

MODELING OF VEHICLE INSURANCE CLAIM USING EXPONENTIAL HIDDEN MARKOV

Irfani Azis^{1 §}, Berlian Setiawaty², I. Gusti Putu Purnaba³

^{1,2,3}Department of Mathematics

Bogor Agricultural University

Meranti Street, Bogor, 16680, INDONESIA

Abstract: The purpose of this research is to model the vehicle insurance claim using exponential hidden Markov model (EHMM). The maximum likelihood method, forward-backward algorithm and expectation maximization (EM) algorithm are used to determine the EHMM parameter. The best model was selected based on Bayesian information criterion (BIC).

AMS Subject Classification: 60J22, 60J28

Key Words: claim, exponential, hidden Markov, vehicle insurance

1. Introduction

Insurance is an agreement in which the insurer binds itself to an insured company, by receiving a premium to provide a replacement due to loss, damage or loss of expected profit, which may occur due to an uncertain event. Vehicle insurance is one type of insurance that protects the vehicle from loss. The cause of loss is unpredictable, so the fund which is paid for a claim by an insurance company is unknown.

Received: 2017-08-24

Revised: 2017-12-29

Published: March 17, 2018

© 2018 Academic Publications, Ltd.

url: www.acadpubl.eu

[§]Correspondence author

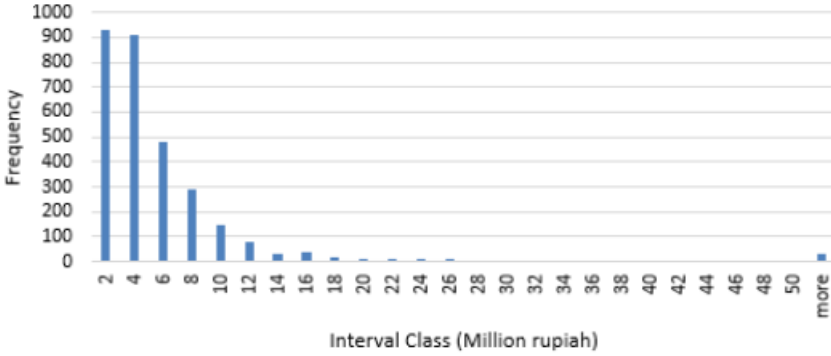


Figure 1: Histogram frequency of vehicle insurance claim of PT XX in 2014.

The greater loss that happened the greater claim would be paid. The loss that happened can be caused by several factors, such as weather condition, traffic condition, driver condition and other factors. Figure 1 shows that the small loss often occurred but huge loss rarely occurred. The distribution of claim can be seen as an exponential distribution or a mix of the exponential distribution. If the cause of loss is assumed to form a Markov chain and is not observed directly, and only data of claim is observed, then the pair of claim and cause of loss can be modeled by hidden Markov model. The purpose of this research is to model of vehicle insurance claim using exponential hidden Markov model.

2. Exponential Hidden Markov Model

Exponential hidden Markov model (EHMM) is a model which consists of pair of stochastic process $\{X_t, Y_t\}_{t \in N}$. $\{X_t\}_{t \in N}$ is unobserved and form a Markov chain. $\{Y_t\}_{t \in N}$ is an observation process which depends on $\{X_t\}_{t \in N}$. $Y_t|X_t$ has the exponential distribution, for all t .

Let $\{X_t\}_{t \in N}$ be the Markov chain with state space $S_X = \{1, 2, 3, \dots, m\}$ and transition probability matrix $A = [a_{ij}]_{m \times m}$, where $a_{ij} = P(X_t = j | X_{t-1} = i) = P(X_2 = j | X_1 = i)$, $a_{ij} \geq 0$ and $\sum_{j=1}^m a_{ij} = 1$. Let $\phi = [\phi_i]_{m \times 1}$ is the initial state probability vector, where $\phi_i = P(X_1 = i)$, $\sum_{j=1}^m \phi_i = 1$.

The distribution of $Y_t | X_t = i, i \in S_X; t \in N$ has an exponential distribution with parameter λ_i . The probability density function of $Y_t | X_t = i$ is $\gamma_{yi} =$

$f(y) = \lambda_i e^{-\lambda_i y}, y > 0$. Suppose that $\lambda = [\lambda_i]_{m \times 1}$.

So EHMM can be identified by parameter $\phi = (m, A, \lambda)$.

