한 국 통 계 학 회 논 문 집
2009, 16권, 4호, 587–594

종속 오차에 대한 분포 변화 검정법

나성룡[1,a]

[a]연세대학교 정보통계학과

요 약

이 논문에서는 선형회귀모형의 오차항에 대한 변화점 검정 문제를 다룬다. 고정 혹은 변동 모형의 독립변수와 약한 종속성을 가지는 오차항을 가정하는 관계로 통상적인 중회귀모형뿐만 아니라 ARMA 등의 시계열 모형까지 본 논문에서 포괄한다고 하겠다. 오차항의 분포 변화를 검정하기 위하여 회귀모형의 잔차에 기초한 확률밀도함수 추정값을 이용한다. 적절한 가정하에서 잔차를 이용한 검정이 실제 오차를 이용한 경우와 동일한 극한 분포를 가짐을 보였다.

주요용어: 변화점, 시계열, 회귀모형, 종속 오차, 강혼합, 확률밀도함수 추정.

1. 서론

변화점에 대한 통계학적인 검토는 수집된 자료의 정확한 분석을 위하여 필수적으로 요구되는 경우가 많다. 여러 이유로 통계자료에 변화점이 존재할 수 있으며, 특히 외적 환경 요소가 특정 시점에 변하는 경우에 시계열자료의 통계적 특성이 변화하는 경우가 많다. 적합된 모형의 변동성이 지나치게 큰 경우에는 변화점의 존재를 고려할 수 있다. 이때 통계적으로 유의한 적절한 수의 변화점은 전체 자료를 안정적인 부분 자료로 분할하는 역할을 하게 된다. 각기 분할된 자료에 가장 적합한 모형을 적합할 수 있고, 이를 통하여 전반적인 변화를 적절히 설명할 수 있게 된다. 변화점은 자료의 비정상성을 설명하는 방법으로 사용되기도 하는데 변화점의 도입으로 정상 시계열이 생성되고 다양한 분석기법이 비로소 가능하게 된다.

Bai (1994)는 ARMA 모형에 대한 분포 변화를 검정하는 문제를 다루면서 잔차에 기초한 순차적 경험확률과정이 Kiefer 확률과정으로 수렴함을 증명하였다. 반면 Ling (1998)과 Na 등 (2006) 등을 보면 잔차의 순차적 경험확률과정의 극한 분포가 모수 추정과 오차의 종속성에 영향을 받음을 알 수 있다. 특히 오차항에 종속성이 존재하면 오차를 직접 이용하는 경우에도 경험확률과정의 극한 성질이 종속성에 영향을 받는다 (가령, Billingsley, 1999). 따라서 분포함수 추정값에 기초한 분포무관 검정법을 구하기가 어렵다.

이러한 분포함수 추정값에 기초한 검정의 단점을 극복하는 방법으로 확률밀도함수 추정값을 이용하는 것을 고려할 수 있다. Takahata와 Yoshihara (1987), Lee와 Na (2002) 등은 시계열자료의 적합도 검정에 확률밀도함수 추정값을 이용한다. 한편 강혼합 성질을 가지는 시계열자료의 분포 변화를 검정하기 위해서 분포함수 대신 확률밀도함수 추정값을 이용하는 것을 Lee와 Na (2004)가 제안했고, 확률밀도함수 추정값에 기초한 검정통계량이 자료의 종속성과 무관한 극한분포를 가짐을 보였다.

이 논문에서는 선형회귀모형에서 분포 변화에 대한 검정 문제를 다룬다. 강혼합 오차항을 가정하고 독립변수에 대하여 고정모형과 변동모형 모두를 가정하는데 AR과 ARMA 모형 등이 포함된다. 오차의 분포 변화를 검정하기 위하여 잔차를 이용하며 종속성의 문제를 해결하기 위하여 확률밀도함수

[1] (220-710) 강원도 원주시 흥업면 매지리 234, 연세대학교 정보통계학과, 부교수. E-mail: nasr@yonsei.ac.kr

추정값을 이용한 검정을 고려한다. 적절한 가정이 성립할 때 잔차를 이용한 검정이 실제 오차를 이용하는 것과 동일한 극한 성질을 가짐을 보임으로써 오차항의 종속성에 무관한 검정이 됨을 보인다. 2절에서 모형과 검정통계량을 설명하고, 이 검정통계량의 극한 성질을 연구한다.

