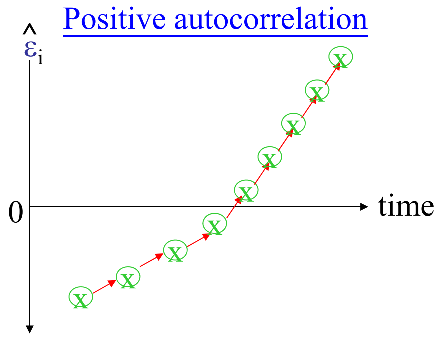
제 11 강

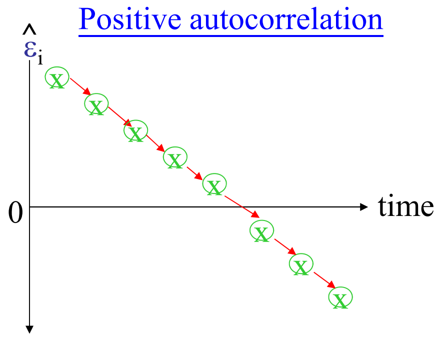
자기상관
Autocorrelation

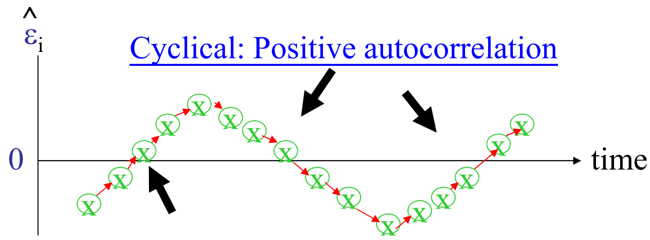
자기상관의 본질

- 시계열자료(time-series data)를 다루는 경우 항상 연속되는 오차항들이 서로 상관되어 있을 가능성이 존재함
- 어떤 특정 시점에서 해당 시점의 오차항은 해당 시점의 충격 뿐 아니라 과거로부터의 충격으로부터 이전된 영향들도 포함함
- 이러한 이전된 영향으로 인해 해당 시점의 충격은 과거의 충격들과 상관될 것이며, 이러한 상황은 오차항들이 상관을 낳게됨
- 이 경우 자기상관이 존재한다고 함
- 양(positive)의 자기상관과 음(negative)의 자기상관이 존재할 수 있음

