

BÖLÜM 10

EN KÜÇÜK KARELERDE PROBLEMLER

GİRİŞ

Önceki bölümlerde anlatılan en küçük kareler regresyon modeli, hataların eklenebilir (modelin sabit-etkiler kısmına) ve σ^2 ortak varyanslı normal dağılıma sahip bağımsız rassal değişken olduğu varsayımına dayanmaktadır. Bu varsayımlara dayanan en küçük kareler tahminlerine sıradan **en küçük kareler** denilmektedir. Bağımsızlık ve ortak varyans varsayımlarının sağlandığı durumlarda, en küçük kareler tahmin edicileri tüm mümkün doğrusal yansız tahmin ediciler arasında en iyi tahmin edici (minimum varyans) olma özelliğine sahiptir. Normallik varsayımı sağlandığında en küçük kareler tahmin edicisi aynı zamanda en yüksek olabilirlik tahmin edicisidir.

NORMAL OLMAMA DURUMU

Artıkların ϵ , normal dağıldığı varsayımı regresyon parametrelerinin tahmini ve toplam değişimin ayrışması için gerekli değildir. Normallik, anlamlılık testleri ve parametre tahminlerinin güven aralıkları için gereklidir. t -testi, F -testi ve ki-kare testi teste konu olan rassal değişkenin normal dağılmasını gerektirir. Benzer şekilde; geleneksel güven aralık tahminleri gerek direk gerek student- t dağılımından dolayı normal dağılıma dayanır.

DEĞİŞEN VARYANSLAR

- Sıradan en küçük karelerde ortak varyans varsayımı önemli bir rol oynamaktadır.
- Bu varsayım, bağımlı değişkene ait her gözlemin eşit miktarda bilgi içerdiğini ifade eder.
- Bundan dolayı, sıradan en küçük karelerdeki her bir gözlem eşit ağırlık alır.
- Diğer yandan; heterojen varyans, bazı gözlemlerin diğerlerinden daha fazla bilgi içerdiğini ima etmektedir.
- Verinin rasyonel kullanımı daha çok bilgi içeren gözlemlere daha fazla ağırlık verilmesini gerektirecektir.

İLİŞKİLİ HATALAR

- Artıklar arasındaki korelasyon birçok kaynaktan dolayı meydana gelebilir.
- Zaman serisinden toplanan verinin korelasyonlu hatalara sahip olması yaygındır; zamandaki bir noktaya ait gözlem ile ilişkilendirilmiş hata hemen bitişik gözlemlerdeki hatalar ile ilişkili olmaya meyillidir.
- Nerdeyse tüm zamana göre değişen fiziksel süreçler seri korelasyon gösterirler.
- Örneğin kömür bacasından kirletici salınımının saatlik ölçümleri, yüksek seri korelasyona sahiptir.
- Aynı bireyler üzerinde yapılan ve tekrar eden ölçümler ile ilgili biyolojik çalışmalarda, örneğin bitki ve hayvan gelişme çalışmaları veya klinik deneyler, genellikle ilişkili hatalara sahip olacaktır.

ETKİLİ VERİ NOKTALARI VE AYKIRI DEĞERLER

- Klasik en küçük kareler yöntemi her gözleme eşit ağırlık verir.
- Fakat, her gözlem çeşitli en küçük kareler sonuçlarında aynı etkiye sahip değildir.
- Örnek olarak; basit bir doğrusal regresyon problemindeki eğim genelde ortalamanın çok uzağında değer alan bağımsız değişken gözlemlerinden etkilenir.
- Diğer veri noktalarından uzaklaştırılan tek bir nokta neredeyse regresyon sonuçları kadar diğer noktaların birleşimi ile aynı etkiye sahiptir.
- Bu gözlemlere etkili noktalar veya yüksek kaldıraç gücü noktaları denilmektedir.

MODEL YETERSİZLİKLERİ

- Sıradan en küçük kareler tahmin edicileri eğer model doğru ise yansızdır.
- Bu tahmin ediciler model eğer birçok farklı yoldan hatalıysa yansız olmayacaktır.
- Örneğin, eğer önemli bir bağımsız değişken modelden dışlanırsa, ortalama artık kare (pozitif yönde) 'nin yanlış bir tahmin edicisidir ve tüm bağımsız değişkenlerin regresyon katsayıları (dışlanan değişken modeldeki tüm bağımsız değişkenlere dik olmadıkça) yanlıdır.
- Bağımsız değişkenin sadece birinci kuvvetini kullanan yaygın doğrusal model Y 'nin her bir bağımsız değişken ile ilişkisinin doğrusal ve her bir bağımsız değişkene etkisinin diğer değişkenlerden bağımsız olduğunu varsaymaktadır.

