

# 금융시장 예측을 위한 시계열자료의 변환기법 융합을 이용한 패턴 모델 결정

전진호\*, 김민수\*\*  
가톨릭관동대학교 경영학과\*, 가톨릭관동대학교 무역학과\*\*

## Determination of Pattern Models using a Convergence of Time-Series Data Conversion Technique for the Prediction of Financial Markets

Jin-Ho Jeon\*, Min-Soo Kim\*\*

Dept. of Business Administration, Catholic Kwan-Dong University\*

Dept. of International Trade, Catholic Kwan-Dong University\*\*

**요 약** 수출주도정책, FTA 체결 및 규제개선 등과 같은 다양한 시장지향적인 정책을 통해 경제시장의 규모가 지속적으로 커졌다. 이에 따라 올바른 의사결정을 위하여 경제시장을 정확하게 분석, 예측하는 문제가 중요한 이슈가 되었다. 경제시장을 표현하는 여러 지표 중 가장 대표적인 주식지표의 정확한 분석 및 의사결정을 위하여 시계열자료의 모델링에 적합한 은닉마아코프모델을 토대로 자료 내에 내재된 예외적인 특징과 잡음을 제거하기 위한 변환기법의 융합모델을 제안하여 모델 추정과 예측 문제에 적용하였으며 그 유효성을 확인하였다. 실험 결과를 통해, 본 연구에서 제안하는 변환조합을 적용하는 모델추정 기법이 유효한 모델 상태 추정 결과를 보여주었으며 실제 코스피지수와 예측의 문제에서도 매우 유사한 운동양태를 확인할 수 있었다.

**주제어** : 금융, 예측, 시계열, 패턴, 융합

**Abstract** Export-led policies, FTA signed and economics of scale through a variety of market-oriented policies, such as regulations to improve market grew constantly. Accordingly, the correct decision making accurately analyze the economics market for decision, a problem has been an important issue in predicting. For accurate analysis and decision-making of the most common indicators of the stock market by proposing a number of indicators of economic transformation techniques were applied to the convergence model combining estimation and forecasts problem confirmed its effectiveness. Experimental result, gave the model estimation method to apply a transform to show the valid combinations proposed model state estimation result was confirmed in a very similar exercise aspect of the physical problem and the KOSPI index prediction.

**Key Words** : Finance, Prediction, Time Series, Pattern, Convergence

Received 16 March 2015, Revised 20 April 2015

Accepted 20 May 2015

Corresponding Author: Min-Soo Kim  
(Catholic Kwan-Dong University)

Email: jhhy@cku.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서론

우리나라 경제시장의 규모는 1960년대 초반부터 지속적으로 이어진 수출주도정책 및 현재 여러 나라들과 체결된 자유무역협정(FTA)과 이를 효과적으로 지원하기 위한 규제개선 등과 같은 다양한 시장지향적인 정책들에 힘입어 그동안 꾸준히 발전해왔다. 사회의 모든 구성원들이 경제문제와 매우 중요한 관계를 맺고 있기 때문에 경제시장의 규모가 커짐에 따라 올바른 의사결정을 위하여 경제시장을 올바르게 분석, 이해 및 예측하는 것은 모든 구성원들에게 중요한 이슈가 되었다. 이에 따라 경제시장을 정확히 분석, 이해 및 올바른 의사결정을 내리 위해서는 경제현상을 잘 반영하는 경제지표들을 통해 경제시장의 정확한 분석과정이 이루어져야 할 것이다. 경제시장을 반영하는 지표로 사용하는 것을 경기지표라고 한다. 경기지표는 주가지수, 환율, 물가지수, 무역 및 국제수지 지표 등이 있다. 이러한 다양한 지표들 중 현대 경제시스템을 대표하는 주식지표는 점차 주식시장의 규모가 급격하게 팽창하고 있으며, 다양한 거시경제 지표들과의 상관관계가 밀접해짐에 따라 실물경제 전체를 표현한다고 하여도 큰 무리는 아닐 것이다[1].

