응용통계연구 제 **13**권 1호, 2000년, pp. 97-104

일차 비선형 시계열 패널자료의 확률계수 동질성 검정

김인규[1) 황선영[2) 이성덕[3)

요 약

본 논문은 m개의 독립적인 일차 비선형 시계열로 구성된 패널자료의 동질성 검정에 대한 연구로서 먼저 일반적인 일차 비선형 시계열의 정상성 조건을 유도하고 이어서 동질성 검정법을 제시하고 연관된 극한분포를 규명하였다. 또한 모의실험을 하여 제안된 검정법의 모의검정력을 구하였다.

주요용어: 확률계수 자기 회귀 모형, 동질성 검정, Wald 통계량, 비선형 시계열 패널자료.

1. 서론

여러개의 독립적인 시계열로 구성된 시계열 패널 자료는 실제 데이터 분석시 자주 나타나고 있다. 예를 들어 여러명(m)의 혈압을 여러시점에 걸쳐 기록하는 경우에 개인별 자료는 서로 독립이라 할 수 있고, 각 개인의 혈압기록은 종속구조를 가진 시계열 자료로 모형화 할 수 있다. 이러한 자료를 분석할 때 먼저 m개 시계열의 동질성을 검정하여 가능하면 m개 시계열을 종합(pooling)하여 연관된 모수를 추정하고 검정 하는 절차를 밟는 것은 모수절약의 원칙에 충실한 시계열분석방법으로 수많은 개개의 시계열(individuals)을 축약할 수 있다. 선형 ARMA 모형의 패널 자료에 대한 동일성 검정은 Anderson(1978)에 의해 연구되었으며 Lee(1993)은 계절 ARMA 모형 패널 자료의 동일성 검정과 연관된 추정문제에 대해 연구한 바 있다. 비선형모형중 고정계수(fixed coefficient)모형의 패널자료는 Hwang과 Basawa(1994)의 연구결과가 보고되어 있다. 본 연구에서는 일차 비선형 시계열 패널자료의 확률계수들의 동질성 검정을 수행하고자 한다. $X_t(i)$를 i번째 ($i = 1, 2, \cdots, m$) "사람" (individual series)의 t시점($t = 1, 2, \cdots, n$)자료라 하고 다음 모형을 고려하자.

$$X_t(i) = F(X_{t-1}(i), Z_t(i); \theta(i)) + \epsilon_t(i) \tag{1.1}$$

여기서 F는 알려진 실수값을 갖는 함수 형태를 가지며 $Z_t(i)$는 $\theta(i)$의 확률성을 부여하는 관측불가능한 $p \times 1$확률벡터이며, $\theta(i)$는 $p \times 1$모수벡터 그리고 $\{Z_t(i)\}$와 $\{\epsilon_t(i)\}$는 각각 $i.i.d$확률벡터와 확률변수로서 평균이 0, 그리고 분산(공분산) $\sum_z$와 σ_ϵ^2을 가진다. 식(1.1)은 기존의 많은 비선형모형을 포함하는 포괄적인 모형이며(c.f Tong(1990), chap 6.) F가 선형인 경우 이식은 일차 확률계수 자기 회귀 모형(random coefficient autoregression model ;

1) (300-715) 대전시 동구 자양동 우송정보대학 전자정보계열 조교수
2) (140-742) 서울시 용산구 청파동 숙명여자대학교 통계학과 부교수
3) (360-763) 충북 청주시 흥덕구 개신동 충북대학교 통계학과 교수
E-mail: sdlee@cbucc.chungbuk.ac.kr

RCA, Nicholls and Quinn(1982))을 나타내며 F를 적절히 선택하므로서 여러 가지 유용한 확률계수 비선형 모형을 만들어 낼 수 있다.

본 논문의 구성은 다음과 같다. 2절에서는 식(1.1)의 정상적 조건(stationality condition)에 대해 알아보도록 한다. 대부분의 시계열(패널)자료의 경우는 추세나 계절성을 가지고 있다. 이때 추세와 계절성을 제거(detrend and deseasonality)하면 정상시계열 $X_t(i)$를 얻을 수 있다.(Brockwell and Davis(1992), ch.1.) 따라서 $X_t(i)$가 정상시계열을 따른다고 해도 일반성을 잃지 않을 것이며 본 연구에서는 정상시계열 모형만을 고려하기로 한다. 3절에서는 m개 시계열의 동질성 검정, 즉,

