http://dx.doi.org/10.14400/JDC.2014.12.7.261

중선형 모형을 이용한 비선형 시계열 패널자료의 동질성검정에 대한 연구

김인규
우송정보대학 컴퓨터정보과

A Study on the Test of Homogeneity for Nonlinear Time Series Panel Data Using Bilinear Models

Inkyu Kim
Dept. of Computer & Information, Woosong Information College

요 약 시계열 모형에서 모수의 수가 많으면 모수추정에 따르는 오차가 커지게 되므로 예측을 하는데 많은 어려움이 있다. 만약 여러개의 시계열 자료들이 동일한 모형에서부터 얻어졌다고 하는 동질성 가설이 채택되면 모수축약을 이룰 수 있고, 더 좋은 예측값을 얻을 수 있다. 비선형 시계열 패널 자료는 각각의 시계열마다 모수들이 있기 때문에 매우 많은 모수가 존재하게되고, 모수의 수가 많으면 모수추정에 따르는 오차가 커지게 되어 예측의 정확도가 떨어지게 된다. 패널내에 존재하는 독립적인 여러 시계열들의 동질성이 만족되면 시계열을 종합하여 모수를 추정하고 검정할 수 있다. m개의 독립적인 비선형 시계열 패널 자료의 동질성 검정을 알아보기 위하여 모형을 설정하고 이 모형에 대한 정상성 조건을 구하였고, 동질성 검정통계량을 유도했으며, 구한 검정 통계량의 극한분포가 χ^2 분포를 따르는 것을 보였다. 실증분석에 있어서는 비선형 시계열 자료중 중선형 시계열 모형의 동질성 검정을 하고, 실제 우리나라 주식자료를 2개의 집단으로 나누어 비선형 시계열 패널 자료의 동질성 검정에 대한 분석을 하였다.

주제어 : 비선형 시계열 패널 자료, 동질성 검정, 검정통계량, 주식자료, 모수추정

Abstract When the number of parameters in the time series model are diverse, it is hard to forecast because of the increasing error by a parameter estimation. If the homogeneity hypothesis which was obtained from the same model about severeal data for the time series is selected, it is easy to get the predictive value better. Nonlinear time-series panel data for each parameter for each time series, since there are so many parameters that are present, and the large number of parameters according to the parameter estimation error increases the accuracy of the forecast deteriorated. Panel present in the time series of multiple independent homogeneity is satisfied by a comprehensive time series to estimate and to test of the parameters. For studying about the homogeneity test for the m independent non-linear of the time series panel data, it needs to set the model and to make the normal conditions for the model, and to derive the homogeneity test statistic. Finally, it shows to obtain the limit distribution according to χ^2 distribution. In actual analysis,, we can examine the result for the homogeneity test about nonlinear time series panel data which are 2 groups of stock price data.

Key words : Non-Linear Time Series Panel Data, Test of Homogeneity, Wald Statistic, Stock Price Data, Parameter Estimation

Received 29 April 2014, Revised 30 May 2014
Accepted 20 July 2014
Corresponding Author: InKyu Kim(Woosong Information College, Dept. of Computer & Information)
Email: ikkim0056@wsi.ac.kr

ISSN: 1738-1916

1. 서론

선형 시계열 모형들은 실제 시계열 자료들에 매우 잘 적합 될 뿐만 아니라 분석을 하는데 있어서 많은 연구가 이루어져 왔음에도 불구하고 선형 시계열 모형들은 강한 비대칭을 나타내는 자료들이나 또는 시간을 역행할 수 없는 자료들에 대해서는 관념적으로 적당하지 않다고 알려져 있다(Tong(1990))[1]. 여러 분야에서 얻어지는 다양한 시계열 자료의 일반적인 도약현상(jump phenomena)과 한계순환(limit cycles)같은 것들은 선형 시계열 모형으로는 잘 설명되어 질 수 없어서 이러한 선형 시계열 모형들의 제한들과 비선형 시계열 모형들의 이점들을 이용한 비선형 시계열 모형들에 관한 많은 연구가 이루어져 왔으나 비선형 모형을 이용한 예측에 관한 연구는 여전히 미흡한 편이다.