본 연구에서는 금융시장의 다양한 여러 지표 중 코스피(KOSPI) 지수를 통해 모델을 추정하고 그 유효성을 확인하고자 한다. 경제시장을 함축적으로 표현하는 코스피 지수는 금융시장에 영향을 미치는 다양한 변수와 예외적인 상황과 같은 불확실한 요소들이 시시각각 발생되어 반영된다. 이러한 복잡한 경제, 금융시장을 최적으로 설명할 수 있는 모델을 추정하기 위하여 아무리 복잡한 현상을 나타내는 자료라도 최소의 자료로서 설명이 가능하다는 오컴의 면도날(Occam's Razor) 원리처럼 시계열자료들의 모델링에 적합한 은닉마야코프모델과 이를 토대로 시계열자료에 내재된 예외적인 특징과 잡음을 제거하기 위한 변환조합 기법을 제안하여 모델을 추정하고 예측문제에 적용하여 모델의 유효성을 확인한다.

본 연구는 제2장 관련연구에서 시계열자료의 분석, 예측을 위한 모델 결정 방법론들을 살펴본다. 제3장에서 시계열자료의 모델추정에 가장 효율적인 은닉마야코프모델을 토대로 자료 내에 내재된 예외적인 특징과 잡음을 제거하기 위한 변환기법으로서 정규화 기법과 이동평균 기법을 설명한다. 제4장에서는 실제 코스피지수를 통해

제안하는 모델추정 방법에 대한 상태 수 추정과 실제와 예측된 운동양태의 유사도 비율을 확인함으로써 제안하는 모델추정 방법론의 유효성을 확인한다. 제5장에서는 연구의 결과와 향후 연구방향을 제시한다.

## 2. 관련 연구

주가지수, 물가지수, 국제수지 지표, 환율 등과 같이 시간의 흐름에 따라 동적인 특징들로 표현되는 시계열자료의 연구영역은 크게 기본적 분석과 기술적 분석의 연구영역으로 구분된다.

기본적 분석은 금융자산의 경제적 가치를 평가하는 것이다. 이전의 많은 연구에서는 금융자산의 가치가 랜덤월을 따른다고 하였다. 하지만 근래의 많은 연구에서는 일부 금융자산 가치가 랜덤월을 따르지 않는다는 많은 연구결과들이 입증되었다[2,3].

기술적 분석방법은 시계열자료의 각 지표에 내재된 특징들을 통해 과거의 운동양태가 미래의 운동양태를 예측하는데 의미 있는 정보를 제공한다는 가정에 바탕을 두고 있다. 이에 따라 본 연구에서는 기술적 분석방법에서 시계열자료의 모델링에 적합한 모델기반의 방법을 통해 유효성을 확인하고자 한다. 자기회귀모델, 이동평균모델, 신경망[4], ARMA(ARIMA)[5], 마야코프모델, 은닉마야코프모델[6,9] 그리고 상호정보[7]를 이용한 방법 등이 있다[1,3].

자기회귀모델과 이동평균모델은 자료들의 시점 사이의 관계를 추정하는 기법이다. 기본 개념은 시계열자료의 한 점에서의 값은 그 시점 이전과 이후의 값과 밀접한 관계를 갖는다는 사실에 근거를 둔다. 이 방법은 자연계나 사회적 현상의 다양한 형태를 모델링하고 예측하는데 자주 사용된다[1,3,8]. 인공신경망은 모델의 구조가 알려져 있다는 것과 모델의 해석이 어렵다는 점으로 일반적인 시계열자료의 모델링에는 유효하지 않다[1,3,4]. ARMA 모델은 계절성, 비정지성 등 다른 요소에 의해 예측에 한정한 역할을 하게 되는 단점이 있다[1,3,5]. 상호정보를 포함한 다른 통계적 방법도 항상 고도의 기술적 제한 조건 및 적용 가능한 환경에 대한 조건을 제시하지 않을 경우 모델링이 매우 제한적이다[1,3,7,11]. 마야코프모델은 계산이 효율적이며 방향성 그래프로 표현되어 이해와 해

석이 쉽다. 하지만 이산적인 시계열자료의 모델링은 적합하고 연속적인 자료에는 적합하지 않다는 제약점을 가지고 있다[1,3,6]. 은닉마아코프모델은 상태들의 확률함수로 이루어져 연속적인 시계열자료의 모델링에 적합하다. 이유는 각 상태에서 특징들에 대한 적합한 확률함수를 사용하여 연속적인 값을 갖는 시간적 특징을 갖는 데이터를 쉽게 처리하며 데이터의 묘사가 쉽기 때문이다 [1,3,5,8].

