

7. ÇOKLU DOĞRUSALLIK SORUNU (MULTICOLLINEARITY)

7.1. Çoklu Doğrusallık Sorunu Nedir?

Çoklu doğrusallık ekonometrik denklemin açıklayıcı değişkenleri arasındaki ilişkinin derecesi ile ilgili bir sorundur. İki tür olarak karımıza çıkabilir. Eğer açıklayıcı değişkenler arasında tam bir ilişki varsa “tam çoklu doğrusallık” sorunu söz konusudur. Eğer açıklayıcı değişkenler tam olmasa da birbiriyle bir ilişki içindeyse “tam olmayan çoklu doğrusallık” sorunu vardır. Aşağıda bu sorunları ayrı ayrı ele alacağız.

7.1.1 Tam Çoklu Doğrusallık Sorunu

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i, \quad (i = 1 \dots n)$$

veya matrislerle $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$

genel doğrusal modelinde X matrisi

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix} \text{açıklayıcı değişken verilerinden oluşmaktadır.}$$

Burada örneğin X_2 ile X_3 arasında tam bir doğrusal ilişki varsa bu değişkenler arasındaki korelasyon katsayısı 1 olacaktır. Bu durumda $\text{rank}(\mathbf{X}) = k$ varsayımı sağlanmayacak, $\text{rank}(\mathbf{X}) < k$ olacaktır. Yine bu sorun $|\mathbf{X}| = 0$ ve $|\mathbf{X}'\mathbf{X}| = 0$ sonucunu doğurduğundan \mathbf{X}^{-1} ve $(\mathbf{X}'\mathbf{X})^{-1}$ bulunamaz ve denklem tahmin edilemez.

Tam çoklu doğrusallık sorunu genellikle verilerin türetilişinde ve/veya kullanımında farkında olmadan yapılan bir yanlışlıktan kaynaklanır. Örneğin

- i) Denklemden bir sabit terim vardır ve açıklayıcı değişkenlerden birisi sabit kalmaktadır.
- ii) Bir açıklayıcı değişken diğer açıklayıcı değişkenlerin doğrusal bileşimi olarak türetilmiştir.
- iii) Kukla değişken tuzağına düşülmüştür.

7.1.2 Tam Olmayan Çoklu Doğrusallık Sorunu

Tam çoklu doğrusallık sorunu açıklayıcı değişkenler arasında aşağıdaki gibi kesin bir doğrusal ilişkinin varlığı anlamına gelir.

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0 \quad (7.1)$$

Tam olmayan çoklu doğrusallık sorunu ise açıklayıcı değişkenlerin tam olmasa da, aşağıdaki gibi, birbiriyle ilişki içinde oldukları anlamına gelir.

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k + v_i = 0 \quad (7.2)$$

Burada v_i rassal hata terimidir.

7.1 ve 7.2 yeniden düzenlendiğinde sırasıyla aşağıdaki gibi yazılabilir.

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \dots - \frac{\lambda_k}{\lambda_2} X_{ki}$$

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \dots - \frac{\lambda_k}{\lambda_2} X_{ki} - \frac{1}{\lambda_2} v_i$$

Diğer bir deyişle, tam çoklu doğrusallıkta açıklayıcı değişkenlerden birisi (örneğin X_2) diğer açıklayıcı değişkenlerin kesin doğrusal bileşimi olarak yazılabilmekte, tam olmayan çoklu doğrusallıkta ise kesin olmayan doğrusal bileşimi olarak yazılabilmektedir.

Tam olmayan çoklu doğrusallık aşağıdaki durumlarda ortaya çıkabilir.

- i) Açıklayıcı değişkenlerin değerlerinin sınırlı olduğu (az değiştiği) bir aralıkta örneklem alınmış olabilir.
- ii) Bazı açıklayıcı değişkenler arasında güçlü bir ilişki olabilir. Örneğin denklemin sağ tarafında hem gelir hem de konut büyüklüğünün yer aldığı bir model düşünelim. Yüksek gelirli aileler genellikle daha geniş evlerde oturuyorsa, burada çoklu doğrusallık sorunu olabilir.
- iii) Açıklayıcı değişkenlerden birisinin aralığı dar olduğu halde denklemde karesinin de yer alması yine bu soruna yol açacaktır.
- iv) Gözlem sayısından daha fazla sayıda açıklayıcı değişken olması da sorunun bir diğer nedeni olabilir.

7.2. Çoklu Doğrusallık EKK Tahmin Edicilerini ve Tahmin Sonuçlarını Nasıl Etkiler?