여러 개의 독립적인 시계열로 구성된 시계열 패널자료는 실제 데이터 분석에서 자주 나타나고 있다. 예를 들어, 여러 명의 혈압을 여러 시점에서 걸쳐 기록하는 경우에 개인별 자료는 서로 독립이라 할 수 있고, 각 개인의 혈압 기록은 종속 구조를 가진 시계열 자료로 모형화 할 수 있다. 이 때에 m개의 독립적인 시계열 모형이 생기는데, 이러한 자료를 분석할 때 먼저 시계열 동질성 검정을 하여 가능하면 m개 시계열을 종합(pooling)하여 연관된 모수를 추정하고 검정하는 절차를 밟는다.

시계열 모형의 모수축약의 원칙은 매우 중요하다. 모수의 수가 많으면 모수추정에 따르는 오차가 커지게 되므로 예측을 정확하게 할 수가 없다. 만약 여러개의 시계열 자료들이 동일한 모형에서부터 얻어졌다고 하는 동질성 가설이 채택되면 모수축약을 이룰 수 있고, 더 좋은 예측값을 얻을 수 있다. 본 논문의 목적은 비선형 시계열 자료 중 중선형 시계열모형의 동질성 검정을하고, 실제 패널자료를 이용해 실증분석을 하고자 한다.

Anderson(1978)의 연구에서는 의사가 여러 환자의 혈압을 여러 날 동안 측정하여 이 환자들의 혈압 간에 동질성이 있는지를 알아보았으며[2], Lee(1993)는 계절성 시계열 패널 자료에 대한 동질성 검정을 연구하였다[3]. Pham과 Tran(1981)은 중선형 모형의 모수들에 대한 수정된 최소제곱 추정량들을 구하였고, 그것들은 강한 일치성을 가지고 있다는 것을 보여주었으며, 점근적인 정규성(asymptotic normality)같은 추정량들의 분포적인 성질에 관해서는 언급하지 않았다[4]. Gaber(2008)의 연구에서는 중선형 시계열 모형은 선형 모형과 백색 잡음을 구분하는 데 사용되어지고, 그리고 일반 선형 모형에서도 사용 할 수있다고 하였으며[5], Abdelouahab B.와, Abdelhakim A.(2010)은 강한 일관성과 Yule-Walker 추정의 점근 정상뿐만 아니라 Wald의 절차를 통해 가설검정을 얻을수 있다고 하였다[6]. Iheanyi S.와 Iwueze, Ohakwe Johnson(2011)의 연구에서는 제곱의 시계열 특성은 보통 시계열 모형의 선형 이동 평균 시계열 모형의 특성과 비교됨을 보여 주었고[7], E. H. Etuk 와 I. A. Iwok(2012)는 AR 과정에 대한 N-차원 벡터의 일반적인 경우에 대한 행렬을 공식화 한 후 화이트 노이즈 제로 지연의 특별한 경우를 연구하였다[8].

본 논문의 구성은 다음과 같다. 제 1장에서는 연구배경 및 목적에 대하여 설명을 하였고, 기존에 연구되어 왔던 문헌들을 알아보았다. 제 2장에서는 비선형 시계열 자료중 중선형 시계열 모형(Bilinear Time Series Model)에 대하여 알아보았고, m개의 독립적인 비선형 시계열 패널자료의 동질성 검정을 알아보기 위하여 모형을 설정하고 이 모형에 대한 추정 방법을 알아보았고, 동질성 검정통계량을 유도했으며, 구한 검정통계량의 극한분포를 알아보았다. 제 3장에서는 실증 분석을 통해 실제 우리나라 8개도 Mumps(항아리 손님)자료를 이용하여 중선형 패널 시계열 모형의 동질성 검정을 하였다.

2. 중선형 모형

2.1 BL(Bilinear)모형

중선형 시계열 모형은 경제학, 생물학, 의약학, 금융 및 재정에 잘 적합 된다. Granger와 Anderson(1978)이 처음 고안하였으며, 이산(discrete) 시간과정 $(X_t)_{t \in z}$은 다음과 같이 중선형 모형 BL(p,q,r,s)으로 정의된다.

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=1}^{q} c_j e_{t-j} + \sum_{m=1}^{r} \sum_{l=1}^{s} b_{ml} X_{t-m} e_{t-l} + e_t \qquad (2.1)$$

여기에서, $(a_{i,}\ 1 \le i \le p)$, $(c_{j,}\ 1 \le j \le q)$ 그리고 $(b_{ml,}\ 1 \le m \le r, 1 \le l \le s)$는 이 모형의 계수이

며, e_t는 평균이 0이고 분산이 σ^2인 iid한 Gaussian 과정을 따른다.