### 3. 시계열자료의 변환기법을 이용한

#### 모델결정

시계열자료에 대한 계산비용 감소와 자료 내에 내포되어 있는 예외적인 특징들과 잡음들을 제거하고 효율적인 모델결정을 위하여 시계열자료들에 대하여 정규화 과정과 이동평균 변화를 적용하여 변환 후 은닉마아코프 모델링을 통해 결정하는 기법을 제안한다[13].

#### 3.1 시계열자료에 대한 변환기법 적용

##### 3.1.1 정규화(Normalization) 변환

정규화 변환을 통하여 주어진 시계열자료가 갖는 요소 값의 절대적인 크기를 무시할 수 있다. 따라서 정규화 변환은 요소 값의 크기는 서로 다르지만 변화하는 패턴이 유사한 시계열자료들을 파악하는데 매우 효율적이다.

시계열자료  $S = (s[i])(0 \leq i < Len(S))$ 를 정규화변환한 시계열자료  $Norm(S)$ 는 식(1)처럼 정의할 수 있다[11,19].

$$s'[i] = \frac{s[i] - \frac{Max(S) + Min(S)}{2}}{\frac{Max(S) - Min(S)}{2}} \quad (1)$$

##### 3.1.2 이동평균(Moving Average)변환

이동평균 변환은 시계열자료의 연속되는  $k$ 개의 요소 값들의 평균값들을 순차적으로 나열하는 변환이다. 이동평균 변환을 통하여 시계열자료에 내포되어 있는 잡음(Noise)의 영향을 제거하여 최소화할 수 있다. 따라서 이동평균 변환은 잡음의 영향 없이 전체적인 변화 경향이 유사한 시계열자료들을 파악하는데 매우 유용하다. 이동평균 계수  $k$ 는 잡음의 영향을 줄이고자 하는 정도에 따

라 적절하게 선택된다. 이동평균 계수가 커짐에 따라 시계열자료 내의 잡음의 영향은 감소한다[2,11,16]. 시계열자료  $S = (s[i])(0 \leq i < Len(S))$ 를 이동평균계수  $k(1 \leq k < Len(S))$ 로 이동평균 변환한 시계열자료  $MV_k(S)$ 는 다음 식(2)과 같이 정의된다[[17,18,20].

$$s_k[j] = \frac{1}{k} + (s[j] + s[j+1] + \dots + s[j+k-1]) \\ = \frac{1}{k} \times \sum_{i=j}^{j+k-1} s[i] \quad (2)$$

#### 3.2 시계열자료에 대한 모델링

은닉마아코프모델은 시계열자료 내에 내포된 은닉상태를 예측하기 위하여 실제적인 관측을 통해서 변화되는 특징들의 확률적 모델링에 적합한 모델이다.

시계열자료의 내포되어 있는 특징들에 대응하는 상태들에 대하여 초기상태확률을 나타내는 하나의 벡터, 상태전이를 나타내는 행렬과 방출확률을 나타내는 행렬로 구성되는 모델로서 확률추정 문제, 최적상태 순서 결정 문제, 모델의 확률을 최대로 하는 매개변수의 추정문제를 다룬다. 첫 째, 특정시점에 특정상태가 발생할 확률추정의 문제는 식(3),(4)를 통해 구할 수 있다[9].

$$\alpha_t(i) = \left( \sum_{j=1}^N \alpha_{t-1}(j) a_{ij} \right) \cdot P_i(O_t) \quad (3)$$

$$\beta_t(i) = \sum_{j=1}^M a_{ij} \cdot P_i(O_t) \cdot \beta_{t+1}(j) \quad (4)$$

두 번째, 주어진 시계열 관측 열에 대한 대응하는 최적의 상태 열을 결정하는 방법에는 동적프로그래밍 기법 중 하나인 비터비(Viterbi) 알고리즘을 적용한다. 세 번째, 모델의 확률을 최대로 하는 매개변수의 추정문제는 새로 추정된 모델과 이전 모델이 관찰 열을 생성하는 확률의 차이가 특정 값 이상이 될 때까지 새로운 모델을 추정해나간다. 모델 매개변수의 갱신은 식(5),식(6),식(7)과 같다.

