Değişen Varyans

* Gauss-Markov varsayımlarından biri hata terimi varyansının sabit olduğunu söylüyordu.

Çoklu Doğrusal Regresyon Modeli $\Rightarrow$ $y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_k X_k + u$

ÇDR7: Gözlemlenemeyen hata terimi u-nun bağımsız değişkenlere (x) göre koşullu varyansı sabittir.

$$\text{(1)} \quad Var(u \mid x_1, x_2, \dots x_k) = \sigma^2$$
$$Var(u \mid X) = \sigma^2 \longrightarrow X: \{x_1, x_2, \dots x_k\}$$

* Bu varsayıma göre gözlemlenemeyen hata terimindeki değişkenlik (varyans) açıklayıcı değişkenle (bağımsız değişken) ilişkili olamaz.

* Yukarıda 1 numaralı denklemle gösterilen sabit varyans varsayımı altında bağımlı değişkenin koşullu varyansı da sabittir.

$$\text{(2)} \quad Var(y \mid x_1, x_2, \dots, x_k) = \sigma^2$$
$$Var(y \mid X) = \sigma^2$$

* Sabit varyans durumuna "homoskedasticity" de denilmektedir.

* Sabit varyans varsayımının sağlanmadığı durumda

* Bu konuda ÇDR7 sabit varyans varsayımının sağlanmadığı durumda SEKK tahmincilerinin özelliklerini, değişen varyansın nasıl test edileceğini ve düzeltileceğini göreceğiz.

* Fakat değişen varyansı derinlemesine incelemeden önce karşılaştırma yapabilmek amacıyla, sabit varyans durumunu grafikle inceleyelim.

* Örnek olarak basit doğrusal regresyon modelinde sabit varyansı ayrıntısıyla inceleyelim.

model $\rightarrow$ $y = B_0 + B_1 X_1 + u$ $\longrightarrow$
- ÇDR1 - ÇDR4 ✓
- ÇDR5 $\Rightarrow E(u \mid X) = 0$ ✓ (sıfır koşullu ort)
- ÇDR6 $\Rightarrow Corr(u_i, u_j \mid X) = 0$ ✓ (otokorelasyon yok)
- ÇDR7 $\Rightarrow Var(u \mid X) = \sigma^2$ ✓ (sabit varyans)
  $\downarrow$
  homoskedastik
  yani hata terimlerinin varyansı X-lerin değerlerine göre değişmiyor, sabit.

ARF $\rightarrow$ $E[y \mid X] = B_0 + B_1 X_1$ $\rightarrow$ X-in lineer bir fonksiyonu yani bağımlı değişkenin ortalaması bağımsız değişkenin değerlerine göre değişiyor.
(3)

$Var(y \mid X) = \sigma^2 \Rightarrow$ X-in lineer bir fonksiyonu değil yani bağımlı değişkenin varyansı bağımsız değişkenin değerlerine göre değişmiyor.

- Hatırlatma

① $E(y|X) = E[B_0 + B_1 X + u|X]$

$= B_0 + B_1 X + E[u|X]$ → GDR5 ⇒ $E[u|X] = 0$

$= B_0 + B_1 X$

sabit varyans için neden gerekli? Daha sonra göreceğiz!

② $Var(y|X) = Var(B_0 + B_1 X + u|X)$

GDR 6 ⇒ $Corr(u_i, u_j|X) = 0$

GDR7 ⇒ $Var(u|X) = \sigma^2$

$= Var(u|X)$

$= \sigma^2$

③ $Var(u|X) = \sigma^2$

Eğer u ve X bağımsız ise

$Var(u) = \sigma^2$ olarak da yazılabilir.

