

Test of Homogeneity for Panel Bilinear Time Series Model

ShinHyung Lee^a · SunWoo Kim^b · SungDuck Lee^{b,1}

^aRegin Cooperation Department, Gyeonggi District Employment and Labor Office

^bDepartment of Information and Statistics, Chungbuk National University

(Received April 25, 2013; Revised May 21, 2013; Accepted May 21, 2013)

Abstract

The acceptance of the test of the homogeneity for panel time series models allows for the pooling of the series to achieve parsimony. In this paper, we introduce a panel bilinear time series model as well as derive the stationary condition and the limiting distribution of the test statistic of the homogeneity test for the model. For the applications study, we use Korea Mumps data from January 2001 to December 2008. Finally, we perform test of homogeneity for the panel data with 8 independent bilinear time series.

Keywords: Panel data, bilinear time series model, homogeneity, Mumps data.

1. 서론

여러 개의 독립적인 시계열들로 구성되어진 패널 시계열 자료를 분석하는 것은 경제 및 금융 시계열뿐만 아니라 의료 및 보건 시계열 자료 분석에서도 매우 빈번하게 발생하는 중요한 통계분석의 한 분야이다. 패널 시계열 자료는 각각의 시계열마다 모수들이 있기 때문에 매우 많은 모수가 존재하게 되고, 모수의 수가 많으면 모수추정에 따르는 오차가 커지게 되어 예측의 정확성도 떨어지게 된다. 패널 내에 존재하는 독립적인 여러 시계열들의 동질성이 만족되면 시계열을 종합하여 모수를 추정하고 검정할 수 있다. 즉 만약 여러개의 자료들이 동일한 모형에서 왔다고 하는 동질성 가설이 채택된다면, 모형 설정에 있어서 모수 절약을 이룰 수 있고 또한 결합된 자료에 의해 보다 나은 모수 추정을 얻을 수 있다. 한편 자료들 사이에 유의한 차이가 존재한다 하더라도 표본 크기가 커지면 대표본하에서 어떤 규칙을 얻을 수도 있을 것이다. 이러한 시계열자료의 동질성검정은 Anderson (1978)의 연구에서 의사가 여러 날에 걸쳐 여러 환자의 혈압을 반복 측정하여 환자들의 혈압 간에 동질성이 있는지를 알아보면서 제안되었다. Basawa 등 (1984)은 ARMA(p, q) 모형을 따르는 시계열 패널 자료에 대한 동질성 검정을 연구하였으며, 계절성 시계열 패널 자료에 대한 동질성 검정에는 Lee (1993)의 연구가 있다.

한편, 많은 연구가 이루어져 온 선형 시계열 모형들은 실제 시계열 자료들에 매우 잘 적합이 되긴 하나 강한 비대칭을 나타내는 자료들이나 시간의 가역성을 만족하지 않는 시계열 자료들에 대해서는 적당

This research was supported by Basic Science Research Program through the national Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A4A01013867).

¹Corresponding author: Professor, Department of Information and Statistics, Chungbuk National University, 52 Naesudong-ro, Heungdeok-gu, Cheongju, Chungbuk 361-763, Korea. E-mail: sdlee@chungbuk.ac.kr

하지 않다고 알려져 있다 (Tong, 1990). 이러한 선형 시계열 모형의 단점을 해결하기 위한 것으로 여러 가지 비선형 모형이 제안되었는데 Granger와 Andersen (1978)은 중선형 시계열모형(Bilinear Time Series Model)을 제안하고, 실제 현상에서의 유용성을 보였는데 특히 의료 및 보건 자료분석에서 매우 유용함을 보였으며, 모수추정 등 여러 가지 통계적 추론에 대한 연구를 수행하였다. 그후 중선형 모형에 대한 연구로는 Subba Rao와 Gabr (1984)이 정규조건하에서 ergodic에 대한 연구를 하였고 Liu (1990)는 일반 중선형 모형하에서의 가역성 조건을 제시하였으며 Pham과 Tran (1981)은 중선형 모형의 모수들에 대한 수정된 최소제곱 추정량들을 구하였고, 그것들이 강한 일치성을 가지고 있다는 것을 보였다.

