5. Hafta

Ekonometrik Modelin Spesifikasyonu

Modelin Büyüklüğünün Yanlış Belirlenmesi

Modelin büyüklüğü ile ilgili olan hatalarda modele gereksiz bir değişken ilave edilmiş olabilir ya da model üzerinde etkisi olan gerekli bir değişken dışlanmış olabilir.

1. Dışlanmış Değişken

İktisat teorisine göre bir malın talebi, malın fiyatına, malı talep eden kişilerin gelirine, rakip ve tamamlayıcı malların fiyatına bağlıdır. Talep modelinin tahmininde bilerek veya bilmeyerek bu değişkenlerden biri veya birden fazlası model dışında bırakılırsa, model eksik tanımlanmıştır ve dışlanmış değişkenin sebep olduğu olumsuzlukları taşır.

Basit olarak doğru modelin aşağıdaki gibi olduğunu;

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$$

ancak herhangi bir nedenle X_2 değişkeninin göz ardı edilerek modelin

$$Y_i = \alpha_0 + \alpha_1 X_{i1} + u_i$$

kurulduğunu varsayalım. X_2 değişkeni dışlanmış değişkendir ve bu durumun doğurduğu sonuçlar ise aşağıdaki gibidir.

Model dışında bırakılan değişken (X_2) modelde yer alan X_1 değişkeni ile ilişkiliyse kısaca aralarındaki korelasyon katsayısı sıfırdan farklı ($r_{X_1X_2} \neq 0$) ise, α_0 ve α_1 parametre tahminleri sapmalıdır. Kısaca;

$$E(\hat{\alpha}_0) \neq \beta_0 \quad \text{ve} \quad E(\hat{\alpha}_1) \neq \beta_1$$

ile gösterilir. Büyük örnek özelliği olarak bilinen örnek sayısı arttırılarak tahmin edilen parametrenin gerçek değerine yaklaşması dışlanmış değişkenin varlığı durumunda geçerli değildir, dışlanmış değişken varsa örnek sayısının arttırılması ile sapma giderilemez, EKK tahmincisi tutarsızdır. Sapmasız olması için $r_{X_1X_2} = 0$ olması gerekir. Ancak söz konusu durumda bile $\hat{\alpha}_1$ sapmasız iken, kesim parametre sapmalıdır.

$$E(\hat{\alpha}_1) = \beta_1 \quad \text{ancak} \quad E(\hat{\alpha}_0) \neq \beta_0$$

Dışlanmış değişkenin sebep olduğu sapmalı parametre tahminleri, doğrusal regresyon modeli için $E(u)=0$ olarak ifade edilen varsayımın gerçekleşmediği anlamına gelmektedir. Hata terimi ile dışlanmış değişken arasındaki korelasyon arttıkça buna bağlı olarak sapma da büyüktür.

Bir değişkeni model dışında bırakılırsa, bundan dolayı anakütle hata terimi varyansı da (σ^2) yanlış tahmin edilecektir.

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-k}$$

Model dışında bırakılan değişken hem $\sum \hat{u}_i^2$ 'nin hem de (*n*-*k*) değerinin yanlış belirlenmesine sebep olmaktadır. Doğru modelde X_1 değşkenine ait paramatrenin varyansı

$$Var(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{\sum x_1^2 (1-r_{12})}$$

iken, dışlanmış değişkenin olduğu modelde X_1 değşkenine ait paramatre $\hat{\alpha}_1$'nin varyansı $Var(\hat{\alpha}_1) = \sigma^2 / \sum x_i^2$ ‘e eşittir ve $\hat{\beta}_1$'in varyansının sapmalı bir tahmincisidir. Dışlanmış değişkenin olduğu modelde tahmin edilmiş parametrelerin standart hataları daha büyükdür.
Dolayısıyla varyansın karekökü standart hatanın kullanıldığı hipotez testleri ve güven aralıkları güvenirliliğini yitirecek, yanıltıcı sonuçlar verecektir. Çünkü kalıntıların kareleri toplamı ve serbestlik dereceleri farklı olmaktadır. Özetlemek gerekirse, modelde dışlanmış bir değişken olması dolayısıyla tahmin edilen parametreler eğilimli ve tutarsız olmaktadır. $\sum \hat{u}_i^2$ ve *(n-k)* serbestlik derecesi yanlış belirlendiğinden anakütle hata terimi varyansı da yanlış tahmin edilmektedir.