Çoklu doğrusallık bir veri sorunu olduğu için EKK tahmin edicileri sorundan etkilenmez. Dolayısıyla EKK tahmin edicileri DESTTE özelliklerini korumayı sürdürürler. Ancak tahmin sonuçları bu sorundan önemli ölçüde etkilenir.

Tam çoklu doğrusallık sorunu olması $|X'X| = 0$ olacağı söylenmişti. Bu durumda $(X'X)^{-1}$ sonsuza gidecektir. Tam olmayan çoklu doğrusallıkta ise çoklu doğrusallığın derecesi arttıkça $|X'X| = 0$ olmasa bile küçülecek ve sifıra yaklaşacaktır. $(X'X)^{-1}$ matrisi ise gittikçe büyüyecektir. Bunun sonucu $\text{Var,Cov}(\hat{\beta}) = \sigma_u^2(X'X)^{-1}$ de gittikçe büyüyen değerler alacaktır.

Böylece t istatistikleri giderek küçüleceğinden aslında bağımlı değişkeni anlamlı olarak açıklayan bağımsız değişkenler, hipotez testlerinde, yanlış olarak, anlamsız bulunacaklardır. Buna karşın R^2 değeri yüksek olabilir.

Sorunun varlığı durumunda ayrıca EKK tahmin edicileri ile standart hataları verilerdeki küçük değişmelere duyarlı olacaktır.

7.3. Çoklu Doğrusallık Sorununun Varlığı Saptanabilir mi?

Çoklu doğrusallık sorununun varlığını araştırmak için aşağıdaki göstergeler incelenebilir. Bu göstergelerin hiç biri çoklu doğrusallık sorununu saptamada tek ölçüt olarak kullanılmamalı, en az iki-üç göstergeye bakılmalıdır.

- 1) Katsayı tahminlerinde *beklenmedik işaret ve/veya büyüklüklerin* bulunması.
- 2) R^2 ile t testleri *çelişkisi*. Açıklayıcı değişken katsayıları için t testleri, katsayıların tümü veya çoğunun anlamsız olduğunu gösteriyor diyelim. Buna rağmen R^2 değeri yüksek (örneğin 0.8'den büyük) ve F testi $R^2=0$ boş hipotezinin reddedildiği sonucunu veriyorsa t testleri ile R^2 arasında bir çelişki vardır. Bu çelişki çoklu doğrusallığın bir göstergesi olabilir.
- 3) *Açıklayıcı değişkenler arası korelasyon katsayılarının yüksekliği*. İki açıklayıcı değişken arasındaki korelasyon katsayısı ($r(X_i, X_j)$) yüksekse, örneğin 0.8'i aşıyorsa) denklemde çoklu doğrusallık olabilir. Ancak tersi doğru değildir: korelasyon katsayılarının düşük çıkması çoklu doğrusallığın olmadığı anlamına gelmez. Çünkü çoklu doğrusallık ikiden fazla değişken arasındaki ilişkiden kaynaklanabilir.
- 4) *Yan regresyonlar*. Çoklu doğrusallık ikiden fazla değişken arasındaki ilişkiden kaynaklanabileceği için bu tür ilişkileri saptamak amacıyla açıklayıcı değişkenlerin tek

tek bağımlı deęişken olarak yer aldığı denklemler tahmin edilerek R^2 deęerleri incelenebilir. $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$, modelinde $X_{ji} = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_{j-1} X_{(j-1)i} + \beta_{j+1} X_{(j+1)i} + \dots + \beta_k X_{ki} + u_i$, ($j=2, 3, \dots, k$) tahmininde R^2 yüksekse (örneğin 0.8'den büyükse) çoklu doğrusallık sorunu var demektir.

7.4. Çoklu Doğrusallık Sorununun Çözümü Var mıdır?

Denklemden çoklu doğrusallık sorunu olması durumunda aşağıdaki çözüm yolları uygulanabilir.

- 1) Denklemden soruna yol açan açıklayıcı deęişkenler biliniyorsa bunlarla ilgili bilgiler başka çalışmalardan alınabilir. Böylece örneğin aralarında ilişki olan açıklayıcı deęişkenler, denkleme doğrusal bileşimleri şeklinde sokulabilir.
- 2) Eğer bulunabiliyorsa ek veriler kullanılabilir veya kesit ve zaman serileri birleştirilerek havuzlanmış veriler kullanılabilir.
- 3) Kullanılan veriler zaman serisi ise ve trend içerdiği için çoklu doğrusallık sorunu ortaya çıkıyorsa serilerin farkı veya yüzde deęişmesi alınarak trendden arındırılabilir.