본 논문에서는 비선형 시계열자료 중 중선형 시계열모형의 동질성 검정을 소개하고, 중선형 시계열모형을 따르는 실제 패널자료의 동질성 검정을 수행하여 실증분석하고자 한다. 2장에서는 중선형 시계열 모형과 칼만-필터를 이용한 모수 추정에 대해 서술하고, 3장에서는 중선형 시계열모형의 동질성 검정에 대해 언급하였다. 4장에서는 실증분석으로 우리나라 8개 도별 Mumps자료를 이용하여 패널 중선형 시계열 모형의 동질성 검정을 수행하고 이 분석방법의 효율성을 입증하였으며, 5장에서는 결론에 대하여 기술하였다.

2. 패널 중선형 시계열 모형

2.1. 패널 중선형 시계열 모형

$X_t(k)$, $t = 0, \dots, n$, $k = 1, \dots, m$ 을 k 번째 대상의 t 시점 자료라고 하면 중선형 모형의 완전한 BL(p, q, r, s)의 패널모형은 다음과 같다.

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

여기서 $e_t(k)$ 는 평균이 0이고 분산이 σ^2 인 정규분포를 따르는 iid 확률변수이다.

본 논문에서는 순수대각(pure diagonal) 중선형 모형인 BL($0, 0, p, p$) 모형을 고려해 본다.

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

여기서 $b_{ii}(k)$ 는 BL(p, q, r, s)의 패널모형에서 $m = l = i$ ($i = 1, 2, \dots, p$)인 $b_{ml}(k)$ 와 같다.

BL($0, 0, p, p$) 모형의 정상성 조건은 상태공간 모형으로 재표현하여 찾을 수 있고, 모수를 추정하기 위해서는 가역성 조건이 필요한데 일반 중선형 모형하에서의 가역성 조건은 Liu (1990)에 의해 다음과 같이 제안되었다.

$$p^2 \sum_{i=1}^p b_{ii}^2(k)\sigma^2 < 1.$$

2.2. 칼만-필터를 이용한 모수추정

$\theta = (b_{11}, b_{22}, \dots, b_{pp})$ 을 모수벡터라고 하자.

최대우도방법에 의하여 모수 벡터 θ 를 구하기 위해 BL($0, 0, p, p$) 모형을 상태 공간 모형으로 표현하고 칼만-필터를 이용하고자 한다.

상태 공간 모형은 시점 t 에서의 상태를 표현하는 상태방정식과 관측되지 않은 상태 벡터와 관측 오차의 함수인 관측방정식으로 구성되는데 BL(0, 0, p) 모형을 다음과 같은 상태공간모형으로 표현한다.

$$\begin{aligned}\xi_{t+1} &= A\xi_t + v_{t+1} && \text{상태방정식,} \\ X_t &= H_{t-1}\xi_t && \text{관측방정식,}\end{aligned}$$

여기서 $H_{t-1} = [1, b_{11}X_{t-1}, b_{22}X_{t-2}, \dots, b_{pp}X_{t-p}]$

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \xi_t = \begin{pmatrix} e_t \\ e_{t-1} \\ \vdots \\ e_{t-p} \end{pmatrix}, \quad v_t = \begin{pmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

단, A 는 $(p+1) \times (p+1)$ 행렬이고, ξ_t 와 v_t 는 각각 $p+1$ 벡터이다.

시간 t 에서 관측값을 $\mathbf{x}_t = (x_1, \dots, x_t)$ 라 하면 칼만-필터 알고리즘은 상태벡터 ξ_{t+1} , $t = 1, \dots, n$ 의 최적 예측값 $\hat{\xi}_{t+1|t}$ 을 반복적으로 구할 수 있다. 먼저 ξ_1 의 예측값인 상태벡터 $\hat{\xi}_{1|0}$ 을 초기화해서 X_1 의 예측값인 $\hat{X}_{1|0} = H_0\hat{\xi}_{1|0}$ 를 구하고, $t = 2, \dots, n$ 에 대해서 $\hat{\xi}_{t+1|t}$ 를 반복해 구한 후 $\hat{X}_{t+1|t} = H_t\hat{\xi}_{t+1|t}$ 식을 이용하여 X_{t+1} 을 예측한다.