2. İlgisiz Değişken

Çok değişkenli regresyon analizinde bir diğer sorun ilgisiz değişken(ler)in modele dahil edilmesidir. Ana kütlede *Y* üzerinde etkisi olmayan bir değişkenin tahmin edilecek modele yer alması bir model kurma hatasıdır ve *aşırı tanımlama* olarak da adlandırılmaktadır.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i$$

Yukarıdaki modelinin parametrelerine uygulanan anlamlılık test sonuçlarına göre β_1 istatistiksel açıdan anlamlı, ancak β_2 istatistiksel açıdan anlamsız ise kısaca $H_1: \beta_1 \neq 0$ kabul edilir iken $H_0 : \beta_2 = 0$ temel hipotezi reddedilemezse, bunun anlamı X_1'in bağımlı değişken Y'nin beklenen değeri üzerinde etkili ancak X_3'ün etkili olmadığıdır. Doğru model aşağıdaki gibidir.

$$Y_i = \alpha_0 + \alpha_1 X_{i1} + u_i$$

İlgisiz değişkenin yer aldığı aşağıdaki model tahmin edilirse;

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}$$

β_0 ve β_1 yine sapmasız ve tutarlıdır. Kısaca,

$$E\left(\hat{\beta}_0\right) = \beta_0 \quad E\left(\hat{\beta}_1\right) = \beta_1$$

ancak $E\left(\hat{\beta}_2\right) = 0$ olduğu sonucunu çıkarabiliriz. Rassal bir değişken olan $\hat{\beta}_2$ bir örneklemde sıfır olmasa bile, rassal örneklemlerde ortalaması sıfırdır.

Bu modelde hata terimi varyansı doğru tahmin edilmiştir. Güven aralığı ve hipotez test süreçleri geçerlidir.

İlgisiz değişken parametrelerin sapmasızlık özelliğini etkilememe beraber, parametrelerin varyans tahminlerinin sapmalı olmasına neden olur. İlgisiz değişken X_3'ün modele dahil edilmesi, β_1 ve β_2 parametre tahmincileri varyansının daha büyük olmasına sebep olacak ve böylece tahmin edilen parametreler etkinlik özelliğini kaybedeceklerdir. Sapmasız tahminciler arasında en küçük varyansa sahip tahminciler etkin tahmin olarak nitelendirilir. İlgisiz değişken durumunda ise tahminciler sapmasız olmasına karşın en artık küçük varyanslı olmadıkları için etkin değildirler. İlgisiz değişkenin varlığı halinde tahmincilerin varyansının büyüdüğünü göstermek için öncelikle EKK tahmincilerinin varyanslarının aşağıdaki gibi olduğu daha önceki derslerimizden hatırlayalım.

Doğru modelde $\hat{\alpha}_1$'in varyans tahmincisi

$$Var\left(\hat{\alpha}_1\right) = \frac{\sigma^2}{\sum x_1^2}$$

iken, model kurma hatasının yapıldığı modelde $\hat{\beta}_1$'in varyans tahmincisi aşağıdaki gibidir.

$$Var\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum x_1^2\left(1 - r_{12}^2\right)}$$

İki varyans tahmincisini birbirine oranlarsak;

$$\frac{Var(\hat{\alpha}_1)}{Var(\hat{\beta}_1)} = \frac{1}{(1-r_{12}^2)}$$

elde edilir. $0 \leq r_{12}^2 \leq 1$ olduğu bilindiğine göre $Var(\hat{\alpha}_1) \geq Var(\hat{\beta}_1)$ sonucuna ulaşılır.

Modele gereksiz bir değişken ilave edildiğinde söz konusu değişkenin anlamlılığının *t* testi ile sınanması gereklidir.

$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_i$ şeklinde oluşturulmuş modeli ele alırsak; X_2 değişkeninin modelde yer alıp almamasına karar verebilmek için, β_2'nin istatistiksel olarak anlamlı olup olmadığının araştırılması gerekir.

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

β_2 parametresi anlamlı ise, X_2 değişkeni modelde yer almalıdır. β_2 anlamsız ise X_2 değişkeni modelde yer almamalıdır.

Örnek : Aşağıdaki modelde yer alan X_2 değişkeninin modelde yer almasının gerekli olup olmadığını α=0,05 analmlılık düzeyine göre sınayınız?

$\hat{Y}_i = -6955,495 + 5,22 X_1 - 4,55 X_2 \qquad R^2 = 0,93 \qquad n = 14$

$t \qquad (-2,35) \quad (11,39) \quad (-0,34)$

X_2 değişkeninin anlamlılığını test etmemiz gerekmektedir.

