

HAFTA 11

Hataların ortaya çıkma nedenleri:

1. Hata varyansı $\hat{\sigma}^2 = \frac{\sum_{t=1}^n \hat{\varepsilon}_t^2}{n-2}$, σ^2 'yi olduğundan daha küçük tahmin edebilir.
2. Bunun sonucunda R^2 olduğundan daha büyük tahmin edilebilir.
3. σ^2 olduğundan daha küçük tahmin edilmese bile, $Var(\hat{\beta}_1)$ ardışık bağımlılığın varlığı altındaki varyans olan $Var(\hat{\beta}_1)_{AB1}$ 'i olduğundan daha küçük tahmin edilebilir. $Var(\hat{\beta}_1)_{AB1}$ 'yi ağırlıklandırılmış EKK varyans tahmini $Var(\hat{\beta}_1)_{AEK}$ ile karşılaştırılsa etkin değildir. Yani; $Var(\hat{\beta}_1)_{AB1} > Var(\hat{\beta}_1)_{AEK}$ dir.
4. Bilinen t ile F testleri artık geçersizdir. Eğer bu testler uygulanırlarsa, tahmin edilen regresyon katsayılarının istatistik bakımından anlamlılıkları konusunda ciddi biçimde yanıltıcı sonuçlar verilebilirler.

Bu önermeleri denemesini yapmak için model

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

ve klasik varsayımlar altında

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^n \hat{\varepsilon}_t^2}{n-2}$$

σ^2 için sapmasız bir tahmin edicisidir. Yani $E(\hat{\sigma}^2) = \sigma^2$ dir.

Ardışık bağımlılığın varlığı altında

$$r = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=1}^n x_t^2}, \quad x \text{ 'lerin ardışık değerleri arasındaki korelasyon katsayısı olmak üzere varyans}$$

tahmininin beklenen değeri

$$E(\hat{\sigma}^2) = \frac{\sigma^2 \{n - [2/(1-\rho)] - 2\rho r\}}{n-2}$$

dir. Eğer hem ρ hem de r artı işaretliyse $E(\hat{\sigma}^2) < \sigma^2$ dir. Buradan hata varyansının gerçek σ^2 'yi olduğundan daha küçük tahmin ettiği görülür.

Monte Carlo simülasyon çalışmasıyla $\beta_0 = 1$ ve $\beta_1 = 0.8$ alınır; basit doğrusal regresyon denklemi

$Y_t = 1.0 + 0.8X_t + u_t$ ve $E(Y_t|X_t) = 1.0 + 0.8X_t$;
ardışık bağımlılığın varlığı için

$$u_t = 0.7u_{t-1} + \varepsilon_t; \quad \varepsilon_t \sim N(0,1), \quad \rho = 0.7$$

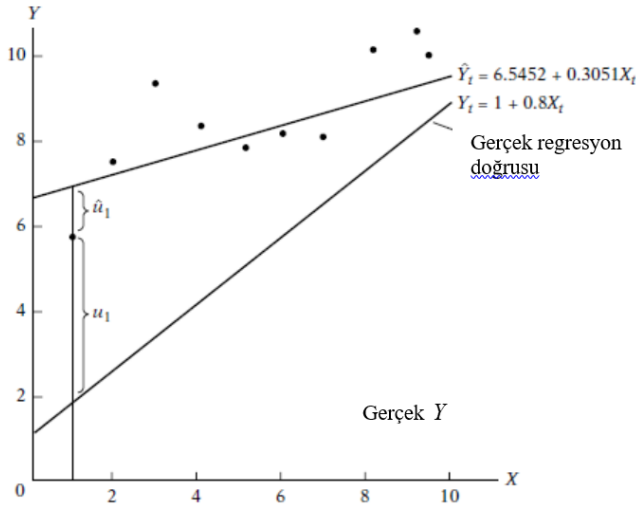
modellerinden (X_t, ε_t) verisi türetilir. Bu veriden elde edilen kestirim modeli