3. Parameter Estimation of Exponential Hidden Markov Model

The parameter $\phi = (m, A, \lambda)$ is estimated using maximum likelihood method (see [1]). Suppose that T is the number of observation, $y = (y_1, y_2, \dots, y_T)$ is the observation data of $\{Y_t\}_{t \in N}$, $x = (i_1, i_2, \dots, i_T)$ is the unobserved data of $\{X_t\}_{t \in N}$, $(x, y) = (i_1, y_1, i_2, y_2, \dots, i_T, y_T)$ is the data of $\{X_T, Y_t\}_{t \in N}$, and $\Phi_m = \{\phi = (m, A, \lambda) : A = [0, 1]^{m^2}, \lambda \in [\epsilon, 1/\epsilon]^m\}$ is the parameter space of EHMM.

Assume that a_{ij} , λ_i , and φ_i are continuous function from R^m to R with $a_{ij}(\phi) = a_{ij}$, $\lambda_i(\phi) = \lambda_i$ and $\varphi_i(\phi) = \varphi_i$, for all $i, j \in S_X$. Let $L_T^{(m)}(\phi)$ be an incomplete likelihood function, that is a probability density function of $y_1, y_2, \dots, y_T$.

$$L_T^{(m)}(\phi) = f_\phi(y) = f_\phi(y_1, \dots, y_T) = \sum_{i_1=1}^m \dots \sum_{i_T=1}^m \varphi_{i_1} \gamma_{y_1 i_1} \prod_{t=2}^T a_{i_{t-1} i_t} \gamma_{y_t i_t}.$$

Let $L_T^{C(m)}(\phi)$ be a complete likelihood function. Where

$$L_T^{C(m)}(\phi) = f_\phi(x, y) = f_\phi(i_1, y_1, \dots, i_T, y_T) = \varphi_{i_1} \gamma_{y_1 i_1} \prod_{t=2}^T a_{i_{t-1} i_t} \gamma_{y_t i_t}.$$

We obtain

$$L_T^{(m)}(\phi) = \sum_x L_T^{C(m)}(\phi).$$

Calculating $L_T^{(m)}(\phi)$ directly when m and T are big, is a complex process. Because of that, we use the forward-backward algorithm to calculate $L_T^{(m)}(\phi)$. The forward-backward algorithm consists of two sections, those are forward algorithm, and backward algorithm. Define forward variabel $\alpha_t(i, \phi, T)$ for $t = 1, 2, \dots, T$ and $i \in S_X$ as follows.

$$\alpha_t(i, \phi, T) = P_\phi(Y_1 = y_1, Y_2 = y_2, \dots, Y_t = y_t, X_t = i).$$

Define backward variable $\beta_t(i, \phi, T)$ for $t = T, T-1, \dots, 1; i \in S_X$ as follows.

$$\beta_t(i, \phi, T) = P_\phi(Y_{t+1} = y_{t+1}, Y_{t+2} = y_{t+2}, \dots, Y_T = y_T | X_t = i).$$

Theorem 1. (see [5]) For any $t = 1, 2, \dots, T$ then

$$L_T^{(m)}(\phi) = \sum_{i=1}^m \alpha_t(i, \phi, T) \beta_t(i, \phi, T).$$

Now, the problem is to find $\phi^* \in \arg \max \ln(L_T^{(m)}(\phi))$. In this case, Expectation Maximization (EM) algorithm is used. EM algorithm is an iterative algorithm which consists of two steps for every iteration, those are E step, and M step. EM algorithm is as follows.

1. For $k = 0$, the initial value $\phi^{(k)}$ is given.
2. (E-step) : Calculate

$$Q(\phi, \phi^{(k)}) = E_{\phi^{(k)}} \left(\ln(L_T^{C(m)}(\phi)) | y \right).$$

Use Theorem 2 for calculating $Q(\phi, \phi^{(k)})$.