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

$t_{hesap} = -0,34 \qquad t_{(12)} = 2,179$

$-t_c = -2,179 < t_{hesap} = -0,34 < t_c = 2,179$ veya $|t_{hesap} = -0,34| < t_c = 2,179$ olduğu için temel hipotez reddedilemez.

X_2 değişkenine ait parametre istatistiksel açıdan anlamsız çıktığı için X_2 değişkeni modele gereksiz giren değişkendir, model dışına alınması gerekir.

Modelin Matematiksel Kalıbı İle İlgili Hatalar

Model kurma hatasına sebep olan diğer bir hata türü ise modelin matematiksel kalıbı ile ilgilidir. Doğrusal olarak kurulması gereken bir modeli;

$$Y_i = \alpha_0 + \alpha_1 X_i + u_i$$

bilerek ya da bilmeyerek kareli olarak kurduğumuzu düşünelim.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

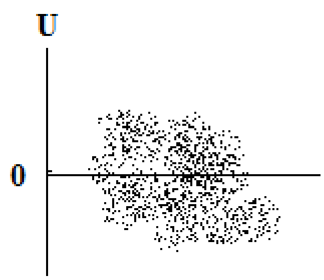
Böyle bir durumda da model kurma hatasına düşmüş sayılırız. Böyle bir hataya düşmemek için değişkenler arasındaki ilişkiye doğru bir şekilde karar vermemiz gerekmektedir.

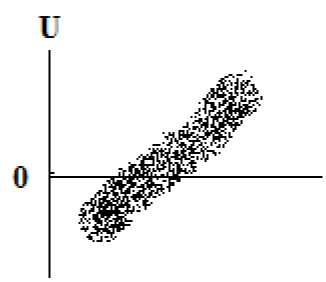
Dışlanmış değişkenler ve yanlış fonksiyonel kalıp için sınamalar

Düzeltilmiş R^2 ($\bar{R}^2$) değerleri yüksek, *t* değerleri anlamsız, parametrelerin işaretleri yanlış ve Durbin Watson istatistiği modelde otokorelasyon olduğunu gösteriyorsa, önemli bir değişken modelden dışlanmıştır ya da modelin matematiksel kalıbı yanlış belirlenmiştir. Bunu sınamak için kullanılan yöntemler aşağıdaki gibidir.

1. **Kalıntıların grafiklerinin incelenmesi:**

Eğer modelden önemli bir değişken dışlanmış ya da matematiksel kalıp yanlış belirlenmişse kalıntıların grafiği gelişigüzel dağılmayacaktır yani belirli bir örüntü sergileyecektir.





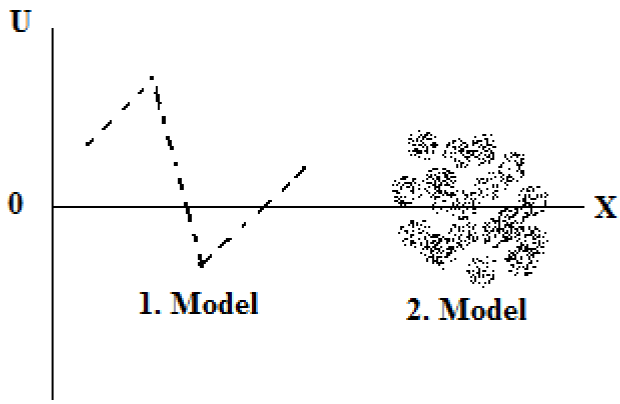
Kalıntılar 0 ekseni etrafında gelişigüzel dağılmalıdır. Eğer böyle değilse kalıntılar birbiri ile ilişkili olacak ve otokorelasyon ortaya çıkacaktır, otokorelasyonun kaynağı da spesifikasyon hatasıdır.

Örnek: Aşağıda verilmiş iki modelden birinin matematiksel formunun yanlış belirlenmiş olduğunu düşünelim.

1) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$

2) $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$

Bu modellere ait kalıntılar aşağıdaki gibidir.



