13주차 강의자료 및 과제 : 시계열 자기상관(교과서 318-365쪽 참고)

1. 오차항 또는 잔차항이 자기상관시 최소제곱추정량(b_2)은 β_2의 최우수선형불편추정량(BLUE)이라고 할 수 없다.

- 시계열 자료를 이용한 다중진회귀모형에 관한 기본가정은 아래와 같다. 이 중 MR4에 대한 기본가정이 깨질경우 우리는 오차항(또는 다중표본회귀모형의 경우 잔차항)이 자기상관되어 있다고 한다. 이때 시계열 자료이므로 오차의 하첨자 표기는 e_i대신 e_t로 대신한다[1].

  MR1: $y_t = \beta_1 + \beta_2 x_{t,2} + ...\beta_k x_{t,k} + e_{t,}\ \ t = 1, ..., N$

  MR2: $E(e_t) = 0$

  MR3: $var(e_t) = \sigma^2$

  MR4: $cov(e_i, e_j) = 0$

  MR5: $x_{t,k}$의 값은 확률적이지 않고 다른 설명변수와 정확히 선형함수 관계에 있지 않다.

  MR6: $y_t \sim N[\beta_1 + \beta_2 x_{t,2} + ...\beta_k x_{t,k}, \sigma^2] \quad \Leftrightarrow \quad e_t \sim N(0, \sigma^2)$

- 최소제곱추정법(또는 통상적 최소제곱추정법; OLS)에 따르면 모든 오차(잔차)의 공분산이 0이며, 즉 오차가 자기상관되어 있지 않다라는 가정하에 추정량(또는 계수)을 산출하였다. 그런데 시계열 표본자료를 이용하여 최소제곱추정법을 적용할 경우 통상적으로 t기의 오차(잔차)와 $t-1$기의 오차(잔차)는 자기상관으로 공분산이 0이 아닐 경우가 흔히 발견된다. 단. 이때 오차항의 기댓값은 0이다. 또 오차의 분산이 모든 관측치에 대해서 동일하다

  · $E(e_t) = 0$
  $var(e_t) = \sigma^2$
  $cov(e_{t,} e_{t-1}) \neq 0$

· 이와 같이 t기의 오차(잔차)와 $t-1$기의 오차가 자기상관이 된 경우 우리는 식(13-1)과 같이 1차자기회귀모형(1차자기상관모형;$AR(1)$)으로 표현한다[2]. 이때 ν_t는 또 다른 오차항으로 기댓값은 0이고, 동분산 조건을 만족하고 동시에 공분산이 0이다. 또 ρ는 또 다른 계수, 우리는 이를 자기상관계수(autocorrelation coefficient)라고 한다.

(13-1) $e_t = \rho e_{t-1} + \nu_t \qquad First\ \ Autoregressive\ \ Model = AR(1)$
$E(\nu_t) = 0$
$var(\nu_t) = \sigma_\nu = constant$
$cov(\nu_{t,} \nu_{t-1}) = 0$

---

1) 통상적으로 횡단면자료의 경우 e_i, e_j, 시계열 자료의 경우 e_t, e_{t+1}로 표기한다.

2) 2차,3차자기회귀모형이 있으나 편의상 이해를 돕기위해 1차로 한정한다.

2. 이와 같이 오차항(표본회귀모형의 경우 잔차항)간 자기상관이 되어 있는 경우에도 불구하고 최소제곱추정법으로 계수(b_2)를 산출하면 어떤 문제가 발생할까?

- 시계열 표본자료를 이용해서 산출한 추정량(b_2)는 확률변수이다. 또 β_2에 대해 선형불편 추정량이다. 그러나 최소분산을 갖지는 않는다. 따라서 오차항간에 자기상관문제가 있는 경우 우리는 최소제곱추정법에 의해 산출한 추정량(b_2)를 β_2의 최우수선형불편추정량 (BLUE)이라고 할 수 없다.

- 이제 보다 적은 분산을 갖는 선형불편추정량을 다시 찾아야 한다. 식 (13-2)의 진회귀모형으로 논의를 전개해보자. 식(13-3)의 경우 t기의 오차와 $t-1$기의 오차가 자기상관이 된 경우이다.

