Hafta #8

* Değişen varyansın düzeltilmesine geçmeden önce modelimizde değişen varyans olup olmadığını test etmemiz gerekir.
* Literatürde çok sayıda DV testi geliştirilmiştir.
* Bu derste yatay-kesit verisinde uygulanan DV testleri üzerinde duracağız.
* Daha çok yaygın olarak kullanılan iki parametrik test üzerinde duracağız: Breusch-Pagan ve White testleri.

DV Testleri

SEKK Tahminlerine Dayanan Testler

1. Breusch-Pagan *
2. White *
3. Goldfeld-Quandt *
4. Glejser *

Maksimum Olabilirlik Yöntemine Dayanan Testler

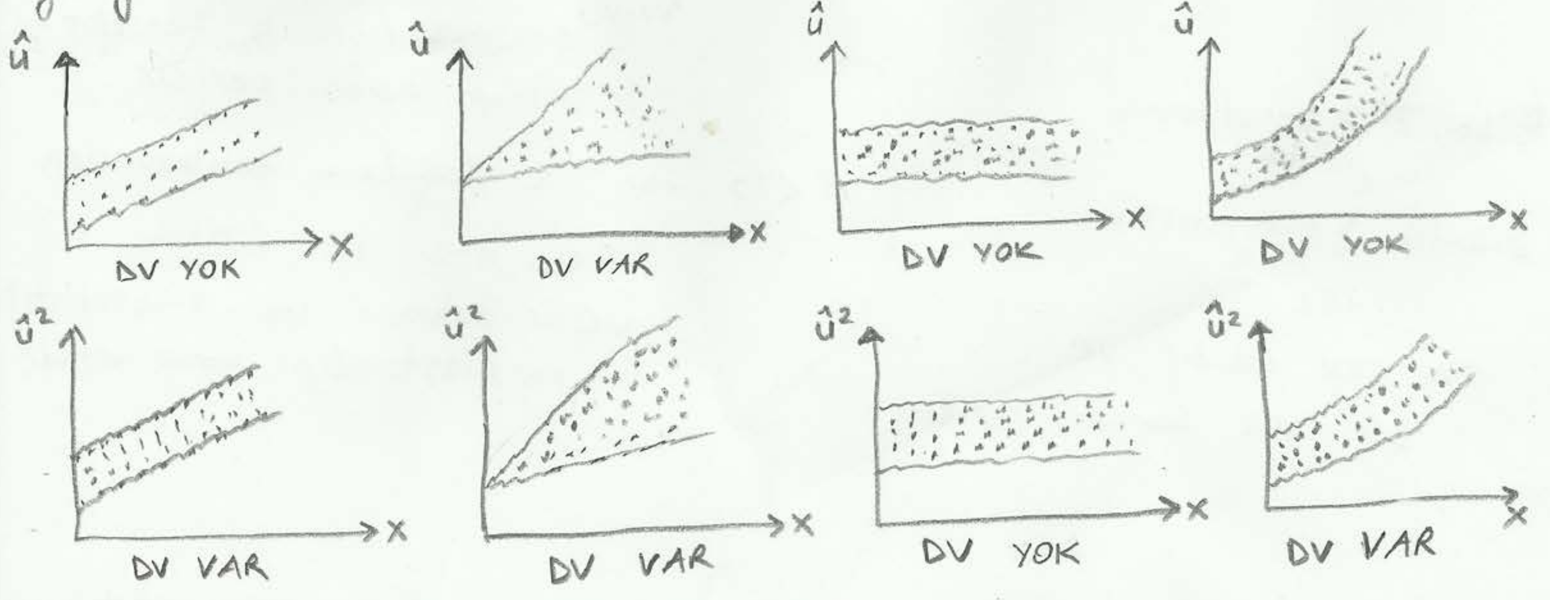
1. LR (Likelihood Ratio)
2. LM (Lagrange Multiplier)
3. Wald

Non-Parametrik Testler

1. Spearman Sıra Korelasyon Testi *
2. Peak testi

* SEKK tahmincilerine dayanan başka testler de vardır: Ramsey Reset, Park ve Harrison-McCabe gibi. Fakat biz yukarıda yazılan 4 DV testini inceleyeceğiz sadece.