2. 분포 변화 검정

2.1. 모형과 확률밀도함수 추정량

다음과 같은 회귀모형을 고려하자.

$$y_t = \mathbf{x}_t'\beta + e_t, \quad t = 1, 2, \ldots, n,$$

$\beta = (\beta_1, \ldots, \beta_p)'$는 $p \times 1$의 회귀모수, $\mathbf{x}_t = (x_{t1}, \ldots, x_{tp})'$는 $p \times 1$의 독립변수, e_t는 약한 종속성을 가지는 오차항, y_t는 종속변수이다. 이때 e_t의 확률밀도함수 g_t에 변화가 있는가를 살피는 것이 이 논문의 분포 변화 검정의 내용이라고 하겠다. g_t에 변화가 없다면 e_t는 정상성을 만족하는 오차항으로 판단한다.

이 논문에서는 강혼합(strong mixing, α-mixing) 오차항을 가정하는데

$$\alpha_k = \sup_{s \ge 1} \sup \left\{ |P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}_1^s, B \in \mathcal{F}_{s+k}^{\infty} \right\}$$

의 강혼합 계수에 대하여 $\lim_{k\to\infty} \alpha_k = 0$이 성립한다. $\mathcal{F}_1^s$, $\mathcal{F}_{s+k}^{\infty}$는 $\sigma(\mathbf{x}_1, e_1, \ldots, \mathbf{x}_s, e_s)$, $\sigma(\mathbf{x}_{s+k}, e_{s+k}, \mathbf{x}_{s+k+1}, e_{s+k+1}, \ldots)$의 두 시그마필드를 나타낸다. 강혼합 성질은 약한 종속성의 일종이며 ARMA 모형 등이 이 성질을 만족한다. 따라서 이 논문은 시계열 오차를 가지는 중회귀모형, 측정 오차를 가지는 자기회귀모형 등을 포괄하고 있다.

변화점 검정에 대한 가설을 다음과 같이 설정한다.

H_0 : 모든 t에 대하여 $g_t = g$

H_1 : 어떤 τ에 대하여 $g_t = g_1$, $t = 1, \ldots, [n\tau]$이고 $g_t = g_2$, $t = [n\tau] + 1, \ldots, n$.

검정에 사용되는 커널 확률밀도함수 추정량은

$$f_n(x) = n^{-1} \sum_{t=1}^{n} K_h(x - e_t), \quad -\infty < x < \infty$$

의 식으로 정의되는데 K는 커널함수, $h = h_n$은 $h \to 0$, $nh \to \infty$을 만족하는 밴드폭이며 $K_h(u)$ $h^{-1}K(u/h)$이다. 확률밀도함수에 대한 커널추정량의 기본적인 성질은 Silverman (1986)과 Bosq (1998) 등을 참고할 수 있다. 한편 순차적 확률밀도함수 추정량은

$$f_{[ns]}(x) = [ns]^{-1} \sum_{t=1}^{[ns]} K_h(x - e_t), \quad 0 \le s \le 1,$$

$$f^*_{n-[ns]}(x) = (n - [ns])^{-1} \sum_{t=[ns]+1}^{n} K_h(x - e_t), \quad 0 \le s \le 1$$

의 식으로 정의되는데, 이들은 각각 g_1, g_2의 추정량으로 이해할 수 있다.

