Finance Analytics
Chapter 7. 시계열자료의 회귀분석

권태연

한국외대 국제금융학과

선형회귀모형의 가정

- 가정 1. 변수 Y와 X의 관계는 선형(Linear)이다. ✓
- 가정 2. X는 확률변수가 아닌 주어진 상수값이다. ✓
- 가정 3. X값이 주어져 있을 때, 오차항의 평균은 0이다. ✓
- 가정 4. X값이 주어져 있을 때, 오차항의 분산은 σ^2로 모든 개체 i에 대해 동일하다 ✓
- (가정 5. 서로 다른 개체간 오차항들은 상관되어 있지 않다. = 자기상관(autocorrelation)이 없다. → , 시계열모형)
- 가정 6. X변수들이 여러개 있을때, X변수들 사이에는 선형관계가 없다. = 다중공선성(Multicollinearity) 문제가 없음을 가정 ✓
- 가정 7. 모형 설정 오류가 없음 ✓
- 가정 8. 오차항은 정규분포를 따름을 가정 ✓

오차항의 자기상관 문제

- 주로 시계열 자료에서 발생
- t검정, F검정의 p-value의 신뢰성에 의심

시계열 자료로 회귀분석이 가능한 이유

- 시계열 자료는 자기상관이 없을 수 없다.
- 그럼에도 시계열 자료를 가지고 회귀분석을 하는 이유?
  → 독립변수에도 자기상관이 있으니까..
- 회귀분석 하고 남은 오차에는 자기상관이 남아있지 않을 수 있다.
  → 그렇다면 가정 위반 아님.
- 회귀분석 하고 남는 오차에 자기상관이 여전히 남아 있다면...
  → 무언가 조치가 필요함.

실제자료 예: 미국 소비함수

- Table6_1.csv
- 미국의 실질소비(C), 실질가처분 소득(DPI), 실질재산(W), 실질이자율(R)
- '실질': 물가상승분이 조정되었음
- 다음의 회귀모형을 고려해보자
- 소비함수 모형:

$$\ln C_t = B_1 + B_2 \ln DPI_t + B_3 \ln W_t + B_4 R_t + u_t$$

- Note! : 아랫첨자 시간을 나타내는 "t" 이용

실제자료 예: 미국 소비함수

```
> model<-lm(lnconsump~lndpi+lnwealth+interest, data=data)
> summary(model)

Call:
lm(formula = lnconsump ~ lndpi + lnwealth + interest, data = data)

Residuals:
Min        1Q    Median       3Q       Max
-0.018441 -0.010001  0.000337  0.007038  0.032579

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.4677120  0.0427780 -10.933 7.33e-15 ***
lndpi        0.8048728  0.0174978  45.998  < 2e-16 ***
lnwealth     0.2012702  0.0175926  11.441 1.43e-15 ***
interest    -0.0026891  0.0007619  -3.529 0.000905 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01193 on 50 degrees of freedom
Multiple R-squared:  0.9996,    Adjusted R-squared:  0.9995
F-statistic: 3.783e+04 on 3 and 50 DF,  p-value: < 2.2e-16
```

