2.1 회귀모형

2.2 다항추세모형

2.3 비선형추세모형

2.4 자기회귀오차모형

2.5 프로그램 및 설명

추세분석

전통적으로 시계열자료를 분석하기 위해 많이 사용되어 온 방법 중의 하나는 관측값 Z_t를 시간의 함수로서 표현하는 방법이다. 즉,

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p + \varepsilon_t \qquad (2-1)$$

와 같이 관측값 Z_t를 시간 t의 다항식으로 나타내는 것이다. (2-1)과 같은 모형을 **선형다항추세모형**(linear polynomial trend model) 또는 **다항추세모형**이라고 부른다.

1장의 [그림 1-1]과 같이 시계열이 일정한 수준을 중심으로 위아래로 움직이는 경우에는 (2-1)에서 $p=0$인 **상수평균모형**(constant mean model)을, [그림 1-2]와 같이 일정한 추세를 갖고 증가하거나 감소하는 경우에는 $p=1$인 **선형추세모형**(linear trend model) 또는 $p=2$인 **2차 추세모형**(quadratic trend model)을 적합시키면 된다. [그림 1-3]과 같이 똑같은 패턴이 일정한 주기를 두고 반복되는 경우에는 **계절추세모형**(seasonal trend model)을 적합시키고, 계절성을 가지며 일정한 선형추세를 따라 움직이는 [그림 1-4]의 경우에는 **선형계절추세모형**(linear and seasonal trend model)을 적합시킨다.

이 외에도 관측값 Z_t를 시간 t의 비선형함수를 이용하여 설명하는 (2-2)와 같은 **비선형추세모형**(nonlinear trend model)을 이용하기도 한다.

$$Z_t = \exp(\beta_0 + \beta_1 t)\,\varepsilon_t \qquad (2-2)$$

비선형추세모형은 선형추세모형과는 달리 시계열이 기하급수적으로 증가하는 양상을 보이거나 비선형적으로 움직일 때 주로 사용되며, 일반적으로 **성장곡선**(growth curve)이라고 부르는 S-곡선 등이 많이 이용된다.

다항추세모형은 다음과 같은 중회귀모형의 특별한 경우로서

$$Z_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + \beta_p X_{tp} + \varepsilon_t \qquad (2-3)$$

독립변수 X_{ti}가 시간 t의 함수라는 점이 특징이다. 따라서 다항추세모형의 분석은

회귀분석의 절차를 똑같이 사용하면 된다. 즉, 모형의 식별 단계에서 주어진 자료가 모형을 적합시키기 위해 필요한 등분산성 등의 가정을 만족시키지 못하는 경우에는 로그변환 등과 같은 분산 안정화 변환 등을 통해 자료를 변환한다. 변환된 자료의 시계열그림을 그려 변환 후의 자료에 적절한 모형을 잠정적으로 선택한다. 모형의 추정 단계에서는 잠정적으로 선택된 모형에 포함된 모수를 주어진 자료를 이용하여 추정한다. 모형의 진단 단계에서는 잔차분석 등을 통해 추정된 모형이 주어진 자료에 잘 적합한지를 진단한다. 최종적으로 확정된 회귀모형은 예측에 사용될 수 있다.

다음 절에서는 종속변수와 일련의 설명변수들 간의 선형관계를 설명해 주는 회귀모형에 대해 설명하기로 한다.

2.1 회귀모형

2.1.1 회귀모형

두 개 이상의 변수들 $Z, X_1, X_2, \cdots, X_p$ 사이의 상호관련성을 (2-4)와 같이 표현한 회귀모형에 의하여 분석하는 방법을 **회귀분석**(regression analysis)이라 한다.

$$Z_t = f(\boldsymbol{X}_t ; \boldsymbol{\beta}) + \varepsilon_t , \; t = 1, 2, \cdots, n \qquad (2-4)$$

(2-4)에서 Z_t는 X_{ti}들의 함수관계로부터 예측되는 변수로서 **종속변수**(dependent variable) 혹은 **반응변수**(response variable)라고 부르며, $f(\cdot)$는 p개의 **독립변수**(independent variable) 혹은 **설명변수**(explanatory variable)라고 부르는 $\boldsymbol{X}_t = (X_{t1}, X_{t2}, \cdots, X_{tp})'$과 p개의 미지의 모수들인 $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$의 함수이다.[1] 회귀모형에서 아래첨자 t는 시간 혹은 관측값 번호를 나타낸다. t가 관측값 번호를 나타내는 경우 t의 값은 단지 관측값들을 구별해 주는 역할만 하는 반면에 t가 시간을 나

1. 상수항이 모형에 포함된 경우는 $\boldsymbol{X}_t = (1, X_{1t}, \cdots, X_{tp})'$이고 $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_p)'$이 된다.