④ $Var(u|X) = \sigma^2$ → GDR5 ⇒ $E(u|X) = 0$

$Var(u|X) = E\left[(u - E[u|X])^2 | X\right]$

$= E[u^2|X]$

$Var(u|X) = E[u^2|X] = \sigma^2$

$Var(u) = \sigma^2 = E(u^2)$

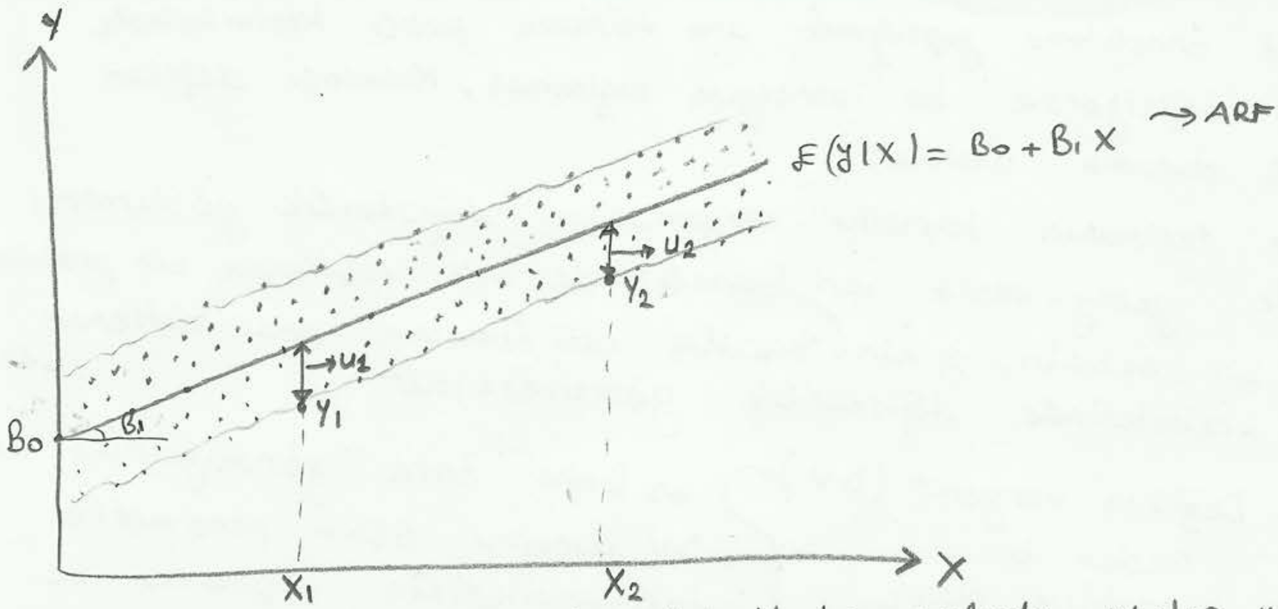
⑤ σ^2 ⇒ hata terimi varyansı

σ ⇒ hata terimi standard sapması

$\hat{\sigma}^2$ ⇒ hata terimi varyans tahmini (Regresyon varyansı)

$\hat{\sigma}$ ⇒ hata terimi standard hatası (Regresyon standart hatası)

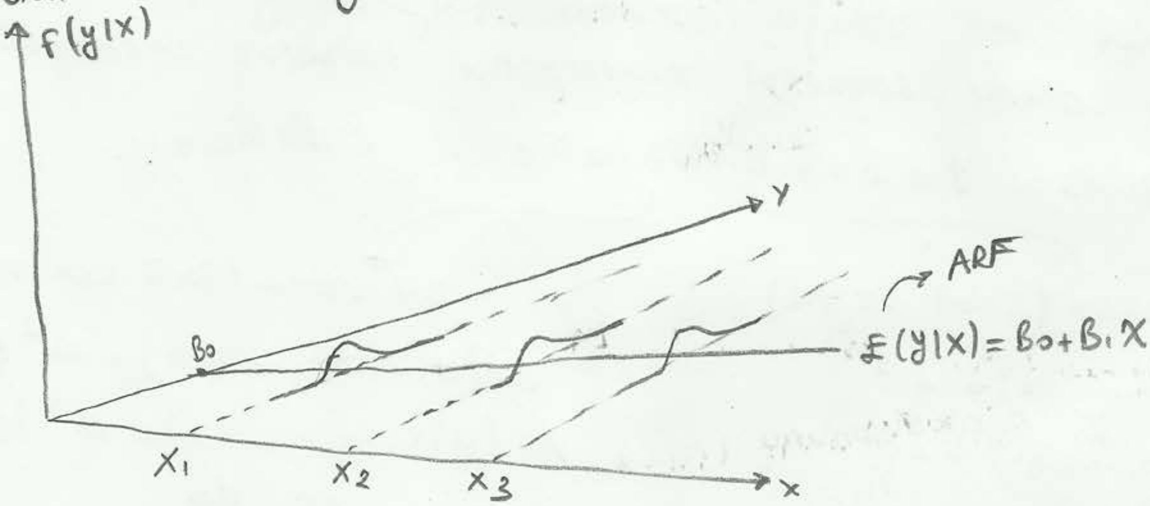
* Şimdi 3 numaralı denklem ile daha önce tanımladığımız ana kütle regresyon fonksiyonunu (ARF) iki farklı grafikle sabit varyans için gösterelim.