Theorem 2. (see [2])

$$\begin{aligned} Q(\phi, \phi^{(k)}) &= E_{\phi^{(k)}} \left(\ln(L_T^{C(m)}(\phi)) | y \right) \\ &= \sum_{i=1}^m \frac{\alpha_1(i, \phi^{(k)}, T) \beta_1(i, \phi^{(k)}, T)}{\sum_{l=1}^m \alpha_t(l, \phi^{(k)}, T) \beta_t(l, \phi^{(k)}, T)} \ln(\varphi_i) \\ &+ \sum_{i=1}^m \sum_{j=1}^m \frac{\sum_{t=1}^{T-1} \alpha_t(j, \phi^{(k)}, T) \gamma_{y_{t+1}, j}^{(k)} a_{ij}^{(k)} \beta_{t+1}(j, \phi^{(k)}, T)}{\sum_{l=1}^m \alpha_t(l, \phi^{(k)}, T) \beta_t(l, \phi^{(k)}, T)} \ln(a_{ij}) \\ &+ \sum_{i=1}^m \frac{\sum_{t=1}^T \alpha_t(j, \phi^{(k)}, T) \beta_t(j, \phi^{(k)}, T)}{\sum_{l=1}^m \alpha_t(l, \phi^{(k)}, T) \beta_t(l, \phi^{(k)}, T)} \ln(\gamma_{y_{ti}}). \end{aligned}$$

Theorem 3. (see [6]) Suppose that

$$\Phi_m = \left\{ \phi = (m, A, \lambda) : A = [0, 1]^{m^2}, \lambda \in [\epsilon, 1/\epsilon]^m \right\}$$

and $\epsilon > 0$ is very small near zero and EHMM parameter continuous assumption is fulfilled, then

- (a) Φ_m is compact set in R^{m^2+m} .

(b) $\ln(L_T^{(m)}(\phi))$ is a continuous function in Φ_m and differentiable in interior Φ_m .

(c) $Q(\hat{\phi}, \phi)$ is a continuous function in $\Phi_m \times \Phi_m$.

3. (M-step) : Find $\phi^{(k+1)}$ which maximizes $Q(\phi, \phi^{(k)})$. That is

$$Q(\phi^{(k+1)}, \phi^{(k)}) \geq Q(\phi, \phi^{(k)}), \forall \phi \in \Phi_m.$$

The existence of $\phi^{(k+1)}$ is guaranteed by Theorem 3. Furthermore, using Theorem 4 we have the parameter $\phi^{(k+1)}$.

Theorem 4. (see [2]) Suppose that $\phi^{(k)} \in \Phi_m$ with $\phi^{(k)} = (m, A^k, \lambda^k)$, then the parameter $\phi^{(k+1)} = (m, A^{(k+1)}, \lambda^{(k+1)})$ which maximizes $Q(\phi, \phi^{(k)})$ in Φ_m can be written as

$$a_{ij}^{k+1} = \frac{\sum_{t=1}^{T-1} \alpha_t(j, \phi^{(k)}, T) \gamma_{y_{t+1}, j}^{(k)} a_{ij}^{(k)} \beta_{t+1}(j, \phi^{(k)}, T)}{\sum_{l=1}^m \alpha_t(l, \phi^{(k)}, T) \beta_t(l, \phi^{(k)}, T)}$$

and

$$\lambda_i^{(k+1)} = \left(\frac{\sum_{t=1}^T \alpha_t(j, \phi^{(k)}, T) \beta_t(j, \phi^{(k)}, T) (y_t)}{\sum_{l=1}^m \alpha_t(l, \phi^{(k)}, T) \beta_t(l, \phi^{(k)}, T)} \right)^{-1}.$$

4. Replace k with $k + 1$ and if $|\ln(L_T^{(m)}(\phi^{(k+1)})) - \ln(L_T^{(m)}(\phi^{(k)}))| > \epsilon$ then go back to 2nd until 4th step. Iteration is stopped when $|\ln(L_T^{(m)}(\phi^{(k+1)})) - \ln(L_T^{(m)}(\phi^{(k)}))| < \epsilon$. In other words $\{\ln(L_T^{(m)}(\phi^{(k+1)}))\}$ converges to $\{\ln(L_T^{(m)}(\phi^*))\}$. Its existence is guaranteed by Theorem 5.

Theorem 5. (see [6]) Suppose that $Q(\hat{\phi}, \phi)$ is continuous function in $\Phi_m \times \Phi_m$ and $\{\phi^{(k)}\}$ is the set of EHMM estimator which is obtained from EM algorithm, then

(a) $\{\ln(L_T^{(m)}(\phi^{(k)}))\}$ is monotonically increasing sequence.

(b) $\exists \phi^* \in \Phi_m \ni \lim_{k \rightarrow \infty} \ln(L_T^{(m)}(\phi^{(k)})) = \ln(L_T^{(m)}(\phi^*))$.