타내는 경우는 관측된 순서의 의미를 갖는다.

(2-4)에서 ε_t는 오차(error)이며, 일반적으로 다음과 같은 가정을 만족한다.

(i) 오차 ε_t의 평균은 0이고 분산은 σ_ε^2이다. 즉, $E(\varepsilon_t) = 0$, $\mathrm{Var}(\varepsilon_t) = \sigma_\varepsilon^2$.

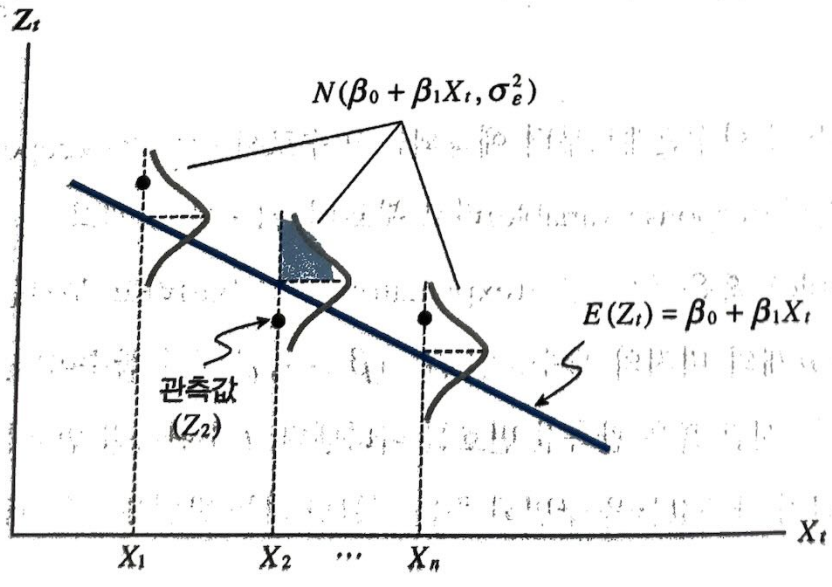
(ii) 오차 ε_t들은 상관관계가 없다. 즉, 모든 t_1과 t_2($t_1 \neq t_2$)에 대하여

$$\mathrm{Cov}(\varepsilon_{t_1}, \varepsilon_{t_2}) = E(\varepsilon_{t_1}\varepsilon_{t_2}) = 0$$

(iii) 오차 ε_t들은 정규분포를 따른다.

즉, 오차 ε_t는 서로 독립이고 $N(0, \sigma_\varepsilon^2)$분포를 따르는 확률오차 (random error)이다. 따라서 회귀모형 (2-4)는 확률모형(probability model)으로 설명변수들 $X_t = (X_{t1}, X_{t2}, \cdots, X_{tp})'$이 주어질 때 Z_t들은 서로 독립이고 평균이 $E(Z_t \mid X_t) = f(X_t; \boldsymbol{\beta})$이고 분산이 σ_ε^2인 정규분포를 따른다. 즉, 설명변수 X_t가 주어진다면 종속변수 Z_t의 값들은 분산 σ_ε^2를 갖는 정규분포의 평균 $E(Z_t \mid X_t)$를 중심으로 확률오차 ε_t 만큼의 차이로 랜덤하게 산포되어 있을 것이다. 이것의 의미는 [그림 2-1]에 설명되어 있다.

그림 2-1 단순선형회귀모형의 그림



다음은 일반적으로 많이 사용되는 회귀모형식들이다.

① 상수평균모형(constant mean model) : $Z_t = \beta_0 + \varepsilon_t$

② 단순선형회귀모형(simple linear regression model) : $Z_t = \beta_0 + \beta_1 X_t + \varepsilon_t$

③ 다중선형회귀모형(multiple linear regression model) :

$$Z_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \cdots + \beta_p X_{tp} + \varepsilon_t \qquad (2-5)$$

④ 2차 선형모형(quadratic linear model) : $Z_t = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + \varepsilon_t$

⑤ 두 개의 독립변수를 갖는 2차 선형모형:

$Z_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_{11} X_{t1}^2 + \beta_{22} X_{t2}^2 + \beta_{12} X_{t1} X_{t2} + \varepsilon_t$

⑥ 지수성장모형(exponential growth model) : $Z_t = \exp(\beta_0 + \beta_1 X_t)\varepsilon_t$

흔히 선형회귀모형에서 선형(linear)은 모수(parameter)에 관해서 선형임을 의미한다. 즉, 위의 모형식 ①~⑤는 모두 모수에 관하여 선형이므로 선형회귀모형이라 한다. 보통 선형회귀모형은 간단히 회귀모형이라고 부른다. ⑥의 모형식은 선형이 아닌 것 같으나 양변에 자연로그를 취하면

$$\ln Z_t = \beta_0 + \beta_1 X_t + \ln \varepsilon_t$$

이다. 이제 $\ln Z_t \equiv Z_t^*, \ln \varepsilon_t \equiv \varepsilon_t^*$라 놓으면

$$Z_t^* = \beta_0 + \beta_1 X_t + \varepsilon_t^*$$

가 되어 모수 β_0와 β_1에 관하여 선형이다. 따라서 ⑥과 같이 본질적으로 선형(변환 후 선형)인 모형의 분석은 선형회귀모형의 분석법에 의해 수행된다.