자기상관의 본질

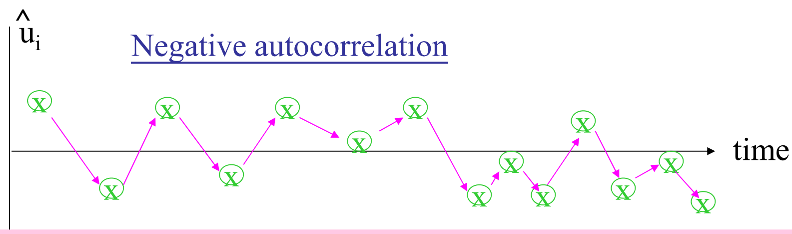




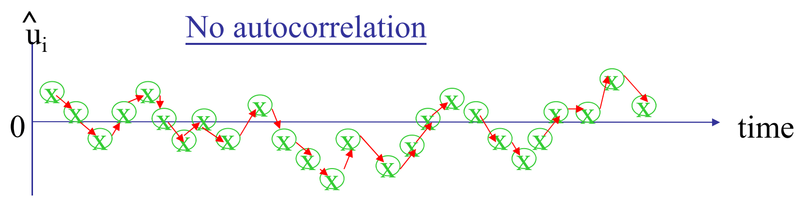


현재의 오차항이 과거의 오차항과 같은 부호를 가지려는 경향이 있음

자기상관의 본질



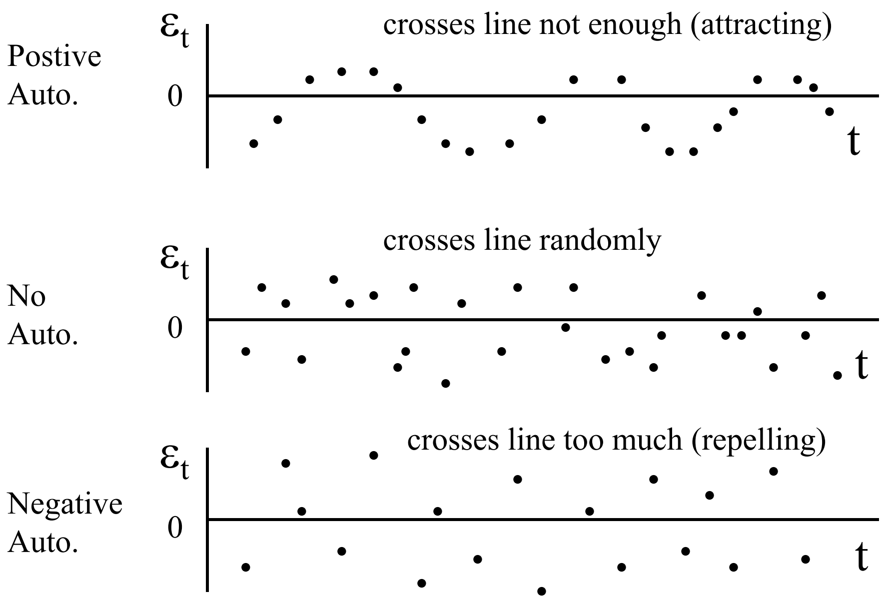
현재의 오차항이 과거와 다른 부호를 가지려는 경향이 있음



현재의 오차항이 과거와는 무관하게 나타남

자기상관의 본질

계량경제학
11.5



자기상관의 본질

계량경제학
11.6

- 표본의 무작위성(randomness)은 서로 다른 관측치들에 대한 오차항들이 상관되어 있지 말아야 함을 의미하나, 자기상관은 이러한 무작위성이 위반됨을 의미함
- 경제 시계열 모형의 오차항에 존재하는 자기상관(Autocorrelation)은 거의 대부분 체계적 패턴임
- 유효성(efficiency, accurate estimation/prediction)을 위해서는 모든 체계적인 정보가 회귀모형에 체화되어 있어야 함

자기상관(=계열상관)의 본질

단순선형모형

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

zero mean:	$E(\varepsilon_\tau) = 0$
homoskedasticity:	$var(\varepsilon_t) = \sigma^2$
nonautocorrelation:	$cov(\varepsilon_t, \varepsilon_s) = 0 \quad t \neq s$
autocorrelation: (serial correlation)	$cov(\varepsilon_t, \varepsilon_s) \neq 0 \quad t \neq s$

자기상관 존재하의 최소제곱 추정량

- 최소제곱추정량은 여전히 선형 불편 추정량이나 유효추정량은 아님
- 표준오차를 계산하기 위한 통상의 공식은 더 이상 정확하지 않으며, 따라서 그에 근거한 신뢰구간이나 가설검정 역시 잘못되게 됨

HAC 표준오차

- HAC(heterskedastisty and autocorrelation consistent) 표준오차 : Newey-West 표준오차라고도 하며, 오차항의 분산과 공분산에 대한 가정이 충족되지 않는 경우 잘못 계산되는 표준오차 값을 대체할 수 있는 값을 계산해줌
  - 이분산성과 자기상관이 동시에 존재하는 경우에도 적용 가능함
  - 계산과정에서 시차선택이 달라짐에 따라 패키지마다 다소 다른 값이 계산될 수 있음
- 자기상관의 구조가 불분명하여 일반최소제곱의 추정이 적절하지 않은 경우 적용 가능함

자기회귀모형(Autoregressive Model)

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

1차 **자기회귀모형 :** $\varepsilon_t = \rho\, \varepsilon_{t-1} + \nu_t$

2차: $\varepsilon_t = \rho_1\, \varepsilon_{t-1} + \rho_2\, \varepsilon_{t-2} + \nu_t$

3차: $\varepsilon_t = \rho_1\, \varepsilon_{t-1} + \rho_2\, \varepsilon_{t-2} + \rho_3\, \varepsilon_{t-3} + \nu_t$

오차항이 1차 자기회귀모형을 따름을 가정할 것임

AR(1) : $\varepsilon_t = \rho\, \varepsilon_{t-1} + \nu_t$

AR(1) 모형

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

$$\varepsilon_t = \underbrace{\rho\, \varepsilon_{t-1}}_{} + \nu_t \qquad \text{where } -1 < \rho < 1$$