(c) If $\lim_{k \rightarrow \infty} \phi^{(k)} = \phi^*$, then ϕ^* is the stationary point of $(L_T^{(m)}(\phi))$.

Suppose that $\hat{\phi} = \phi^{(k+1)}$ is obtained in the last iteration of EM algorithm, so $\hat{\phi}$ is EHMM parameter estimation, where $\hat{\phi} = (\hat{m}, \hat{A}, \hat{\lambda})$. Bayesian information criterion (BIC) is used for finding state $\hat{m}$ (see [4]). The selection of $\hat{m}$ is based

on $\hat{m}$ which maximizes $\ln(L_t^{(m)}(\phi)) - (\ln T)d_m$, where $\ln(L_t^{(m)}(\phi))$ is the value of likelihood function and d_m is the number of estimation parameter, so

$$BIC = \ln L_T^{(m)}(\phi) - \frac{(\ln T)d_m}{2}.$$

The number of estimation parameter in EHMM is $(m^2 + m)$, so the formula of BIC as follows

$$BIC = \ln L_T^{(m)}(\phi) - \frac{(\ln T)(m^2 + m)}{2}.$$

Furthermore to validate of the model, a set of estimation data $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T$, is generated, based on $\hat{\phi}$, where

$$\hat{y}_t = E_{\hat{\phi}}(Y_t = y_t | Y_1^{t-1}) = \frac{1}{L_{t-1}^m(\hat{\phi})} \int_0^\infty y_t L_t^m(\hat{\phi}) dy_t,$$

and error is calculated by mean absolute percentage error (MAPE), where

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100\%.$$

Interpretation of MAPE is as follows.

Table 1: Interpretation of MAPE [3]

MAPE%	Interpretation
≥ 10	Highly accurate forecasting
10 – 20	Good forecasting
20 – 50	Reasonable forecasting

4. Programming Algorithm of Exponential Hidden Markov Model

Suppose that $y = \{y_1, y_2, \dots, y_T\}$ is an observation set and EHMM parameter $\phi = (m, A, \lambda)$ will be estimated. The algorithm which is used is as follows

1. i.) Input T the number of data and data $y_1, y_2, \dots, y_T$.
- ii.) Set $m = 2, k = 0, l = 1$.

2. i.) For $k = 0$, choose the initial value of $\phi^{(k)} = (m, A^{(k)}, \lambda^{(k)})$, where $A^{(k)}$ and $\lambda^{(k)}$ are generated randomly, $A^{(k)} = [a_{ij}^{(k)}]_{(m \times m)}$, $\lambda^{(k)} = [\lambda_i^{(k)}]_{(m \times 1)}$ and fulfill $\sum_{j=1}^m a_{ij}^{(k)} = 1$, $\lambda_i^{(k)} \geq 0$ and $i = 1, 2, \dots, m$.
- ii.) Generate $\varphi^{(k)}$ value which fulfill $(\varphi^{(k)})^T = (\varphi^{(k)})^T A^{(k)}$.
- iii.) Generate $\gamma_{y_l i_l}^{(k)}$ from $Y_l | X_l = i_l$.

3. Determine $\epsilon = 10^{(-1)}$ as bound of the likelihood iteration, that is

$$|\ln(L_T^{(m)}(\phi^{(k+1)})) - \ln(L_T^{(m)}(\phi^{(k)}))| < \epsilon.$$

4. Calculate Forward and Backward function.
5. (E-Step) Calculate $L_l(\phi^{(k)})$

$$L_l(\phi^{(k)}) = \sum_{i=1}^m \alpha_l(i, \phi^{(k)}, l) \beta_l(i, \phi^{(k)}, l).$$

6. (M-Step) Calculate parameter estimation value $\phi^{(k+1)} = (m, A^{(k+1)}, \lambda^{(k+1)})$, where

$$a_{ij}^{k+1} = \frac{\sum_{t=1}^{T-1} \alpha_t(j, \phi^{(k)}, T) \gamma_{y_{t+1}, j}^{(k)} a_{ij}^{(k)} \beta_{t+1}(j, \phi^{(k)}, T)}{\sum_{l=1}^m \alpha_t(l, \phi^{(k)}, T) \beta_t(l, \phi^{(k)}, T)}$$

$$\lambda_i^{(k+1)} = \left(\frac{\sum_{t=1}^T \alpha_t(j, \phi^{(k)}, T) \beta_t(j, \phi^{(k)}, T) (y_t)}{\sum_{l=1}^m \alpha_t(l, \phi^{(k)}, T) \beta_t(l, \phi^{(k)}, T)} \right)^{-1}.$$