2.1.2 최소제곱법에 의한 모수추정

회귀모형 (2-4)에서 모수 $\boldsymbol{\beta}$를 추정하기 위한 **최소제곱법**(method of least squares)은 오차제곱합

$$S(\boldsymbol{\beta}) = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} [Z_t - f(\boldsymbol{X}_t ; \boldsymbol{\beta})]^2$$

을 최소로 하는 $\boldsymbol{\beta}$를 찾는 방법이다. 최소제곱법에 의해 구한 $\boldsymbol{\beta}$의 추정값을 **최소제곱추정량**(least squares estimator ; LSE)이라고 하며, $\hat{\boldsymbol{\beta}}$으로 표시하기로 한다.

예를 들면, 상수평균모형 $Z_t = \beta_0 + \varepsilon_t$에서 β_0의 LSE는 오차제곱합

$$S(\beta_0) = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} (Z_t - \beta_0)^2$$

을 최소로 하는 β_0로, 다음의 **정규방정식**(normal equation)

$$\frac{\partial S(\beta_0)}{\partial \beta_0} = 2 \sum_{t=1}^{n} (Z_t - \beta_0)(-1) = 0$$

으로부터

$$\hat{\beta}_0 = \frac{1}{n} \sum_{t=1}^{n} Z_t = \overline{Z} \qquad (2-6)$$

이다. 다음으로 단순회귀모형 $Z_t = \beta_0 + \beta_1 X_t + \varepsilon_t$에서 β_0와 β_1의 LSE는 오차제곱합

$$S(\beta_0, \beta_1) = \sum_{t=1}^{n} \varepsilon_t^2 = \sum_{t=1}^{n} [Z_t - (\beta_0 + \beta_1 X_t)]^2$$

을 최소로 하는 β_0와 β_1이다. 다음의 정규방정식

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_0} = 2 \sum_{t=1}^{n} [Z_t - (\beta_0 + \beta_1 X_t)](-1) = 0$$

$$\frac{\partial S(\beta_0, \beta_1)}{\partial \beta_1} = 2 \sum_{t=1}^{n} [Z_t - (\beta_0 + \beta_1 X_t)](-X_t) = 0$$

을 β_0와 β_1에 관하여 풀면 β_0와 β_1의 LSE는

$$\hat{\beta}_0 = \overline{Z} - \hat{\beta}_1 \overline{X}$$

$$\hat{\beta}_1 = \frac{\sum_{t=1}^{n}(X_t - \overline{X})Z_t}{\sum_{t=1}^{n}(X_t - \overline{X})^2}$$

가 된다. 단, $\overline{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t$이고 $\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$이다.

모든 회귀모형은 (2-5)의 중회귀모형으로 표현할 수 있다. 중회귀모형을 행렬의 형태로 나타내면 다음과 같다.

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2-7}$$

단, $\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}$, $\boldsymbol{X} = \begin{pmatrix} 1 & X_{11} & X_{12} \cdots X_{1p} \\ 1 & X_{21} & X_{22} \cdots X_{2p} \\ \vdots & \vdots & \vdots \quad \ddots \vdots \\ 1 & X_{n1} & X_{n2} \cdots X_{np} \end{pmatrix}$, $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$, $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$

이고, 오차벡터 $\boldsymbol{\varepsilon}$는 평균

$$E(\boldsymbol{\varepsilon}) = \begin{pmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_n) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{0}$$

과 대칭 공분산행렬

$$\mathrm{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \begin{pmatrix} \mathrm{Var}(\varepsilon_1) & \mathrm{Cov}(\varepsilon_1, \varepsilon_2) & \cdots & \mathrm{Cov}(\varepsilon_1, \varepsilon_n) \\ & \mathrm{Var}(\varepsilon_2) & \cdots & \mathrm{Cov}(\varepsilon_2, \varepsilon_n) \\ & & \ddots & \\ \text{대칭} & & \cdots & \mathrm{Var}(\varepsilon_n) \end{pmatrix} = \sigma_\varepsilon^2 I$$

를 갖는 다변량 정규분포를 따른다.