

## 확장된 베이지안 정보기준을 이용한 경기지표의 변동성 분석 연구

전진호  
가톨릭관동대학교 경영학과

### A Study on the Volatility Analysis of Economic Indicators Using Extended Bayesian Information Criteria

Jin-Ho Jeon

Dept. of Business Administration, Catholic Kwan-Dong University

**요약** 우리나라 뿐만 아니라 세계경제는 산업화시대, 정보화시대를 거쳐 4차 산업혁명을 바라보는 현재까지 다양한 시장 친화적인 정책들을 지속적으로 수행하며 새로운 경제질서를 모색하고 있다. 그 결과 경제시장의 규모가 지속적으로 커짐과 동시에 시장 변화에 영향을 미치는 요인도 다양해졌다. 이에 따라, 많은 기업의 의사결정자들에게는 급격하게 동적으로 변화하는 경제시장에서 정확하고 효과적인 의사결정을 위해 경제시장을 정확하게 분석, 예측하는 문제가 의사결정자의 역량으로서 중요한 이슈가 되었다. 본 연구에서는 복잡 다양한 경제시장 환경의 정확한 분석 및 예측을 위한 최적화된 모델링의 결정을 위해 기존의 제한된 정보기준에서 확장된 정보기준을 적용하여 예측 모형의 정확도와 유용성을 향상시키고자 한다. 실험을 통해, 본 연구에서 적용된 확장된 정보기준의 유용성을 검증하기 위해 KOSPI 자료를 대상으로 실험한 본 연구와 기존의 연구를 비교한다. 실험 결과, 확장된 정보기준을 적용하는 것이 기존의 정보기준을 사용하는 것보다 향상된 예측정확도를 보여 유용성을 확인하였다.

**Abstract** The global economy, including Korea, has continuously searched for various market-friendly policies and new economic systems in pursuit of the forth industrial revolution. As a result, economic markets have grown, and factors affecting markets have diversified. Therefore, as for many company's decision makers, it has become an important issue to analyze and forecast markets accurately and effectively for rapid and appropriate decision making. In this study, we aim to improve the accuracy and validity of forecast models by applying extended information criteria in existing restricted information criteria to determine optimized modeling for the accurate analysis and prediction of complex market environments. In order to verify the practical use of the extended information criteria adopted in this study, we compare this study employing KOSPI data with previous studies. Experimental results show that applying extended information criteria is more accurate than using the existing information criteria.

**Keywords** : Bayesian, Information Criteria, Economic Indicators, Volatility, Forecast

### 1. 서론

우리나라뿐만 아니라 세계경제는 산업화시대, 정보화시대를 거쳐 4차 산업혁명을 바라보는 현재까지 다양한

시장 친화적인 정책들을 지속적으로 수행하며 새로운 경제 질서를 모색하고 있다. 그 결과 경제시장의 규모가 지속적으로 커짐과 동시에 시장에 영향을 미치는 변화의 요인도 다양해졌다. 이에 따라, 다양한 경제영역에서 많

본 논문은 가톨릭관동대학교 연구과제로 수행되었음.

\*Corresponding Author : Jin-Ho Jeon(Catholic Kwan-Dong Univ.)

Tel: +82-33-649-7362 email: jhgy@cku.ac.kr

Received February 9, 2017

Revised (1st March 2, 2017, 2nd April 6, 2017)

Accepted April 7, 2017

Published April 30, 2017

은 의사결정자들이 정확한 의사결정을 위해 시장을 정확하게 분석, 예측하는 문제가 중요한 이슈가 되었다.

경제시장의 환경을 나타내는 다양한 지표들이 있으며 이 지표들은 주로 시계열 데이터의 형태로 표현되고 있다. 시계열로 표현되는 경제 환경을 분석, 예측모형에는 여러 가지가 있지만 양적인 예측에 사용되는 모형들은 크게 두 가지로 살펴볼 수 있다[1]. 첫 번째 유형은 과거 자료가 어떤 기간 동안 일정한 경향이나 양상을 보이면 그러한 경향이나 규칙성은 미래에도 계속된다는 가정 아래 분석, 미래 값을 예측하는 모형이며 두 번째 유형은 설명모형으로 예측하고자하는 종속변수가 하나 또는 다수의 독립변수와 원인-결과관계를 가진다는 설정 하에서 이 관계가 미래에도 변하는 않는다는 가정 아래 미래 값을 예측하는 모형이다.