실제자료 예: 미국 소비함수

- 추정결과를 해석해 보자.
- 탄력성: 가처분소득, 부
- 반-탄력성: 실질이자율
- 적합도

오차의 자기상관 탐색방법

오차의 자기상관 탐색방법

1. 잔차의 time plot
2. 잔차의 Autocorrelation function
3. 더빈왓슨 검정

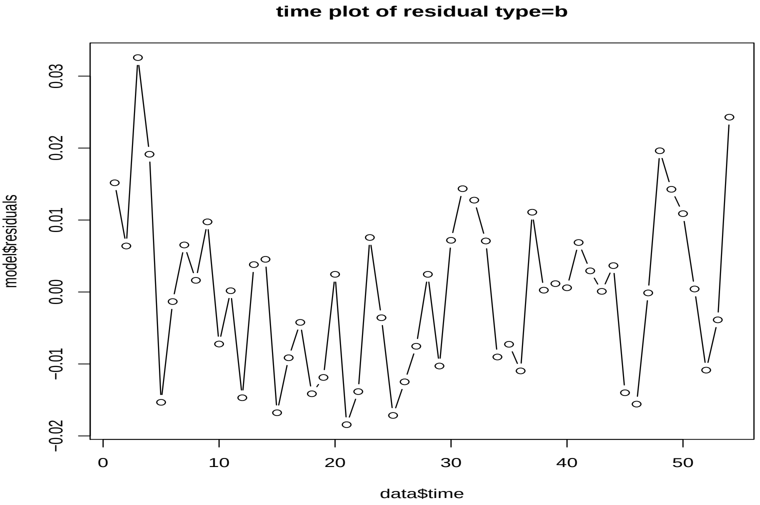
오차의 자기상관의 탐색방법 1: 잔차의 time plot

- 잔차의 time plot을 이용한다.
- 다음의 R program을 실행해 보자

```
data$time<-seq(1947:2000)   #time 변수 만들기#
par(mfrow=c(2,2))
plot(model$residuals~model$fitted,main="residual-fitted")
plot(model$residuals~data$time,main="time plot of residual")
plot(model$residuals~data$time, type="l", main="time plot of residual type=l")
plot(model$residuals~data$time, type="b", main="time plot of residual type=b")
```

오차의 자기상관의 탐색방법 1: 잔차의 time plot

자기상관을 갖는 잔차의 time plot: 시소(see-saw)형태의 잔차의 time plot



오차의 자기상관의 탐색방법 2: 잔차의 자기상관도표 (Autocorrelation function)

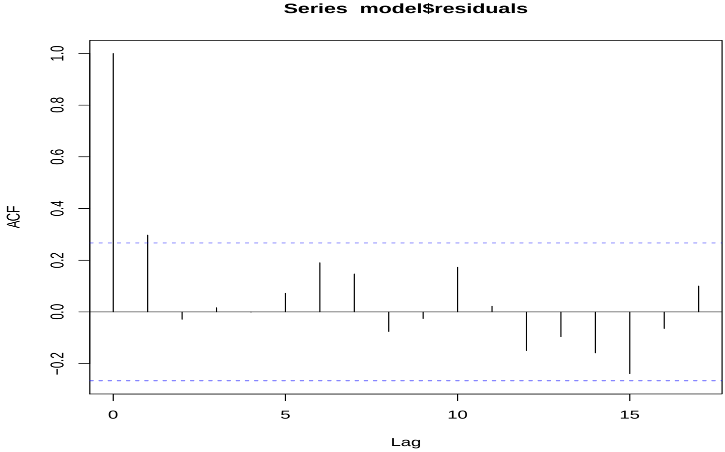
다음과 같은 R-code로 잔차의 자기상관계수들을 구할 수 있다.

```
acf(model$residuals)
```

잔차의 상관도표는

- 잔차의 자기상관 계수
- 즉 $corr(e_t, e_{t-1})$, $corr(e_t, e_{t-2})$, $corr(e_t, e_{t-3})$,...들을 표시한다.
- 각 상관계수의 유의성도 함께 표시된다.

오차의 자기상관의 탐색방법 2: 잔차의 자기상관도표 (Autocorrelation function)



오차의 자기상관의 탐색방법 3: 더빈왓슨 검정 (Durbin-Watson d-statistics)

$$d = \frac{\sum_{t=2}^{t=n}(e_t - e_{t-1})^2}{\sum_{t=1}^{t=n} e_t^2}$$

- $corr(e_t, e_{t-1}) = 0$ 이면, d는 2에 가까운 값을 갖는다.
- $corr(e_t, e_{t-1}) = 1$ 이면, d는 0에 가까운 값을 갖는다.
- $corr(e_t, e_{t-1}) = -1$ 이면, d는 4에 가까운 값을 갖는다.

✓ DW test는 H0: d=2에 대한 유의성 검정을 한다.

- 귀무가설을 기각하면, 즉 p-value가 0.05보다 작으면?
  → 오차항의 자기상관 문제 있음!!!
- 귀무가설을 기각하지 않으면, 즉 p-value가 0.05보다 크면?
  → 오차항의 자기상관 문제 없음!!!

오차의 자기상관의 탐색방법 3: 더빈왓슨 검정 (Durbin-Watson d-statistics)

다음과 같은 dwtest()함수로 R을 이용한 잔차의 DW test를 실시할 수 있다.

```
> install.packages("lmtest")
> library(lmtest)
>  model<-lm(lnconsump~lndpi+lnwealth+interest, data=data)
> dwtest(model)

Durbin-Watson test

data:  model
DW = 1.2892, p-value = 0.0009445
alternative hypothesis: true autocorrelation is greater than 0
```

오차의 자기상관 해결방법

오차의 자기상관 해결방법

1. 독립변수 추가: 시계열 가변수 추가
2. 독립변수의 과거값을 새로운 독립변수로 추가
3. 변수변환 -차분
4. 다른 회귀계수 추정방법
5. 시계열 모형
    a. 자기회귀모형(Autoregressive model- AR model)
       자기회귀 시차분포 모형(Autoregressive Distributed Lag- ARDL Model)
       자기상관오차 회귀모형 (Regression model with autoregressive error)
    b. 그 외 시계열 모형 (MA, ARMA, ARIMA,.....)