<u>Not:</u> Bu testlerden önce yapılması gereken, bağımsız değişkenlerin hangisinden şüpheleniliyorsa o değişken ve artıkların grafik yoluyla incelenmesidir.



1 Spearman Sıra Korelasyon Testi

* Parametrik olmayan bir testtir.
* Dağılımla ilgili herhangi bir varsayıma ihtiyaç duymaz çünkü örnek gözlemlerinin değeri yerine sırasını kullanır.
* Artıkların sırası ve bağımsız değişkenlerin sıraları arasındaki ilişkiyi inceler.

$H_0: \rho_s = 0$ DV YOK (sabit varyans)

$H_1: \rho_s \neq 0$ DV VAR

$y = \beta_0 + \beta_1 X + \varepsilon \longrightarrow$
- SEKK yöntemi ile $\hat{\varepsilon}$ hesapla
- mutlak değer hesapla $\Rightarrow |\hat{\varepsilon}|$
- $|\hat{\varepsilon}|$ ve X-i değerlerine göre ayrı ayrı küçükten büyüğe sıra numarası ver.
- Sıra numarası verilirken değişkenlerin aldıkları değerler birden fazla tekrarlanıyorsa bu değerler ile ilgili sıra numarasının ortalaması alınır.

$d_i =$ Sıra $X_i -$ Sıra $|\hat{\varepsilon}_i|$

$$\rho_s = 1 - 6\left[\frac{\sum d_i^2}{n(n^2-1)}\right]$$

ρ_s: sıralar arasındaki korelasyon

$-1 \leq \rho_s \leq 1$

anlamlılık test edilir α anlamlılık düzeyi ile birlikte.

$$t = \frac{\rho_s\sqrt{n-2}}{\sqrt{1-\rho_s^2}}$$

$n > 30$ ise normal dağılım tablosundan $Z_{\alpha/2}$ ile kritik değer bulunur ve t-istatistiği ile karşılaştırılıp karar verilir.

$n < 30$ ise t-dağılım tablosundan $t_{\alpha/2, n-2}$ ile kritik değer bulunur ve t-istatistiği ile karşılaştırılıp karar verilir.

t-kritik > t-istatistik
ya da
z-kritik > t-istatistik
ise
H_0 red edilir ve
DV VAR kararı verilir.

* Bu test modeldeki her değişken için ayrı ayrı uygulanabilir.
* Çoklu doğrusal regresyonda her değişken için ayrı ayrı test yapılır. Sabit varyans sonucuna varabilmek için her bağımsız değişken için Ho temel hipotezini red etmemiz gerekir.

Parametrik DV Testleri için Önemli Notlar

* Parametrik DV testleri GDR1 - GDR6 varsayımlarının geçerli olduğunu dolayısıyla SEKK parametre tahmincilerinin sapmasız ve tutarlı olduğunu varsayar.
* Test etmek istediğimiz hipotez GDR7 yani sabit varyans varsayımının geçerli olup olmadığıdır.

model ⇒ $y = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_k X_k + u$

(1)
$$H_0: Var(u|X) = \sigma^2 \quad \text{sabit varyans}$$
$$H_1: Var(u|X) \neq \sigma^2 \quad \text{değişen varyans}$$

* Temel hipotez aşağıdaki gibi de yazılabilir

$$H_0: E(u^2|X) = E(u^2) = \sigma^2$$

* Bu nedenle sabit varyans için bağımsız değişkenler hata teriminin karesiyle ilişkisiz olmalıdır. Değişen varyans testleri bu ilişkinin olup olmadığını test eder.

(1) Breusch-Pagan Testi

* DV-nin fonksiyonel formuna (daha sonra açıklanacak) bağlı olmayan asimtotik (büyük örneklem) testidir.
* White testi gibi DV-ye yol açtığı düşünülen değişken sayısı birden fazla ise kullanılabilir. Fakat DV için genel bir test olarak düşünüldüğünden dolayı DV-ye neden olan değişken sayısına bakılmaksızın kullanılır.

model ⇒ $y = B_0 + B_1 X_1 + B_2 X_2 + \ldots + B_k X_k + u$ (2) → GDR1- GDR6 ✓, GDR 7 → test edilecek

* Eğer 1 numaralı denklem kümesinde belirtilen temel hipotez yanlışsa, u^2-nin koşullu beklenen değeri x-lerin herhangi bir fonksiyonu olabilir.