$$a'_{ij} = \frac{\sum_{t=1}^L \alpha_t(i) a_{ij} P_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{t=1}^L \sum_{j=1}^M \alpha_t(i) a_{ij} P_j(O_{t+1}) \beta_{t+1}(j)} \quad (5)$$

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

$$\sigma'_{ik} = \sqrt{\frac{\sum_{t=1}^L \alpha_t(i) \cdot \beta_t(i) \cdot (O_t^k - \mu'_{ik})^2}{\sum_{t=1}^L \alpha_t(i) \beta_t(i)}} \quad (7)$$

시계열자료들에 대하여 잘 분석하고 설명할 수 있는 모델의 결정 시 모델 상태 수 결정은 주어진 시계열자료가 불완전한 경우에는 정확한 해를 얻는 것은 어렵다. 그러므로 정확한 결과 값에 수렴하지만 계산비용이 상대적으로 큰 단점이 있는 근사기법들을 적용한다. 위와 같은 단점을 극복하기 위하여 본 연구에서는 최우도 원리의 확장에 바탕을 두고 있는 정보기준(Information Criterion)을 적용한다. 이 정보기준은 정확성은 상대적으로 정확한 값에 미치지 못하지만 계산비용을 많이 줄이는 효율적인 기법이다[11]. 다음 식(8),(9)과 같다.

$$\log P(X|M) \approx \log P(X|\hat{\theta}, M) + \log P(\hat{\theta}|M) + \frac{d}{2} \log(2\pi) - \frac{1}{2} \log |A| \quad (8)$$

$$\log P(M|X) \approx \log P(X|M, \hat{\theta}) - \frac{d}{2} \log N \quad (9)$$

식(8),(9)에서  $d$ 는 모델구성의 매개변수 수,  $N$ 은 자료들의 수,  $\hat{\theta}$ 는 모델에 대한 한계우도의 매개변수이다. 식(9)에서 첫 번째 항은 시계열자료를 잘 설명할 수 있는 모델을 선호하는 우도(Likelihood)항이며 두 번째 항은 매개변수의 수를 작게 하여 모델을 일반화시키는 항이다. 다시 식(10)으로 표현할 수 있다.

$$IC = -2 \ln(\text{최우도}) + (\text{매개변수수}) \cdot \ln(n) \quad (10)$$

위의 식을 통해 모델의 상태 수가 증가할수록 우도값이 증가하지만 반대로 상태의 수 증가에 따라 계산비용 또한 증가한다. 두 항의 상쇄를 통해 결정되는 종료기준을 갖는 순차적 탐색 기준을 통해 가장 높은 기준 값을 갖는 상태 수가 모델의 유효한 상태수를 나타내는 것이다[3,14,15].

## 4. 실험

시계열자료의 패턴모델 추정을 위한 자료로 실제의 경제지표 자료 중 하나인 2012년, 2013년 그리고 2014년 3개년의 코스피(KOSPI) 지수를 선정하였다.

실험은 수집된 코스피 자료를 통해 시계열자료의 특징을 결정짓는 상태 수 추정하고 추정된 상태 수를 통해 모델을 결정하고 실제의 자료와 모델을 통한 미래예측 결과의 유사성 측정을 통해 제안하는 모델의 유효성을 확인한다. 모델의 상태 수 추정과 모델결정은 은닉마아코프모델, 정보기준을 이용한 은닉마아코프모델 그리고 변환조합과 정보기준을 이용한 은닉마아코프모델 이렇게 세 가지 기법을 이용하여 추정하였다.

### 4.1 표본 길이에 따른 모델 상태 수 추정

주어진 코스피 지수 자료들에 대해 마아코프모델, 정보기준을 이용한 마아코프모델 그리고 제안하는 변환조합을 이용한 자료에 대한 모델 상태 수와 모델 결정 추정의 유효성을 확인하고자 한다. 정보기준의 유효성은 이전의 여러 연구들을 통해 확인할 수 있다[3,10].

2013년 1월2일부터 10주(50일)의 표본길이를 통해 10일에서 50일까지 10일 구간을 설정하여 자료 내에 내재된 특징의 상태 수를 추정하였다.