본 연구에서는 복잡 다양한 시장 환경의 정확한 분석 및 예측을 위해 관측되는 시계열 데이터의 최적모델링 결정을 위한 기준들을 살펴보고 실제의 코스피(KOSPI) 데이터를 대상으로 모델링을 통한 분석, 예측실험을 통해 각 기준들의 유용성을 확인하고자 한다.

제2장 관련 연구에서는 경제시장의 상황을 표현하는 다양한 경기지표들의 주요 형태인 시계열 데이터에 대한 분석, 예측을 위한 모형으로 은닉 마코프모델에 대하여 살펴본다. 제3장에서 시계열 데이터의 모델 결정에 가장 효율적인 모델 구조(상태 수)를 결정지를 근사기법을 살펴본다. 제4장에서는 기존에 연구되었던 다양한 근사기법들을 토대로 실제의 코스피(KOSPI) 데이터의 최적 모델링 추정에 적용하여 시계열 데이터의 분석과 예측을 위한 변동성 운동양태의 예측을 통해 유효성을 확인한다. 제5장에서는 연구의 결과와 미진한 부분에 대한 향후 진행되어야 할 연구방향을 제시한다.

## 2. 은닉 마코프모델(Hidden Markov Model)

시계열 데이터의 변동성을 분석, 예측하는 연구는 크게 기본적인, 기술적 분석 영역으로 볼 수 있다. 본 연구에서는 시계열 데이터의 모델링에 적합한 모델기반 방법의 기술적 방법을 통해 유효성을 확인하고자 한다[2, 3, 4]. 다양한 모델기반의 방법 중 은닉 마코프모델은 데이터 내에 내재된 관측이 불가능한 프로세스를 예측하기 위하여 관측이 가능한 다른 프로세스를 통해 변화되는 통계

적인 특징들을 확률적으로 모델링하며 주가, 환율, 음성 인식, 온라인 필기 인식 등의 시계열 데이터 즉, 시간에 따라 변해가는 동적인 특징들로 표현되는 데이터와 데이터의 파라미터의 수가 많을수록 모델링이 잘되는 특징을 가지고 있다[2].

모델  $\lambda$ 는  $M$ 개의 상태들,  $s = (s_1, s_2, \dots, s_M)$ 을 갖으며,  $K$ 시간의 특징들로 묘사되는 3개의 확률집합  $\{S, A, B\}$ 로 표현된다.

$S = \{S_1, \dots, S_N\}$ 는 상태 수가  $N$ 개인 상태 집합을,  $A = [a_{ij}]$ 는 상태전이확률 행렬을,  $B = \{b_i(x)\}$ 는 출력확률 집합으로  $b_i(x)$ 는 다음과 같이 정의된 상태  $S_i$ 에 연관된 확률이다.

$$b_i(x) = \sum_{l=1}^L w_{il} N(x, \mu_{il}, \sum_{il}) \quad (1)$$

$N(x, \mu_{il}, \sum_{il})$ 은 정규분포이고,  $\mu_{il}$ 는  $i$ -번째 상태의  $l$ -번째 믹스처의 평균이고,  $\sum_{il}$ 은 공분산,  $w_{il}$ 은 가중치이다. 각 믹스처에는  $D$ -차원의 특징벡터가 있고, 모든 상태에서는  $L$ 개의 믹스처로 가정한다[5, 12].

이와 같이 3개의 확률 집합으로 정의되며 실제로 모델을 활용하기 위해서는 3가지의 주요 문제를 해결해야 한다. 첫 번째, 확률추정의 문제로서 연속적인 데이터의 특징시점에 특정상태가 발생될 확률로서 식(2), 식(3)을 통해 구할 수 있다[2].

$$\alpha_t(i) = \left( \sum_{j=1}^N \alpha_{t-1}(j) a_{ij} \right) \cdot P_i(O_t) \quad (2)$$

$$\beta_t(i) = \sum_{j=1}^M a_{ij} \cdot P_i(O_t) \cdot \beta_{t+1}(j) \quad (3)$$

두 번째, 관측되어진 연속적인 데이터에서 최적의 상태 순서를 결정하는 문제로서 비터비(Viterbi) 알고리즘을 적용한다[7].