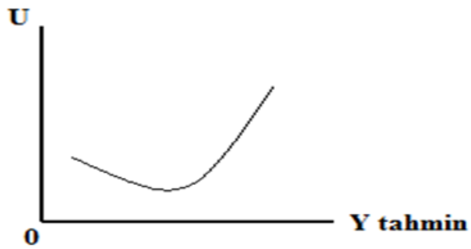
1.modelde kalıntılar hem 0'dan farklı değerler izlemekte hem de belli bir örüntü sergilemektedir. Hâlbuki istenen durum kalıntıların gelişigüzel dağılmasıdır. 2.modeldeki kalıntılar ise 0 etrafında rastgele bir dağılım izlemektedir. Böyle bir durumda 2. modelin matematiksel kalıbının doğru olduğu söylenebilmektedir.

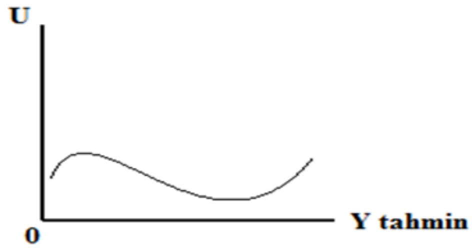
2. Ramsey'in Reset Testi:

Ramsey model kurma hatalarının varlığının tespiti için genel bir yöntem önermiştir. Maliyet üretim ilişkisini dikkate aldığınızda aslında bu ilişkinin doğrusal olmadığı bilinmektedir. Ancak maliyet (Y_i) üretimin (X_i) doğrusal bir fonksiyonu imiş gibi ele alınır ve aşağıdaki model tahmin edilirse araştırmacı tarafından model kurma hatası yapılır.

$Y_i = \beta_0 + \beta_1 X_i + u_i$

Yukarıdaki fonksiyonel forma verilen maliyet fonksiyonu tahmin edilir ve kalıntıları (u_i) hesaplanır. Hesaplanan kalıntıları $\hat{Y}_i$ 'nin değerlerine göre çizer ve kalıntıların grafiğinin bir örüntü sergiledikleri tespit edilirse model kurma hatası varlığının araştırılması gerekir. Maliyet örneğinde kalıntıların $\hat{Y}_i$ 'nin değerlerine göre grafiği çizilirse, kalıntıların önce bir maksimum daha sonra bir minimum yaparak bir örüntü sergiledikleri görülebilir. $\sum \hat{u}_i = 0$ ve $\sum \hat{u}_i \hat{Y}_i = 0$ sıfır olmasına rağmen, kalıntı ortalamaları $\hat{Y}_i$ 'ye bağlı olarak değişir Bu yukarıdaki modele $\hat{Y}_i$ bağımsız değişken olarak dahil edilirse belirginlik katsayısının (R^2) yükseleceğidir . R^2 'deki bu artış istatistiksel olarak anlamlı ise doğrusal maliyet fonksiyonu yanlış kurulmuştur. Reset testinin uygulama aşamaları aşağıdaki gibidir.





Temel ve alternatif hipotezler aşağıdaki gibidir.

$H_0 : \beta_2 = \beta_3 = 0$ Model doğru kurulmuştur.

$H_1 : \beta_2 \neq \beta_3 \neq 0$ Model yanlış kurulmuştur

Tahmin edilen modelden bağımlı değişken Y_i 'nin tahmini değerleri ($\hat{Y}_i$) hesaplanır.

Yukarıdaki regresyon modeli için $\hat{u}_i$ ile $\hat{Y}_i$ arasında eğrisel bir ilişki (max ve min yapar) olduğu için $\hat{Y}_i^2$ ve $\hat{Y}_i^3$ bağımsız değişkenlerinin de yer aldığı yeni bir regresyon modeli kurulur.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 \hat{Y}_i^2 + \beta_3 \hat{Y}_i^3 + u_i$$

İlk modelin (eski model) ve ikinci modelin (yeni model) belirginlik katsayıları hesaplanır, R^2_{eski}, R^2_{yeni}. R^2 'deki artışın istatistiksel anlamlılığı aşağıdaki *F* testi ile araştırılır. (*Not*: Modele ilave edilen değişkenlerin R^2 'nin değerini arttıracağı için $R^2_{yeni} > R^2_{eski}$ olduğunu hatırlayın.)

$$F = \frac{\left(R^2_{yeni} - R^2_{eski}\right) / v}{\left(1 - R^2_{yeni}\right) / n - k} \sim F_{(v, n-k)}$$

Burada *v* modele ilave edilen değişken sayısı, *k* ise yeni modeldeki parametre sayısıdır.

