnowska
Institute of Econometrics
Warsaw School of Economics

Binary Hidden Markov Models in Analysis of the Results of Business Surveys[1]

Abstract

In recent years it turned out that Markov Switching Models (MSM) are very useful in analyses of macroeconomic time series. In most papers continuous state space models are considered. However, for the purpose of analyzing time series from business surveys discrete state space models are more suitable. Therefore, we use Hidden Markov Models (HMM), which can be treated as kind of MSM. In particular, we focus on binary HMM to demonstrate their efficacy in inference based on business survey results.

In this analysis data base of industry business surveys, carried out by the Research Institute of Economic Development of the Warsaw School of Economics is used.

Key Words: Hidden Markov models, discrete-valued time series, business surveys

1. Introduction

During the last two decades HMM have been widely used in modeling sequence of dependent random variables. It is worth emphasizing that the concept of this type of process is due to (Bather, 1965), (Baum, 1966), (Petrie, 1969), and that HMM were firstly applied in speech recognition problems (Levinson,1983). Now HMM seem to be very useful and popular in modeling phenomena in the area of biology (Michalek, 2001), genetics (Barndorff, 2001), meteorology (MacDonald, 2000), geophysics (MacDonald, 2000), medicine (Leroux, 1992) demography (MacDonald, 2000). They are also used in analysis of problems connected with image recognition.

In the area of econometric modeling a brilliant development of HMM-based methods is due to J.D. Hamilton. He investigated and successfully applied a particular kind of a HMM, a HMM with normal distribution, and introduced the term of Markov

[1] A more detailed version of this paper was published in *Roczniki Kolegium Analiz Ekonomicznych*, zeszyt 13/2004, Szkoła Główna Handlowa, Warszawa 2004

Switching Model (MSM) to econometrics. Nowadays extensions of primary model undoubtedly reach beyond the basic idea connected with HMM (Hamilton, 2002).

Binary HMM were applied in many areas (McDonald, 2000), but hardly ever in econometrics . In this paper we demonstrate that binary HMM could be successfully applied in modeling economic time series, and for this purpose we include an example.

2. Definitions

Hidden Markov Models are partially observable stochastic processes with a discrete time parameter.

There is a few, in general equivalent definitions of HMM (Elliot et al.,1995), (Bickel et.al., 1996), (McDonald, 2000), (MacKay, 2002). For our purposes we introduce the following definition.

Definition 1. A stochastic process $\{(X_n, Y_n), n \geq 0\}$ is called a hidden Markov model if it satisfies the following conditions.

1. The process $\{X_n, n \geq 0\}$ (unobservable) is a time homogeneous Markov chain with a finite state space S_X and transition matrix $P = [p(i,j)]_{i,j \in S_X}$. The process $\{X_n, n \geq 0\}$ is called a underlying Markov chain.

2. Given $(X_0, X_1, ..., X_n)$, random variables $Y_0, Y_1, ..., Y_n$, are independent with the conditional distribution of $Y_i, i = 0,1,...,n$, depending on X_i only, and given by $P(Y_i \leq y | X_i = x_i) = H(y; \theta_{x_i})$, where $H(.; \theta_{x_i})$ is a distribution function indexed by parameters $\theta_{x_i} \in \Theta$. Set of value of random variables Y, which we denote by S_Y, is called a signal space.

In the case of discrete signal space it is convenient to use emission matrix $\Pi = \left[P(Y_i = k | X_i = j)\right]_{k \in S_Y, j \in S_X}$.

One can estimate parameters of the model using maximum likelihood method. For large class of HMM, maximum likelihood estimators are known to be consistent and asymptotically normal (Baum1966), (Leroux, 1992), (Bickel1998).

For computing the maximum likelihood estimator the expectation-maximization (EM) algorithm can be applied (MacLachlan,1996).

One could think about extending this model to include underlying Markov chain with a longer memory. There are some examples of applying models with a hidden process being second-order Markov chain (MacDonald, 2000).

We use the following definition of a second-order Markov model, which is consistent with definition 1.

Definition 2. A process $\{(X_n, Y_n), n \geq 1\}$ is called a basic second-order Markov model if it satisfies following conditions.