7. i.) If $|\ln(L_T^{(m)}(\phi^{(k+1)})) - \ln(L_T^{(m)}(\phi^{(k)}))| > \epsilon$, go back to 4th until 6th step and replace k by $k + 1$. Otherwise go to the next step. Replace l by $l + 1$ and return to 2nd step, where $l = 1, 2, \dots, T - 1$.
- ii.) Generate set of estimation data $\hat{Y}_l$.
- iii.) Verify that $\left| \frac{\hat{Y}_l - Y_l}{Y_l} \right| < 7\%$, otherwise replace l to $l + 1$, where $l = 1, 2, \dots, T - 1$ and return to 1st step.
- iv.) Calculate BIC value.
- v.) Calculate error value of estimation data by MAPE. If $\text{MAPE} > 30\%$ then return to 2nd step.
8. i.) For determining optimum m , do the 1st step until 8th step for every state $m = 2, 3, \dots, 6$ with different T .
- ii.) Chose m^* which maximize BIC.

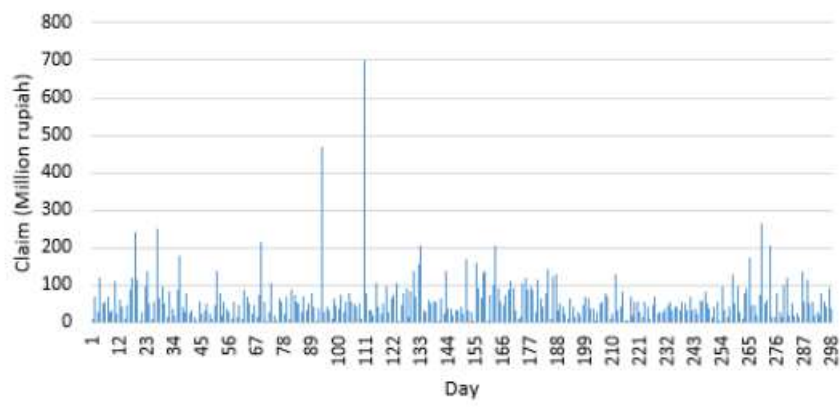


Figure 2: Vehicle Insurance Claim for each Day in 2014.

5. Application of EHMM on Vehicle Insurance Claim

This research uses data of vehicle insurance claim in one of Indonesia insurance company in 2014. The data can be seen in Figure 2. The data of claim is called observation data. Suppose that Y_t is the observation data, X_t is the unobserved data, and t is t^{th} day, where $t = 1, 2, \dots, T$ and $T = 298$ days. Parameter estimation of the model is exercised for 2-states until 6-states. The last 10 data are used for testing estimation data or prediction data. Based on the programming result, log-likelihood value and BIC value is obtained for every state with 288 days of observation data which can be seen in Table 2.

Table 2: Log-likelihood value and BIC value for every state with 288 days observation data

<i>m</i> - state	Loglikelihood value	BIC value
2 - state	-1472.046	-1489.035
3 - state	-1469.229	-1503.207
4 - state	-1472.037	-1528.667
5 - state	-1472.062	-1557.006
6 - state	-1471.996	-1590.918

Based on Table 2, if the number of states increase, then the BIC value will decrease. Afterwards interval observation will be determined starting from a

week, a month, and 3 months. Where assumption of a week is 6 day and a month is 26 days. Based on programming log-likelihood value and BIC value is obtained for every observation interval and state can be seen in Table 3 and Table 4.

Table 3: Log-likelihood value for every state and observation interval

m - state	Loglikelihood value		
	1 week	1 month	3 month
2 - state	-30.659	-137.790	-389.556
3 - state	-30.659	-137.790	-389.557
4 - state	-30.659	-137.790	-389.556
5 - state	-30.660	-137.790	-389.557
6 - state	-30.461	-137.790	-389.557

Table 4: BIC value for every state and observation interval

m - state	BIC value		
	1 week	1 month	3 month
2 - state	-36.034	-147.564	-402.626
3 - state	-41.409	-157.339	-415.697
4 - state	-48.577	-170.371	-433.123
5 - state	-57.536	-186.662	-454.907
6 - state	-68.088	-206.210	-481.047

The best model is chosen by the highest BIC value. Based on Table 2, EHMM with state $m = 2$ has the highest BIC value, so the optimum parameter estimation is state $m = 2$. Based on Table 3 and Table 4, if the observation interval increase then log-likelihood value and BIC value will decrease. So the optimum number of state and observation interval is state $m = 2$ and 1 week observation interval.