중선형 시계열 모형의 동질성 검정은 수리적으로 간단한 BL(0,0,1,1)모형을 이용하여 검정할 것이다. 이 모형은 순수대각(pure diagonal) 중선형 과정인 BL(0,0,p,p)모형에서 p=1인 경우이므로 순수대각 중선형 모형에 대한 설명을 하고자 한다.

다음 식은 간단한 수리적 구조식을 가지고 있는 BL(0,0,p,p)인 순수대각(pure diagonal)중선형과정이다.

$$X_t = \sum_{i=1}^{p} b_{ii} X_{t-i} e_{t-i} + e_t \tag{2.2}$$

여기에서, e_t 는 평균이 0이고 분산이 σ^2인 iid한 Gaussian 과정을 따른다.

Liu(1990)은 식(2,1)에 대한 일반 중선형 모형하에서의 가역성을 다음과 같이 정의 하였다[9].

$$p^2 \sum_{i=1}^{p} b_{ii}^2 \sigma^2 < 1 \tag{2.3}$$

2.2 BL모형의 동질성검정

2.2.1 모형설정

$\{X_t(k), t=0,\dots,n, k=1,\dots,m\}$을 k번째 대상의 t시점 자료라고 하고, 중선형 모형의 완전한 BL(p,q,r,s)의 패널모형은 다음과 같다.

$$X_t(k) = \sum_{i=1}^{p} a_i(k) X_{t-i}(k) + \sum_{j=1}^{q} c_j(k) e_{t-j}(k) + \sum_{m=1}^{r}\sum_{l=1}^{s} b_{ml}(k) X_{t-m}(k) e_{t-l}(k) + e_t(k) \tag{2.4}$$

위의 모형에서 $\{e_t(k)\}$는 평균이 0이고 분산이 σ^2인 정규분포를 따르는 iid확률변수이다. 본 논문에서는 간단한 순수대각(Pure Diagonal) 중선형 모형인 BL(0,0,p,p)모형을 아래와 같이 고려해본다.

$$X_t(k) = \sum_{i=1}^{p} b_{ii}(k) X_{t-i}(k) e_{t-i}(k) + e_t(k) \tag{2.5}$$

2.2.2 추정방법

BL(0,0,p,p)모형에서의 모수$\theta = (b_{11}, b_{22}, \dots, b_{pp})$라고 하자. 최대우도 방법에 의하여 모수 θ의 벡터를 구할 수 있다. $(p+1)\times 1$벡터를 다음과 같이 정의한다.

$$\xi = \begin{pmatrix} e_t \\ e_{t-1} \\ \vdots \\ e_{t-p} \end{pmatrix}, \quad \upsilon_t = \begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}' \tag{2.6}$$

그리고 $H_{t-1} = [1, b_{11}X_{t-1}, b_{22}X_{t-2}, \dots, b_{pp}X_{t-p}]$이므로 $X_t = H_{t-1}\xi_t$가 된다.

$\chi_t = (x_1, \dots, x_t)$라고 하면, θ의 로그우도함수(log-likelihood function)는 다음과 같다.

$$L(X;\theta) = -\frac{n}{2} log(2\pi) - \frac{1}{2}\sum_{t=1}^{n} \log(\widehat{M}_{t|t-1}) - \frac{1}{2}\sum_{t=1}^{n} \frac{(x_t - \widehat{X}_{t|t-1})^2}{\widehat{M}_{t|t-1}} \tag{2.7}$$

위의 식(2.7)에서 $\widehat{X}_{t|t-1} = E[X_t | X_{t-1}]$과 $\widehat{M}_{t|t-1} = E[(X_t | X_{t-1})^2]$은 Kalman Filter 알고리즘을 이용하여 구할 수 있다.