X_t 는 평균이 $\hat{X}_{t|t-1} = E[X_t|\mathbf{x}_t]$ 이고, 분산이 $\hat{M}_{t|t-1} = E[(X_t - \hat{X}_{t|t-1})^2]$ 인 Gaussian 분포를 따른다 (Hamilton, 1994). 따라서 θ 의 로그우도함수는 다음과 같다.

$$L(X; \theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^n \log(\hat{M}_{t|t-1}) - \frac{1}{2} \sum_{t=1}^n \frac{(X_t - \hat{X}_{t|t-1})^2}{\hat{M}_{t|t-1}},$$

여기서 $\hat{X}_{t|t-1}$ 과 $\hat{M}_{t|t-1}$ 은 위의 칼만 필터 알고리즘을 이용하여 구할 수 있다.

3. 중선형 시계열 모형의 동질성 검정

패널 중선형 시계열 모형의 모수 벡터에 대한 동질성검정을 위하여 BL(0, 0, p)에서 $p = 1$ 인 경우, 즉 BL(0, 0, 1, 1) 모형을 고려해 본다.

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

BL(0, 0, 1, 1) 모형에서 m 개의 시계열 모수의 동질성검정에 대한 가설은 다음과 같다.

$$H : b(1) = b(2) = \dots = b(m).$$

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

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

이 된다.

이 가설에 대한 검정 통계량으로서 Wald 검정통계량 T_n 은 다음과 같다.

$$T_n = n\hat{\psi}' \left[C\hat{\Gamma}^{-1}C' \right]^{-1} \hat{\psi},$$

여기서

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

$$C = \begin{bmatrix} I & -I & 0 & \cdots & 0 \\ I & 0 & -I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ I & 0 & 0 & \cdots & -I \end{bmatrix} : (m-1) \times m \text{행렬},$$

$$\Gamma = -n^{-1} \left(\frac{\partial^2 L}{\partial \psi_r \partial \psi_s} \right), \quad r, s = 1, 2, \dots, k,$$

여기서 $\hat{\Gamma}$ 은 Γ 에 ψ 들의 최우추정치를 대입함으로써 얻을 수 있다.

m 개의 시계열이 서로 독립이므로 다음과 같은 결과를 얻을 수 있다.

$$L \begin{pmatrix} [\hat{b}(1) - b(1)] \\ \vdots \\ [\hat{b}(m-1) - b(m)] \end{pmatrix} \rightarrow N(0, \Gamma^{-1}). \quad (3.1)$$

위 식을 이용하면 귀무가설 H_0 하에서 T_n 의 극한분포는 자유도가 $(m-1)$ 인 카이제곱분포를 따름을 알 수 있고, 유의수준 α 에서의 기각역은 다음과 같다.

$$T_n \geq \chi_{(m-1)}^2(\alpha). \quad (3.2)$$

동질성 검정이 기각되지 않으면 m 개 시계열자료의 평균값을 계산, 한 개의 시계열로 종합하여 모수를 추정하는 것이 타당하다.

4. 실증연구

4.1. 자료 소개

본 논문에서는 한국질병관리 본부에서 수집한 16개 시도의 2001년 1월부터 2008년 12월까지 한국의 월별 Mumps(유행성 이하선염)자료를 8개 도(경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남)로 재분류하여 분석에 사용하였다. 이 자료는 연속형 자료가 아닌 빈도자료로 포아송 분포를 따르는 변수로 고려될 수 있으나 각 도별 모집단 전체 인구에 비해 걸린 환자수가 매우 적은 비율을 차지하므로 포아송 분포의 정규근사는 타당하다. 이에 평균과 표준편차를 계산해 자료를 표준화하였다.

다음 Table 4.1은 우리나라 8도에 관한 Mumps 시계열자료의 일부로 본 논문에 사용되는 자료의 구조는 각 계열(series)의 크기가 96이고 패널의 수가 8개로 구성되어 있다.

우리나라 8도에 대한 원 시계열도를 각각 그려보면 Figure 4.1, Figure 4.2와 같다. 각 지역의 원 시계열도가 시간에 따라 추세를 가지는 경향이 있고 자기상관함수(ACF)를 검토하고 단위근 검정을 실시한 결과 원 자료가 12개월 주기의 계절성을 보여 분산을 안정화 시킨 후 12차 차분을 하여 정상성을 만족시켰다.