The random component in time period *t*

Carryover from the random error in the previous period

A "new" shock to the level of the economic variable

Autocorrelation coefficient

AR(1) 모형

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \quad \varepsilon_t = \rho\, \varepsilon_{t-1} + \nu_t$$

$$\text{where } -1 < \rho < 1$$

$$\boxed{E(\nu_t) = 0} \quad \boxed{var(\nu_t) = \sigma_\nu^2} \quad \boxed{cov(\nu_t, \nu_s) = 0, t \neq s}$$

These assumptions about ν_t imply the following about ε_t :

$$E(\varepsilon_t) = 0 \qquad cov(\varepsilon_t, \varepsilon_{t-k}) = \sigma_\varepsilon^2 \rho^k \quad \text{for } k > 0$$

$$var(\varepsilon_t) = \sigma_\varepsilon^2 = \frac{\sigma_\nu^2}{1-\rho^2} \qquad corr(\varepsilon_t, \varepsilon_{t-k}) = \rho^k \quad \text{for } k > 0$$

Homoskedasticity

AR(1) 모형

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \quad \varepsilon_t = \rho\,\varepsilon_{t-1} + \nu_t$$

where $-1 < \rho < 1$

$$\varepsilon_t = \rho\,\varepsilon_{t-1} + \nu_t$$
$$\varepsilon_{t-1} = \rho\,\varepsilon_{t-2} + \nu_{t-1}$$

$$\Rightarrow \varepsilon_t = \rho^2\,\varepsilon_{t-2} + \nu_t + \rho\nu_{t-1}$$

$$\Downarrow$$

$$\varepsilon_t = \nu_t + \rho\nu_{t-1} + \rho^2\nu_{t-2} + \rho^3\nu_{t-3} + \ldots$$

AR(1)하의 일반최소제곱추정

AR(1) : $\varepsilon_t = \rho\,\varepsilon_{t-1} + \nu_t$

substitute in for ε_t

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

$$y_t = \beta_1 + \beta_2 x_t + \overbrace{\rho\,\varepsilon_{t-1} + \nu_t}$$

Now we need to get rid of ε_{t-1}

(continued) $\longrightarrow$

일반최소제곱추정

$$y_t = \beta_1 + \beta_2 x_t + \rho\, \varepsilon_{t-1} + \nu_t$$

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

$$\varepsilon_t = y_t - \beta_1 - \beta_2 x_t$$

$$\varepsilon_{t-1} = y_{t-1} - \beta_1 - \beta_2 x_{t-1}$$

lag the errors once

$$y_t = \beta_1 + \beta_2 x_t + \rho(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \nu_t$$

(continued) →

일반최소제곱추정

$$y_t = \beta_1 + \beta_2 x_t + \rho(y_{t-1} - \beta_1 - \beta_2 x_{t-1}) + \nu_t$$

$$\rightarrow y_t = \beta_1 + \beta_2 x_t + \rho y_{t-1} - \rho\,\beta_1 - \rho\,\beta_2 x_{t-1} + \nu_t$$

$$\rightarrow \underbrace{y_t - \rho y_{t-1}}_{y_t^*} = \beta_1 \underbrace{(1-\rho)}_{x_{1t}^*} + \beta_2 \underbrace{(x_t - \rho x_{t-1})}_{x_{2t}^*} + \nu_t$$

$$y_t^* = \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \nu_t$$

$$t = 2, 3, \ldots. T$$

일반최소제곱추정

$$y_t^* = y_t - \rho y_{t-1} \qquad x_{1t}^* = (1-\rho) \qquad x_{2t}^* = x_t - \rho x_{t-1}$$

$$y_t^* = \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \nu_t \qquad t = 2, 3, \ldots .T$$

최소제곱으로 이 모형을 추정함에 있어의 문제:

1. 변환된 변수들을 만들어 냄에 있어 시차 (lagged) 변수를 사용함으로써 관측치 하나를 날려버리게 되어 T−1의 관측치만으로 모형을 추정
2. ρ 의 값을 모름. 그것을 추정하기 위한 방법을 찾아야 함

일반최소제곱추정

첫번째 관측치의 복구

- Adding $y_1 = \beta_1 + \beta_2 x_1 + \varepsilon_1$ to the estimation?
  (That is, $y_1^* = y_1$, $x_{11}^* = 1$, $x_{21}^* = x_1$)
  - However, recovering the 1st observation this way and applying least squares is **not** the **best linear unbiased** estimation method.
  - Efficiency is lost because the variance of the error associated with the 1st observation is not equal to that of the other errors.
  - This is a special case of the heteroskedasticity problem except that here all errors are assumed to have equal variance except the 1st error.