* Yukarıdaki grafikte nokta ile gösterilenler aslında y_i-ler yani gözlemlediğimiz bağımlı değişken değerleridir. $\{x_1, x_2\}$, $\{y_1, y_2\}$ ve $\{u_1, u_2\}$ ise birinci ve ikinci gözlem için sırasıyla bağımsız değişken, bağımlı değişken ve hata terimi değerleridir.

* Görüldüğü üzere hem Y hem de u -da, x-lerin değerleri artıkça ortalamadan sapmanın değerinde (yani ∠) bir değişiklik gözlemlenmemektedir. Kısacası sabit varyans vardır.

* Yukarıdaki grafiği y-nin x-e göre koşullu yoğunluk fonksiyonu ile de gösterebiliriz.



* Yoğunluk fonksiyonu Y-lerin nerelerde daha çok kümeleştiğini gösteriyor. X_1, X_2, X_3 ise bağımsız değişkenin farklı değerlerini temsil ediyor.

* Görüldüğü üzere Y bağımlı değişkeni her bir X değeri için aynı yoğunlukta ARF eğrisi etrafında kümelenmiş. Kısacası sabit varyansı işaret ediyor.

* fakat örnekleme yaptığımız ana kütlenin farklı kesimlerinde varyans değişiyorsa bu varsayım sağlanmaz. Kısacası değişen varyans durumu oluşur.

* Hata teriminin koşullu varyansının değişkenlik göstermesi özellikle yatay-kesit veri analizinde sık rastlanan bir problemdir. Bunun bir sebebi y-nin koşullu dağılımının ana kütlenin farklı kesimlerinde değişkenlik göstermesidir.

↳ Değişen varyans (DV) neden bir problem olarak görülüyor?

⇒ Daha önce gördüğünüz üzere DV durumu SEKK parametre tahmincilerinin sapmasız ve tutarlı olmasını engellemez!

Peki problem NE? SEKK parametre tahmincileri DV durumunda etkin olmaz yani parametre tahmincilerinin varyansları minimum olmaz → olması gereken değerden büyük de olabilir küçük de.

← Diğer bir ifadeyle SEKK parametre tahmincilerinin varyansı DV durumunda sapmalı olur.

* DV-nin neden olduğu problemleri daha sonra ayrıntısıyla inceleyeceğiz.

* DV-nin var olduğu durumda varyansın gösteriminin nasıl olacağını çoklu doğrusal regresyonu kullanarak inceleyelim.

model ⇒ $y = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_K X_K + u$

DV Yoksa (sabit varyans)

$$Var(u \mid x_1, x_2, \ldots, x_K) = \sigma^2$$
$$Var(y \mid x_1, x_2, \ldots, x_K) = \sigma^2$$

ya da

$$\begin{cases} Var(u \mid X) = \sigma^2 \\ Var(y \mid X) = \sigma^2 \end{cases}$$

↳ Bazen indeksle de gösterilir

$$Var(u_i \mid X) = \sigma^2$$
$$Var(y_i \mid X) = \sigma^2$$

DV Varsa (Değişen varyans)

$$Var(u \mid x_1, x_2, \ldots, x_K) = \sigma^2 h(x)$$
$$Var(y \mid x_1, x_2, \ldots, x_K) = \sigma^2 h(x)$$

ya da

$$\begin{cases} Var(u \mid X) = \sigma^2 h(x) \\ Var(y \mid X) = \sigma^2 h(x) \end{cases}$$

Bazen bu şekillerde de gösterilebilir.

$$Var(u_i \mid X_i) = \sigma_i^2$$
$$Var(u_i \mid X_i) = \sigma^2 h_i$$

Not:

① $X: \{x_1, x_2, x_3, \ldots, x_K\}$

② $h(x)$ ve h_i bağımsız değişken x-lerin bir fonksiyonudur.

③ $h(x) > 0$ ve $h_i > 0$ ⇒ her X için ⇒ neden?