<Table 1> State Number Estimation by Sample Size

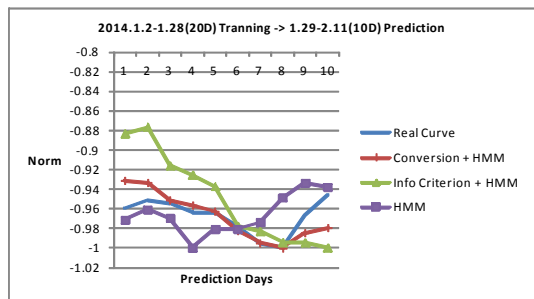
| Length<br>Model          | 10day<br>(2 W) | 20day<br>(4 W) | 30day<br>(6 W) | 40day<br>(8 W) | 50day<br>(10 W) |
|--------------------------|----------------|----------------|----------------|----------------|-----------------|
| HMM                      | 8(7)           | 7(5)           | 4(4)           | 4(4)           | 4(4)            |
| Infor Criterion<br>+ HMM | 7(6)           | 6(5)           | 4(4)           | 4(4)           | 4(4)            |
| Conversion<br>+ HMM      | 6(5)           | 4(4)           | 4(4)           | 4(4)           | 4(4)            |

<Table 1>은 은닉마아코프모델, 정보기준을 이용한 은닉마아코프모델 그리고 변환조합과 정보기준을 적용한 은닉마아코프모델의 각 모델별로 주어진 시계열자료 길이에 내포되어 있는 특징의 상태 수 추정결과를 보여 주고 있다. 주어진 시계열자료의 길이가 30일에서 50일 내에서는 모두 정확하게 상태 수를 추정하는 것을 알 수 있다. 이는 주어진 자료의 길이가 충분하다면 자료 내에 내포되어 있는 규칙성을 충분히 반영할 수 있기 때문으로 생각된다. 하지만 시계열자료의 길이가 상대적으로

짧은 20일 내의 구간에서는 상태 수 추정이 부정확함을 알 수 있다. 이유는 주어진 자료의 길이가 상대적으로 짧을 경우 내재되어 있는 규칙성의 횡수를 정확히 반영하기 어렵기 때문이다[13]. 주어진 시계열자료의 매 특징시점에 관측되어지는 값마다 특징 점으로 보게 되어 상태 수가 장기간의 자료에서 보다 많게 추정된다. 하지만 위의 결과 중에서 자료의 길이가 20일 경우 중에서 변환조합을 이용한 은닉마야코프모델의 경우 정확한 상태 수를 추정한 것을 확인할 수 있다. 이는 본 연구에서 제안하는 변환조합을 이용하는 은닉마야코프모델의 경우 자료 내에 내재되어 있는 예외적인 특징들과 잡음들을 제거하고 효율적인 상태 수 추정이 이루어지는 것으로 볼 수 있다.

#### 4.2 실제자료와 추정된 모델을 통한 예측된 운동양태의 비교

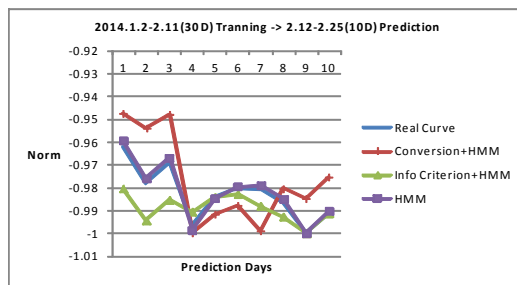
[Fig. 1, 2, 3, 4]는 앞의 실험에서 주어진 자료길이에 따라 추정된 특징들의 상태 수를 통해 모델을 생성한 후 각 모델별로 향후 10일을 예측하여 모델별로 실제의 주가 운동양태와 비교한 결과를 보여주고 있다.