두 순차적 확률밀도함수의 차이로

$$d_n(s, x) = \left(\frac{nh}{f_n(x)\|K\|^2} \right)^{\frac{1}{2}} \frac{[ns]}{n} \left(\frac{n - [ns]}{n} \right) \left(f_{[ns]}(x) - f^*_{n-[ns]}(x) \right), \quad 0 \le s \le 1$$

로 정의하는데, 여기에서 $\|K\|^2 = \int K^2(u)du < \infty$이고 $f_n(x) = 0$이면 $d_n(s,x) = 0$이다. 여러 개의 $x_1,\ldots,x_m$에 대하여 이 값들을 계산하고 이들의 최대값이 일정 수준을 넘어서면 H_0를 기각하고 분포변화를 받아들인다. Lee와 Na (2004)에 식 (2.1)의 극한성질이 잘 유도되어 있다. 즉 오차항이 강혼합 조건을 만족할 때

$$(d_n(s,x_1),\ldots,d_n(s,x_m)) \xrightarrow{w} \left(\bar{B}_1(s)\ldots,\bar{B}_m(s)\right), \quad 0 \le s \le 1 \tag{2.1}$$

의 약 수렴이 서로 독립인 브라운다리 $\bar{B}_1(s),\ldots,\bar{B}_m(s)$에 대하여 성립한다. 식 (2.1)의 결과를 이용하면 기각역을 쉽게 구할 수 있다.

2.2. 잔차에 기초한 검정

회귀모형의 오차는 관측이 불가능한 것이 일반적이며 잔차를 대신 사용할 수 있다. 회귀계수 β의 일치추정량 $\hat{\beta} = (\hat{\beta}_1,\ldots,\hat{\beta}_p)'$에 대하여 잔차는

$$\hat{e}_t = y_t - \mathbf{x}_t'\hat{\beta} = y_t - \hat{\beta}_1 x_{t1} - \cdots - \hat{\beta}_p x_{tp}, \quad t = 1,\ldots,n$$

의 식으로 정의된다. 잔차 $\hat{e}_t$에 기초한 순차적 확률밀도함수 추정량은

$$\hat{f}_{[ns]}(x) = [ns]^{-1}\sum_{t=1}^{[ns]} K_h(x-\hat{e}_t), \quad 0 \le s \le 1,$$

$$\hat{f}^*_{n-[ns]}(x) = (n-[ns])^{-1}\sum_{t=[ns]+1}^{n} K_h(x-\hat{e}_t), \quad 0 \le s \le 1$$

의 식으로 정의되며, 차이 확률과정의 정의는

$$\begin{aligned}\hat{d}_n(s,x) &= \left(\frac{nh}{\hat{f}_n(x)\|K\|^2}\right)^{\frac{1}{2}} \frac{[ns]}{n}\left(\frac{n-[ns]}{n}\right)\left\{\hat{f}_{[ns]}(x) - \hat{f}^*_{n-[ns]}(x)\right\} \\ &= \left(nh\hat{f}_n(x)\|K\|^2\right)^{-\frac{1}{2}}\left\{\sum_{t=1}^{[ns]} K\left(\frac{x-\hat{e}_t}{h}\right) - \frac{[ns]}{n}\sum_{t=1}^{n} K\left(\frac{x-\hat{e}_t}{h}\right)\right\}\end{aligned}$$

이다.

먼저 $g_t(x_i) > 0$인 $x_1,\ldots,x_m$을 고정하자. 검정통계량은

$$T_n = \max_{1\le i\le m}\sup_{0\le s\le 1}\left|\hat{d}_n(s,x_i)\right|$$

이며 T_n의 값이 크면 H_0를 기각하게 된다. 기각역은 $\hat{d}_n(s,x_i)$의 극한분포를 이용해서 정한다.

오차항 e_t가 정상성을 만족하고, 따라서 $H_0 : g_t = g$가 성립할 때, $\hat{d}_n(s,x_i)$에 대하여 식 (2.1)의 약 수렴 성질이 성립하기 위한 조건을 서술한다.