1. The process $\{X_n, n \geq 0\}$ (unobservable) is a second–order time homogeneous Markov chain with a finite state space S_X .
2. The same as in the definition 1.

It can be proved that process $\{(X_{n-1}, X_n) n \geq 1\}$ is a Markov chain. In the sequel this process is called underlying extended Markov chain.

In the above model the conditional distribution of random variable Y_n depends only on the state of the hidden Markov chain at time *n*. It is easy to think of a little bit parsimonious relation between unobservable and observable variable, where the conditional distribution of random variable Y_n depends on the state underlying process at time *n* and time *n-1* as well.

Definition 3. A process $\{(X_n, Y_n), n \geq 1\}$ is called a second-order hidden Markov model if the process $\{(X_{n-1}, X_n), Y_n); n \geq 1\}$ is a HMM.

The state space of process $\{Y_n, n \geq 1\}$ is called a signal space of the second-order HMM.

For a discrete signal space it is convenient to use the following notation: $P(Y_n = j | X_n = i, X_{n-1} = l) = \pi_{j(il)}$.

3. Binary Hidden Markov Models in business survey analysis

In this section we show that binary HMM can be applied to analyze certain time series obtained from business surveys.

We deal with the business survey on private industrial enterprises. This survey is carried out monthly by the Institute of the Economic Development of the Warsaw School of Economics (RIED). We purposely choose the question on the production level as this question was changed in March 1997. Initially, respondents were asked to assess changes in the production level (above normal, normal, below normal) relative to a normal level for the given month. Since March 1997 changes in the output have been assessed in comparison with a preceding month. As a consequence of this change two time periods need be compared: January 1993-February 1997 (in the sequel this time interval is called period 1) and March 1997-May 2001(we refer to this period as to period 2). Although the survey under study has been conducted by the RIED since

1986, we are forced to exclude replies prior to January 1993 as they seem to be untypical, owing to the transformation of Polish economy. Furthermore, to preserve consistency of our results, we decided to limit the range of the research up to May 2001, which gives two periods of close lengths.

Below we present the transformation rules that, in addition, allow for an easy economic interpretation. Moreover, they enable us to determine whether the above-mentioned change of question has been reflected in responses.

Let B_t denotes a percentage balance of answers on the production level. In the sequel we use following notation.

$$\Delta B_t = B_{t+1} - B_t,$$

$$\Delta\Delta B_t = \Delta B_{t+1} - \Delta B_t.$$

It seems that simple information about signs of ΔB_t and $\Delta\Delta B_t$ in certain time intervals could be crucial in assessing business activity. Suggested interpretation of relevant signs in the two periods under study is given in table 1.

Table1. Interpretation of the signs B_t, ΔB_t, $\Delta\Delta B_t$

	Assessment of business activity	
	Period 1	Period 2
Bt>0	---	Growth
Bt<0	---	Downturn
ΔBt>0	Growth	Growth at increasing rate or downturn at decreasing rate
ΔBt<0	Downturn	Growth at decreasing rate or downturn at increasing rate
ΔΔBt>0	Growth at increasing rate or downturn at decreasing rate	---
ΔΔBt<0	Growth at decreasing rate or downturn at increasing rate	---

Based on the available series of $B_t, \Delta B_t, \Delta\Delta B_t$, we constructed three binary series for each period under study.

$$balance_t = \begin{cases} 0 & \text{for } B_t \geq 0 \\ 1 & \text{for } B_t < 0 \end{cases}$$

$$delta_t = \begin{cases} 0 & \text{for } \Delta B_t \geq 0 \\ 1 & \text{for } \Delta B_t < 0 \end{cases}$$

$$scdelta_t = \begin{cases} 0 & \text{for } \Delta\Delta B_t \geq 0 \\ 1 & \text{for } \Delta\Delta B_t < 0 \end{cases}.$$

We fitted HMM of second-order (defined in previous section) to thus derived binary series. In this paper we deal with underlying second-order Markov chain with state space $S_X = \{A, B\}$.