★ u^2'nin X'ler ile doğrusal ilişkili olduğu varsayılırsa aşağıdaki model test için kullanılabilir.

u^2 tüm X'ler üzerine regres edilir $\Rightarrow$ $u^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + v$ → bu modeldeki hata

↳ yardımcı regresyon

★ Bu model çerçevesinde hipotezler aşağıdaki gibi yazılabilir.

$$H_0: \overbrace{\alpha_1 = \alpha_2 = \dots = \alpha_k = 0}^{DV\ YOK} \quad vs \quad H_1: \overbrace{H_0 \text{ doğru değil.}}^{DV\ VAR}$$

★ Boş hipotez altında $E(u^2|X) = \alpha_0$, sabit bir sayı, olmaktadır.

★ u gözlemlenemediğinden modelden elde edilen kalıntılar $\hat{u}$ kullanılarak F ya da LM (Lagrange Multiplier) testi yapılabilir.

★ 2. denklemdeki modeli tahmin ettikten sonra kalıntıların karesinin tüm X'ler üzerine regresyonunu kurarak:

$$\hat{u}^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + v \rightarrow \underline{R^2_{\hat{u}} \text{ çek}}$$

F ve LM testlerini bu model çerçevesinde hesaplayabiliriz.

$$F\text{-ist} = \frac{R^2_{\hat{u}} / k}{(1 - R^2_{\hat{u}}) / (n-k-1)} \Rightarrow F_{k,\, n-k-1} \Rightarrow F \text{ dağılımı yapar.}$$

$sd1 = k$ (pay)

$sd2 = n-k-1$ (payda)

$$LM\text{-ist} = n R^2_{\hat{u}} \Rightarrow \chi^2_k \Rightarrow \text{ki-kare dağılımı yapar}$$

$sd = k$

$\boxed{\text{Not: } n = \text{gözlem sayısı} \ ; \ k: \text{bağımsız değişken sayısı}}$ → yardımcı regresyondaki

★ α: anlamlılık seviyesinde eğer

$F\text{-ist} > F_{\alpha,\, k,\, n-k-1}$ (F-kritik)

ya da

$LM\text{-ist} > \chi^2_{\alpha, k}$ (kritik değer)

} Ho red edilir ve DV sonucuna varılır. Aksi halde Ho red edilemez ve sabit varyans sonucuna varılır.

★ Yukarıda bahsettiğimiz testin LM versiyonuna Breusch-Pagan değişen varyans testi adı verilir.

② White Test

* DV için yaygın olarak kullanılır.

* DV için regresyonda kaç değişken varsa hepsi ele alınır. Aynı Breusch - Pagan testindeki gibi.

* DV-ye neden olan değişkenin bilinmemesi durumunda uygun olan bir testtir. Aynı Breusch - Pagan testi gibi.

* SEKK tahminlerine dayanan parametrik asimtotik (büyük örneklem) testidir. Aynı Breusch - Pagan test gibi.

* Ekonometri I dersinde Gauss - Markov varsayımlarının tümünün sağlanması halinde SEKK standart hatalarının ve test istatistiklerinin asimptotik olarak geçerli olacaklarını görmüştük.

* Bu, sabit varyans varsayımının, daha zayıf şu varsayımla yer değiştirebileceği anlamına gelir: "u^2, tüm bağımsız değişkenlerle, x_J, onların kareleriyle, x_J^2, ve çapraz çarpımlarıyla, $x_J x_h$, $J \neq h$, ilişkisizdir.

* Bu varsayım White (1980) değişen varyans testinin temelini oluşturmaktadır.

* Varyanstaki değişkenlik x_J'lerle doğrusal olmayan bir şekilde ilişkiliyse White testi bunu yakalayabilir.