* Şimdi basit doğrusal regresyonu kullanarak DV durumunu grafik ile incelemeye çalışalım. GDR1 - GDR4 ✓

model => $y = B_0 + B_1 X + u$

GDR5 => $E(u|X) = 0$ ✓ (sıfır koşullu ortalama)

GDR6 => $Corr(u_i, u_j|X) = 0$ ✓ (otokorelasyon yok)

GDR7 => $\cancel{Var(u|X) = \sigma^2}$ ⊗

→ $Var(u|X) = \sigma^2 h(X)$ → Değişen varyans

↓

heteroskedastik yani hata terimlerinin varyansı x-lerin değerlerine göre değişiyor, sabit değil

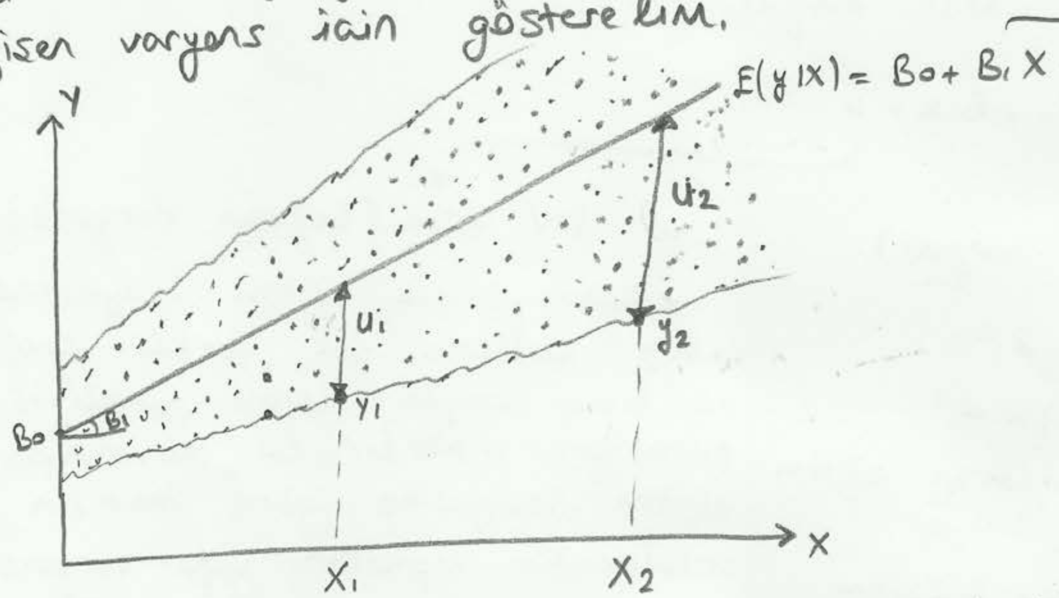
ARF => $E(y|X) = B_0 + B_1 X$ (4)

→ X-in lineer bir fonksiyonu yani bağımlı değişkenin ortalaması bağımsız değişkenin değerlerine göre değişiyor

$Var(y|X) = \sigma^2 h(x)$ => x-in lineer bir fonksiyonu yani bağımlı değişkenin varyansı bağımsız değişkenin değerlerine göre değişiyor.

* Değişen varyans durumuna "heteroskedasticity" de denilmektedir.

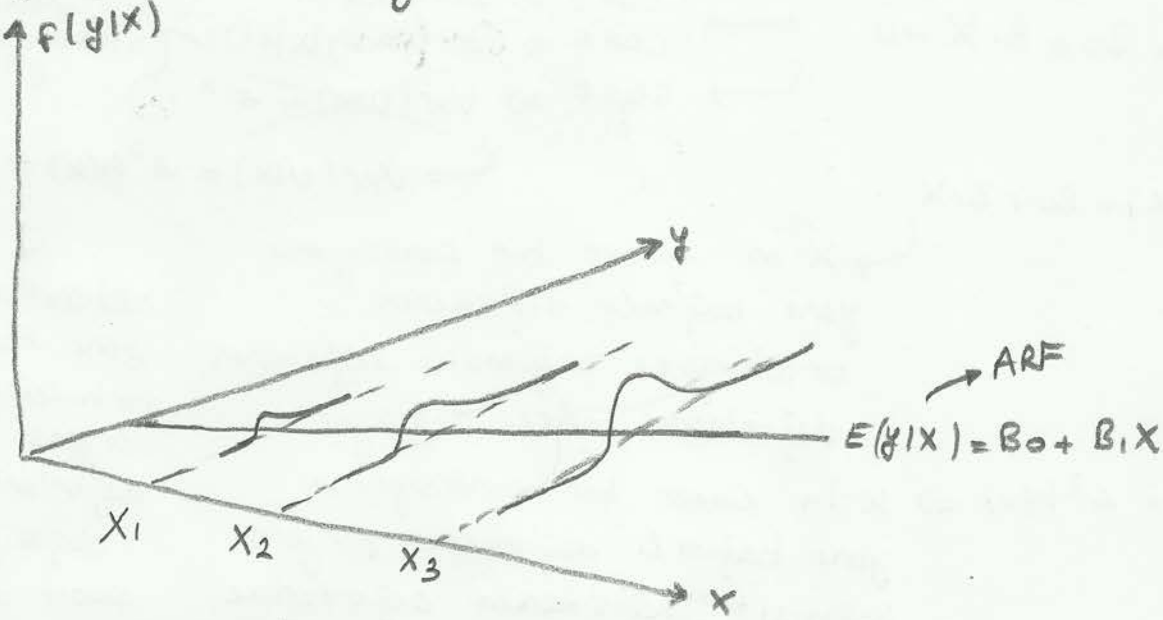
* Şimdi 4 numaralı denklem ile daha önce tanımladığımız ana kütle regresyon fonksiyonunu (ARF) iki farklı grafikle değişen varyans için gösterelim.