일반최소제곱추정

첫번째 관측치의 복구

첫번째 관측치는 원래의 모형에 적합(fit)되어야 함

$$y_1 = \beta_1 + \beta_2 x_1 + \varepsilon_1$$

with error variance: $var(\varepsilon_1) = \sigma_\varepsilon^2 = \sigma_v^2 / (1-\rho^2)$.

이를 변형된 변수들을 이용한 추정에 포함시킬 수 있으나, 다른 관측치들과 같은 오차항의 분산을 가지도록 변형해야만 함

Note: The other observations all have error variance σ_v^2.

일반최소제곱추정

첫번째 관측치의 복구

$$y_1 = \beta_1 + \beta_2 x_1 + \varepsilon_1$$

Multiply through by $\sqrt{1-\rho^2}$ to get:

$$\sqrt{1-\rho^2}\, y_1 = \sqrt{1-\rho^2}\, \beta_1 + \sqrt{1-\rho^2}\, \beta_2 x_1 + \sqrt{1-\rho^2}\, \varepsilon_1$$

$$\Rightarrow \quad y_1^* = \beta_1 x_{11}^* + \beta_2 x_{21}^* + \varepsilon_1^*$$

The transformed error $\varepsilon_1^* = \sqrt{1-\rho^2}\, \varepsilon_1$ has variance σ_v^2 .

이 변형된 첫번째 관측치가 다른 (T-1) 관측치들에 추가되어 T개의 관측치들을 완전히 복원하는 것이 가능

일반최소제곱추정

ρ의 추정

오차항 ε_t들의 값을 안다면 다음을 추정할 수 있음

$$\varepsilon_t = \rho\, \varepsilon_{t-1} + \nu_t$$

우선 최소제곱 추정을 통해 다음을 추정함

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$$

이 추정으로부터의 잔차를 구함

$$\hat{\varepsilon}_t = y_t - b_1 - b_2 x_t$$

일반최소제곱추정

ρ의 추정

다음을 최소제곱추정으로 추정함

$$\hat{\varepsilon}_t = \rho\, \hat{\varepsilon}_{t-1} + \nu_t$$

최소제곱추정량은 다음과 같이 주어짐:

$$\hat{\rho} = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=2}^{T} \hat{\varepsilon}_{t-1}^2}$$

일반최소제곱 추정

복원된 t=1관측치와 앞서 추정한 $\hat{\rho}$ 를 이용하여 y_t^*, x_{1t}^*, x_{2t}^* , $t = 1, 2, \ldots.T$ 의 값을 계산하여 다음의 식을 추정할 수 있음

$$y_t^* = \beta_1 x_{1t}^* + \beta_2 x_{2t}^* + \varepsilon_t^* \quad t = 1, 2, \ldots.T$$

$$(\varepsilon_1^* = \sqrt{1-\rho^2}\varepsilon_1 \,,\quad \varepsilon_t^* = \nu_t \;\; t \geq 2)$$

→ $\hat{\beta}_1, \hat{\beta}_2$: 일반최소제곱 추정치

Prais-Winsten(Cochrane-Orcutt) Estimator

$\hat{\rho}$ 의 추정은 efficient하지 않은 OLS에서 얻어진 잔차로부터 이루어짐 → 개선의 여지가 있음

$\hat{\varepsilon}_t' = y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t$: 앞서 얻은 GLS추정치로부터 잔차를 다시 계산함

$\hat{\varepsilon}_t' = \rho\hat{\varepsilon}_{t-1}' + \nu_t$:새로운 잔차값을 이용하여 ρ를 다시 추정함

새로운 ρ에 대한 추정치를 이용하여, 다시 GLS를 얻음 → 이 과정은 반복될 수 있음→ Prais-Winsten(Cochrane Orcutt) 추정

자기상관의 탐지

DW 검정

H_o: $\rho = 0$ vs. H_1: $\rho > 0$, ($\rho \neq 0$ or $\rho < 0$)

대부분의 경제 자료에 대한 응용에 있어서
자기상관은 양의 자기상관의 형태로 나타남

더빈-왓슨 검정통계량
(The Durbin-Watson Test statistic), d는 :

$$d = \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2}$$

자기상관의 탐지

DW 검정

d 는 근사적으로 $\hat{\rho}$ 와 다음과 같은 관련을 가짐:

$$d \approx 2(1-\hat{\rho})$$

When $\hat{\rho} = 0$, the Durbin-Watson statistic is $d \approx 2$.