* Testin adımları Breusch - Pagan testine benzer. Sadece ikinci adımdaki yardımcı regresyonda x-lerin kareleri ve çapraz çarpımları eklenir.

* Bu testte bağımsız değişken sayısı, k, arttıkça yardımcı regresyondaki serbestlik derecesi azalmaktadır. Bu nedenle anlatımı basitleştirmek için $k=3$ varsayalım.

$k=3$ iken

ana model → $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$ ③ → GDR1 - GDR6 ✓

GDR7 → test edilecek.

→ $\hat{u}$'leri çek ve $\hat{u}^2$-leri hesapla

- $\hat{u}^2$-leri tüm x-ler, x-lerin kareleri ve çapraz çarpımları üzerine regres edip yardımcı regresyonu oluştur.

$$\hat{u}^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + \alpha_6 X_3^2 + \alpha_7 X_1 X_2 + \alpha_8 X_1 X_3 + \alpha_9 X_2 X_3 + v$$

yardımcı model → ④ → $\underline{R^2_{\hat{u}} \text{ çek}}$

v → yardımcı modeldeki hata terimi

$H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_q = 0 \longrightarrow$ DV YOK (sabit varyans)

H_1: Ho doğru değil → yani yukarıdaki eşitliklerden en az biri doğru değil → DV VAR

★ Yukarıdaki hipotezi test etmek için aynen Breusch-Pagan testte olduğu gibi F ve LM testi kullanabiliriz.

$$F\text{-ist} = \frac{R^2_{\hat{u}} / k}{(1-R^2_{\hat{u}})/(n-k-1)} \Longrightarrow F_{k,\,n-k-1} \Rightarrow F \text{ dağılımı yapar.}$$

sd1 = k (pay)

sd2 = n-k-1 (payda)

5

$$LM\text{-ist} = nR^2_{\hat{u}} \Rightarrow \chi^2_k \Rightarrow \text{ki-kare dağılımı yapar}$$

sd = k

$\boxed{\text{Not: } n = \text{gözlem sayısı} \; ; \; k = \text{bağımsız değişken sayısı}}$ → yardımcı regresyondaki

Önemli: örneğimizdeki modelde k=3 iken yardımcı regresyonda k=9 oldu.

★ α = anlamlılık seviyesinde eğer

$F\text{-ist} > F_{\alpha,k,n-k-1}$ (F-kritik)

ya da

$LM\text{-ist} > \chi^2_{\alpha,k}$ (kritik değer)

} Ho red edilir ve DV sonucuna varılır. Aksi halde Ho red edilmez ve sabit varyans sonucuna varılır.

★ Ana modelde k=3 iken Breusch-Pagan ve White testlerinin yardımcı regresyonundaki k-yi karşılaştırdığımızda, White testinde k=9 iken, Breusch-Pagan testinde k=3-dür.

★ Örneğin ana modelde k=6 olduğunda, White testi için oluşturulan yardımcı regresyondaki k=27 olacaktır. Bu da serbestlik kaybına yol açar. → serbestlik derecesi azalır. Bu nedenle White testinin yukarıdaki versiyonunun zayıf tarafı budur.

★ Bu nedenle uygulamada genellikle yardımcı modelde daha az bağımsız değişken içeren (yani daha az k) test versiyonu kullanılır.

★ Yardımcı regresyonda X-lerin kareleri ve çapraz çarpımlarını açıkça yazmak yerine, ana modelin tahmininden elde edilen $\hat{y}$ değerleri ve bunun kareleri kullanılabilir.