$$\hat{Y}_t = 6.5452 + 0.3051X_t$$

$$S_{\hat{\beta}_1} : 0.6153 \quad 0.0992$$

$$t : 10.6366 \quad 3.0763$$

$$r^2 = 0.6419 \quad \text{ve} \quad \hat{\sigma}^2 = 0.8114$$



$\rho = 0$ alınarak similasyon çalışması yenilenirse, kestirim denklemini

$$\hat{Y}_t = 2.5345 + 0.6145X_t$$

$$S_{\hat{\beta}_1} : 0.6796 \quad 0.1087$$

$$t : 3.7910 \quad 5.6541$$

$$r^2 = 0.7997 \quad \text{ve} \quad \hat{\sigma}^2 = 0.9752$$

Bu kestirim denklemini gerçek regresyon modelini daha iyi yansıtmaktadır.

$\rho = 0.7$ alındığında $\hat{\sigma}^2 = 0.8114$

$\rho = 0$ alındığında $\hat{\sigma}^2 = 0.9752$

$\hat{\beta}_0$ ve $\hat{\beta}_1$ 'nin standart hatalarının arttığı görülür.

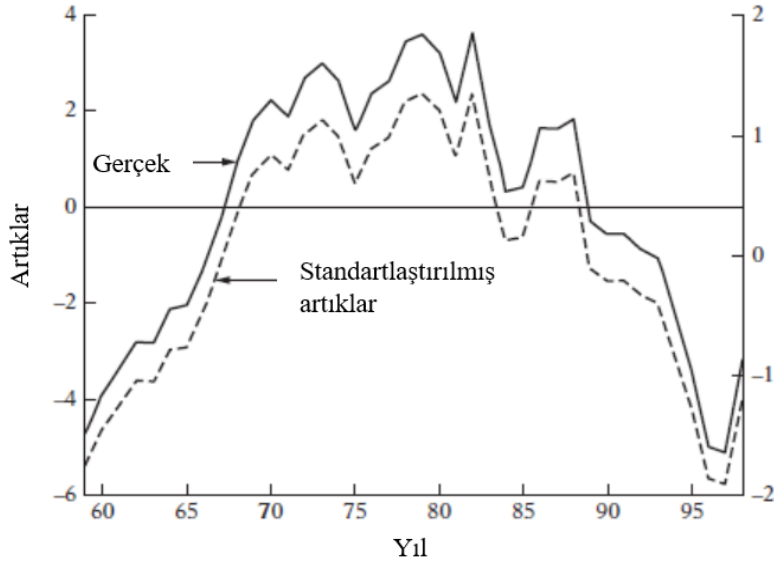
Ardışık bağımlılığın var olup olmadığını aramak:

– **Çizim Yöntemi:**

Hata terimleri $\hat{\varepsilon}_t$ 'lerin çizimleri yalnız ardışık bağımlılık konusunda değil, sabit varyans ve model kurma sapması ya da model yetersizliği konularında yararlı bilgiler verir. Bazı çizim yöntemleri:

– Zaman sıralı çizim

– Standartlaştırılmış hataların zamana göre çizimi



Dizilim sınaması: (Geary sınaması) (Parametrik olmayan bir yöntem)

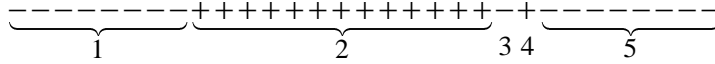
n = toplam gözlem sayısı = $n_1 + n_2$

n_1 = + işaretli artıkların sayısı

n_2 = - işaretli artıkların sayısı

k = dizilim sayısı

Örneğin; artıklar $\hat{\varepsilon}_t$,



$k = 5$ dizilim var.

Hataların normal dağılıma sahip olduğu varsayımı altında

H_0 : Hatalar ardışık bağımsız

H_0 hipotezinin doğruluğu altında $n_1 > 10$ ve $n_2 > 10$ varsayımıyla dizilimlerin sayısı asimptotik olarak beklenen değeri ve varyansı

$$E(k) = \frac{2n_1n_2}{n_1 + n_2} + 1,$$

$$Var(k) = \sigma_k^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

olmak üzere normal dağılım gösterir. Eğer rasgelelik önsavı ileri sürülecekse, bir problemde bulunan dizilim sayısı k 'nın %95 güvenle $E(k) \pm 1.96\sigma_k$ arasında olması beklenir.