F istatistiğinin hesaplanan değeri verilen α anlamlılık düzeyinde $F_c = F_{(v,n-k)}$ 'den büyükse, H_1 alternatif hipotez kabul edilir. Bu sonuca göre belirginlik katsayısındaki artış istatistiksel olarak anlamlıdır ve yeni model maliyet ile üretim arasındaki ilişkinin doğru fonksiyonel kalıbıdır.

Örnek

10 gözlem kullanılarak aşağıda verilen maliyet fonksiyonu tahmin edilmiştir. göre model kuma hatasının varlığını araştıralım.

$$\hat{Y}_i = 24.7747 + 2.9415X_i \qquad R^2 = 0.88$$
$$Se(\hat{\beta}_i) \quad (4.84) \quad (0.83)$$

Ramsey'in Reset testine göre model kuma hatasının varlığını araştıralım.

Öncelikle $\hat{Y}_i$ 'ler hesaplanır ve $\hat{Y}_i$ 'lerinde yer aldığı aşağıdaki yeni regresyon modeli tahmin edilir.

$$\hat{Y}_i = 41.906 + 5.642X_i - 0.852\hat{Y}_i^2 + 00004\hat{Y}_i^3 \qquad R^2 = 0.97$$
$$Se(\hat{\beta}_i) \quad (4.84) \quad (0.83) \quad (0.0075) \quad (0.00002)$$

Temel ve alternatif hipotezler aşağıdaki gibidir.

$$H_0 : \beta_2 = \beta_3 = 0$$
$$H_1 : \beta_2 \neq \beta_3 \neq 0$$

Test istatistiğinin değeri hesaplanır.

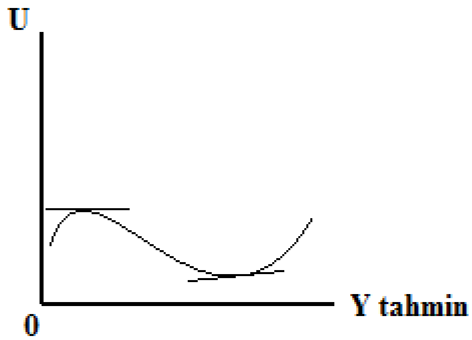
$$F = \frac{\left(R^2_{yeni} - R^2_{eski}\right)/v}{\left(1 - R^2_{yeni}\right)/n-k} = \frac{(0.97 - 0.88)/2}{(1-0.97)/10-4} = 9$$

0.05 anlamlılık düzeyinde $F_c = F_{(2,6)} = 5.14$'dir. $F = 9 > F_c = 5.14$ olduğu için model kurma hatası yapıldığı kabul edilir, çünkü $\hat{Y}_i^2$ ve $\hat{Y}_i^3$'ün modele dahil edilmesi ile belirginlik katsayısındaki artış istatistiksel açıdan anlamlıdır. Yeni regresyonun veriler uyumu daha iyidir.

Örnek

$$\hat{Y}_i = 166,467 + 19,93X_i \qquad R^2 = 0,84 \qquad n = 10$$

Modelin tahmininden elde edilmiş $\hat{Y}_i$ değerleri hesaplanmışt ve $\hat{Y}_i$ değerleri ile kalıntıların grafiği çizilmiştir.



$\hat{Y}_i$ 'nin kareli ve küplü terimleri modele ilave ediilir ve model aşağıdaki gibi yeniden tahmin edilir.

Not :Eğer grafik parabol olmuş olsaydı sadece $\hat{Y}_i$ 'nin kareli değerlerini ilave etmemiz yeterli olacaktı.

$\hat{Y}_i = 2140,72 + 476,6X_i - 0,09\hat{Y}_i^2 + 0,001\hat{Y}_i^3 \qquad R^2 = 0,99$

Temel ve alternatif hipotezler aşağıdaki gibidir.

$H_0 : \beta_2 = \beta_3 = 0$

$H_1 : \beta_2 \neq \beta_3 \neq 0$

Test istatistiğinin değeri hesaplanır.

$$F = \frac{\left(R^2_{yeni} - R^2_{eski}\right)/v}{\left(1 - R^2_{yeni}\right)/n-k} = \frac{(0.99 - 0.84)/2}{(1 - 0.99)/10 - 4} = 284,4.$$

0.05 anlamlılık düzeyinde $F_c = F_{(2,6)} = 5.14$'dir. $F = 284,4 > F_c = 5.14$ olduğu için model kurma hatası yapıldığı kabul edilir, çünkü $\hat{Y}_i^2$ ve $\hat{Y}_i^3$'ün modele dahil edilmesi ile belirginlik katsayısındaki artış istatistiksel açıdan anlamlıdır. Model doğrusal olmamalıdır.Yeni regresyonun veriler uyumu daha iyidir.