* Yukarıdaki grafikte nokta ile gösterilenler aslında yi'ler yani gözlemlediğimiz bağımlı değişken değerleridir. $\{X_1, X_2\}$, $\{y_1, y_2\}$ ve $\{U_1, U_2\}$ ise birinci ve ikinci gözlem için sırasıyla bağımsız değişken, bağımlı değişken ve hata terimi değerleridir.

* Görüldüğü üzere hem Y hem de U-da, X-lerin değerleri arttıkça ortalamadan sapmanın değerinde (varyans) bir artış gözlemlenmektedir. Kısacası değişen varyans vardır.

* Yukarıdaki grafiği Y-nin X-e göre koşullu yoğunluk fonksiyonu ile de gösterebiliriz.



* Yoğunluk fonksiyonu Y-lerin nerelerde daha çok kümeleştiğini gösteriyor. X_1, X_2, X_3 ise bağımsız değişkenin farklı değerlerini temsil ediyor.

* Görüldüğü üzere Y bağımlı değişkeni her bir X değeri için artan yoğunlukta ARF eğrisi etrafında kümelenmiş. Kısacası değişken varyansı işaret ediyor.

ÖRNEK: $Maaş = B_0 + B_1X + u$

DV Yoksa (sabit varyans)

$$\begin{cases} Var(u \mid Eğitim) = \sigma^2 \\ Var(maaş \mid Eğitim) = \sigma^2 \end{cases}$$

→ u ve maaşın varyansı eğitim seviyesine göre değişmiyor.

$E[Maaş \mid Eğitim] = B_0 + B_1 Eğitim$

→ ortalama maaş eğitim ile değişiyor

* Fakat bu durum gerçekçi olmayabilir.

DV Varsa (Değişen Varyans)

* Daha yüksek eğitim seviyesine sahip kişilerin çok farklı uğraşları ve buna bağlı olarak farklı iş fırsatları olabilir. Bu durum da eğitim seviyesine göre maaşın ortalamadan sapmasını yani varyansını arttırır. Ama düşük eğitimlilerin opsiyonu sınırlıdır ve daha çok asgari ücretli işlerde çalışırlar. Bu da düşük varyans demektir. Kısacası değişen varyans vardır.

$$Var(u \mid Eğitim) = \sigma^2 h(Eğitim)$$

$$Var(Maaş \mid Eğitim) = \sigma^2 h(Eğitim)$$

Sabit Voryons vs Değişen Voryon Durumunda SEKK Parametre Tahmincileri ve Voryensları

* Şimdi SEKK parametre tahmincilerinin varyanslarını sabit voryons ve değişen voryons durumlarında basit doğrusal regresyonu kullanarak karşılaştıralım.

daha sonra Çoklu doğrusal regresyon içinde gösterileceği için böyle verildi.

Basit Doğrusal Regresyon

model $\Rightarrow$ $y = B_0 + B_1 X + u$ $\longrightarrow$ GDR 1 - GDR 4 ✓

GDR5 $\Rightarrow$ $E(u|X) = 0$ ✓

indeks ile model $\Rightarrow$ $y_i = B_0 + B_1 X_i + u_i$ (5)

GDR6 $\Rightarrow$ $Corr(u_i, u_j | X) = 0$ ✓

GDR7 $\Rightarrow$ karşılaştırma durumuna göre değiştireceğiz.