When $\hat{\rho} = 1$, the Durbin-Watson statistic is $d \approx 0$.

검정통계량 **d** 의 확률분포가 설명변수들의 값에 의존하기 때문에 그 임계값에 대한 표는 일률적으로 제시될 수 없음. (많은 패키지들은 해당 d값에 대한 p값을 제시해주시 않음.)

Reject H_o if p-value < α, the significance level.

자기상관의 탐지

DW 검정

p-value을 계산해주는 통계패키지가 없을 경우, 경계점정(bounds test)로 알려진 검정방법을 통해 부분적으로 문제를 해결할 수 있음

- 두 개의 다른 검정통계량 d_L 과 d_U 은 $d_L < d < d_U$ 이고, 그 분포들이 설명변수들에 의존하지 않음 (표로 제시되어 있음)

  If $d \le d_L^c$, reject H_0.

  - If $d \ge d_U^c$, accept H_0.
  - If $d_L^c < d < d_U^c$, the test is inconclusive.

자기상관의 탐지

Lagrange Multiplier (Breusch-Godfrey) 검정

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t = \beta_1 + \beta_2 x_t + \rho\varepsilon_{t-1} + v_t$$

$$\Rightarrow y_t = \beta_1 + \beta_2 x_t + \rho\hat{\varepsilon}_{t-1} + v_t \qquad \text{(a)}$$

$$y_t = \hat{y}_t + \hat{\varepsilon}_t = b_1 + b_2 x_t + \hat{\varepsilon}_t$$

$$\Rightarrow b_1 + b_2 x_t + \hat{\varepsilon}_t = \beta_1 + \beta_2 x_t + \rho\hat{\varepsilon}_{t-1} + v_t$$

$$\Rightarrow \hat{\varepsilon}_t = (\beta_1 - b_1) + (\beta_2 - b_2) x_t + \rho\hat{\varepsilon}_{t-1} + v_t$$

$$= \gamma_1 + \gamma_2 x_t + \rho\hat{\varepsilon}_{t-1} + v_t \qquad \text{(b)}$$

자기상관의 탐지

Lagrange Multiplier (Breusch-Godfrey) 검정

Use a *t*- or *F*-test to test the significance of the coefficient of $\hat{\varepsilon}_{t-1}$ in (a) or (b)→ 동일한 결과를 낳음

(b)의 경우 γ_1, γ_2 의 기대값은 0이며, $T \times R^2$ 값을 통계치로 사용하는 χ^2검정의 사용이 가능함 (자유도 = 1) : LM 검정

- Using the first observation requires $\hat{\varepsilon}_0$
  - Set $\hat{\varepsilon}_0 = 0$
  - Or, drop the first observation.
- DW test is an exact valid in finite samples while LM test is an approximate large sample test
  - This approximation occurs because ε_{t-1} is replaced by $\hat{\varepsilon}_{t-1}$
- DW test can only be applied to the AR(1) while LM test can be extended to the test of higher order autocorrelation.

자기상관 존재시 예측

오차항에 자기상관에 존재할 경우, 이전 시긴의 오차항은 미래의 오차항을 예측하는데 도움을 줌

다음기에 대한 최우수(the best) 예측치, y_{T+1} 는

$$\hat{y}_{T+1} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+1} + \hat{\rho}\,\tilde{e}_T$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are generalized least squares estimates and $\tilde{e}_T$ is given by:

$$\tilde{e}_T = y_T - \hat{\beta}_1 - \hat{\beta}_2 x_T$$

h기 이후에 대한 최우수 예측치는

$$\hat{y}_{T+h} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+h} + \hat{\rho}^h \tilde{e}_T$$

Assuming $|\hat{\rho}| < 1$, the influence of $\hat{\rho}^h \tilde{e}_T$ diminishes the further we go into the future (the larger h becomes).