오차의 자기상관 해결방법

1. 시계열 가변수를 추가한 시계열 회귀모형

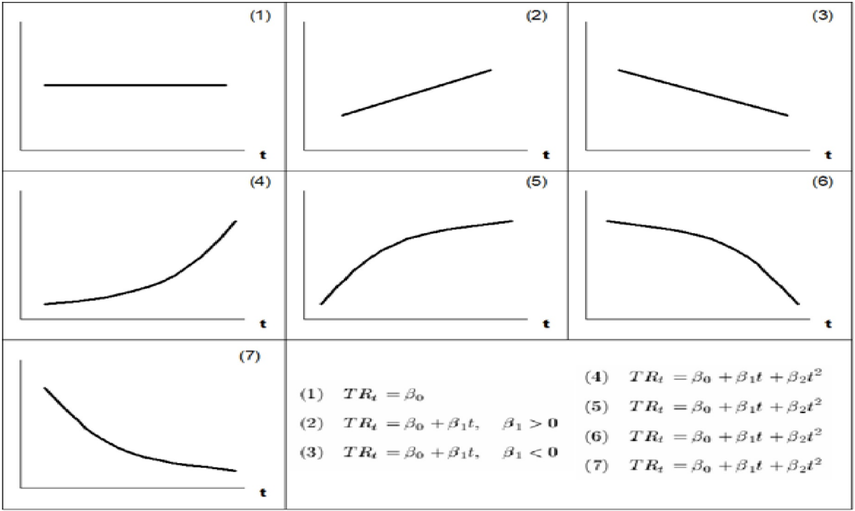
시계열 회귀분석

- 독립변수로 시간, 계절을 설명하는 가변수와 순환요인을 설명하는 삼각함수를 사용하여 회귀모형을 적합

1. 추세를 이용한 시계열 회귀분석

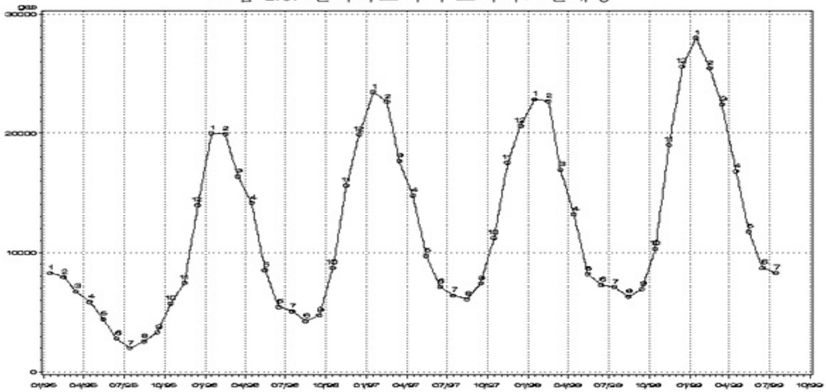
   : 모형에 독립변수로 t, t^2등의 변수를 추가한다.
2. 계절을 반영하는 가변수(dummy variable)을 모형에 독립변수로 추가한다.
3. 순환효과를 설명하는 삼각함수항을 추가한다.

추세요인

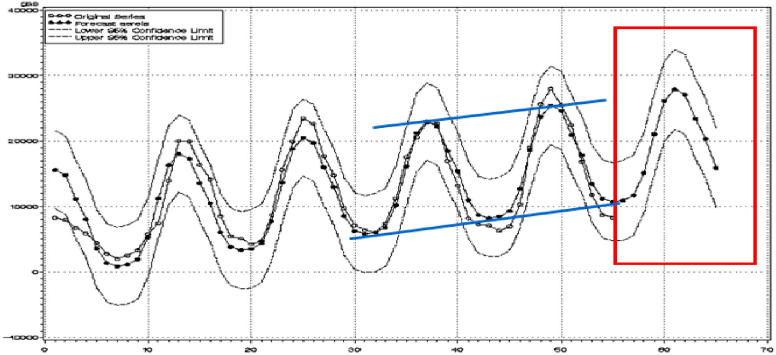

(1)
(2)
(3)
(4)
(5)
(6)
(7)
t
(1) $TR_t = \beta_0$
(2) $TR_t = \beta_0 + \beta_1 t, \quad \beta_1 > 0$
(3) $TR_t = \beta_0 + \beta_1 t, \quad \beta_1 < 0$
(4) $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$
(5) $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$
(6) $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$
(7) $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$