* Karşılaştırmayı basitleştirmek için sadece $\hat{B}_1$ ve onun varyansını inceleyelim. Benzer sonuçlar $\hat{B}_0$ için de geçerli olacaktır.

$$\hat{B}_1 = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} = \frac{\sum(x_i-\bar{x})\,y_i}{\sum(x_i-\bar{x})^2} \quad (6)$$

* 5. denklemi 6. denklemdeki son eşitlikte yerine koyalım.

$$\hat{B}_1 = \frac{\sum(x_i-\bar{x})\,y_i}{\sum(x_i-\bar{x})^2} = \frac{\sum(x_i-\bar{x})(B_0+B_1X_i+u_i)}{\sum(x_i-\bar{x})^2}$$

$$= \frac{B_0\sum(x_i-\bar{x}) + B_1\sum(x_i-\bar{x})\,x_i + \sum(x_i-\bar{x})\,u_i}{\sum(x_i-\bar{x})^2}$$

$\sum(x_i-\bar{x}) = 0$

$\sum(x_i-\bar{x})x_i = \sum(x_i-\bar{x})^2$

$$\boxed{\hat{B}_1 = B_1 + \frac{\sum(x_i-\bar{x})\,u_i}{\sum(x_i-\bar{x})^2}} \quad (7)$$

* 7. denklemin X'e göre koşullu olarak beklenen değerini alalım ve $\hat{B}_1$ parametre tahmincisinin sapmalı olup olmadığını bulalım.

$$E(\hat{B}_1|X) = E\left[B_1 + \frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}\right]$$

$\Rightarrow$ GDR5 $\Rightarrow$ $E(u|X) = 0$

$$= B_1 + \frac{\sum\left[(x_i-\bar{x})\,\underbrace{E[u_i|X]}_{=0}\right]}{\sum(x_i-\bar{x})^2}$$

$E[\hat{B}_1|X] = B_1$

Kısacası $\hat{B}_1$ parametre tahmincisi sapmasızdır. Sapmasızlığı kanıtlarken GDR7 varsayımını kullandık mı?

★ Kısacası sabit varyan varsayımı (GAR7) sağlansın ya da sağlanmasın SEKK parametre tahmincileri sapmasızdır. Yani DV sapmasızlığı etkilemez.

★ Şimdi de her iki durumda SEKK parametre tahmincilerinin tutarlı olup olmadığına bakalım.

Tutarlılık = $n \to \infty$ iken parametre tahmincilerinin değeri gerçek parametre değerine yakınsar.

(8) $\underset{n \to \infty}{\text{plim}}\ \hat{B}_1 = B_1$ ise tutarlıdır.

★ 7. denklemdeki $\hat{B}_1$ parametre tahmincisini kontrol edelim.

$$\underset{n \to \infty}{\text{plim}}\ \hat{B}_1 = B_1 + \underbrace{\frac{\Sigma(x_i-\bar{x})u_i}{\Sigma(x_i-\bar{x})^2} = 0}_{n \to \infty}$$

$n\uparrow \to \Sigma(x_i-\bar{x})^2 \uparrow$

$\underset{n \to \infty}{\text{plim}}\ \hat{B}_1 = B_1 \Rightarrow$ $\hat{B}_1$ parametre tahmincisi sabit varyans varsayımı olsun ya da olmasın tutarlıdır.

★ Kısacası sabit varyans varsayımı (GAR7) sağlansın ya da sağlanmasın SEKK parametre tahmincileri tutarlıdır. Yani DV tutarlılığı etkilemez.

★ Şimdi de her iki durumda SEKK parametre tahmincilerinin varyanslarını inceleyip etkin olup olmadıklarına bakalım.

★ Önce 7. denklemdeki $\hat{B}_1$ parametre tahmincisinin varyansını X-e göre koşullu olarak hesaplayalım.

$$Var(\hat{B}_1 | X) = Var\left(B_1 + \frac{\Sigma(x_i-\bar{x})u_i}{\Sigma(x_i-\bar{x})^2} \Big| X\right)$$

B_1 sabit parametredir. ve toplam olarak yazılmış bu nedenle varyansı etkilemez

$$= Var\left(\frac{\Sigma(x_i-\bar{x})u_i}{\Sigma(x_i-\bar{x})^2} \Big| X\right)$$

$$= \frac{\cancel{Var}\ 1}{\left(\Sigma(x_i-\bar{x})^2\right)^2} Var\left(\Sigma(x_i-\bar{x})u_i | X\right)$$

X-e göre koşullu olduğu için X-in fonksiyonu olan $\Sigma(x_i-\bar{x})^2$ sabit terim görevi görür. Çarpım durumunda olduğu için karesi ile varyans dışına çıkar.