(A1) 어떤 $\gamma > 3$에 대하여 강혼합 계수가 $\alpha_k = O(k^{-\gamma})$.

(A2) $n^{1/2}(\hat{\beta}-\beta) = O_P(1)$.

(A3) g가 두번 미분가능하고

$$\|g\|_\infty = \sup_x g(x) < \infty, \qquad \|g''\|_\infty = \sup_x |g''(x)| < \infty.$$

(A4) e_{t_1}, e_{t_2}의 결합밀도함수 g_{t_1,t_2}와 $e_{t_1}, e_{t_2}, e_{t_3}, e_{t_4}$의 결합밀도함수 g_{t_1,t_2,t_3,t_4}에 대하여

$$\sup_{t_1<t_2} \|g_{t_1,t_2}\|_\infty < \infty, \qquad \sup_{t_1<t_2<t_3<t_4} \|g_{t_1,t_2,t_3,t_4}\|_\infty < \infty.$$

(A5) $\sup_t E(\|\mathbf{x}_t\|^2) = \sup_t E(\sum_{i=1}^p x_{ti}^2) < \infty.$

(A6) 커널함수 K는 대칭인 확률밀도함수이고

$$\|K\|_\infty < \infty, \quad \lim_{u\to\infty} uK(u) = 0, \quad \int_{-\infty}^{\infty} u^2 K(u)du < \infty,$$

$$\int_{-\infty}^{\infty} |uK'(u)|du < \infty, \quad \int_{-\infty}^{\infty} |u^2 K'(u)|du < \infty, \quad \|K''\|_\infty < \infty.$$

(A7) $c_n \to \infty$이고 모든 $\delta > 0$에 대하여 $c_n n^{-\delta} \to 0$인 c_n에 대하여 $h = h_n = c_n n^{-1/5}$.

(A8) $E(K'((x - e_t)/h)|\mathbf{x}_t) = O(h^2)$.

(A1)은 ARMA모형 등에 의하여 만족된다. (A2)는 자기회귀오차를 가정하는 회귀모형, ARMA모형 등의 회귀계수 추정에서 성립하고, 측정오차를 포함하는 ARMA모형에서도 최소제곱추정량에 대하여 성립한다 (가령, Chanda, 1995). 고정된 범위의 고정모형 혹은 정상성을 만족하는 확률모형의 독립변수는 (A5)을 쉽게 만족시킴을 알 수 있다. (A7)의 c_n으로는 $\log n$ 등이 가능하며 확률밀도함수의 추정에서 $n^{-1/5}$는 최적의 차수로 알려져 있다. 커널함수 K의 대칭성에서 (A8)의 가정은 어렵지 않다.

앞의 조건이 성립하고 e_t가 정상성을 가진다면

$$\sup_{0\le s\le 1} (nh)^{-\frac{1}{2}} \left| \sum_{t=1}^{[ns]} \left\{ K\left(\frac{x-\hat{e}_t}{h}\right) - K\left(\frac{x-e_t}{h}\right) \right\} \right| = o_P(1) \tag{2.2}$$

과

$$\hat{f}_n(x) - f_n(x) = o_P(1) \tag{2.3}$$

이 성립한다. 확률밀도함수 추정량의 수렴속도가 $n^{-1/2}$보다 느리다는 잘 알려진 사실을 고려하면 $\hat{\beta}$의 $\sqrt{n}$-일치성을 가정하는 조건 (A2)에서 식 (2.2)와 (2.3)의 결과가 유도될 수 있음은 어느 정도 기대할 수 있다. 이때

$$\begin{aligned}\hat{d}_n(s,x) - d_n(s,x) = &\left\{nh\hat{f}_n(x)\|K\|^2\right\}^{-\frac{1}{2}} \left[\sum_{t=1}^{[ns]} \left\{ K\left(\frac{x-\hat{e}_t}{h}\right) - K\left(\frac{x-e_t}{h}\right) \right\} - \frac{[ns]}{n} \sum_{t=1}^{n} \left\{ K\left(\frac{x-\hat{e}_t}{h}\right) - K\left(\frac{x-e_t}{h}\right) \right\} \right] \\ &+ f_n(x)^{\frac{1}{2}} \left\{ \hat{f}_n(x)^{-\frac{1}{2}} - f_n(x)^{-\frac{1}{2}} \right\} \cdot d_n(s,x)\end{aligned}$$

의 식과 식 (2.2)와 (2.3)의 결과로 다음의 정리를 얻는다.