[Fig. 1] Prediction Pattern by 20 Data Length

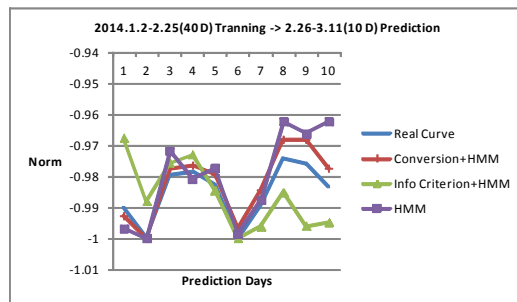
[Fig. 1]은 2014년 1월2일에서 1월28일의 20일 짧은 기간의 훈련자료로 은닉마야코프모델, 정보기준을 적용한 은닉마야코프모델 그리고 제안한 변환조합을 적용한 은닉마야코프모델의 각 세 모델을 생성한 후 각 모델별로 향후 10일을 예측한 주가의 운동양태이다. 실제 운동양태는 예측 1일차부터 8일차까지 지속적인 하락하는 운동양태를 보이며 마지막에 2일 동안 상승하는 운동양태를 보이고 있다. 하지만 은닉마야코프모델과 정보기준을 적용한 마야코프모델은 8일차까지 등락을 반복하며 마지막에 하락하는 불규칙적인 운동양태를 보이고 있다. 이는

상대적으로 모델생성을 위한 훈련자료가 상대적으로 짧은 경우 자료에 내재되어 있는 규칙성을 정확히 반영하기 어렵기 때문이다. 하지만 변환조합을 적용한 은닉마야코프모델은 실제운동양태와 비슷한 운동양태를 보이고 있다. 이는 제안하는 변환조합을 적용하는 모델이 정규화와 이동평균 기법을 통해 자료 내에 섞여있는 예외적인 특징과 잡음들을 효과적으로 제거하기 때문이다.



[Fig. 2] Prediction Pattern by 30 Data Length

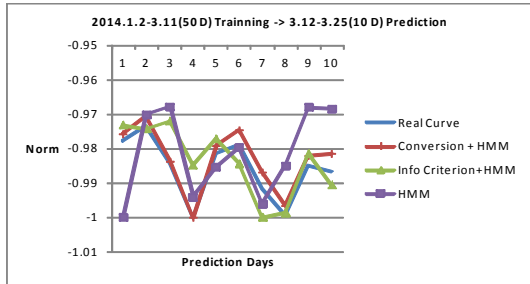
[Fig. 2]는 2014년 1월 2일에서 2월 11일까지 30일의 훈련자료로 각 모델을 생성한 후 생성된 모델들을 통해 예측한 결과를 보여주고 있다. 앞의 상태 추정 실험에서 살펴본 것처럼 단기의 훈련자료보다 상대적으로 충분한 실험 자료의 길이를 통해 예측한 운동양태와 실제의 운동양태는 초반 하락, 중반 상승과 하락 그리고 마지막 9일차에 상승하는 유사한 운동양태를 모두 보여주고 있으나 변환조합을 적용한 모델에서 좀 더 유사한 운동양태 결과를 보여주고 있다.



[Fig. 3] Prediction Pattern by 40 Data Length

[Fig. 3]은 40일의 훈련자료로 모델 생성과 예측한 운동양태를 보여주고 있다. 실제 운동양태와 예측한 운동양태 모두에서, 1일에서 2일까지는 하락, 3일에서 6일은

상승하여 하락하는 봉우리 모양의 운동양태를 그리고 6 일에서 10일은 급격한 상승과 하락을 보이는 유사한 운동양태를 보여주고 있다.



[Fig. 4] Prediction Pattern by 50 Data Length

[Fig. 4]는 50일의 훈련자료로 모델 생성과 예측한 운동양태를 보여주고 있다. 실제 운동양태가 예측기간동안 상승과 하락을 반복하는 운동양태를 보여주고 있다. 은닉마아코프모델과 정보기준을 적용한 모델은 상승과 하락을 반복하는 전체적인 운동양태는 유사하지만 시간차는 차이를 보이고 있다. 하지만 변환조합을 적용한 모델은 상승과 하락 및 운동양태의 시간차에서도 유사한 운동양태를 보여주고 있다.

<Table 1>과 [Fig. 1, 2, 3, 4]을 종합해 보면 상태 수 결정과 모델결정 후 향후 운동패턴 예측 모두 30일, 40일, 50일의 자료길이에서 실제의 코스피지수의 운동패턴과 유사한 예측 운동양태 결과를 보여준다.