$$H : \theta(1) = \theta(2) = \cdots = \theta(m) \tag{1.2}$$

절차에 대해 연구하도록 한다. 한가지 언급하고 싶은 것은 식(1.2)가 성립한다고 해서 m개 시계열이 동일한 것은 아니며, 모수 $\theta(i)$의 확률성을 나타내는 $Z_t(i)$가 남아 있어 i번째 개인만의 특성(특히 조건부 분산의 이질성)을 유지할 수 있다. $\theta(i)$는 어떤 면에서 확률계수의 기대값이며 이런 의미에서 식(1.1)에서의 검정(1.2)는 실용적이라 하겠다. 따라서 우리는 식(1.2)검정을 동일성(equality)검정이라 하지 않고 동질성(homogeneity)검정이라 한다. 4절에서는 연구된 검정 통계량의 검정력을 계산하기 위하여 모의실험을 실시한다.

2. 모형의 정상성 조건

식(1.1)에서 i를 없앤 다음과 같은 모형을 먼저 고려하자.

$$X_t = F(X_{t-1}, Z_t; \theta) + \epsilon_t \tag{2.1}$$

$\{Z_t\}$와 $\{\epsilon_t\}$의 $i.i.d$성질에 의해 $\{X_t\}$는 실수값을 갖는 마코프과정(real valued Markov process)이고 $P(A,x)$가 정상추이확률(stationary transition probability)을 나타낼 때, 즉, 실수집합 R상의 보렐셋 A에 대해 다음이 성립한다.

$$\begin{aligned}(P(A,x) &= P(X_t \in A | X_{t-1} = x) \\ &= P(F(X_{t-1}, Z_t; \theta) + \epsilon_t \in A | X_{t-1} = x) \qquad (2.2) \\ &= P(F(x, Z_t; \theta) + \epsilon_t \in A).\end{aligned}$$

Tweedie(1975)는 일반적인 마코프과정의 정상성 및 에르고딕성(ergodicity)의 조건에 대해 연구한 바 있으며, 이 결과를 이용해 식(2.1)의 정상성 조건을 유도하기 위해 다음의 조건을 고려한다.

$(C.1)$ (i) F는 연속함수(*real valued continuous function*)이다.
(i i) $\{\epsilon_t\}$는 확률밀도함수 $f_\epsilon(\cdot)$를 가지며 $E|\epsilon_t| < \infty$이다.
(iii) $-\infty < x \neq 0 < \infty$에 관한 함수

$$\tau(x) = |x|^{-1}E[\ |F(x, z_t; \theta)|\ ] \tag{2.3}$$

는 유계(bounded)함수이며, $|x| \geq x_0$인 모든 x에 대해 $\tau(x) < 1$인 x_0가 존재한다.

참 고 : 1. 조건(iii)의 $E(\cdot)$는 Z_t에 관한 기대값이며따라서 확률계수 Z_t가 없는 경우에는 기대값이 필요치 않으며 x_0는 모수 θ에 의존할 수 있다.
2. 기존의 선형 $AR(1)$모형 : $X_t = \theta X_{t-1} + \epsilon_t$에서는 $\tau(x) = |\theta|$이므로 조건(iii)은 잘 알려진 $|\theta| < 1$이 된다.

정리 2.1 조건 (c.1)하에서 얻은 식(2.1)의 시계열 $\{X_t\}$는 정상(stationary)이며 에르고딕(ergodic)다.

증명: Tweedie(1975)의 정상성 조건은 다음의 3가지로 구성되어 있다.
(1) X_t는 적절한 실수집합 R의 σ-finite measure ϕ에 대해 ϕ-irreducible이다.
(2) 정상추이확률 $P(A, x)$는 x에 대해 강연속(strongly continuous)이다.
(3) 다음조건을 만족시키는 양수 α와 β가 존재한다.

$$\gamma_x = E[|X_t| \mid X_{t-1} = x] - |x|, \ -\infty < x < \infty \tag{2.4}$$

일 때 $|x| > \alpha$인 모든 x에 대해 $\gamma_x \leq -\beta$이며 $|x| \leq \alpha$인 모든 x에 대해 γ_x는 위로유계(bounded above)이다.

먼저(1)과 (2)에 대해서는, 조건 (C.1)의 (i)과 (ii)에 의해서 ϕ를 Lebesque-measure로 선택하면 $\{X_t\}$는 ϕ-irreducible(Tweedie(1975)참조)하며, 정상추이확률 $P(A, x)$는 식(2.2)로부터

$$P(A, x) = E[P\{F(x, Z_t; \theta) + \epsilon_t \in A | Z_t\}]$$

이므로 모든 $Z_t = z_t$에 대해

$$P\{\epsilon_t \in A - F(x, z_t; \theta)\}$$

가 x에 대해 연속임을 증명하면 된다. 이는 F가 연속함수이고 $\{\epsilon_t\}$가 확률밀도함수를 가짐으로 자명하다.