2.2.3 검정방법

BL(0,0,p,p)에서 p=1인 경우, 즉 BL(0,0,1,1)모형을 고려해 본다.

$$X_t(k) = b(k) X_{t-1}(k) e_{t-1}(k) + e_t(k) \tag{2.8}$$

여기에서 m개의 시계열 동질성검정에 대한 가설은 다음과 같다.

$$H: b(1) = b(2) = \dots = b(m) \tag{2.9}$$

<Wald statistic>

위의 동질성 가설에서 새로운 기호인 $\psi_k = b(k) - b(k+1), k = 1, \dots, m-1$으로 가설을 다시 표현하면,

$$H: \psi_1 = \dots = \psi_{m-1} = 0 \tag{2.10}$$

이 된다.

이 가설에 대한 검정 통계량으로써 Wald 검정통계량 Qn은 다음과 같다.

$$Q_n = n\hat{\psi}'[C\widehat{T^{-1}}C']^{-1}\hat{\psi} \tag{2.11}$$

여기에서,

$$\hat{\psi} = \begin{pmatrix} \hat{\psi}_1 \\ \vdots \\ \hat{\psi}_{m-1} \end{pmatrix} = \begin{pmatrix} \hat{b}(1) - \hat{b}(2) \\ \vdots \\ \hat{b}(m-1) - \hat{b}(m) \end{pmatrix} : (m-1) \times 1 \text{ 벡터}$$

그리고

$$C = \begin{pmatrix} I_p & -I_p & \cdots & 0 \\ I_p & & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ I_p & & 0 & \cdots & 0 \end{pmatrix} : (m-1)p \times m \text{ 행렬}$$

2.2.4 극한분포

귀무가설 H하에서 Tn의 극한분포는 자유도가 (m-1)p인 카이제곱 분포를 따른다. 따라서 유의수준 α 에서의 기각역은 다음과 같다.

$$\{Q_n \geq \chi^2_{(m-1)p}(\alpha)\} \tag{2.12}$$

동질성 검정이 기각되지 않으면 m개 시계열자료를 종합하여 추정하는 것이 타당하다.

2.3 모의검정력 계산

식(2.11)에서 제안된 검정통계량 Q_n 의 검정력을 모의실험을 통해 알아보도록 하자. 논의의 간편성을 위해 (m=2)개의 독립적인 시계열로 구성된 자료로서 각 시계열은 정상 일차 중선형 모형을 따르도록 하였다(p=1). 길이 n=500인 독립적인 두 개의 모의 시계열 자료를 생성시켜 모수를 구하였다. 이제 검정 통계량 Q_n은 다음과 같다.

$$Q_n = 500\hat{\phi}'(C\widehat{T^{-1}}C')^{-1}\hat{\phi}. \tag{2.13}$$

모의실험결과 1000번중 유의수준 α=0.05인 χ^2_1 = 3.843을 넘는 비율을 계산하여 <Table.1>과 같은 결과를 얻었다.

〈Table 1〉 Simulation statistical power(%) : $\theta(2)$ = 0, $\delta = \theta(1)$

σ^2 \ δ	0	0.01	0.03	0.05	0.07	0.1	0.2	0.3	0.4	0.5	0.6	0.7
0	5.3	5.6	7.4	12.5	19.6	32.9	89.7	100	100	100	100	100
0.05	5.2	7.5	8.6	14.7	20.3	33.2	86.8	99.7	100	100	100	100
0.1	5.4	8.3	11.6	16.1	20.9	36.2	87.2	100	100	100	100	100

<Table 1>으로부터 우리는 δ값이 커질수록 검정력이 매우 좋아짐을 알 수 있었다. 그리고 α=0.05에서 δ=0인 경우 유의수준의 값이 σ^2이 커질수록 약간 커짐을 알 수 있었고, 이 유의수준값은 아주 만족할 만한 값이라고 할 수 있다. m개의 독립적인 비선형 시계열로 구성된 패널자료의 동질성 검정을 하기 위하여 Wald통계량을 제안하였고, 그 극한분포가 χ^2분포에 따름을 보였으며, 모의실험을 통하여 검정력이 만족할 만하였다. 따라서 주가 수익률 자료와 같은 패널자료의 분석시 먼저 m개의 시계열의 동질성검정을 본 논문의 검정법을 이용하여 수행하여, 가능하면 m개의 시계열을 종합(pooling)하여 연관된 모수를 추정하면 모수축약의 원칙에 충실한 분석을 할 수 있을 것이다.