Table 4.1. Structure of Korea Mumps Panel data

	2001.1	2001.2	2001.3	...	2008.12
경기(GG)	88	32	85	...	310
강원(GW)	3	4	3	...	9
충북(CB)	0	0	0	...	15
충남(CN)	2	0	11	...	3
전북(JB)	18	5	11	...	1
전남(JN)	21	16	7	...	5
경북(GB)	16	1	10	...	26
경남(GN)	3	2	8	...	31

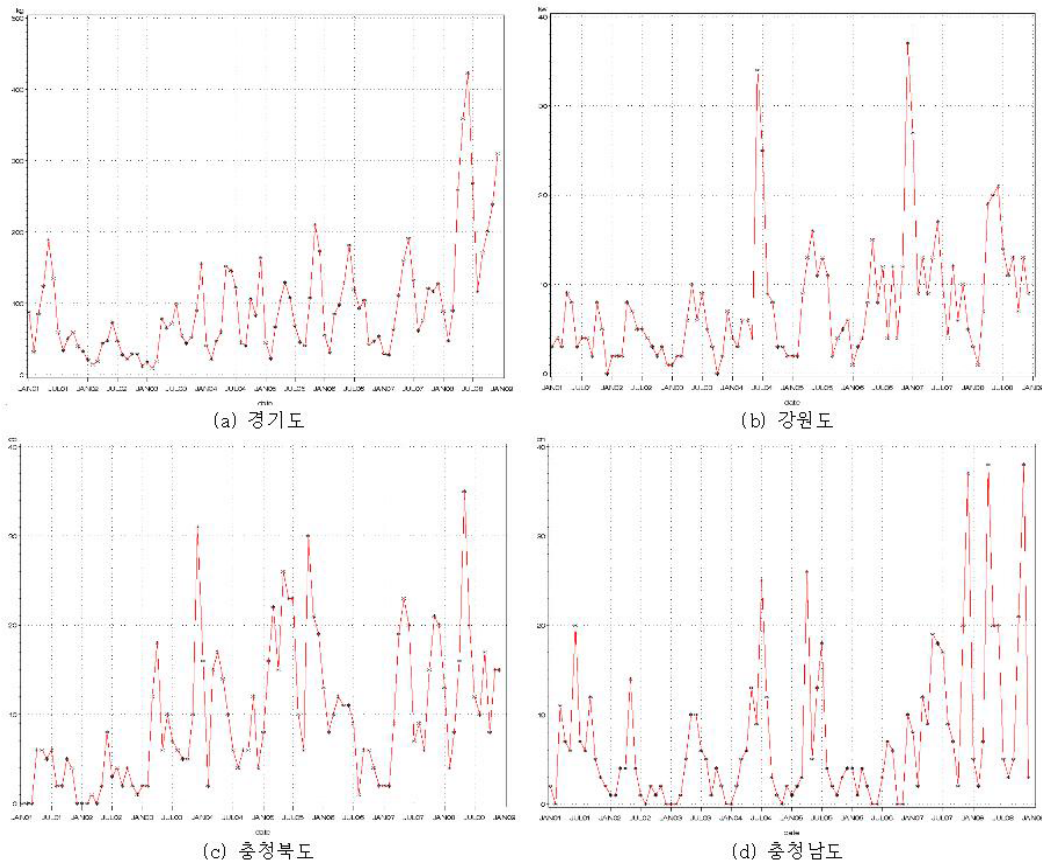


Figure 4.1. Time series plot of Korea 8do Mumps data

다음 Table 4.2는 확률적 추세가 있는지 알아보기 위한 단위근 검정의 결과이다. 8개 변수 모두 유의수준 1%에서 유의하므로 12차 차분이 필요함을 알 수 있다. 시계열 정상화를 위해 분산 안정화 변환 후 12차 차분을 하였고, 12차 차분 후의 단위근 검정에서는 유의수준 1%에서 단위근은 존재하지 않아 정상 시계열이 되었음을 확인하였다.