(13-2) $y_t = \beta_1 + \beta_2 x_t + e_t$

(13-3) $e_t = \rho e_{t-1} + \nu_t$

· 오차항 사이에 1기 시차를 두고 자기상관 관계를 갖고 있다. 따라서 식(13-3)을 기초로 할 경우 e_{t-1}, e_{t-2}, e_{t-3}는 식 (13-4)~(13-6)과 같이 전개할 수 있다.

(13-4) $e_{t-1} = \rho e_{t-2} + \nu_{t-1}$

(13-5) $e_{t-2} = \rho e_{t-3} + \nu_{t-2}$

(13-6) $e_{t-3} = \rho e_{t-4} + \nu_{t-3}$

· 식 (13-4)~(13-6)을 식 (13-3)에 대입하고 또 시차를 $t-\infty$ 같이 확장할 경우 식 (13-7)과 같이 무한등비수열이 된다.

(13-7) $e_t = \nu_{t-1} + \rho\nu_{t-2} + \rho^2\nu_{t-3} + \rho^3\nu_{t-4} + \rho^4\nu_{t-5} + \cdots\cdots$

· 따라서 오차의 분산은 식(13-8)과 같다.

(13-8) $var(e_t) = \sigma_e^2 = \dfrac{\sigma_\nu^2}{1-\rho^2}$

· 따라서 자기상관이 없는 경우 자기상관계수(ρ)는 0이다. 이 경우 b_2에 대한 최소제곱추정량의 분산은 식(13-9)과 같다. 그러나 자기상관이 있는 경우 자기상관계수(ρ)는 0이 아니다. 이 경우 b_2에 대한 최소제곱추정량의 분산은 식(13-10)와 같다.

(13-9) $var(b_2) = \dfrac{\sigma_\nu^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$ (13-10) $var(b_2) = \dfrac{\dfrac{\sigma_\nu^2}{1-\rho^2}}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

·식(13-9)와 식(13-10)의 분산 중 어느 것이 더 클까? 자기상관이 있는 경우 b_2의 분산이 더 크다. 물론 표준오차 $se(b_2)$도 더 크다. 따라서 t값이 작아진다.

따라서 자기상관이 있는데도 불구하고 이를 무시하고 최소제곱추정법으로 계수를 구한 후 t검정으로 실시할 경우 식 (13-9)에 의해서 b_2의 분산이 작다고, 표준오차 $se(b_2)$도 작다고, 따라서 t값이 더 크다고 오판할 수 있다. 최종적으로는 계수가 통계적으로 유의하다고 과장할 오류가 있다.

- 만일 자기상관이 있는 경우에도 최소제곱추정법으로 계수를 산출하고, 분산과 표준오차를 추정해서 신뢰구간을 추정하거나 가설검정을 실시하면 어떻게 될까? 답은 오류를 범하게 된다.
  - 표준오차($se(b_2)$)를 더 작게 계산함으로써 신뢰구간이 더 좁아지게 된다. 또한 t값이 더 커진다. 따라서 귀무가설을 기각할 확률이 더 높아 진다. 즉, 부정확한 신뢰구간 추정이나 부정확한 검정통계량 값을 피해야 한다.

- 따라서 일정방법에 의해 관측치에 대해 가중치를 줌으로써 이분산을 동분산으로 바꾼 후 유효한(efficient) 분산과 표준오차를 산출해서 올바른 신뢰구간을 추정하거나 가설검정을 실시해야 한다. 우리는 이와 같은 방법을 일반최소제곱추정법(generalized least squares method)라한다.

13주차 과제

다음은 1차자기회귀모형에 한정한 질문이다

1. 오차항의 분산이 다음과 같다. 이를 증명하시오.

$$var(e_t) = \sigma_e^2 = \frac{\sigma_\nu^2}{1-\rho^2}$$

2. 오차항간에 1차 자기상관이 있는 경우 부정확한 신뢰구간 추정이나 부정확한 검정통계량 값을 피해야 한다. 이를 해결하기 위한 코크레인-오르커트(Cochrane-Orcutt)방법과 그리고 프라이스-윈스텐(Prais-Winsten)방법에 대해 설명하시오.(교과서 356~357쪽 참조)

3. 더빈-왓슨(Dubin-Watson)검정에 대해 설명하시오.(교과서 358~361쪽 참조)