계절 가변수

그림 2.6: 전라북도지역 도시가스판매량



계절 가변수



$$\hat{z}_t = 13890 + 204.5t + 1511.1D_{1,t} + 522.8D_{2,t} - 3378.8D_{3,t} - 6650.9D_{4,t} - 11299D_{5,t}$$

$$- 13753D_{6,t} - 14450D_{7,t} - 14413D_{8,t} - 13806D_{9,t} - 10614D_{10,t} - 4914.2D_{11}$$

순환요인-삼각함수

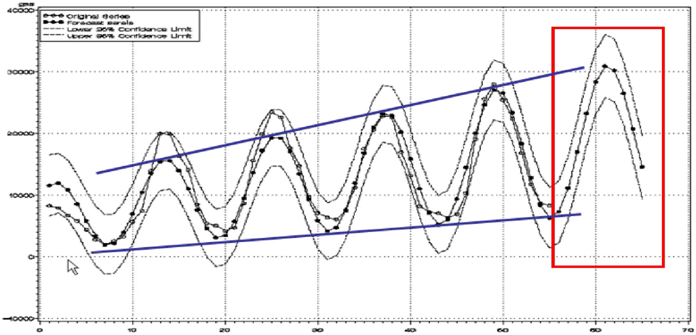
고정계절변동의 경우 삼각함수를 이용한 모형

1. $z_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{L}\right) + \beta_3 \cos\left(\frac{2\pi t}{L}\right) + \epsilon_t$
2. $z_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{L}\right) + \beta_3 \cos\left(\frac{2\pi t}{L}\right) + \beta_4 \sin\left(\frac{4\pi t}{L}\right) + \beta_5 \cos\left(\frac{4\pi t}{L}\right) + \epsilon_t$

확산계절변동의 경우 삼각함수를 이용한 모형

1. $z_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{L}\right) + \beta_3 t \sin\left(\frac{2\pi t}{L}\right) + \beta_4 \cos\left(\frac{2\pi t}{L}\right) + \beta_5 t \cos\left(\frac{2\pi t}{L}\right) + \epsilon_t$
2. $z_t = \beta_0 + \beta_1 t + \beta_2 \sin\left(\frac{2\pi t}{L}\right) + \beta_3 t \sin\left(\frac{2\pi t}{L}\right) + \beta_4 \cos\left(\frac{2\pi t}{L}\right) + \beta_5 t \cos\left(\frac{2\pi t}{L}\right) + \beta_6 \sin\left(\frac{4\pi t}{L}\right) + \beta_7 t \sin\left(\frac{4\pi t}{L}\right) + \beta_8 \cos\left(\frac{4\pi t}{L}\right) + \beta_9 t \cos\left(\frac{4\pi t}{L}\right) + \epsilon_t$

순환요인-삼각함수



$$\widehat{z}_t = 6331.4 + 205.9t + 3724.7\sin\left(\frac{2\pi t}{12}\right) + 55.5t\sin\left(\frac{2\pi t}{12}\right) + 3540.6\cos\left(\frac{2\pi t}{12}\right) + 102.1t\cos\left(\frac{2\pi t}{12}\right)$$

오차의 자기상관 해결방법

2. 독립변수의 과거값을 새로운 독립변수로 추가

독립변수의 과거값을 새로운 독립변수로

$$\ln C_t = B_1 + B_2 \ln DPI_t + B_3 \ln W_t + B_4 R_t + B_2 \ln DPI_{t-1} + B_3 \ln W_{t-1} + B_4 R_{t-1} + u_t$$

- 종속변수의 과거값 $\ln C_{t-1}$을 새로운 독립변수로 넣는것은? –¿ 다른문제, 다른가정에 문제발생, 시계열모형

오차의 자기상관 해결방법

3. 차분

자기상관 해결방법: 변수변환- 차분

오차항이 위와 같이 1차 자기상관(1 계자기상관)을 가지고 있는 경우 다음과 같은 1차 차분으로 자료를 변형하여 오차항의 자기상관 문제를 해결할 수 있다.