3. Lagrange Çarpanı (LM) Testi

Lagrange çarpanı testi model kurma hatsının tespiti için kullanılan Ramsey'in Reset testine alternatif bir testtir ve aşağıdaki gibi uygulanır.

Öncelikle tahmin edilen $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ modelinin kalıntıları hesaplanır.

$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$ modeli doğru ise $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$'dan hesaplanan kalıntılar X_i^2 ve X_i^3 ile ilişkili olması gerekir.

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ modelinden hesaplanan kalıntılar ($\hat{u}_i$) için bütün değişkenlerin yer aldığı aşağıdaki yardımcı regresyon modeli tahmin edilir.

$$\hat{u}_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \alpha_3 X_i^3 + v_i$$

Temel ve alternatif hipotezler aşağıdaki gibidir.

$$H_0 : \alpha_2 = \alpha_3 = 0$$
$$H_1 : \alpha_2 \neq \alpha_3 \neq 0$$

Tahmin edilen ($\hat{\hat{u}}_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 X_i^2 + \hat{\alpha}_3 X_i^3$) regresyon modelinin belirginlik katsayısı (R^2) hesaplanır. Büyük örneklerde gözlem sayısı n ile R^2 'nin çarpımı kısıt sayısına eşit serbestlik derecesi ile ki-kare dağılımına uygunluk göstermektedir.

Bu bağlamda Lagrange testi için test istatistiği aşağıdaki gibidir.

$$nR^2 \sim \chi^2_{sd}$$

Ki-kare istatistiğinin hesaplanan değeri verilen α anlamlılık düzeyinde $\chi_c = \chi_{sd} = \chi_{kısıt\ sayısı}$ 'dan büyükse, H_1 alternatif hipotez kabul edilir. Kısıtlı model, kısıtsız model lehine reddedilir.

Örnek

Model kurma hatasının varlığının araştırmak için Ramsey'in Reset testini uyguladığımız aşağıdaki örneğe aynı amaç için Lagrenge çarpanı testini uygulayacağız.

$$\hat{Y}_i = 24.7747 + 2.9415 X_i \qquad R^2 = 0.88$$
$$Se(\hat{\beta}_i) \quad (4.84) \quad (0.83)$$

Öncelikle kalıntılar hesaplanır ve kalıntıların bağımlı değişken olduğu yeni regresyon modeli tahmin edilir.

$$\hat{\hat{u}}_i = 41.906 + 5.642 X_i - 0.852 X_i^2 + 00004 X_i^3 \qquad R^2 = 0.985$$
$$Se(\hat{\beta}_i) \quad (4.84) \quad (0.83) \quad (0.0075) \quad (0.00002)$$

Temel ve alternatif hipotezler aşağıdaki gibidir.

$$H_0 : \alpha_2 = \alpha_3 = 0$$
$$H_1 : \alpha_2 \neq \alpha_3 \neq 0$$

Lagrange testi için test istatistiğinin değeri hesaplanır.

$nR^2 = 10 \times 0.985 = 9.85$

0.01 anlamlılık düzeyinde $\chi_c = \chi_2 = 9.21$'e eşittir. $nR^2 = 9.85 > \chi_2 = 9.21$ olduğu için alternatif hipotez kabul edilir. Her iki test sonucu doğrusal maliyet modelinin yanlış matematiksel kalıp olduğunu göstermiştir.

4.Durbin Watson istatistiği:

$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{u}_i$ şeklinde tahmin edilmesi gereken bir modeli gerekli bir değişekn dışlanarak

$$Y_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{u}_i$$

şeklinde kurulmuş olabilir. Model kurma hatası olup olmadığını tespit etmek için daha önce otokorelasyonun test edilmesi yöntemlerinde anlatılmış olan Durbin-Watson testi ile sınamak mümkündür. Bunun için takip edilmesi gereken aşamalar aşağıdaki gibidir:

- 2.model tahmin edilir ve kalıntıları hesaplanır.
- Kalıntılar X_2 değişkeninin değerlerine göre küçükten büyüğe doğru sıralanır.

Diğer bir ifade ile X_2 değişkeni ile kalıntılar birlikte sıralanır.

- Temel ve alternative hipotezler

  H_0: Model doğru kurulmuştur.