Having in mind possible interpretation of Markov chain states, we select symbol A to denote a state $i \in S$ satisfying the following inequality:

$$P(Y_t = 1 | X_t = i, X_{t-1} = i) > P(Y_t = 1 | X_t = j, X_{t-1} = j), \text{ where } j \neq i.$$

We expected that states (A,A) and (B,B) of the extended Markov chain would produce probability density significantly different from the discrete uniform density. Furthermore, states (A,B) and (B,A) were expected to generate densities „very similar" to the uniform one. The underlying reason for the above expectations was as follows. The pieces of a sample path of second order Markov chains such as: A,A,A,.. or B,B,B,... should correspond with some kind of stabilization in economy, while the pieces A,B,A,B, ... should be treated as a sign of destabilization.

At the same time we considered possibility of using first order-binary hidden Markov models with a two-state state space of the underlying Markov chains. Therefore, two models were fitted to each set of data, giving a total of 12 models.

To compare the models we used Bayesian Information Criterion (BIC). As a consequence, the model that minimizes BIC=-2L+klogn, where L is the value of the likelihood function, k is the number of independent parameters, and n is the number of observations, has been selected as the best (Schwarz, 1978), (MacDonald , 2000).

In the sequel we use the following notation. *P(X)* and Π(X) denote estimates of transition matrix and emission matrix of second-order hidden Markov model fitted to time series X (for example, *X=balance1* means the binary time series corresponding to period 1 and transformation given by the relation *balance)*.

The estimation results show that second-order models are preferred by BIC (table 2).

Table 2. BIC values

	Balance 1	*Balance 2*	*Delta1*	*Delta2*	*Scdelta1*	*Scdelta 2*
HMM1	78.312	84.114	85.412	79.799	68.442	90.578
HMM2	69.904	69.147	67.442	72.31	68.93	68.23

The estimates for all models are given in appendix. Here we present and discuss only the most interesting or typical.

For time series balance1 results are as follows.

$$P(balance1) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0,2175 & 0,7825 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0,149 & 0,8709 & 0 & 0 \end{bmatrix}$$

$$\Pi(balance1) = \begin{matrix} \\ 0 \\ 1 \end{matrix} \begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \\ \left[\begin{matrix} 0,3212 \\ 0,6788 \end{matrix} \right. & \begin{matrix} 0,1308 \\ 0,8692 \end{matrix} & \begin{matrix} 1,0000 \\ 0,0000 \end{matrix} & \left. \begin{matrix} 1,0000 \\ 0,0000 \end{matrix} \right] \end{matrix}.$$

Estimates of transition probabilities in matrix P(balance1) suggest that in the underlying second-order Markov chain there are no pieces of sample path of type B,B. Therefore, it seems that the states have no economic interpretation. Obviously, in the case of time series balance1 this result is not surprising (see table 1). Unfortunately, one can draw similar conclusions from estimates concerning time series *balance2*.

$$P(balance2) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0,2539 & 0,7461 \\ 0,4015 & 0,5984 & 0 & 0 \\ 0 & 0 & 0,2583 & 0,7497 \\ 0,4371 & 0,5628 & 0 & 0 \end{bmatrix}.$$

A relatively low transition probability from state (A,B) to state (B,B) seems to exclude economic interpretation as well. Despite of lack of interpretation, some interesting conclusions can be drawn. For example, it is worth noticing that the underlying extended Markov chain moves from state (A,B) and (B,B) with an almost equal probability. Comparing second row of matrix P(balance 2) with the forth one leads to a similar conclusion. Therefore, one can think that model with first-order underlying Markov chain is more suitable. It turns out, however, that a longer memory of model is nested in relation between unobservable variable and observable one. This is evidenced from the following estimates

$$\Pi(balance2) = \begin{matrix} \\ 0 \\ 1 \end{matrix} \begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \\ \left[\begin{matrix} 0,4083 \\ 0,5917 \end{matrix} \right. & \begin{matrix} 0,2718 \\ 0,7282 \end{matrix} & \begin{matrix} 0,5851 \\ 0,4149 \end{matrix} & \left. \begin{matrix} 0,7936 \\ 0,2064 \end{matrix} \right] \end{matrix}.$$