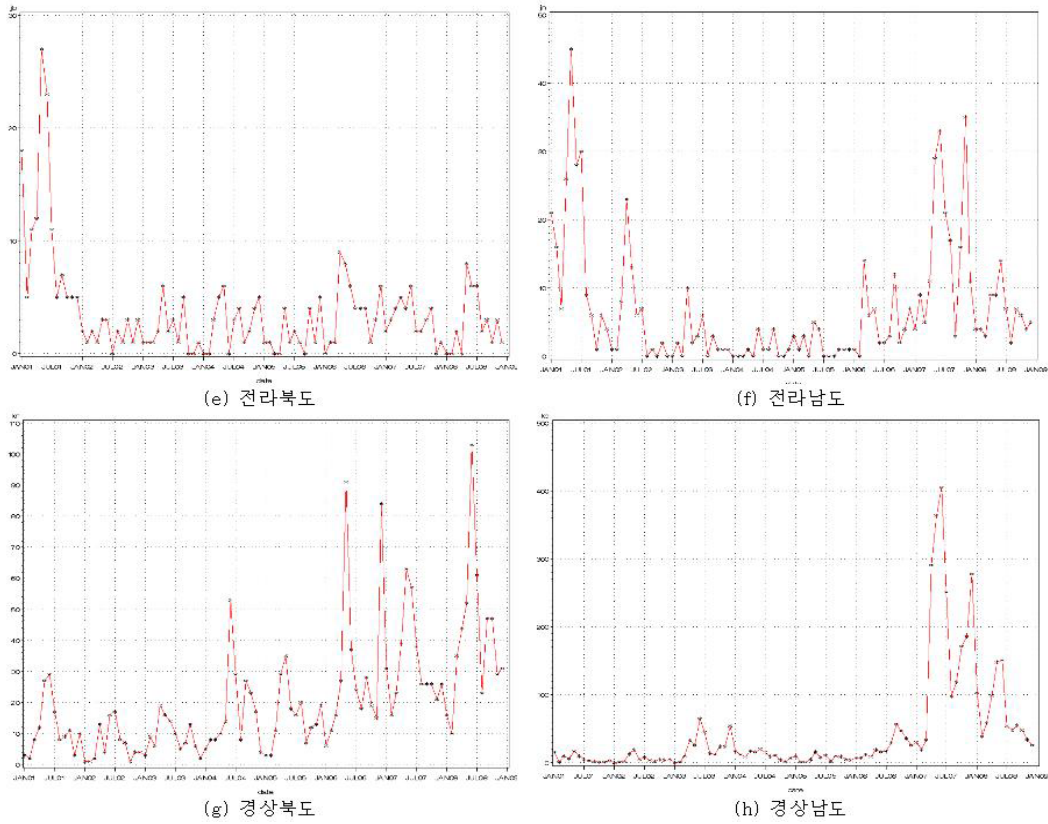


Figure 4.2. Time series plot of Korea 8do Mumps data

Table 4.2. Unit root test

	GG	GW	CB	CN	JB	JN	GB	GN
Tau	1.948	-2.256	-2.908	-2.970	-3.052	-2.246	-1.873	-3.089
p-value	0.309	0.188	0.049	0.042	0.035	0.192	0.343	0.031

4.2. 동질성 검정

다음의 가설에 대해 우리나라 8도 Mumps자료의 동질성 검정을 수행하였다.

$$H_0 : b(1) = \dots = b(8),$$

$$H_1 : 적어도 하나는 다르다.$$

다음의 Table 4.3은 SAS/IML을 이용하여 각 지역에 대한 BL(0, 0, 1, 1) 모형의 8개 모수의 추정값과 동질성 검정을 위한 Wald 통계량 $T_n = n\hat{\psi}'[C\hat{\Gamma}^{-1}C']^{-1}\hat{\psi}$ 이다.

위의 Table 4.3을 보면 유의수준 5%에서의 기각역에 대해 $T_n \leq \chi^2_{(k-1)p}(\alpha)$ 를 만족하므로 지역별로 중선형 시계열 모형의 모수가 같다는 귀무가설을 기각하지 못한다. 따라서 우리나라 8도에 대한 Mumps 발병 모형은 동질성을 만족한다고 할 수 있다.

Table 4.3. Parameter estimates and Wald test statistics

	$\hat{b}(k)$	Wald 통계량	기각역
GG	-0.0864		
GW	-0.1068		
CB	-0.1157		
CN	-0.1156	4.0798	14.0671
JB	-0.0957		
JN	0.0635		
GB	0.0671		
GN	0.0611		

Table 4.4. Comparison of SSE

	개별모형	패널모형
GG	1.9613	1.9385
GW	5.2142	3.6640
CB	3.8479	2.2869
CN	2.9936	2.0957
JB	2.3207	1.6398
JN	0.8807	1.8805
GB	2.0871	3.2451
GN	1.0739	2.3991
SSE	20.3794	19.1496