세 번째, 모델의 확률을 최대로 하는 매개변수 추정의 문제는 바움-웰치(Baum-Welch) 알고리즘을 이용하여 새로 추정된 모델과 이전 모델이 관찰 열을 생성하는 확률의 차이가 특정 값 이상이 될 때까지 새로운 모델을 추정해나간다. 모델 매개변수의 갱신은 식(4), 식(5), 식

(6)과 같다[2, 10].

$$a'_{ij} = \frac{\sum_{t=1}^L \alpha_t(i) a_{ij} P_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^L \sum_{j=1}^M \alpha_t(i) a_{ij} P_j(O_{t+1}) \beta_{t+1}(j)} \quad (4)$$

$$\mu'_{ik} = \frac{\sum_{t=1}^L \alpha_t(i) \cdot \beta_t(i) \cdot O_t^k}{\sum_{t=1}^L \alpha_t(i) \beta_t(i)} \quad (5)$$

$$\sigma'_{ik} = \sqrt{\frac{\sum_{t=1}^L \alpha_t(i) \cdot \beta_t(i) \cdot (O_t^k - \mu'_{ik})^2}{\sum_{t=1}^L \alpha_t(i) \beta_t(i)}} \quad (6)$$

### 3. 근사기준을 이용한 은닉 마코프모델 (HMM)의 최적화 결정

시계열 데이터에 대한 은닉 마코프모델(HMM) 구조의 최적화 연구는 다양한 방법으로 진행되어지고 있다. 최적 구조 결정에 있어 모델 감소(reduction) 전략과 모델 확장(expansion) 전략과 같은 방법들이 연구되어 왔다[7]. 하지만 관측된 시계열 데이터가 완전치 못한 경우에는 정확한 해를 얻는 것이 어렵다. 그러므로 다양한 효율적인 근사 추정기법들을 이용한다[2, 6]. 추정하는 모델의 적합도에 대한 비교지표로서 AIC(Akaike Information Criteria), BIC(Bayesian Information Criteria) 등이 있으며 모델의 복잡성과 로그 우도의 최적화에 기반을 둔 추정 방법이다. AIC는 표본의 크기에 의존하여 샘플의 크기가 클 때 좀 더 복잡한 모델을 선호한다. 이러한 이유로 샘플의 수가 클 때 AIC보다 표본의 크기를 반영하는 단순한 모델을 선호하는 베이지안정보기준(BIC) 기준을 적용하여 추정하고자 한다.

#### 3.1 모델 구조(상태 수) 결정을 위한 기준

##### 3.1.1 최대우도(Maximum Likelihood) 기준

우도함수  $L$ 은  $L(a|X) = f(X|a)$ ,  $a \in A$  and  $X \in S$  으로 정의하며  $L(a|X)$ 의 값이 최대가 되는 파라미터  $a$ 의 평균값  $\mu(x)$ 를 구하는 것이 최대 우도 기준이다.  $\mu(x)$ 는 파라미터  $a$ 의 최대 우도 추정이 된다.  $L(a|X)$ 의 최대 값은, 자연로그함수  $\ln$ 이 증가할 때 이 값이 존재한다면

$\ln[L(a|X)]$ 의 최대 값이다.  $\ln[L(a|X)]$ 함수는 로그 우도 함수이며 많은 경우에 우도 함수보다 더 쉽게 연산 할 수 있기 때문에 많이 쓰인다. 이 방법은 학습에 사용되어 질 관측된 데이터의 양이 충분한 경우에 높은 추정 성능을 보여준다[5].

##### 3.1.2 베이지안 정보기준(Bayesian Information Criteria : BIC)

라플라스 근사로부터 유도되며 많은 양의 데이터에 대하여,  $P(\theta|X, M) \propto P(X|\theta, M) P(\theta|M)$ 는 다변량 가우시안 분포로서 근사되어진다.