It is worth to note that analysis of other models leads to similar conclusions. For example, we have

$$P(delta1) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0{,}2425 & 0{,}7548 \\ 0{,}3679 & 0{,}6321 & 0 & 0 \\ 0 & 0 & 0{,}2435 & 0{,}7565 \\ 0{,}3599 & 0{,}6424 & 0 & 0 \end{bmatrix}$$

$$\Pi(delta1) = \begin{matrix} \\ 0 \\ 1 \end{matrix} \begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \\ \left[0{,}3599 \right. & 0{,}3179 & 0{,}5191 & \left. 0{,}8186 \right] \\ \left[0{,}6401 \right. & 0{,}6821 & 0{,}4809 & \left. 0{,}1814 \right] \end{matrix}$$

Let us mention that estimates for P(balance2), П(balance2) and P(delta1), П(delta1) are very closed, on the contrast with estimates for P(balance1) P(balance1). Comparing P(scdelta1), П(scdelta1) and P(delta2), П(delta2) one can see similar relation. It seems to confirm our supposition that binary HMM enabled us to detect above-mentioned change of the question. Unfortunately our statements are far from mathematical accuracy. Statistical inference based on the likelihood ratio test in the case of such small sample requires farther investigations.

4. Conclusions

General conclusion are as follows:

1. The long-memory models seem to be useful in business survey analysis.
2. The binary HMMs are sensitive tool for investigation of business survey results. They are able to detect change of the question in survey.
3. The states of the underlying chain seem not to have economic interpretation.
4. Considered HMM of the second order enabled us to draw essential conclusions about the way of the questions interpretation by the respondents.
5. This is worth considering models with much longer memory by replacing ordinary underlying chain by MTD.
6. The still opened problem are finite sample properties of the Maximum Likelihood Estimator in Hidden Markov Models.

References

Barndorff-Nielsen, O. E., D. Cox and C. Klüppelberg (2001), Complex Stochastic Systems, Chapmann&Hall /CRC.

Bather, J. A. (1965), Invariant Conditional Distributions, *Ann. Math. Statist.* 36, pp. 829-846.

Baum, L. E. and T. Petrie (1966), Statistical Inference for Probabilistic Functions of Finite State Markov Chains, *Ann. Math. Statist.*,37, pp. 1554-1563.

Bickel, P. J., Y. Ritov and T. Ryden (1998), Asymptotic Normality of the Maximum Likelihood Estimator for General Hidden Markov Models, *The Annals of Statistics* 26, No 4 , pp. 1614-1635.

Bickel, P. J. and Y. Ritov.(1996), Inference in Hidden Markov Models I: Local Asymptotic Normality in the Stationary Case, *Bernoulli* 2(3) , pp. 199-228.

Elliot, R. J., L. Aggoun and J. B. Moore (1995), Hidden Markov Models: Estimation and Control, Springer New York.

Hamilton, J. D. and B. Raj (2002), Advances in Markov –Switching Models, Springer-Verlag.

Leroux, B. G. (1992), Maximum-likelihood Estimation for Hidden Markov Models, *Stochastic Processes and Their Applications,* 40, pp. 127-143.

Levinson, S., L. Rabiner and M. Sondhi (1983), An Introduction to the Application of the Theory of Probabilistic Functions of Markov Process to Automatic Speech Recognition, *The Bell System Technical Journal*, 62(4), pp. 1035-1074.

MacDonald, I. L . and W. Zucchini (2000), Hidden Markov and Other Models for Discrete – valued Time Series, Chapmann&Hall/CRC.

MacKay, R. J. (2002), Estimating the Order of a Hidden Markov Model. *The Canadian Journal of Statistics, 30*(4), pp. 573-589.

MacLachlan, G. J. and K. Thriyambakam (1996), The EM Algorithm and Extensions, Willey&Sons.

Michalek, S., M. Wagner and J. Timmer (2001), Finite Sample Properties of the Maximum Likelihood Estimator and of Likelihood Ratio Test in Hidden Markov Models, *Biometrical Journal* 43 (7), pp. 863-879.

Petrie, T. (1969), Probabilistic Functions of Finite State Markov Chains, *The Ann. Math. Statist.* 40(1), pp. 97-115.