  H_1: Model yanlış kurulmuştur.
- Sıralanmış olan bu kalıntılardan hareketle DW istatistiği elde edilir.

$$DW = \frac{\sum_{t=2}^{t=n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n} (\hat{u}_t)^2}$$

- DW değeri DW tablosuna göre anlamlı ise model yanlış kurulmuştur.

Örnek

$Y_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_1 + \hat{\alpha}_2 X_1^2 + \hat{u}_i$ Kısıtlı Model

$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_1^2 + \hat{\beta}_3 X_1^3 + \hat{u}_i$ Kısıtsız model

Öncelikle kısıtlı modeli EKK ile tahmin ediyoruz.

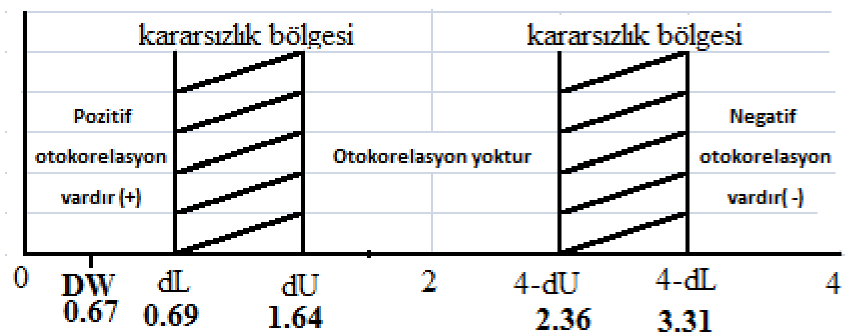
$$\hat{Y}_i = 222,383 - 8,02X_1 + 2,54X_1^2 \qquad R^2 = 0,92 \quad n = 10$$
$$t \quad (9,468) \quad (-0,818) \quad (2,925)$$

X_1 ve X_1^2 değerleri yerine koyulup $\hat{Y}_i$ değerleri elde edilir. $(Y_i - \hat{Y}_i)$ den kalıntılar elde edilir. Kalıntılar modelden dışlanmış olan X_1^3 değişken değerlerine göre sıralanır. Sıralanmış olan kalıntılar kullanılarak DW istatistiği hesaplanır.

$DW = 0.67$

H_0 : Model doğru kurulmuştur.

H_1 : Model yanlış kurulmuştur.



Modelde pozitif otokorelasyonun bulunması H_0 hipotezinin reddedildiğini göstermektedir. Dolayısıyla model kurma hatası söz konusudur.

Değişkenlerdeki Ölçme Hataları

1. Bağımlı Değişkende Ölçme Hataları

Bağımlı değişken Y'de herhangi bir nedenle ölçme hatası varsa yani Y , Y^* olarak hesaplanmışsa:

$$Y_i^* = \alpha_0 + \alpha_1 X_i + u_i$$

şeklinde oluşturulmuş bağımlı değişkeninde ölçme hatası olan bir model olduğunu düşünelim. Buna bağlı olarak doğru model;

$$Y_i = (\alpha_0 + \alpha_1 X_i + u_i) + e_i = Y_i^* + e$$

olacaktır. Model yeniden düzenlenirse;

$$Y_i = \alpha_0 + \alpha_1 X_i + (u_i + e_i)$$

$$Y_i = \alpha_0 + \alpha_1 X_i + v_i$$

haline dönüşecektir. Yukarıda görüldüğü üzere bağımlı değişkende bir ölçme hatası olduğunda bu durum sadece hata terimi etkilemektedir. Hata terimi “e” kadar büyümektedir. Bağımlı değişkendeki bu ölçme hatası parametrelerin eğilimsizlik ve tutarlılık özelliklerini etkilememektedir. Fakat hata teriminin varyansı olduğundan büyük tahmin edilmektedir. Bu nedenle hesaplanan varyanslar etkin değildir.

2. Bağımsız değişkende Ölçme Hataları

Bağımsız değişkende ölçme hatalarının varlığı halinde bağımsız değişken ile hata terimleri ilişkidir ki; bu durumda klasik doğrusal regresyon modelinin temel varsayımlarından $E(u_i X_i) = 0$ gerçekleşmemiştir. Bu durumda tahmin edilen parametreler sapmalıdır. Örnek sayısı sonsuza giderken bile tahminler ana kütle değerine yaklaşmazlar ve dolayısıyla tutarsız tahminlerdir. Bağımsız değişkendeki ölçme hatası model için önemli bir sorun teşkil etmektedir.