$P(X|\theta, M) P(\theta|M) \equiv e^{g(\theta)}$ 로 정의하면,

$$g(\theta) = \log(P(X|\theta, M) P(\theta|M)), \quad (7)$$

$g(\theta)$ 를 최대화시키는  $\theta$ 를  $\hat{\theta}$ 로 정의하면, 이 파라미터의 조합은  $P(\theta|X, M)$ 를 최대화시키므로 최대사후확률(MAP)이라 부른다.  $g(\theta)$ 에 대해 2차 테일러 다항식으로 근사시키면  $\hat{\theta}$ 는 식(8)와 같이 성립된다.

$$g(\theta) \approx g(\hat{\theta}) - \frac{1}{2} (\theta - \hat{\theta})^T A (\theta - \hat{\theta}), \quad (8)$$

여기서,  $(\theta - \hat{\theta})^T$ 는 행벡터  $(\theta - \hat{\theta})$ 의 트랜스포즈이고  $A$ 는  $\hat{\theta}$  조합에서  $g(\theta)$ 에 대한 음의 Hessian이다.  $e^{g(\theta)}$ 과 식(7)를 이용하면 다음의 식(9)을 얻는다.

$$\begin{aligned} P(X|\theta, M) P(\theta|M) &= e^{g(\theta)} \\ &\approx e^{\left\{ g(\hat{\theta}) - \frac{1}{2} (\theta - \hat{\theta})^T A (\theta - \hat{\theta}) \right\}} \\ &= e^{g(\hat{\theta})} e^{-\frac{1}{2} (\theta - \hat{\theta})^T A (\theta - \hat{\theta})} \\ &= P(X|\hat{\theta}, M) P(\hat{\theta}|M) e^{-\frac{1}{2} (\theta - \hat{\theta})^T A (\theta - \hat{\theta})} \end{aligned} \quad (9)$$

식(9)를  $\Theta$ 에 대하여 적분하여 식(10)으로 전개한다.

$$\begin{aligned} \int P(X|\theta, M) P(\theta|M) d\theta &= \int P(X, \theta|M) d\theta = P(X|M) \\ &= \int P(X|\hat{\theta}, M) P(\hat{\theta}|M) e^{-\frac{1}{2} (\theta - \hat{\theta})^T A (\theta - \hat{\theta})} d\theta \\ &= P(X|\hat{\theta}, M) P(\hat{\theta}|M) \cdot (2\pi)^{\frac{d}{2}} / \sqrt{|A|} \end{aligned}$$

$$P(X | M) = P(X | \hat{\theta}, M) P(\hat{\theta} | M) \cdot (2\pi)^{\frac{d}{2}} / \sqrt{|A|} \quad (10)$$

식(10)에 로그를 취하여 식(11)로 전개한다.

$$\log P(X|M) \approx \log P(X|\hat{\theta}, M) + \log P(\hat{\theta}|M) + \frac{d}{2} \log(2\pi) - \frac{1}{2} \log |A| \quad (11)$$

$$\log P(M | X) \approx \log P(X | M, \hat{\theta}) - \frac{d}{2} \log N \quad (12)$$

식(12)에서  $d$ 는 파라미터의 수를  $N$ 은 데이터의 수를  $\hat{\theta}$ 는 모델  $M$ 의 한계우도(ML)의 파라미터 구성이다. 식(12)에서 첫 번째 항은 데이터를 가장 잘 설명하는 항이며 데이터를 잘 설명할수록, 즉 파라미터가 많을수록 큰 값을 갖게 되는 우도 값이다. 두 번째 항은 파라미터의 수를 적게 하여 모델을 일반화시키는 항으로 파라미터의 수를 적게 하여 한계우도의 값을 크게 한다[8, 15]. 하지만 사전확률은 자유 파라미터 개수에만 제한되고, 모델을 정의하는 각각의 파라미터 유형에 따라 별도의 고려를 하지 않는다, 은닉 마코프모델에는 동질적이지 못한 파라미터 집합이 존재하기 때문에 이러한 제한점은 부족하다[8, 9, 11, 13].