Karar Kuralı:

Eğer $E(k) - 1.96\sigma_k < k < E(k) + 1.96\sigma_k$ ise H_0 hipotezi red edilemez. Aksi halde k bu sınırların dışındaysa H_0 hipotezi red edilir.

Örnek: Artıkların işaretlerine bakıldığında $n_1 = 14$ ve $n_2 = 18$ ise

$$E(k) = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(14)(18)}{14+18} + 1 = 16.75,$$

$$Var(k) = \sigma_k^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} = \frac{2(14)(18)(2(14)(18) - 14 - 18)}{(14+18)^2(14+18-1)} = 7.49395$$

$$\Rightarrow \sigma_k = 2.7375$$

%95 güvenle dizilim sayısı için güven aralığı $16.75 \pm 1.96(2.7375) \Rightarrow (11.3845, 22.1155)$ bulunur. Dizilim sayısı $k=5$ olduğuna göre bu aralığın dışına düşmektedir. O halde %95 güvenle H_0 hipotezi reddedilir.

Eğer n_1 ya da n_2 'den birisi 20'den küçükse Swed ile Eisenhart gözlemlerin rasgele sıralamaları durumunda beklenen dizilim sayılarının eşik değerlerini veren özel çizelgeler geliştirmişlerdir.

Durbin-Watson d sınaması:

Ardışık bağımlılığı bulmak için kullanılan en yaygın sınamadır. Durbin-Watson d istatistiği

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

basitçe ardışık artıkların fark kareleri toplamının artık kareler toplamı SSE'ye oranıdır. Bu d istatistiğinin payında $n-1$ tane gözlem vardır. Çünkü ardışık farklar alınırken bir gözlem kaybolur.

d- İstatistiğin gerisinde yatan varsayımlar:

1. Orijinden geçen regresyon modelinde olduğu gibi β_0 terimi yoksa SSE'yi bulmak için regresyonun sabit terimle bir kez daha bulunması gerekir.
2. X 'ler olasılıklı değildir ya da yinelenen örneklerde değişmezler.
3. u_t hata terimleri 1. dereceden

$$u_t = \rho u_{t-1} + \varepsilon_t$$

modelinden türetilmiştir.

4. Regresyon modeli bağımlı değişkenin gecikmeli değer(ler)ini açıklayıcı değişken olarak almaz. Demek ki bu sınama;

$Y_{t-1} = Y_t$ 'nin bir dönem gecikmeli değeri olmak üzere

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t$$

modellerine uygulanamaz. Böyle modellerde **ardışık bağımlı modeller** denir.

5. Verilerde eksik gözlem yoktur. d istatistiği eksik gözlemleri hesaba katmaz.

Durbin-Watson d istatistiği:

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

$\sum_{t=1}^n \hat{u}_t^2$ ile $\sum_{t=2}^n \hat{u}_{t-1}^2$ arasında bir gözlemlilik bir fark olmasından dolayı yaklaşık eşit alınabilir.

Öyleyse;

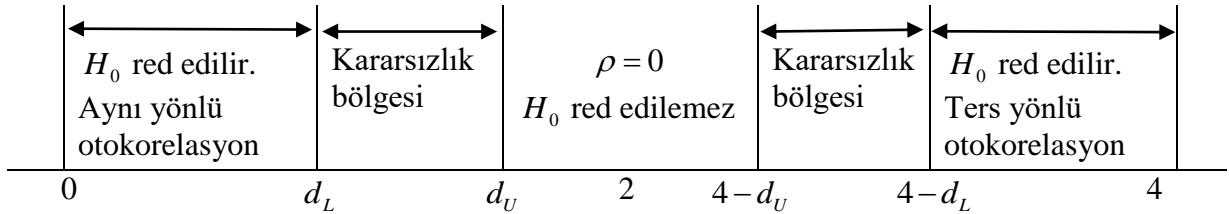
$$d \approx 2 \left[1 - \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2} \right]$$

dır. $\hat{\rho} = \frac{\sum_{t=2}^n \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^n \hat{u}_t^2}$ alınırsa, $d = 2(1 - \hat{\rho})$ eşit olur.