정리 1. 조건 (A1)–(A8)이 성립함을 가정하자. 그러면 H_0이 성립할 때 $g(x_i) > 0$, $i = 1, \ldots, m$을 만족하는 임의의 $x_1, \ldots, x_m$에 대하여

$$\left(\hat{d}_n(s,x_1), \ldots, \hat{d}_n(s,x_m)\right) \xrightarrow{w} \left(\bar{B}_1(s), \ldots, \bar{B}_m(s)\right), \quad 0 \le s \le 1,$$

이 서로 독립인 브라운다리 $\bar{B}_1(s), \ldots, \bar{B}_m(s)$에 대하여 성립한다.

따라서, 기각역을 결정하기 위한 검정통계량의 분포는 다음과 같이 정리할 수 있다.

정리 2. 위 정리 1과 동일한 조건에서

$$T_n \xrightarrow{d} \max_{i=1,\ldots,m} \sup_{0\le s\le 1} \left|\bar{B}_i(s)\right|$$

이 성립한다.

정리 2의 $\bar{B}_i(s)$의 독립성에 의하여 유의수준 α에서 기각역은

$$P\left(\sup_{0\le s\le 1} \left|\bar{B}_1(s)\right| \ge C_\alpha\right) = 1-(1-\alpha)^{\frac{1}{m}}$$

을 만족하는 C_α에 대하여 $T_n \ge C_\alpha$로 정하면 된다. 즉 브라운다리의 최대값 분포의 분위수를 이용해서 간단하게 결정할 수 있다.

3. 결론

지금까지 선형회귀모형에서의 오차항에 대한 분포변화를 검정하는 방법을 살펴보았다. 특히 오차항이 강혼합의 종속성을 가지는 경우를 가정하였는데 기존의 분포함수 추정에 기초한 방법이 잘 적용되지 않는 경우이다. 시계열자료의 분포변화 검정을 위하여 Lee와 Na (2004)가 제안한 확률밀도함수 추정을 이용한 방법을 이 모형에 적용하였다. 회귀계수 추정에 적절한 일치성이 있다면 잔차를 이용하는 것이 실제 오차를 이용하는 것과 극한분포에서는 차이가 없음을 증명하였다.

확률밀도함수 추정량을 이용한 검정통계량은 오차의 종속성과 무관한 극한분포를 가지는 장점이 있다. 앞에서 기각역은 손 쉽게 주어짐을 살펴보았는데, 적절한 수와 위치의 비교점을 택하여 검정을 실시하면 된다. 또한 잔차에 기초한 방법은 실제 관측값에 높은 상관관계가 존재하는 경우에 관측값을 직접 이용하는 것보다는 효율이 좋음이 알려져 있다.

이 논문이 다루고 있는 모형은 다양한 모형을 포괄하고 있다. 일반적인 중회귀모형이 표현될 뿐 아니라 독립변수를 $\mathbf{x}_t = (y_{t-1},\ldots,y_{t-p})'$로 설정하고 IID의 ϵ_t에 대하여 $e_t = \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}$의 설정을 한다면 ARMA 모형도 표현이 가능하다. 그 밖에도 측정오차를 포함하는 자기회귀모형 등이 표현 가능하다.