유사도 측정을 위해 위의 각 실험을 통해 실제 코스피 지수에서 나타나는 운동양태와 제안하는 모델을 통한 예측 운동양태의 유사도 측정은 두 자료 모두 정규화한 후 일별차이를 고려하는 평균제곱오차(Mean Square Error)를 적용하여 실제와 예측 두 운동양태 수치 간 오차의 크기로서 확인하였다. 두 자료 사이의 평균치 0.5를 기준으로 평균치 임계점보다 낮을수록 유사도가 큰 것으로 평균치 임계점보다 높을수록 비유사도가 큰 것으로 나타낸다.

<Table 2> Prediction Rate the Similar Pattern by Model

| Length<br>Model          | 20 D-<br>10 D   | 30 D-<br>10 D   | 40 D-<br>10 D   | 50 D-<br>10 D   |
|--------------------------|-----------------|-----------------|-----------------|-----------------|
| HMM                      | 20(13)=<br>0.65 | 20(15)=<br>0.75 | 20(15)=<br>0.75 | 20(14)=<br>0.70 |
| Infor Criterion<br>+ HMM | 20(12)=<br>0.60 | 20(14)=<br>0.70 | 20(13)=<br>0.65 | 20(13)=<br>0.65 |
| Conversion<br>+ HMM      | 20(15)=<br>0.75 | 20(16)=0.8<br>0 | 20(16)=<br>0.80 | 20(15)=<br>0.75 |

<Table 2>는 은닉마아코프모델, 정보기준을 적용하여 추정된 은닉마아코프모델, 변환조합을 적용하여 추정된 은닉마아코프모델의 각 모델들을 통해 운동양태를 20회씩 예측하여 평균제곱오차를 기준으로 실제 코스피지수의 운동양태와 각 모델들을 통한 예측 운동양태의 유사패턴을 확인한 비율을 보여주고 있다. 세 모델 모두에서 모델추정을 위해 사용된 코스피지수의 길이가 30, 40 일에서는 20일, 50일보다 상대적으로 높은 유사한 운동양태 비율을 보여주고 있다. 그러나 20일과 50일의 코스피 자료 길이에서는 운동양태의 유사도가 상대적으로 낮게 나왔다. 이것은 실험에 주어진 자료의 길이가 상대적으로 짧은 경우에는 자료 내에 내포된 규칙성을 반영하기 어렵고 또한 자료의 길이가 긴 경우에는 경제에 미치는 불규칙적이고 다양한 예외적인 변수들이 많이 반영되어 추가형성에 내재된 결과로 생각된다.

본 연구에서 제안하는 변환조합을 적용하여 추정된 모델을 통해 예측한 유사도 비율이 은닉마아코프모델과 정보기준을 적용하여 추정된 모델을 적용하였을 때 보다 높은 상대적 높은 예측률을 보여주고 있다. 이 실험을 통해 단기예측을 위한 모델추정에 주어진 코스피 지수에 내포된 예외적인 특징과 잡음을 제거에 적용할 수 있는 유효한 모델로 고려된다.

## 5. 결론

본 연구에서는 시계열자료의 모델에 적합한 은닉마아코프모델을 토대로 시계열자료 내에 내재된 예외적인 특징과 잡음을 제거하여 경제지표 중 가장 큰 축인 주식지수의 올바른 분석 그리고 의사결정에 적용을 위한 모델 추정 기법을 제안하였으며 실험을 통해 모델의 추정과 예측의 유효성을 확인하였다. 첫째, 모델추정에 사용된 자료의 길이가 상대적으로 짧거나 긴 경우에는 운동양태의 유사도 비율이 낮아지는 것을 확인하였으며 둘째, 기존의 두 모델보다 제안하는 모델추정 방법이 모든 자료의 길이에서 운동양태의 유사도 비율이 높게 나왔다.

경제시스템 각 영역에서 발생하는 다양한 시계열자료의 효율적인 모델추정을 위하여 본 연구에서 제안하는 모형 결정 기법을 적용한다면 주어진 각 영역의 문제해결의 목표를 극대화할 수 있을 것으로 기대되며 향후 더

다양한 부분의 일반화된 시계열자료에 적용 가능한 연구를 통하여 일반적이고 유연한 분석, 예측모형을 세울 수 있는 부분으로 연구영역의 확대가 필요할 것이다.