### 3.1.3 은닉 마코프모델 특성을 이용한 확장된 베이지안 정보기준(HMM-BIC)

은닉 마코프모델 구조에 적합한 선택 기준을 유도하기 위해 베이지안 정보기준으로부터 시작하여 모델구조에 대한 사전확률  $P(M)$ 을 통해 다음 식(13)을 정의할 수 있다.

$$\begin{aligned} EBIC(M) &= \log P(X|\hat{\theta}, M) + \log P(\hat{\theta}|M) \\ &\quad + \frac{d}{2} \log(2\pi) - \frac{d}{2} \log N + \frac{1}{2} \log P(M) \\ &= BIC(M) + \log P(M) + \log P(\hat{\theta}|M) + \frac{d}{2} (2\pi) \end{aligned} \quad (13)$$

확장된 은닉 마코프모델 - 베이지안 정보기준은 베이지안 정보기준 선택기준에다 은닉 마코프모델의 구조를 반영하는데 보다 적합한 항으로 구성된다.  $P(M)$ 을 계산하는데 은닉 마코프모델 구조를 이용함으로써 확장된 은닉 마코프 - 베이지안 정보기준(HMM-BIC)은 모델

구조 추정에 보다 적합한 선택기준이 될 수 있다[5].

## 4. 실험

경기지표의 변동성 분석을 위한 모형 추정을 위한 실험으로 실제 경기지표의 하나인 코스피(KOSPI)의 2016년 데이터를 선정하여 기준에 주로 음성인식과 온라인 필기체 인식에 적용하였던 것과 차별점을 갖고자 한다. 실험은 관측된 코스피 데이터를 통해 모델 구조(상태 수) 추정에 대한 실험과 추정된 구조(상태 수)를 통한 실제 변동성 양태와 예측된 변동성 양태를 비교함으로써 모형결정에 대한 기준들의 유용성을 확인하고자 한다. 모델 구조(상태 수) 추정을 위해서 은닉 마코프 툴킷 (Hidden Markov Model Tool Kit : HTK)을 사용하였다 [14].

### 4.1 자료길이에 따른 기준별 구조(상태 수) 추정

시계열 데이터의 변동성 분석을 위한 기술적 모형의 구조(상태 수)를 제시된 기준별로 유효성을 확인하기 위하여 코스피의 2016년 1월 4일부터 5월 31일까지 100일(20주) 데이터를 20일, 40일, 60일, 80일, 100일의 데이터 길이 구간을 통해 모델의 구조(상태 수)를 추정해보았다.

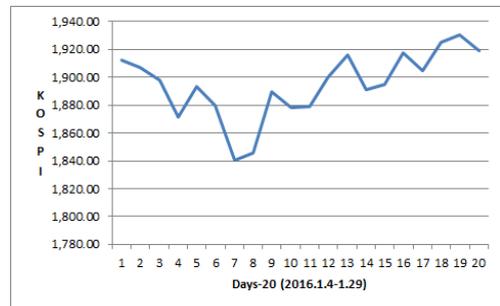


Fig. 1. Number of State nested in the 20 Days' Data

Fig. 1은 2016년 1월 4일부터 1월 29일까지 데이터 길이 20일의 운동양태를 보여주고 있으며 상태1(1일차, 19일차), 상태2(7일차), 상태3(4일차, 10, 11일차), 상태4(13일, 16일차), 상태5(5일, 9일, 14일차)처럼 5개의 상태를 내포하고 있는 것을 확인할 수 있다.

아래 Fig. 2는 2016년 1월 4일부터 3월 3일까지 데이터 길이 40일의 운동양태를 보여주고 있으며 상태1(1일

차), 상태2(3일, 7일, 20일, 39일차), 상태3(11일차, 18일차, 29일차, 34일차), 상태4(15일차, 27일차)처럼 4개의 상태를 내포하고 있는 것을 확인할 수 있다.

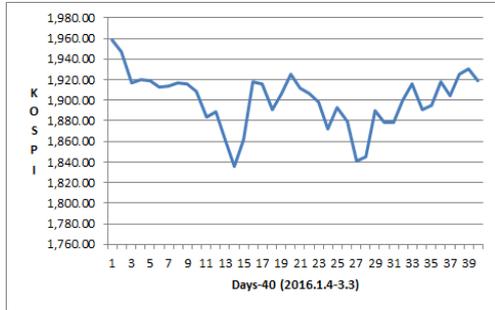


Fig. 2. Number of State nested in the 40 Days' Data

위의 같이 2016년 1월 4일부터 3월 31일까지 데이터 길이 60일의 운동양태에서는 5개의 상태, 4월 29일까지 데이터 길이 80일의 운동양태에서는 5개의 상태, 5월31일까지 데이터 길이 1000일의 운동양태에서는 5개의 상태를 내포하고 있다.