실제 검정에서는 m과 비교점 $x_1,\ldots,x_m$의 선택에 신중을 기하여야 한다. 선택된 비교점 중에 변화점이 존재해야만 성공적인 검정이 이루어진다. 그러나 m이 크거나 그로 인해 x_i들이 지나치게 가까우면 $\hat{d}_n(s,x_i)$들의 유한 표본에서의 종속성으로 인하여 검정의 정확도가 떨어질 수 있다. 하지만 정규분포 등과 같이 일반적으로 다루어지는 확률밀도함수의 형태를 고려해보면 m의 값은 3 또는 4이면 충분하고 x_i들의 선택은 크게 문제가 되지 않을 것으로 판단된다. 즉 분포 기저가 연속이면 어떤 x_i를 택해도 되는데 자료의 $1/m$ 분위점을 이용하는 것이 적절할 것이다.

부록

이제 식 (2.2)과 (2.3)에 대한 증명을 하자. 테일러 전개를 하면 $(x-e_t)/h$와 $(x-\hat{e}_t)/h$ 사이의 $\zeta_t(x)$에 대하여

$$(nh)^{-\frac{1}{2}}\sum_{t=1}^{[ns]}\left\{K\left(\frac{x-\hat{e}_t}{h}\right)-K\left(\frac{x-e_t}{h}\right)\right\} = (nh)^{-\frac{1}{2}}\sum_{t=1}^{[ns]}\left\{\frac{e_t-\hat{e}_t}{h}K'\left(\frac{x-e_t}{h}\right)+\frac{(e_t-\hat{e}_t)^2}{2h^2}K''(\zeta_t(x))\right\}$$

$$= n^{-\frac{1}{2}}h^{-\frac{3}{2}}\left(\hat{\beta}-\beta\right)'\sum_{t=1}^{[ns]}\left\{\mathbf{x}_tK'\left(\frac{x-e_t}{h}\right)-E\mathbf{x}_tK'\left(\frac{x-e_t}{h}\right)\right\} + n^{-\frac{1}{2}}h^{-\frac{3}{2}}\left(\hat{\beta}-\beta\right)'\sum_{t=1}^{[ns]}E\mathbf{x}_tK'\left(\frac{x-e_t}{h}\right)$$
$$+ \frac{1}{2}n^{-\frac{1}{2}}h^{-\frac{5}{2}}\sum_{t=1}^{[ns]}\left\{\left(\hat{\beta}-\beta\right)'\mathbf{x}_t\right\}^2 K''(\zeta_t(x))$$

의 식이 성립한다. 강혼합 확률과정의 재구성기술(reconstruction technique) (Doukahn, 1994)과 콜모고로프 최대부등식을 이용해서

$$\begin{aligned}
&\sup_{0\le s\le 1}\left|n^{-\frac{1}{2}}h^{-\frac{3}{2}}\left(\hat{\beta}-\beta\right)'\sum_{t=1}^{[ns]}\left\{\mathbf{x}_tK'\left(\frac{x-e_t}{h}\right)-E\mathbf{x}_tK'\left(\frac{x-e_t}{h}\right)\right\}\right| \\
&\le n^{-\frac{1}{2}}h^{-\frac{3}{2}}\left\|\hat{\beta}-\beta\right\|\cdot\sum_{j=1}^{p}\sup_{0\le s\le 1}\left|\sum_{t=1}^{[ns]}\left\{x_{tj}K'\left(\frac{x-e_t}{h}\right)-Ex_{tj}K'\left(\frac{x-e_t}{h}\right)\right\}\right| \\
&= O_P\left(n^{-\frac{1}{2}}h^{-\frac{3}{2}}n^{-\frac{1}{2}}(nh)^{\frac{1}{2}}\right) \\
&= o_P(1)
\end{aligned}$$