$$Var(\hat{B}_1 | X) = \frac{1}{\left(\sum(x_i-\bar{x})^2\right)^2} Var\left(\sum(x_i-\bar{x}) u_i | X\right)$$

Bu sadeleştirmeyi yapmak için özel bir varyans kuralı kullandık.

$$= \frac{1}{\left(\sum(x_i-\bar{x})^2\right)^2} \sum(x_i-\bar{x})^2 Var(u_i|X)$$

Kural: Eğer a_i sabit bir sayı ve u_i-ler ikili olarak ilişkisiz ise yani $Corr(u_i, u_j) = 0$ ise

$$Var\left(\sum a_i u_i\right) = \sum a_i^2 Var(u_i)$$

$$= \frac{\sum(x_i-\bar{x})^2 Var(u_i|X)}{\left(\sum(x_i-\bar{x})^2\right)^2}$$

Dikkat: hata terimleri arasında otokorelasyon olmaması işte bu noktada devreye giriyor. Varyans formülünün sadeleştirilmesinde işe yarıyor.

Not: formüldeki $a_i = x_i - \bar{x}$

$$\boxed{Var(\hat{B}_1 | X) = \frac{\sum(x_i-\bar{x})^2 \sigma_i^2}{\left(\sum(x_i-\bar{x})^2\right)^2}} \quad (9)$$

* Eğer sabit varyans varsayımı sağlandıysa yani eğer $Var(u|X) = \sigma^2$ ise 9. denklem sadeleştirilebilir.

$$Var(\hat{B}_1 | X) = \frac{\sigma^2 \sum(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2}$$

$$(10) \quad \boxed{Var(\hat{B}_1) = \frac{\sigma^2}{\sum(x_i-\bar{x})^2}} \Rightarrow \text{sabit varyans varsa}$$

* Eğer DV varsa yani eğer $Var(u_i|x_i) = \sigma_i^2$ ise, 9. denklem olduğu gibi kalır.

$$(11) \quad \boxed{Var(\hat{B}_1) = \frac{\sum(x_i-\bar{x})^2 \sigma_i^2}{\left(\sum(x_i-\bar{x})^2\right)^2}} \Rightarrow \text{Değişen varyans varsa}$$

* Kısacası DV durumunda SEKK parametre tahmincilerinin varyansı minimum değildir. Yani varyans ve standard sapmalar sapmalıdır. Bu nedenle parametre tahmincileri etkin değildir. Yani en iyi sapmasız parametre tahmincileri değildir. → DESTE değildir.
* Bu nedenle t ve F testleri de artık güvenilir değildir.

* Yukarıda basit doğrusal regresyon için yazdığımız varyans denklemlerini (10. ve 11.) şimdi çoklu doğrusal regresyon için de yazalım.

Çoklu Doğrusal Regresyon

model ⇒ $y = B_0 + B_1 x_1 + B_2 x_2 + \dots + B_k X_k + u$

model indeks ile yazılışı ⇒ $y_i = B_0 + B_1 X_{i1} + B_2 X_{i2} + \dots + B_k X_{ik} + u$

• (12) $$\boxed{Var(\hat{B}_J) = \frac{\sigma^2}{SST_J(1-R_J^2)} = \frac{\sigma^2}{SSR_J}}$$ ⇒ sabit varyans durumu

Hatırlatma: • $SST = SSE + SSR$
• $R^2 = \frac{SSE}{SST}$
• $SSR = SST(1-R^2)$

• (13) $$\boxed{Var(\hat{B}_J) = \frac{\sum \hat{r}_{ij}^2 \sigma_i^2}{SSR_J^2}}$$ ⇒ değişen varyans durumu

Hatırlatma:
• $SST_J = \sum(x_{ij} - \bar{x}_J)^2$
• R_J^2 ⇒ X_J'yi diğer tüm bağımsız değişkenler üzerine regres ettiğimizde modelin R^2'si
• SSR_J ⇒ X_J'yi diğer tüm bağımsız değişkenler üzerine regres ettiğimizde modeldeki SSR
• $\hat{r}_{iJ}$ ⇒ X_J'yi diğer tüm bağımsız değişkenler üzerine regres ettiğimizdeki i. artık.

* Görüldüğü gibi DV olduğunda parametre tahmincilerinin hesaplanması zorlaşıyor.

* Yukarıdaki 10, 11, 12 ve 13 numaralı denklemlerdeki varyans henüz tahmin edilen varyans değil. Eğer tahmin etmek istersek $\hat{\sigma}^2$ ya da $\hat{u}_i^2$'yi kullanabiliriz.