Table 1. State Estimation by the Data Length

Days Criterion	20 D	40 D	60 D	80 D	100D
ML	5(3)	4(3)	5(4)	5(5)	5(5)
BIC	5(4)	4(3)	5(5)	5(5)	5(5)
HMM-BIC	5(4)	4(4)	5(5)	5(5)	5(5)

Table 1은 데이터 길이에 따른 기준별 상태 수의 추정결과를 보여주고 있다. 20일의 데이터 길이에서 내재된 5개의 상태를 최대우도(ML) 기준, 베이지안 정보기준(BIC) 그리고 은닉 마코프모델 - 베이지안 정보기준(HMM-BIC) 모두 잘못된 상태 수를 추정하였으며 이유는 짧은 데이터 길이에서는 운동양태의 변동 규칙성을 정확히 반영하기 어렵기 때문으로 생각된다. 최대우도(ML) 기준에서는 데이터 길이 80일과 100일에서, 베이지안 정보기준(BIC)에서는 60일, 80일, 100일에서, 은닉 마코프모델 - 베이지안 정보기준(HMM-BIC)에서는 40일부터 정확히 추정하였다. 위의 추정 결과를 보면 은닉 마코프모델 - 베이지안 정보기준이 데이터 길이가 상대적으로 짧더라도 다른 두 기준들보다 정확한 상태 수 추정함을 보이는 것으로 계산비용을 줄임으로서 그 효용성이 확인되었다.

#### 4.2 기준별 구조(상태 수) 추정을 통한 변동성 분석

Fig. 3, 4, 5, 6은 2016년 1월 4일부터 시작하여 40일, 60일, 80일, 100일의 데이터 길이로 최대우도(ML) 기준, 베이지안 정보기준, 은닉 마코프모델 - 베이지안 정보기준의 각 기준별 모델을 생성한 후 각 모델별로 향후 10일을 예측한 변동성 그래프를 보여주고 있다.

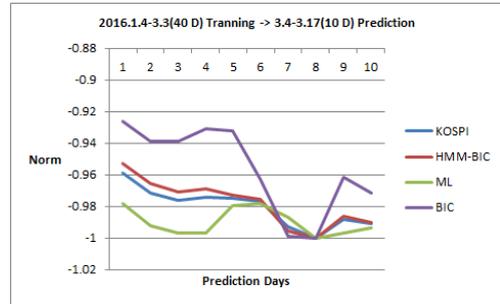


Fig. 3. Prediction Patterns by 40 Day' Data Length

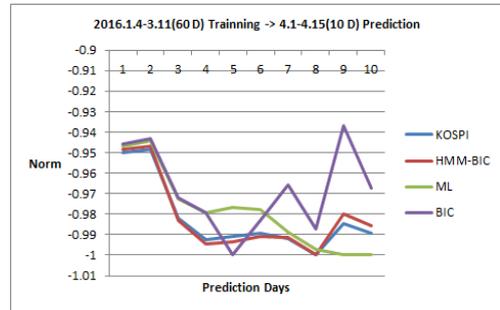


Fig. 4. Prediction Patterns by 60 Day' Data Length

Fig. 3은 40일의 데이터 길이로 향후 10일을 예측한 그래프로서 실제의 운동양태는 1일부터 5일까지 적은 폭으로 하락 후 상승을 6일부터 10일까지 큰 폭의 하락 후 상승을 보여주고 있다. 각 기준별 예측된 변동성을 보면 유사한 패턴을 보여주고 있으며 특히 은닉 마코프모델 - 베이지안 정보기준(HMM-BIC)을 통한 예측 변동성은 실제의 운동양태와 거의 일치하는 것을 보여주고 있다.

Fig. 4는 60일의 데이터 길이로 향후 10일을 예측한 변동성 그래프로서 1일부터 5일까지 급격한 하락 후 점진적 상승과 하락을 반복하는 운동양태를 보여주고 있으며, Fig 3에서와 마찬가지로 실제의 운동양태(KOSPI-블루)와 은닉 마코프모델 - 베이지안 정보기준(HMM-BIC,

레드)의 변동성 양태가 거의 일치하는 것을 보여주고 있다.