Neden? $\hat{y}$ → X-lerin fonksiyonu $\hat{y}^2$ ise → $X_j X_h$ ve X_j^2-lerin fonksiyonu

* Bu versiyon için yine anlatımı basitleştirmek amacıyla $k=3$ varsayalım.

ana model $\longrightarrow$ $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u$ $\longrightarrow$ GDR1 - GDR6 ✓

$\longrightarrow$ GDR7 $\rightarrow$ test edilecek

⑥ $\rightarrow$ $\hat{u}$'leri çek ve $\hat{u}^2$'leri hesapla

$\rightarrow$ $\hat{y}$'leri çek ve $\hat{y}^2$'leri hesapla

yardımcı model $\longrightarrow$ $\hat{u}^2 = \alpha_0 + \alpha_1 \hat{y} + \alpha_2 \hat{y}^2 + v$

$\rightarrow$ yardımcı modeldeki hata terimi

$H_0: \alpha_1 = \alpha_2 = 0 \rightarrow$ DV YOK (sabit varyans)

H_1: H_0 doğru değil $\rightarrow$ yani yukarıdaki eşitliklerden en az biri doğru değil $\longrightarrow$ DV VAR

* Yukarıdaki hipotezi test etmek için, önceki modelde
olduğu gibi F ve LM testi kullanılabilir, 5. denklem
kümesinde verilen formüller ve yönergelere bakın.
ana modeldeki X sayısı
* Bu versiyondaki White testinde
sadece 2 kısıt vardır (yani $k=2$). Böylece,
ne olursa olsun, serbestlik derecesi kaybı burada
testin orijinal versiyonundaki söz konusu değildir.
söz kaybı bu versiyonda

③ Goldfeld - Quandt Test

parametrik bir testtir.
* Küçük örnekleme uygulanan modelde yer alan bağımsız değişken-
* σ_i^2'nin (yani $Var(u|X) = \sigma_i^2$) olduğu varsayılmaktadır.
lerden biri ile pozitif ilişkili bilinmesi durumunda
* DV-ye neden olan bağımsız değişkenlerin bir testtir.
uygulanması daha uygun olan çoklu doğrusal regresyon yerine
* Anlatımı basitleştirmek için testi uygulamaya çalışalım.
basit doğrusal regresyonu kullanarak

model $\longrightarrow$ $y = \beta_0 + \beta_1 X + u$ $\longrightarrow$ GDR1-GDR6 ✓

$\longrightarrow$ GDR7 test edilecek

• X ile σ_i^2 arasında (+) pozitif ilişki olup olmadığı test edildiği için, aslında bu testin mantığı büyük X değerleri büyük artıklara ($\hat{u}$) neden olur mantığına dayanır.

$\hookrightarrow$ $\hat{u}$ aslında σ_i'nin temsilcisidir.

★ Bu testi uygularken öncelikle X bağımsız değişkeninin büyüklüğüne göre (küçükten büyüğe doğru) gözlemler sıralanır. Daha sonra ortadaki "p" kadar gözlem atılır ve iki alt gözlem grubu oluşturulur.

★ p ne çok küçük ne de çok büyük olmalıdır.

- literaturde $p = n/4$ ya da $p = n/5$ kullanılır.

n = gözlem sayısı

★ P belirlendikten sonra ilk alt grup ve ikinci alt grup için regresyon modeli ayrı ayrı tahmin edilerek artıkların kareleri toplamı (SSR) hesaplanır.

SSR → residual sum of squares $= \sum \hat{u}_i^2$

- İlk alt grup için artık kareleri toplamı $= SSR_1 = \sum \hat{u}_{i1}^2$
- İkinci alt grup için artık kareleri toplamı $= SSR_2 = \sum \hat{u}_{i2}^2$

★ Hipotez testi aşağıdaki gibi oluşturulur:

H_0 : DV YOK (sabit varyans)

H_1 : DV VAR

★ Yukarıdaki hipotezi test etmek için kullanmamız gereken test istatistiği aşağıda verilmiştir.

$$F\text{-istatistik} = \frac{SSR_2}{SSR_1} = \frac{\sum \hat{u}_{i2}^2}{\sum \hat{u}_{i1}^2} \Rightarrow F_{sd1,\, sd2} \Rightarrow F \text{ dağılımı yapar}$$

$$sd1 = \left(\frac{n-p}{2}\right) - k - 1 = n_1 - k - 1 \text{ (pay)}$$

$$sd2 = \left(\frac{n-p}{2}\right) - k - 1 = n_2 - k - 1 \text{ (payda)}$$

Not: n = gözlem sayısı (tüm gözlem); n_1 = birinci alt grup gözlem sayısı

n_2 = ikinci " " " "

p = seçilip atılan gözlem sayısı

k = bağımsız değişken sayısı

★ α = anlamlılık seviyesinde

$F\text{-ist} > F_{\alpha, sd1, sd2}$ (F-kritik) ise ⟹ H_0 red edilir ve DV sonucuna varılır. Aksi halde H_0 red edilemez ve sabit varyans sonucuna varılır.