Ardışık bağımlılığa ilişkin hipotezler:

$H_0 : \rho = 0$ otokorelasyon (ardışık bağımlılık) yoktur.

$H_1 : \rho \neq 0$ otokorelasyon (ardışık bağımlılık) vardır.



H_0 : Aynı yönlü ardışık bağımlılık yoktur

H_0^* : Ters yönlü ardışık bağımlılık yoktur

$\hat{\rho} = 0$ ise $d = 2 \Rightarrow$ 1. dereceden ardışık bağımlılık yok

$\hat{\rho} = +1$ ise $d = 0 \Rightarrow$ aynı yönlü ardışık bağımlılık var

$\hat{\rho} = -1$ ise $d = 4 \Rightarrow$ ters yönlü ardışık bağımlılık var

anlamındadır.

Aşamalar:

- Kestirim denkleminde hata terimleri \hat{u}_t 'lar bulunur.
- Durbin-Watson d istatistiği bulunur.
- Durbin-Watson (DW) tablosundan $n =$ gözlem sayısı, $k =$ açıklayıcı değişken sayısı olmak üzere d_L ve d_U değerleri bulunur.
- Karar kuralları:

H_0 hipotezi	Eğer	Karar
Aynı yönlü ardışık } bağımlılık yoktur }	$0 < d < d_L$ $d_L \leq d \leq d_U$	Red Karar yok
Ters yönlü ardışık } bağımlılık yoktur }	$4 - d_L < d < 4$ $4 - d_U \leq d \leq 4 - d_L$	Red Karar yok
Ardışık bağımlılık yok (Ne aynı yönlü, ne de ters yönlü)	$d_U \leq d \leq 4 - d_U$	Red edilemez

Durbin-Watson istatistiğinin kullanılmadığı durumlar:

- Model sabitsiz terimsiz ise (orijinden geçen regresyon doğrusu)
- Bağımsız X açıklayıcı değişkenleri stokastikse (olasılıklı)
- Otokorelasyon değeri 1'den büyükse
- Zaman serisinde ara yıllar noksan ise
- Modelde açıklayıcı değişken olarak gecikmeli açıklanan değişken varsa

Örnek: 50 gözlemlili, 4 açıklayıcı değişkenli bir regresyon modelinde tahmin edilen Durbin-Watson istatistiği $d = 1.43$ olsun. Durbin-Watson tablosundan $d_L = 1.38$ ve $d_U = 1.72$ bulunur. Tahmin edilen d istatistik değeri kararsızlık bölgesindedir. Aynı ya da ters yönlü bir ilişki olup olmadığı söylenemez. Böyle kararsızlık durumlarında uyarlanmış d istatistiği kullanılır.

$$\left. \begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0 \end{array} \right\} \Rightarrow \begin{array}{l} d < d_U \text{ ise } \alpha \text{ anlamlılık düzeyinde } H_0 \text{ red edilir.} \\ \Rightarrow \text{ Aynı yönlü ilişki var.} \end{array}$$

$$\left. \begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho < 0 \end{array} \right\} \Rightarrow \begin{array}{l} (4 - d) < d_U \text{ ise } \alpha \text{ anlamlılık düzeyinde } H_0 \text{ red edilir.} \\ \Rightarrow \text{ Ters yönlü ilişki var.} \end{array}$$

$$\left. \begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{array} \right\} \Rightarrow \begin{array}{l} d < d_U \text{ veya } (4 - d) < d_U \text{ ise } 2\alpha \text{ anlamlılık düzeyinde } H_0 \text{ red edilir.} \\ \Rightarrow \text{ Aynı ya da ters yönlü ilişki var.} \end{array}$$

Örneğe dönersek, $d < d_U$ ($1.43 < 1.72$) olduğundan H_0 red edilir. 1. dereceden ardışık bağımlılık vardır. Bir başka yolda d kararsız bölgede ise ki-kare testi uygulanır.