의 결과를 얻는다. 또한 (A8) 등의 조건을 이용하면

$$\begin{aligned}
\sup_{0\le s\le 1}\left|n^{-\frac{1}{2}}h^{-\frac{3}{2}}\left(\hat{\beta}-\beta\right)'\sum_{t=1}^{[ns]}E\mathbf{x}_tK'\left(\frac{x-e_t}{h}\right)\right| &\le n^{-\frac{1}{2}}h^{-\frac{3}{2}}\left\|\hat{\beta}-\beta\right\|\max_{1\le t\le n}\left\|E\mathbf{x}_tK'\left(\frac{x-e_t}{h}\right)\right\|\cdot n \\
&= O_P\left(n^{-\frac{1}{2}}h^{-\frac{3}{2}}n^{-\frac{1}{2}}h^2 n\right) \\
&= o_P(1)
\end{aligned}$$

이 성립함을 보일 수 있다. 마지막으로

$$\begin{aligned}
\sup_{0\le s\le 1}\left|\frac{1}{2}n^{-\frac{1}{2}}h^{-\frac{5}{2}}\sum_{t=1}^{[ns]}\left\{\left(\hat{\beta}-\beta\right)'\mathbf{x}_t\right\}^2K''(\zeta_t(x))\right| &\le \frac{1}{2}n^{-\frac{1}{2}}h^{-\frac{5}{2}}\left\|\hat{\beta}-\beta\right\|^2\|K''\|_\infty\sum_{t=1}^{n}\|\mathbf{x}_t\|^2 \\
&= O_P\left(n^{-\frac{1}{2}}h^{-\frac{5}{2}}n^{-1}n\right) \\
&= o_P(1)
\end{aligned}$$

이 성립하므로 식 (2.2)의 증명이 완료된다. 그리고 식 (2.2)와

$$\hat{f}_n(x)-f_n(x) = (nh)^{-1}\sum_{t=1}^{n}\left\{K\left(\frac{x-\hat{e}_t}{h}\right)-K\left(\frac{x-e_t}{h}\right)\right\}$$

의 식에서 (2.3)이 성립함은 자명하다.

참고 문헌

Bai, J. (1994). Weak convergence of the sequential empirical processes of residuals in ARMA models, *Annals of Statistics,* **22**, 2051–2061.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edition, John Wiley & Sons, New York.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*, 2nd edition, Springer, New York.

Chanda, K. C. (1995). Large sample analysis of autoregressive moving-average models with errors in variables, *Journal of Time Series Analysis,* **16**, 1–15.

Doukahn, P. (1994). *Mixing: Properties and Examples,* Springer, New York.

Lee, S. and Na, S. (2002). On the Bickel-Rosenblatt test for first-order autoregressive models, *Statistics and Probability Letters,* **56**, 23–35.

Lee, S. and Na, S. (2004). A nonparametric test for the change of the density function in strong mixing processes, *Statistics and Probability Letters,* **66**, 25–34.

Ling, S. (1998). Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models, *Annals of Statistics,* **26**, 741–754.

Na, S., Lee, S. and Park, H. (2006). Sequential empirical process in autoregressive models with measurement errors, *Journal of Statistical Planning and Inference,* **136**, 4204–4216.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis,* Chapman and Hall, London.

Takahata, H. and Yoshihara, K. (1987). Central limit theorems for integrated square error of nonparametric density estimators based on absolutely regular random sequences, *Yokohama Mathematics Journal,* **35**, 95–111.

2009년 5월 접수; 2009년 6월 채택

Test for Distribution Change of Dependent Errors

Seongryong Na[1,a]

[a]Department of Information and Statistics, Yonsei University

Abstract

In this paper the change point problem of the error terms in linear regression models is considered. Since fixed or stochastic independent variables and weakly dependent errors are assumed, usual multiple regression models and time series models including ARMA are covered. We use the estimates of probability density function based on residuals in order to test the distribution change of the unobserved errors. Under some mild conditions, the test using the residuals is proved to have the same limiting distribution as the test based on true errors.

Keywords: Change point, time series, dependent error, strong mixing, estimation of probability density function.

[1] Associate Professor, Department of Information and Statistics, Yonsei University, 234 Maeji, Heungup, Wonju, Gangwon 220-710, Korea. E-mail: nasr@yonsei.ac.kr