이제 (3)을 증명하기 위해서 다음식을 생각해 보자.

$$\begin{aligned} \gamma_x &= E[|F(x, Z_t; \theta) + \epsilon_t|] - |x| \\ &\leq E|F(x, Z_t; \theta)| + E|\epsilon_t| - |x| \\ &= |x||x|^{-1}E|F(x, Z_t; \theta)| - |x| + E|\epsilon_t| \\ &= |x|\{\tau(x) - 1\} + E|\epsilon_t| \end{aligned} \tag{2.5}$$

따라서 조건(iii)을 이용해 α를 다음과 같이 선택하면, 즉

$$\alpha = \max[x_0, (\beta + E|\epsilon_t|)/(1 - \tau(x_0))] \tag{2.6}$$

Tweedie의 조건(3)을 증명할 수 있다.

앞으로는 함수 F가 (C.1)조건을 만족한다고 가정하도록 한다.

3. 동질성 검정통계량

주어진 자료가 $\{X_t(i); t = 1, 2, \cdots, n.\ i = 1, 2, \cdots, m\}$가 총 nm개의 관측치로 구성되어 있는 경우, i번째 시계열자료 $\{X_1(i), \cdots, X_n(i)\}$의 우도함수를 $L_i(\theta(i))$라 표시할 때 전체자료에 대한 우도함수 L은

$$L = \prod_{i=1}^{m} L_i(\theta(i)) \tag{3.1}$$

으로 표현된다.

이제 i를 고정시키고 $\{X_1(i), \cdots, X_n(i)\}$를 이용해서 $\theta(i)$를 추정해보자. 초기치 $X_1(i)$를 고정시키면 대수 우도함수(log-likelihood function) $l_i(\theta(i)) = log[L_i(\theta(i))]$는 다음과 같다.

$$l_i(\theta(i)) = \sum_{t=2}^{n} log[g(x_t(i))|x_{t-1}(i); \theta(i)] \tag{3.2}$$

여기서 g는 $x_{t-1}(i)$가 주어진 경우 $X_t(i)$의 조건부 확률밀도함수로서 다음과 같다.

$$\begin{aligned} &g[x_t(i)|x_{t-1}(i); \theta(i)] \\ &= \int f_\epsilon(x_t(i)) - F(x_{t-1}(i), Z_t(i); \theta(i))\ dM(Z_t(i)) \end{aligned} \tag{3.3}$$

윗 식에서 M은 $Z_t(i)$의 분포함수이며 $\epsilon_t(i)$의 확률밀도함수는 $f_\epsilon(\cdot)$으로 표시하였다. 식(3.2) 대수우도함수의 ($\theta(i)$에 관한)일차미분값을 $S_i(\theta(i)); p \times 1$벡터, 그리고 이차 미분값에 마이너스 부호를 붙여서 얻은 헤시안 행렬을 $H_i(\theta(i)); p \times p$행렬로 표시하자. Hwang과 Basawa(1993)은 식(1.1)을 따르는 확률계수 비선형시계열에서 $\theta(i)$의 one-step MLE $\hat{\theta}(i)$를 S_i와 H_i의 함수형태로 제시하였으며 이는 $S_i(\theta(i)) = 0$을 풀어서 얻은 최대우도추정량(MLE)와 동일한 극한분포를 가짐을 증명하였으며 정규분포가 아닌경우 MLE의 유도가 불가능하거나 혹은 매우 어려운 반면 one-step MLE는 S_i와 H_i의 함수형태로서 비교적 쉽게 얻을 수 있음을 서술하였다. 또한 n이 무한대로 접근할 때, one-step MLE (와 MLE)의 극한분포는 다음과 같음을 제시하였다.

$$\sqrt{n}\ [\hat{\theta}(i) - \theta(i)] \xrightarrow{d} N(0, \Gamma_i^{-1}(\theta(i))). \tag{3.4}$$

여기서 $\Gamma_i(\theta(i)) = Plim[n^{-1}H_i(\theta(i))]$: $p \times p$행렬이며, 이것의 존재성은 에르고딕 정리(ergodic theorem)에 의해 확인된다.

m개의 시계열이 서로 독립이므로 다음 결과를 얻을 수 있다.