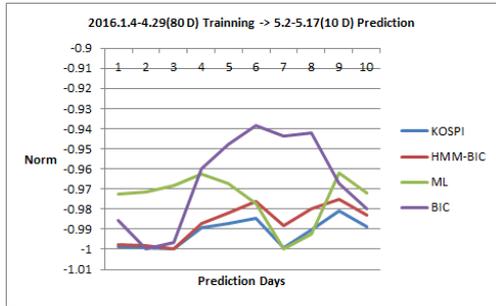


Fig. 5. Prediction Patterns by 80 Day' Data Length

Fig. 5는 80일의 데이터 길이로 향후 10일을 예측한 변동성 그래프로서 모든 곡선의 상승과 하락의 변곡점의 변동양태가 유사한 패턴을 보이고 있지만 최대우도(ML) 기준을 통한 예측의 운동양태는 다른 두 기준과 변곡점의 위치 일수가 차이가 있음을 볼 수 있다.

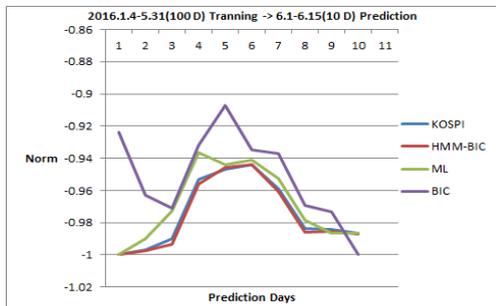


Fig. 6. Prediction Patterns by 100 Day' Data Length

Fig. 6도 Fig. 5와 마찬가지로 최대우도(ML) 기준을 통한 예측의 운동양태는 다른 두 기준과 변곡점의 위치 일수가 차이가 있음을 볼 수 있다.

아래의 Table 2는 각 기준별로 데이터 길이에 따라 모델을 추정하여 30회씩 예측하여 실제 코스피 지수와 기준별 예측된 변동성 패턴을 평균제곱오차를 기준으로 유사성을 확인한 비율을 보여주고 있다. 세 기준 모두에서 데이터 길이가 40일과 100일보다 상대적으로 60일과 80일에서 유사한 변동성 패턴을 나타내고 있다. 40일과 100일에서 상대적으로 낮게 나타난 이유는 데이터의 길이가 짧을 경우에는 데이터 내에 내포된 규칙성을 반영하기 어렵고 또한 데이터 길이가 긴 경우에는 불규칙적

이고 다양한 예외적인 변수들이 많이 반영되어 내재된 결과로 생각된다[1].

Table 2. Prediction Rate for the Similar Pattern by the Data Length

Days Criterion	40 Days	60 Days	80 Days	100 Days
ML	30(18)= 0.6	30(19)= 0.63	30(20)= 0.66	30(19)= 0.63
BIC	30(19)= 0.63	30(20)= 0.66	30(21)= 0.7	30(20)= 0.66
HMM-BIC	30(22)= 0.73	30(22)= 0.73	30(23)= 0.76	30(22)= 0.73

이렇게 Fig. 3, 4, 5, 6과 Table 2의 결과를 보면 본 연구에서 적용하는 은닉 마코프모델 - 베이지안 정보기준(HMM-BIC)을 통한 변동성 예측이 다른 두 기준인 베이지안 정보기준(BIC)과 최대우도(ML) 기준을 적용하였을 때보다 적은 수의 파라미터를 사용해도 상대적으로 유사한 변동성 패턴과 예측률을 보여주고 있다.

## 5. 결론

본 연구에서는 날로 예측 불가능해지는 시장 환경에서 정확한 의사결정을 위하여 경제영역의 시계열 데이터를 통해 최적화된 모델링 결정을 위한 기준들을 살펴보고 또한 실제의 코스피(KOSPI) 데이터를 대상으로 기준별로 분석, 예측을 위한 변동성 패턴을 확인하였다. 실험 결과, 본 연구와 많은 연구에서 제안된 은닉 마코프모델 - 베이지안 정보기준(HMM-BIC)을 통한 모델의 구조 추정과 변동성 예측이 실제의 코스피 데이터의 적용에서도 다른 두 기준인 베이지안 정보기준(BIC)과 최대우도(ML) 기준을 적용하였을 때보다 상대적으로 유사한 변동성 패턴과 예측률을 보여주는 것을 확인하였다. 하지만 단기간의 학습 데이터를 통해 단기 예측만을 살펴본 점과 복잡한 경제 환경에 더욱 민감한 중, 소형주에도 일반화 적용하기에는 불안정한 알고리즘으로 생각이 된다. 향후 연구에서는 장기 예측뿐만 아니라 업종별, 기업별, 중, 소형주 등에서도 적용 가능한 모델링 영역을 확대하여 일반화가 가능하다면 해당 영역의 주어진 문제에 대하여 최적의 의사결정을 할 수 있을 것으로 기대되며 연구영역의 확대가 필요할 것이다.