Modelde yer alan bağımsız değişkenlerde herhangi bir sebeple ölçme hatası olduğunu düşünelim. Ölçme hatası yapılan bağımsız değişken X^* ile gösterilirse;

$$X_i = X_i^* + e \qquad \text{ve} \qquad X_i^* = X_i - e$$

olmaktadır. Buna bağlı olarak model düzenlenirse;

$$Y_i = \alpha_0 + \alpha_1 X_i^* + u_i$$

$$Y_i = \alpha_0 + \alpha_1 (X_i - e) + u_i$$

$$Y_i = \alpha_0 + \alpha_1 X_i - \alpha_1 e + u_i$$

$$Y_i = \alpha_0 + \alpha_1 X_i + (u_i - \alpha_1 e)$$

$$Y_i = \alpha_0 + \alpha_1 X_i + v_i$$

haline dönüşmektedir. Açıkça görüldüğü gibi α_1 ’den dolayıhata terimleri ile bağımsız değişken ilişkilidir ve

$$COV(X_i u_i) \neq 0$$

Bu da EKK yönteminin önemli bir varsayımını çiğnenmesi anlamına gelmektedir. EKK tahmin edicileri hem eğilimli hem de tutarsız olmaktadır.

Model Seçme kriterleri

1. Düzeltilmiş Belirginlik Katsayısı

Regresyon modelinin veriler uyumunun iyiliğinin bir ölçüsü olan belirginlik katsayısı R^2 'nin aşağıdaki tanımlandığını biliyoruz.

$$R^2 = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

R^2, sıfır ile bir değerleri arasında yer alır ve 1'e ne kadar yakın ise regresyon doğrusunun verilere uyumu o kadar daha iyidir. Ancak daha önceki derslerimizden de bildiğimiz gibi belirginlik katsayısının bazı dezavantajları da vardır. Bunları kısaca hatırlayalım.

Bağımlı değişkenin tahmin edilen değerinin ($\hat{Y}_i$) gözlemlenen (teorik) değere (Y_i) ne kadar yakın olduğu anlamındaki örneklem içi uyumu ölçer. Örneklem dışı gözlemler için bir garantisi yoktur.

İki ya da fazla alternatif modelin R^2 ölçüsüne göre karşılaştırılabilmesi için bağımlı değişkenin aynı biçimde ifade edilmesi gerekir.

Modele giren her yeni değişken belirginlik katsayısının değerini arttırır. Ancak bu her zaman regresyon doğrusunun verilere uyumunun arttığı anlamına gelmez.

Modele giren her değişkenin R^2 'nin değerini yükseltmesine karşın, R^2, $(n-k)/(k-1)$ serbestlik derecesi ile yeniden düzenlenerek düzeltilmiş $\bar{R}^2$ ölçüsüne ulaşılır.

$$\bar{R}^2 == 1 - \frac{\sum \hat{u}_i^2/(n-k)}{\sum y_i^2/(n-1)} = 1 - \left(1 - R^2\right)\frac{n-1}{n-k}$$

Yukarıdaki formülü göre $\bar{R}^2$ 'nin R^2 'den küçük olduğu ve daha fazla değişken eklemenin düzeltilmiş R^2 tarafından cezalandırılacağı görülmektedir. Alternatif modellerin seçiminde bağımlı değişkenin aynı olması durumunda düzeltilmiş R^2 yol göstericidir.

2. Akaike Bilgi Kriteri (AIC)

Akaike kriterine göre modele giren her değişken bir cezaya tabii tutulur.

$$AIC = \left(\frac{2k}{n}\right) + \ln\left(\frac{\sum \hat{u}_i^2}{n}\right)$$

Burada yine k =parametre sayısı, n gözlem sayısı iken $2k/n$ ceza teridir. AIC kriteri modele giren bağımsız değişkeni $\bar{R}^2$ 'ye göre daha fazla cezalandırılır. Alternatif modellerin seçiminde AIC değeri küçük olan model tercih edilir.

3. Schwarz Bilgi Kriteri (SIC)

Schwarz bilgi kriteri aşağıdaki gibi tanımlanmaktadır.

$$SIC = \left(\frac{k}{n}\right)\ln n + \ln\left(\frac{\sum \hat{u}_i^2}{n}\right)$$

Burada ceza terimi $(k/n)\ln n$ 'dir. alternatif modellerin seçiminde Schwarz bilgi kriteri küçük olan tercih edilir.