$$\sqrt{n} = \begin{pmatrix} [\hat{\theta}(1) - \theta(1)] \\ \vdots \\ [\hat{\theta}(m) - \theta(m)] \end{pmatrix} \xrightarrow{d} N(0, \Gamma^{-1}) \tag{3.5}$$

여기서 위의 식(3.4)와의 일관성을 위하여 좀더 직접적으로 Γ^{-1}는 $mp \times mp$행렬로써 다음과 같이 주어진다.

$$\Gamma^{-1} = Diag[\Gamma_1^{-1}(\theta(1)), \cdots, \Gamma_m^{-1}(\theta(m))].$$

이제 m개 시계열의 검정을 위한 가설(1.2)를 고려하자. 즉,

$$H : \theta(1) = \cdots = \theta(m)$$

이 가설은 새로운 기호 $\psi_i = \theta(i) - \theta(i+1), i = 1, \cdots, m-1$를 이용하면 다음가설과 동치이다.

$$H^* : \psi_1 = \cdots = \psi_{m-1} = 0 \tag{3.6}$$

한편 이 가설에 대한 검정통계량으로써 Wald 통계량 T_n은 다음과 같이 정의된다.

$$T_n = n\hat{\psi}'[C\hat{\Gamma}^{-1}C']^{-1}\hat{\psi} \tag{3.7}$$

여기서

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

그리고,

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

식(3.5)를 이용하면 귀무가설 H^*(식(3.6)) 하에서 T_n의 극한분포는 자유도가 (m-1)p인 카이제곱분포임을 쉽게 알 수 있다. 따라서 유의수준 α에서의 기각역은 다음과 같다.

$$\{T_n \geq \chi^2_{(m-1)p}(\alpha)\} \tag{3.8}$$

동질성 검정이 기각되지 않으면 m개 시계열자료를 종합해서 θ에 관한 추정을 하는 것이 타당하며 제안된 Wald 통계량 T_n은 적당한 조건하에서 우도비 검정통계량 그리고 스코어 검정통계량과 귀무가설하에서 동일한 극한분포를 가진다.(c.f. Hwang 과 Basawa(1993))

4. 모의검정력 계산

본 절에서는 식(3.7)에서 제안된 검정통계량 T_n의 검정력을 모의실험을 통해 알아본다. 논의의 간편성을 위해(m=2)개의 독립적인 시계열로 구성된 자료로서 각 시계열은 정상 일차 확률계수 자기회귀(RCA)모형을 따르도록 하였다.(p=1) :

$$X_t = (\theta + Z_t)X_{t-1} + \epsilon_t \tag{4.1}$$

여기서 ϵ_t는 $N(0,1)$로부터 얻은 난수이며 모형의 정상성 조건 $\theta^2 + EZ_t^2 < 1$을 고려하여 $\{Z_t\}$는 $N(0, 0.2)$를 따르도록 하였다. 따라서 가능한 θ는 $\theta^2 < 0.8$ 이다.

이제 $i = 1, 2$에 대해

$$X_t(i) = (\theta(i) + Z_t(i))X_{t-1}(i) + \epsilon_t(i)$$

에서 $\theta(i)$의 최대우도 추정량 $\hat{\theta}(i)$는 다음과 같이 구할 수 있다.

$$\hat{\theta}(i) = \frac{\sum_{t=2}^{n} X_t(i)X_{t-1}(i)/[0.2X_{t-1}^2(i)+1]}{\sum_{t=2}^{n} X_{t-1}^2(i)/[0.2X_{t-1}^2(i)+1]} \ , \ i = 1, 2. \tag{4.2}$$

또한 $\hat{\theta}(i)$의 극한분포의 분산 $\Gamma_i(\theta(i))$는 다음과 같다.

$$\Gamma_i^{-1}(\theta(i)) = E\left[\frac{X_1^2(i)}{[0.2X_1^2(i)+1]}\right] \tag{4.3}$$

$\theta(2) = 0$인 경우 $\theta(1)$ 과 $\theta(2)$의 차이인 $\delta = \theta(1) - \theta(2) = \theta(1)$ 을 변화시키면서 길이 $n = 500$인 독립적인 두 개의 모의시계열자료를 생성시켜 식(4.2)를 이용하여 $\hat{\theta}(1)$과 $\hat{\theta}(2)$을 구하였다. 이제 검정통계량 T_n은

$$T_n = 500\hat{\psi}'(C\hat{\Gamma}^{-1}C')^{-1}\hat{\psi}$$

이며 여기서 $\hat{\psi} = \hat{\theta}(1) - \hat{\theta}(2)$, $C = (1, -1)$ 그리고 $\hat{\Gamma}$은 Γ 의 일치추정량(strongly consistent estimator)로서 에르고딕 정리와 식(4.3)을 이용하여 다음과 같이 계산하였다.