3. 실증 분석

3.1 자료 소개

IMF 이전의 우리나라 주식자료를 2개의 집단으로 나누어 비선형 시계열 패널 자료의 동질성 검정에 대한 분석을 하였다. 첫 번째 집단은 대우증권, 대신증권, 삼성증권, 현대증권, LG증권의 1996년 1월 3일부터 1996년 12월 27일까지 각각의 293일에 대한 주식값의 1일 주가 변동율$\left(= \dfrac{\text{오늘종가} - \text{어제종가}}{\text{어제종가}}\right)$에 대한 5개 증권회사의 자료이고, 두 번째 집단은 삼성물산, 현대 종합무역상사, SK상사, (주)대우, LG상사의 1996년 1월 3일부터 1996년 12월 27일까지 각각의 293일에 대한 주식값의 1

일 주가 변동율에 대한 5개 종합무역상사의 자료이다. 즉, 첫 번째 두 번째 모두 각 계열(series)의 개수가 293개이고, 패널의 수는 5개인 자료이다.

3.2 동질성 검정

1일 주가 변동율의 동질성 검정에 대한 귀무가설은

$$H_0:\ \theta(1) = \cdots \theta(m), \quad m = 1, \cdots, 5$$

이고, 대립가설은

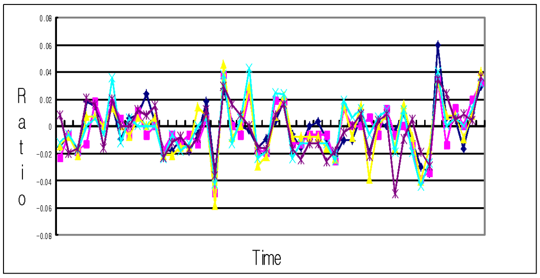
H_1: 적어도 하나는 다르다.

이다. Wald 통계량 $Q_n = n\hat{\phi}'\left[C\widehat{\Gamma^{-1}}C'\right]^{-1}\hat{\phi}$ 을 구한 결과는 다음과 같다.

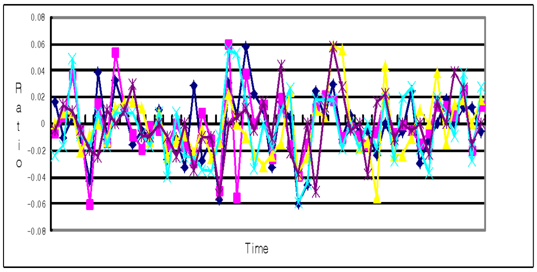
$$Q_n(1) = 0.011404 ,$$

$$Q_n(2) = 0.006699 .$$

여기서 $Q_n(1)$은 5개 증권회사에 대한 결과이고, $Q_n(2)$는 5개 종합무역상사에 대한 결과이다. 이 값은 유의 수준 $\alpha = 0.05$ 에서 둘다 동질성 귀무가설을 기각하지 못한다. 그러므로 5개 증권회사 주식들과 5개 종합무역 상사 주식들 각각에 대한 모형은 동질성을 갖는다고 애기 할 수 있다. 아래의 [Fig. 1]과 [Fig. 2]는 각각 5개 증권회사와 5개 종합무역상사의 1일 주가변동률에 대한 자료를 그래프로 나타낸 것이다. 원래 자료는 n=293인데, 모든 자료를 그래프로 나타내면 자료가 너무 많아 동질성을 확인하기 어려워 처음 50개의 자료만 가지고 그래프로 나타냈다. 이 그림들을 보고 우리는 5개 증권회사와 5개 종합무역상사의 각각의 1일 주가변동률은 서로 동질성을 갖는다는 것을 알 수 있다. [Fig. 1]과 [Fig. 2]에서 보는 바와 같이 5개 증권회사의 1일 주가변동률에 대한 동질성이 5개 종합무역상사의 1일 주가 변동률에 대한 동질성보다 강하다는 것을 알 수 있다. 또한 5개 증권회사의 1일 주가 변동률에 대한 Wald 통계량 값(0.011404)이 5개 종합무역상사의 1일 주가 변동률에 대한 Wald 통계량 값(0.006699)보다 크기 때문에 5개 증권회사의 1일 주가 변동률에 대한 동질성이 더 강하다는 것을 알 수 있다. 증권회사의 주가 수익률은 주식시장의 거래량에 따라서 수익이 좌우되는데, 주식시장의 침체 장세에서는 모든 증권회사의 주가가 같이 하락하고, 활황 장세에서는 같이 상승하기 때문에 5개 증권회사의 1일 주가 수익률에 대한 동질성이 강한 것으로 나타난다. 그러나 종합무역상사의 주가 수익률은 증권회사의 주가 수익률보다는 동질성이 조금 떨어지는데 그 이유는 예를 들면, 반도체가 수출이 잘 되면 반도체를 생산하는 회사(현대종합무역상사, 삼성물산, LG상사)들은 매출이 늘어 주가가 상승하지만, 그에 비해 반도체를 생산하지 않는 회사((주)대우, SK상사)들은 반도체를 생산하는 회사들보다 주가가 상승하지 못하기 때문이다.