다음의 Table 4.4는 각 지역들의 Mumps 발병을 각각의 중선형모형으로 예측한 값과 귀무가설 하에서 종합(pooling)한 패널 중선형모형으로 예측한 값을 실제값과 비교하여 예측오차를 구한 표이다. 즉, $\hat{\epsilon}_t(l) = X_{t+l} - \hat{X}_t(l)$ 을 구하였고, 여기에서 X_{t+l} 은 $t+l$ 시점에서 실제값이며 $\hat{X}_t(l)$ 은 l 시차 후의 예측값으로 그 차이가 예측오차(prediction error)가 된다. 여기에서는 예측력 비교를 위해 오차제곱합(Sum of Squared Error; SSE)을 이용하여 비교하였다.

$$SSE = \sum_{k=1}^m \sum_{t=1}^n \hat{\epsilon}_{t,k}(l).$$

Table 4.4를 보면 개별모형에서의 예측오차 값보다 패널모형으로 종합(pooling)한 모형으로 예측하였을 때 더 좋은 결과를 얻을 수 있음을 알 수 있다. 즉, 동질성을 만족하는 시계열은 패널모형으로 자료를 처리하는 것이 모수추약에 충실할 뿐만 아니라 보다 좋은 예측값을 얻을 수 있다.

5. 결론

시계열 모형에서 모수의 수가 많으면 모수추정에 따르는 오차가 커지게 되므로 정확한 예측이 어려운데 여러 시계열자료들에 대해 동질성이 만족되면 모수를 추약할 수 있고, 더 좋은 예측값을 얻을 수 있다. 본 논문에서는 중선형 패널 시계열자료의 동질성검정을 위하여 Wald 통계량을 제안하였고, 그 극한분포가 정상성과 에르고딕 조건하에서 분포를 따름을 보였다. 실증분석에서는 중선형 시계열자료의 동질성 검정을 위해 우리나라 8도 Mumps 자료를 이용하여 검정한 결과 8개 지역의 발병 추이가 같음을 보였다.

References

- Anderson, T. W. (1978). Repeated measurements on autoregressive process, *Journal of the American Statistical Association*, **73**, 371–378.
- Basawa, I. V., Billard, L. and Srinivasan, R. (1984). Large-sample tests of homogeneity for time series models, *Biometrika*, **71**, 203–206.
- Granger, C. W. J. and Andersen, A. P. (1978). *An Introduction to Bilinear Time Series Model*, Vandenhoeck and Ruprecht, Gottingen.
- Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton.
- Lee, S. D. (1993). Test of homogeneity for a panel of seasonal autoregressive processes, *Journal of Korean Statistical Society*, **22**, 125–132.
- Liu, J. (1990). A note on causality and invertibility of a general bilinear time series model, *Advances in Applied Probability*, **22**, 247–250.
- Pham, T. D. and Tran, L. T. (1981). On the first order bilinear time series model, *Journal of Applied Probability*, **18**, 617–627.
- Subba Rao, T. and Gabr, M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Model*, Springer-Verlag, New York.
- Tong, H. (1990). *Non-linear Time Series*, Oxford University Press, Oxford.

패널 중선형 시계열 모형의 동질성 검정

이신형^a · 김선우^b · 이성덕^{b,1}

^a경기도용노동지청 지역협력과, ^b충북대학교 정보통계학과

(2013년 4월 25일 접수, 2013년 5월 21일 수정, 2013년 5월 21일 채택)

요약

패널 시계열자료 분석에서 모수축약의 원칙에 충실하기 위해서 동질성 검정을 수행한다. 본 논문에서는 독립적인 중선형 시계열 패널 자료의 동질성 검정을 수행하기 위하여 먼저 중선형 시계열 모형의 정상성 조건을 구하고 최우추정량과 동질성 검정통계량과 극한 분포를 이끌어내며, 실증분석으로 우리나라 8도의 Mumps 패널자료를 이용해 8개 지역의 발병 추이에 대한 동질성 검정을 수행한다.

주요어: 중선형 시계열모형, 동질성 검정, 극한분포, Mumps 패널자료.

이 논문은 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2012R1A1A4A01013867).

¹교신저자: (361-763) 충북 청주시 흥덕구 내수동로 52, 충북대학교 정보통계학과, 교수.

E-mail: sdlee@chungbuk.ac.kr