(10') $$\boxed{Var(\hat{B}_1) = \frac{\hat{\sigma}^2}{\sum(x_i-\bar{x})^2}}$$

(11') $$\boxed{Var(\hat{B}_1) = \frac{\sum(x_i-\bar{x})^2 \hat{u}_i^2}{\left(\sum(x_i-\bar{x})^2\right)^2}}$$

Basit Doğrusal Regresyon

(12') $$\boxed{Var(\hat{B}_J) = \frac{\hat{\sigma}^2}{SSR_J}}$$

(13') $$\boxed{Var(\hat{B}_J) = \frac{\sum \hat{r}_{iJ}^2 \hat{U}_i^2}{SSR_J^2}}$$

Çoklu Doğrusal Regresyon

DV YOK | DV VAR

Değişen Varyansın Sonuçları

① Parametre tahmin edicileri sapmasızdır. Ayrıca asimptotik olarak da sapmasızdır.

② Parametre tahmin edicileri doğrusaldır.

③ Parametre tahmin edicilerinin varyansları sapmalıdır.

④ Bu nedenle parametre tahmin edicileri etkin değildir. Ayrıca asimptotik olarak da etkin değildir. Yani n↑ bile (gözlem sayısı artsa) parametre tahmincileri yine de etkin olmaz.

Not: Çoklu Doğrusal Bağıntı (ÇDB) durumunda n↑ ise parametre tahmincileri etkin olmaya varyans düşer ve parametre tahmincileri etkin olmaya yakınlaşır. Fakat DV durumunda n↑ ile sorun çözülemez.

- Bu nedenle daha önceden varyansları hesaplamak için kullanılan formüller (örnek: 10 ve 12) ile varyans hesaplandığında bu değerler gerçek varyanslardan büyük veya küçük olur.
- Bu durum da güven aralığı tahminlerini, t ve F testlerini etkiler. Geleneksel olarak kullanılan t ve F testleri t ve F dağılımı bile yapmaz, n↑ bile.

⑤ Parametre tahmin edicileri hala tutarlıdır.

Sonuç: DV parametre tahmin edicilerinin sapmasız ve tutarsız olmasına neden olmaz. Fakat etkinlik sağlanamaz.

Değişen Varyansın Nedenleri

★ DV daha çok yatay-kesit verilerinde görülür.
★ DV daha çok modelin yapısından veya veriden kaynaklanır.

① Model tanımlama hatası (gerekli değişkenin kullanılmaması)

Doğru model $Y = B_0 + B_1 X_1 + B_2 X_2 + \varepsilon$

$U = B_2 X_2 + \varepsilon$

Kullanılan Model $Y = B_0 + B_1 X_1 + U$

varyans formulündeki kovaryans terimi nerede?

$$Var(U|X_2) = Var(B_2 X_2 + \varepsilon | X_1)$$
$$= B_2^2 \underbrace{Var(X_2|X_1)}_{=?} + \underbrace{Var(\varepsilon|X_1)}_{=\sigma^2}$$

⇒ kısacası DV olacak X_i'e bağlı olarak değişecek

Sıfır olur mu? Genellikle hayır.

(2) Katsayıların zamana (zaman serisinde) veya birime (yatay-kesit) göre değişmesidir.

Zaman serisi

Doğru Model $\Rightarrow$ $y_t = B_0 + B_1 X_{1t} + B_{2t} X_{2t} + \varepsilon_t$ $\rightarrow$ B_2 zamana göre değişiyor

Kullanılan Model $\Rightarrow$ $y_t = B_0 + B_1 X_{1t} + B_2 X_{2t} + u_t$ $\rightarrow$ B_2 sabit

yani
$$\begin{cases} B_{2t} = B_2 + v_t \\ u_t = X_{2t} . v_t + \varepsilon_t \end{cases}$$

$$Var(u_t | X) = \underbrace{Var(X_{2t} . v_t | X)}_{\text{genellikle sabit olmaz.}} + Var(\varepsilon_t | X)$$

Yatay Kesit

$$Tüketim = B_0 + B_1 Gelir + \varepsilon$$

- gelir arttıkça marjinal tüketim değişecektir

$$\frac{\partial Tüketim}{\partial Gelir} = B_1$$ $\Rightarrow$ genelde sabit değildir. Bu nedenle kukla değişken kullanarak veriyi homojen alt gruba bölebiliriz.

(3) Veri toplamasında yapılan hatalar.

(4) Bağımlı değişkende yapılan ölçme hataları.