Appendix

$$P(balance1) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0{,}2175 & 0{,}7825 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0{,}149 & 0{,}8709 & 0 & 0 \end{bmatrix}$$

$$\Pi(balance1) = \begin{matrix} \\ 0 \\ 1 \end{matrix} \begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \\ \left[0{,}3212 \right. & 0{,}1308 & 1{,}0000 & \left. 1{,}0000 \right] \\ \left[0{,}6788 \right. & 0{,}8692 & 0{,}0000 & \left. 0{,}0000 \right] \end{matrix}$$

$$P(balance2) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0{,}2539 & 0{,}7461 \\ 0{,}4015 & 0{,}5984 & 0 & 0 \\ 0 & 0 & 0{,}2583 & 0{,}7497 \\ 0{,}4371 & 0{,}5628 & 0 & 0 \end{bmatrix}$$

$$\Pi(balance2) = \begin{matrix} \\ 0 \\ 1 \end{matrix} \begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \\ \end{matrix} \begin{bmatrix} 0{,}4083 & 0{,}2718 & 0{,}5851 & 0{,}7936 \\ 0{,}5917 & 0{,}7282 & 0{,}4149 & 0{,}2064 \end{bmatrix}$$

$$P(delta1) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0{,}2425 & 0{,}7548 \\ 0{,}3679 & 0{,}6321 & 0 & 0 \\ 0 & 0 & 0{,}2435 & 0{,}7565 \\ 0{,}3599 & 0{,}6424 & 0 & 0 \end{bmatrix}$$

$$\Pi(delta1) = \begin{matrix} 0 \\ 1 \end{matrix} \overset{\begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \end{matrix}}{\begin{bmatrix} 0{,}3599 & 0{,}3179 & 0{,}5191 & 0{,}8186 \\ 0{,}6401 & 0{,}6821 & 0{,}4809 & 0{,}1814 \end{bmatrix}}$$

$$P(delta2) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0{,}2535 & 0{,}7465 \\ 0{,}5839 & 0{,}4161 & 0 & 0 \\ 0 & 0 & 0{,}2686 & 0{,}7314 \\ 0{,}5809 & 0{,}4191 & 0 & 0 \end{bmatrix}$$

$$\Pi(delta2) = \begin{matrix} 0 \\ 1 \end{matrix} \overset{\begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \end{matrix}}{\begin{bmatrix} 0{,}2933 & 0{,}5962 & 0{,}8093 & 0{,}4282 \\ 0{,}7067 & 0{,}4038 & 0{,}1907 & 0{,}5718 \end{bmatrix}}$$

$$P(scdelta1) = \begin{matrix} (A,B) \\ (A,A) \\ (B,B) \\ (B,A) \end{matrix} \begin{bmatrix} 0 & 0 & 0{,}2984 & 0{,}7016 \\ 0{,}5775 & 0{,}4225 & 0 & 0 \\ 0 & 0 & 0{,}2991 & 0{,}7009 \\ 0{,}5743 & 0{,}4257 & 0 & 0 \end{bmatrix}$$

$$\Pi(scdelta1) = \begin{matrix} 0 \\ 1 \end{matrix} \overset{\begin{matrix} (A,B) & (A,A) & (B,B) & (B,A) \end{matrix}}{\begin{bmatrix} 0{,}6113 & 0{,}3835 & 0{,}4201 & 0{,}8516 \\ 0{,}3887 & 0{,}6165 & 0{,}5799 & 0{,}1484 \end{bmatrix}}$$

$$P(scdelta2)=\begin{matrix}(A,B)\\(A,A)\\(B,B)\\(B,A)\end{matrix}\begin{bmatrix}0 & 0 & 0,2923 & 0,7077\\ 0,4682 & 0,5138 & 0 & 0\\ 0 & 0 & 0,2802 & 0,7198\\ 0,4664 & 0,5336 & 0 & 0\end{bmatrix}$$

$$\Pi(scdelta2)=\begin{matrix} \\ 0\\1\end{matrix}\begin{matrix}(A,B) & (A,A) & (B,B) & (B,A)\\ \left[\begin{matrix}0,3295 \\ 0,6705\end{matrix}\right. & \begin{matrix}0,5022\\0,4978\end{matrix} & \begin{matrix}0,8328\\0,1672\end{matrix} & \left.\begin{matrix}0,6785\\0,3215\end{matrix}\right]\end{matrix}$$