[Fig. 1] The volat ility ratio of stock price of 5 securities companies for 50 days



[Fig. 2] The volatility ratio of stock price of 5 general trading companies for 50 days

4. 결론

모수의 수가 많으면 모수추정에 따르는 오차가 커지게 되므로 예측은 정확하게 할 수가 없다. 만약 여러 개의 시계열자료들이 동일한 모형에서부터 얻어졌다고 하는 동질성 가설이 채택되면 모수축약을 이룰 수 있고, 더 좋은 예측 값을 얻을 수 있다. 여러 개의 독립적인 비선형 시계열로 구성된 패널자료의 동질성검정을 하기 위하여 Wald통계량을 제안하였고, 실증분석을 통하여 계수들의 동질성이 만족되었음을 보였다. 중선형 시계열자료의 동질성 검정을 위해 우리나라 주식자료를 2개의 집단

으로 나누어 비선형 시계열 패널 자료의 동질성 검정에 대한 분석을 한 결과 각각에 대한 모형은 동질성을 갖는다는 것을 증명하였다.

본 논문에서는 비선형 시계열 패널자료에 대한 동질성 검정은 각 패널 간에 서로 독립이라고 가정을 하고 동질성검정을 하였으나, 각 패널간에 독립이 아닌 경우 즉, 각 패널간에 서로 상관관계(correlation)가 존재하는 경우의 동질성검정에 대한 앞으로의 연구가 필요하다.

REFERENCES

[1] Tong, H., Non-linear Time Series, Oxford University Press, Oxford, 1990.

[2] Anderson, T. W., Repeated mearsurements on autoregressive process. Journal of American Stat. Association, Vol. 73, pp. 371-378, 1978.

[3] Lee, S. D., Test of Homogeneity for a Panel of Seasonal Autoregressive Processes, Journal of Korean Statistical Society, 22, pp. 125-132, 1993.

[4] Pham, T. D. and Tran, L. T., On the first order bilinear time series model, Journal of Allied Probability, 18, 617-627, 1981.

[5] M. M. Gaber, On the Third Order Moment Structure and Bispectral Ananysis of Some Bilinear Time Series, Journal of Time Series Analysis, Vol. 9, pp. 11-20, 2008.

[6] Abdelouahab B., Abdelhakim A., Yule-Walker type estimators in periodic bilinear models: strong consistency and asymptotic normality, Statistical Methods and Applications, Vol. 19, pp. 1-30, 2010.

[7] Iheanyi S. Iwueze, Ohakwe Johnson, Covariance analysis of the squares of the purely diagonal bilinear time series models, Brazilian Journal of Probability and Statistics, Vo. 1, 2011.

[8] Etuk E. H. and Iwok I. A., Zero-lag white noise vector bilinear autoregressive time series models, American Journal of Scientific and Industrial Research, Vol. 3, pp. 86-93, 2012.

[9] Liu, J., A note on causality and invertibility of a general bilinear time series model, Adv. Appl. Prob, 22, pp. 247-250, 1990.

김 인 규(Kim, Inkyu)

- 1986년 5월 : University of Akron 수학과(이학사)
- 1987년 12월 : University of Georgia 통계학과(이학석사)
- 1999년 2월 : 충북대학교 전자계산학과(이학박사)
- 1992년 9월 ~ 현재 : 우송정보대학 컴퓨터정보과 교수

- 관심분야 : 시계열분석, 신경망, 프로그램 언어
- E-Mail : ikkim0056@wsi.ac.kr