Parameter estimation value which is obtained from 1 week observation interval with state $m = 2$ are transition probability matrix A , initial transition probability vector φ , and exponential parameter estimation vector λ , where the number of 1 week observation data is 283rd until 288th day of claim data.

Transition probability matrix A is obtained as follows.

$$A = \begin{pmatrix} 0.098241 & 0.901759 \\ 0.967669 & 0.032331 \end{pmatrix}.$$

Based on matrix A, if the previous claim is caused by state 1, then the claim would happen again by state 1 with probability value 0.098241. If the previous claim is caused by state 2, then the claim would happen again by state 1 with probability value 0.967669. If the previous claim is caused by state 1, then the claim would happen again by state 2 with probability value 0.901759. If the previous claim is caused by state 2, then the claim would happen again by state 2 with probability value 0.032331.

Initial transition probability matrix was obtained as follows

$$\varphi = \begin{pmatrix} 0.517629 \\ 0.482371 \end{pmatrix}.$$

φ shows that, the claim which is caused by state 1 has parobability value 0.517629 and the claim which is caused by state 2 has probability value 0.482371. Exponential parameter estimation vector λ as follows

$$\lambda = \begin{pmatrix} 0.012115 \\ 0.026525 \end{pmatrix}.$$

λ shows that, the claim that will be happened by state 1 has rate 0.012115 million rupiahs per day. Afterward, the claim that will be happened by state 2 has rate 0.026525 million rupiahs per day. Furthermore the set of estimation data is generated by the best obtained parameter. Estimation data and Observation data can be seen in Figure 3

The MAPE value is obtained is 19.55%, which interprets that estimation data is a good forecasting. Observation data and estimation data of Figure 3 are presented in Table 5.

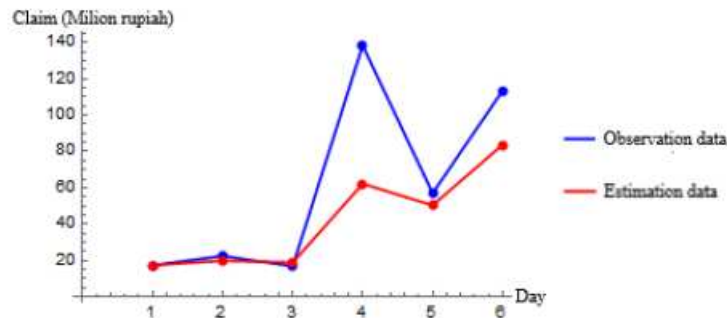


Figure 3: Comparison of observation data and estimation data

Table 5: Comparison of observation data and estimation data for every single day

Day	Observation data (Million rupiah)	Estimation data (Million rupiah)
1	17.300000	17.300000
2	22.500000	19.900000
3	16.719750	18.839918
4	138.443220	62.197811
5	57.004763	50.383258
6	113.626870	83.678178

References

[1] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Series B: Journal of the Royal Statistical Society*, **39**, No. 1 (1977), 1-38.

[2] M. Firmansyah, B. Setiawaty, I.G.P. Purnaba, Parameter estimation of exponential hidden Markov model and convergence of its parameter estimator sequence, *St. Petersburg Polytechnic University Journal* (2017), Submitted.

[3] T.Y. Hu, W.M. Ho, Travel time prediction for urban networks: The comparisons of simulation-based and time-series models, In: *proceedings 17th ITS World Congress*, Busan, South Korea, 2010.

[4] D. Lam, S. Li, D.C. Wunsch, Hidden Markov model with information criteria clustering and extreme learning machine regression for wind forecasting, *Journal of Computer Science and Cybernetics*, **30**, No. 4 (2014), 361-376.

[5] L.R. Rabiner, B.H. Juang, An introduction of hidden Markov models, *IEEE Assp Magazine*, doi: 0740-7467/86/0100-0004 (1986).

- [6] C.F.J. Wu, On the convergence properties of the EM algorithm, *The Annals of Statistics*, **11**, No. 1 (1983), 95-103.