1. 종속변수 y의 1차 차분값 $\Delta y = y_t - y_{t-1}$
2. 독립변수 x의 1차 차분값 $\Delta x = x_t - x_{t-1}$
3. 변형된 새로운 종속변수 Δy와 Δx들로 회귀모형 적합

$$\Delta \ln C_t = B_1 + B_2 \Delta \ln DPI_t + B_3 \Delta \ln W_t + B_4 \Delta R_t + u_t$$

자기상관 해결방법: 변수변환- 차분

다음과 같은 diff()함수로 차분된 변수를 구할 수 있다.

```
> ynew<-data$lnconsump[2:dim(data)[1]]
- data$lnconsump[1:dim(data)[1]-1]
> ynew<-diff(data$lnconsump, difference=1)
> data$ynew<-ynew
다음에 오류가 있습니다‘$<-.data.frame‘(‘*tmp*‘, "ynew", value = c(
replacement has 53 rows, data has 54
```

마지막 라인에서 오류가 발생하는 이유는?

자기상관 해결방법: 변수변환- 차분

다음과 같은 diff()함수로 차분된 변수를 구할 수 있다.

```
>  ynew<-diff(data$lnconsump, difference=1)
> lndpinew<-diff(data$lndpi, difference=1)
> lnwealthnew<-diff(data$lnwealth, difference=1)
> interestnew<- diff(data$interest, difference=1)
> model2<- lm(ynew~lndpinew+lnwealthnew+interestnew)
> summary(model2)

Call:
lm(formula = ynew ~ lndpinew + lnwealthnew + interestnew)

Residuals:
Min         1Q    Median        3Q       Max
-0.027777 -0.006015 -0.001278  0.008555  0.021556

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0070456  0.0033955   2.075   0.0433 *
lndpinew    0.7148135  0.0816885   8.750 1.39e-11 ***
lnwealthnew 0.0782675  0.0381738   2.050   0.0457 *
interestnew 0.0007339  0.0008010   0.916   0.3640
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01078 on 49 degrees of freedom
Multiple R-squared:  0.6453,    Adjusted R-squared:  0.6236
F-statistic: 29.72 on 3 and 49 DF,  p-value: 4.303e-11
```

자기상관 해결방법: 변수변환- 차분

- 차분 후 오차항의 자기상관 문제가 해소 되었는지 확인하여 보자.

```
> dwtest(model2)

Durbin-Watson test

data:  model2
DW = 1.8968, p-value = 0.3457
alternative hypothesis: true autocorrelation is greater than 0
```

자기상관 해결방법: 변수변환- 차분

- 차분 후 적합된 회귀계수의 해석: Δx가 한단위 증가할때 Δy가 회귀계수 (y의 단위) 만큼 증가한다.
- (예) 다른 독립변수의 변화량이 일정할때, 실업률의 전년도 대비 증가량이 1% 더 증가한다면, 신규주택 건설량의 전년대비 증가량은 157*1000건 감소한다.

Recall: log-transformation :

- 회귀모형의 변수변환은 분석적 측면(가정의 성립)뿐만 아니라 모형의 해석적 측면(회귀계수의 의미) 역시도 고려되어야 함에 유의하자.

차분된 값이 의미를 갖는 경우 예제

$$\text{주식의 수익율} = \frac{y(t) - y(t-1)}{y(t-1)}$$

$$\frac{y(t) - y(t-1)}{y(t-1)} \approx \ln \frac{y(t)}{y(t-1)} = \ln(y) - \ln(y(t-1))$$

즉 주식의 수익률은 log변환한 주식 가격의 1차 차분값으로 구해진다.

- log변환한 이유?
- 1차 차분한 이유?
- 위의 이유를 "선형회귀모형"에서는 가정들 (등분산성, 오차의 자기상관문제)에 관련하여 설명하였으나
- "시계열 모형"에서는 → 시계열 자료의 안정성 (Stationary process) 으로 설명함.

오차의 자기상관 해결방법

4. 회귀계수 추정방법 변경- FGLS

오차의 자기상관 해결방법

5. 시계열 모형

자기상관 해결방법 5
: 자기상관을 고려한 시계열 모형의 적합

1. 자기회귀모형(Autoregressive model- AR model)
2. 자기회귀 시차분포 모형(Autoregressive Distributed Lag- ARDL Model)
3. 자기상관오차 회귀모형 (Regression model with autoregressive error)
4. 그 외 시계열 모형 (MA, ARMA, ARIMA,.....)