DOĞRUSAL BAĞINTI PROBLEMİ

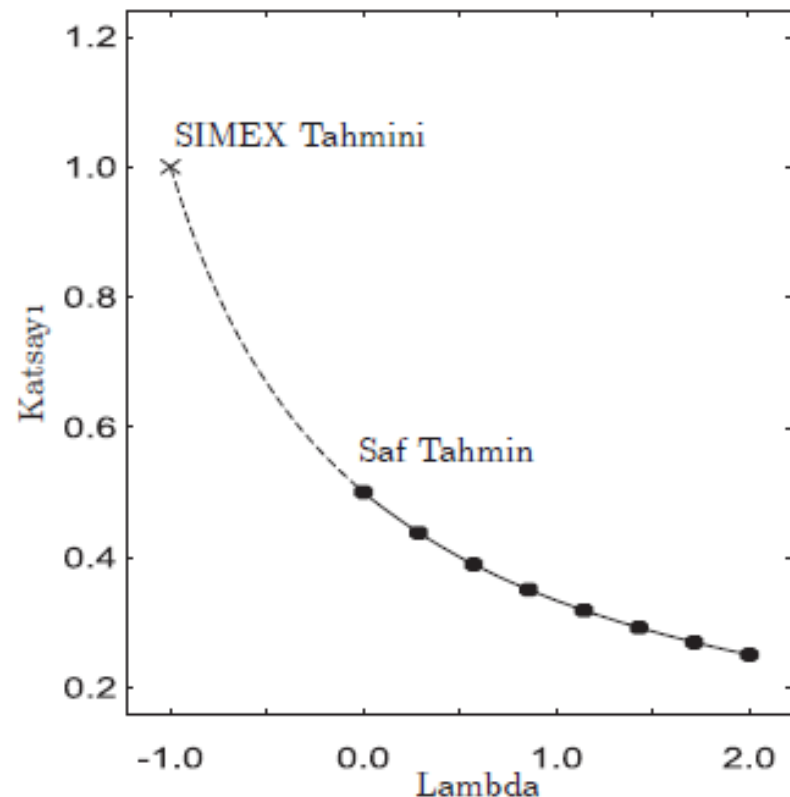
X sonuçlarının tekilliği X sütunlarının doğrusal fonksiyonları tam olarak sıfır vektörüne eşit olduğunda meydana gelir. Bu tür durumlar en küçük kareler analizi uygulandığında açıkça ortaya çıkar, çünkü biricik $(X'X)^{-1}$ yoktur. Daha sıkıntılı ve zor bir durum ise; matrisin tekil olmaya yakın olduğu, vektörlerin doğrusal fonksiyonunun yaklaşık sıfır olduğu durumda ortaya çıkar. Gereksiz bağımsız değişkenler -aynı bilgi farklı formlarda ifade edilmiştir- X 'in hemen hemen tekil olmasına yol açacaktır. Üzerinde çalışılan, sistem içinde karşılıklı bağımlı yakın ilişkili değişkenler X 'de yakın tekilliğe yol açabilir.

BAĞIMSIZ DEĞİŞKENLERDEKİ HATALAR

Orjinal modelde bağımsız değişkenlerin hatasız ölçüldüğü varsayılmış ve bunların regresyon modelinde sabit sayılar oldukları düşünülmüştür. Hatalı-değişkenli modelde, bağımsız değişkenlerin gerçek değerleri ölçüm hataları ile gizlenmiştir. Dolayısıyla, gözlemlenen X_i ;

$$X_i = Z_i + U_i, \quad (10.1)$$

burada Z_i ; X_i 'nin gözlemlenmemiş gerçek değeri olup, U_i ise; ölçüm hatasını göstermektedir. U_i hatasının ortalaması sıfır ve varyansı σ_U^2 olduğu varsayılmaktadır. Örneğin, pişirme süresi içerisinde bir fırındaki ısının etkisini ölçmek üzere yapılan bir denemede, gözlemlenen ısı fırındaki gerçek ısıdan farklı olabilir. Fuller (1987) bağımsız değişkenin hatalı olarak ölçülmüş olduğu birçok örnek vermektedir.



ŞEKİL 10.1. Eğim tahminlerinde $(1 + \lambda)\sigma_V^2$ büyüklüğünün ölçme hata etkisinin jenerik bir çizimidir. Sıradan en küçük kareler metodu tahminleri $\lambda = 0$ 'da belirir ve SIMEX tahmini $\lambda = -1$ 'e doğru bir ekstrapolasyondur.