$$\begin{aligned} \hat{\Gamma} &= Diag(\hat{\Gamma}_1, \hat{\Gamma}_2) \\ \hat{\Gamma}_i &= \textstyle\sum_{t=2}^{n}\{\frac{X_{t-1}^2(i)}{[0.2X_{t-1}^2(i)+1]}\}/499 \ , i = 1, 2. \end{aligned} \tag{4.4}$$

이 과정을 1000번 반복(replication)하여 1000번중 $\chi_1^2(0.05) = 3.843$을 넘는 비율을 계산하여 표 4.1과 같은 모의 검정력 결과를 얻었다.

표 4.1: 모의검정력(%) : $\theta(2) = 0$, $\delta = \theta(1)$

$\sigma^2 \backslash \delta$	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

표 4.1로부터 δ값이 커질수록, 즉 $\theta(1)$ 과 $\theta(2)$의 차이가 점점 커질수록 검정력이 좋아짐을 알 수 있을 뿐만 아니라 경험적 유의수준도 적절히 잘 조절되는 것으로 보여 우수한 검정결과를 얻을 수 있었다.

5. 결론

시계열 모형을 연구함에 있어서 모수축약의 원칙은 회귀분석에 못지 않게 중요한 원칙이다. 모수의 수가 많으면 모수추정에 따르는 오차가 커지게 되므로 예측을 정확하게 할 수가 없게 된다. 여러 개의 시계열 자료들이 동일한 모형으로부터 얻어졌다고 하는 동질성 가설이 채택 된다면 모형설정에 있어 모수축약을 이룰 수 있고, 또한 결합된 자료에 의해 보다 나은 모수추정, 나아가 더 좋은 예측값을 얻을 수 있을 것이다.

본 논문은 여러 개의 독립적인 비선형 시계열로 구성된 패널자료의 동질성 검정을 하기 위하여 Wald 통계량을 제안하였고 그 극한분포가 정상성과 에르고딕 조건하에서 χ^2분포를 따름을 보였으며 모의실험을 통하여 정상 일차 확률계수모형들이 동질성 검정을 수행한 결과, 경험적 유의수준 및 검정력이 모두 우수함을 보였다.

참고문헌

[1] Anderson, T.W.(1978). Repeated measurements on autoregressive processes. *Journal of American Stat. Association*, Vol.73, 371-378.

[2] Brockwell, P.J. and Davis, R.A.(1992). *Time Series: Theory and Methods*, second ed. Springer, New York.

[3] Hwang, S.Y. and Basawa, I.V. (1993). Asymptotic optimal inference for a class of nonlinear time series. *Stochastic Processes and Their Applications*, Vol.46, 91-113.

[4] Hwang, S.Y. and Basawa, I.V. (1994). Large sample inference based on multiple observations from nonlinear autoregressive processes. *Stochastic processes and Their Applications*, Vol.49, 127-140.

[5] Lee, S.D. (1993). Test of homogeneity for a panel of seasonal autoregressive processes. *Journal of the Korean Statistical Society*, Vol 22, No. 1, 125-132.

[6] Nicholls, D.F. and Quinn, B.G. (1982). *Random Coefficient Autoregressive Models : An Introduction*, Lecture Notes in Statistics, Vol.11, Springer, New York.

[7] Tong, h. (1990). *Nonlinear Time Series*, Oxford.

[8] Tweedie, R.L. (1975). Suficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stochastic Processes and Their Applications*, Vol.3, 385-403.

[1999년 7월 접수, 1999년 12월 채택]

Homogeneity Test of Random Coefficient for the First Order Nonlinear Time Series Panel Data

Inkyu Kim[1) Sunyoung Hwang[2) Sungduck Lee[3)

Abstract

A Large sample test of homogeneity for a panel of more than two first order nonlinear time series processess is derived and its limiting distribution is found. We present the results of a Monte Carlo simulation for the proposed test.

Keywords: Random coefficient autoregression model; Homogeneity test; Wald Statistic; Nonlinear time series panel data.

1) Assistant Professor, Dept. of Computer Science, Woosong Information College.
2) Associate Professor, Dept. of Statistics, Sookmyung womens Univ.
3) Professor, Dept. of Statistics, Chungbuk National Univ. E-mail: sdlee@cbucc.chungbuk.ac.kr