## REFERENCES

- [1] J. Jeon and 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, pp305-311. 2013.
- [2] S. Shin, "Evaluation Exchange Rate of Artificial Neural Network and Moving Average Method", *Finance Research*, Vol. 9, no. 1, pp103-135, 1995.
- [3] J. Jeon and 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.
- [4] J. Jeon., "A Study on Determining Prediction Models using Model-based Clustering of Time Series Data", *Dankook Univ Ph. D*, 2007.
- [5] Y. Cho and G. Lee., "A Study on Improving Prediction Accuracy by Modeling Multiple Similar Time Series", *The Institute of Webcasting, Internet and Telecommunication*, Vol. 10, no. 6, pp 137-143, 2010.
- [6] L. Rabiner., "A Tutorial on Hidden Markov Models and selected applications in speech recognition," *Proc. of IEEE*77, pp.257-286, 1989.
- [7] A. Sorjamaa, et al., "Methodology for long-term prediction of time series," *Neurocomputing*, pp178-186. Elsevier, 2007.
- [8] Y. Cho and G. Lee., "Prediction on Clusters by using Information Crriterion and Multiple Seeds", *The Institute of Webcasting, Internet and Telecommunication*, Vol. 10, no. 6, pp 145-152, 2010.
- [9] M. Siddiqi, J. Gordon and W. Moore., "Fast State Discovery for HMM Moel Selection and Learning," *In Proc. Int'l Conference on Artificial Intelligence and Statistics*, 2007.
- [10] J. Jeon and m. Kim., "A study of criterion for efficient clustering estimation of temporal data", *The Institute of Webcasting, Internet and Telecommunication*, Vol. 11, no. 5, pp 139-144, 2011.
- [11] A. Sorjamaa, et al., "Methodology for Long-Term Prediction of Time Series," *Neurocomputing*, pp178-186. Elsevier, 2007.
- [12] L. Jessica, K. Eamonn, L. Stefano and C. Bill., "A Symbolic representation of time series, with implication for streaming algorithms", *8th ACM SIGMOD Workshop on Research Issues in DMKD*, 2003.
- [13] Y. Byungki and F. Christos., "Fast Time Sequence Indexing for Arbitrary Lp Norms", *In The VLDB Journal*, pp 385-394, 2000.
- [14] P. Cheeseman and J. Stutz, "Bayesian Classification" *Kluwer Academic Publishers*, Vol 70. pp117-126, 1996.
- [15] Heckerman, D., Geiger, D., and Chekering, D. M. "A tutorial on learning with bayesian networks," *machine Learning* 20, pp.197-243, 1995.
- [16] M. Kendall, *Time Series*, 2nd Edition, Charles Griffin and Company, 1979.
- [17] C. Chatfield, *The Analysis of Time-Series: An Introduction*, 3rd Edition, Chapman and Hall, 1984.
- [18] D. Rafiei and A. Mendelzon, "Similarity-Based Queries for Time Series Data," *In Proc Int'l Conf. on Management of Data, ACM SIGMOD*, pp13-24, 1997.
- [19] R. Agrawal et al., "Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time Series Databases," *In Proc. Int'l Conference on Very Large Databases, VLDB*, pp, 490-501, 1995.
- [20] W. K. Loh, S. W. Kim, and K. Y. Whang, "Index Interpolation: An Approach for Subsequence Matching Supporting Normalization Transform in Time-Series Databases, 2000.

전 진 호(Jeon, Jin Ho)



- 1990년 3월 ~ 1994년 2월 : 관동대학교 경영학과 경영학사
- 1996년 3월 ~ 1998년 8월 : 명지대학교 경영정보학과 경영학석사
- 2000년 3월 ~ 2007년 2월 : 단국대학교 컴퓨터과학 이학박사
- 2009년 9월 ~ 현재 : 가톨릭 관동대학교 경영학과 조교수

- 관심분야 : 지능형시스템, IT전략 및 정책
- E-Mail : jhgy@cku.ac.kr

김 민 수(Kim, Min Soo)



- 1990년 3월 ~ 1997년 2월 : 관동대학교 무역학과 경영학사
- 1997년 3월 ~ 1999년 2월 : 명지대학교 무역학과 경영학 석사
- 1999년 3월 ~ 2004년 2월 : 명지대학교 무역학과 경영학 박사
- 2009년 9월 ~ 현재 : 가톨릭 관동대학교 무역학과 조교수

- 관심분야 : 국제경영
- E-Mail : mskim@cku.ac.kr