## References

- [1] Y. Kim., "Time Series Analysis and Prediction", *Free Academy*, 1990.
- [2] J. Jeon, m. Kim., "A Study of Economic Indicator Prediction Model using Dimensions Decrease Techniques and HMM", *The Journal of Digital Policy & Management*, vol. 11, no 10, pp. 305-311, 2013.
- [3] S. Shin., "Evaluation Exchange Rate of Artificial Neural Network and Moving Average Method", *Finance Research*, vol. 9, no. 1, pp. 103-135, 1995.
- [4] J. Jeon, m. Kim., "A Study on Prediction the Movement Pattern of Time Series Data using Information Criterion and Effective Data Length", *The Institute of Webcasting, Internet and Telecommunication*, vol. 13, no. 1, pp. 101-107, 2013.  
DOI: <http://doi.org/10.7236/JIIBC.2013.13.1.101>
- [5] M. Park, J. Ha, "HMM Topology Optimization using HBIC and BIC\_Anti Criterion", *The Journal of KIUSE*, vol. 30, no 9, pp. 867-875, 2003.
- [6] J. Jeon, M. Kim, "Determination of Pattern Models using a Conversions of Time-Series Data Conversion Technique for the Prediction of Financial Markets", *The Journal of Digital Convergence*, vol. 13, no. 5, pp. 237-244, May, 2015.  
DOI: <https://doi.org/10.14400/JDC.2015.13.5.237>
- [7] J. Jeon., "A Study on Determining Prediction Models using Model-based Clustering of Time Series Data", *Dankook Univ Ph. D*, 2007.
- [8] A. Sorjamaa, et al., "Methodology for Long-Term Prediction of Time Series", *Neurocomputing*, pp. 178-186. Elsevier, 2007.
- [9] P. Cheeseman, J. Stutz, "Bayesian Classification", *Kluwer Academic Publishers*, vol. 70. pp. 117-126, 1996.
- [10] L. Rabiner., "A Tutorial on Hidden Markov Models and selected applications in speech recognition", *Proc. of IEEE77*, pp. 257-286, 1989.  
DOI: <https://doi.org/10.1109/5.18626>
- [11] D. Heckerman, D. Geiger, D. Chekering, "A tutorial on learning with bayesian networks", *machine Learning 20*, pp. 197-243, 1995.
- [12] J. Ha, B. Alain, S. Jayasheree, "Use of Model Prior for HMM Topology Optimization", *The 4-th Korea-China Joint Symposium on Information Technology for Oriental Language Processing and Pattern Recognition*, Nov, 2001.
- [13] D. Li, A. Biem, J. Subrahmonia, "HMM Topology Optimization for Handwriting Recognition", *ICASSP*, 2001.
- [14] <http://htk.eng.cam.ac.uk/>
- [15] Y. Jung, J. Jeon, "A Fusion of the Period Characterized and Hirarchical Bayesian Techniques for Efficient Cluster Analysis of Time Series Data", *The Journal of Digital Convergence*, vol. 13, no. 7, pp. 169-175, May, 2015.  
DOI: <https://doi.org/10.14400/JDC.2015.13.7.169>

전진호(Jin-Ho Jeon)

[정회원]



- 1998년 8월 : 명지대학교 대학원 경영정보학과 (경영학석사)
- 2007년 2월 : 단국대학교 대학원 컴퓨터과학 (이학박사)
- 2009년 9월 ~ 현재 : 가톨릭관동대학교 경영학과 교수

<관심분야>

지능형시스템, 경영정보시스템, IT정책 및 전략