* Bu teste hangi değişkenin DV-ye neden olduğuna karar vermek için bağımsız değişkenlere ayrı ayrı uygulanabilir. Böylece sabit varyansın geçerli olup olmadığı ve geçersiz ise hangi bağımsız değişkenin neden olduğu bulunabilir.

Glajser Testi

* SEKK tahmin artıklarına dayanan asimptotik (büyük örneklem) testidir.
* DV, hata terimi ve bağımsız değişken arasındaki farklı fonksiyonel formlar ile açıklanmaya çalışılır. DV kalıbı mümkün tüm fonksiyonel formlar denenerek bulunmaya çalışılır.
* Diğer testlerden farklı olarak σ_i^2 temsilcisi olarak ana modelden elde edilen $|\hat{u}_i|$ kullanılır.

ana model $\rightarrow$ $y = B_0 + B_1 X_1 + B_2 X_2 + u$ ⑥ $\longrightarrow$ GDR1 ve GDR6 ✓

GDR7 $\rightarrow$ test edilecek

$\rightarrow$ $\hat{u}_i$-leri çek ve $|\hat{u}_i|$-leri hesapla.

Farklı Fonksiyonel Formlardaki Yardımcı Regresyonlar

$$|\hat{u}| = \alpha_0 + \alpha_1 X_1 + \varepsilon$$

$$|\hat{u}| = \alpha_0 + \alpha_1 \sqrt{X_1} + \varepsilon$$

$$|\hat{u}| = \alpha_0 + \alpha_1 \frac{1}{X_1} + \varepsilon$$

$$|\hat{u}| = \alpha_0 + \alpha_1 \frac{1}{\sqrt{X_1}} + \varepsilon$$

- Başka fonksiyonel formlarda da yardımcı regresyon kullanılabilir.
- Yardımcı regresyonlardaki X_1, DV-ye yol açtığı düşünülen bağımsız değişkendir. Bu yönüyle Goldfeld-Quandt testine benzerken, white ve Breusch-Pagan testlerinden ayrılır (tek bir bağımsız değişken olduğundan)
- Yukarıdaki fonksiyonel kalıpların hepsinin denemesi şart olmayıp bir veya bir kaç tanesi ayrı ayrı denenebilir.

* Yardımcı regresyonlarda genelde tek bir bağımsız değişken kullanılmasının temel sebebi DV'yi gidermenin daha kolay olmasıdır.
* Yardımcı regresyonlardan denenen modellerin t, R^2, F ve SSR-leri karşılaştırılarak en uygun kalıp seçilir. Böylece DV kalıbı elde edilmiş olur, ve DV'yi ortadan kaldırmak için bu kalıp kullanılır.

★ 6. denklemde verilen modelde, DV-yi test etmek için aşağıdaki fonksiyonel form yardımcı regresyon için seçilmiş olsun,

$$|\hat{u}| = \alpha_0 + \alpha_1 X_1 + \varepsilon$$

- α_0 ve α_1 parametrelerin istatistiki olarak anlamlı ya da anlamsız olmasına göre DV kararı verilir.
- Not = α_0 ve α_1 için anlamlılık testi t-testi ile yapılır.
- Eğer t-testleri sonucunda
  - + $\alpha_0 = 0$ ve $\alpha_1 = 0$ ise DV YOK (sabit varyans)
  - + $\alpha_0 = 0$ ve $\alpha_1 \neq 0$ ise DV VAR
    Bu durumda DV kalıbı (formu) $\sigma_i^2 = \alpha_1^2 X_{i1}^2$ olur
  - + $\alpha_0 \neq 0$ ve $\alpha_1 \neq 0$ ise DV VAR
    Bu durumda DV kalıbı (formu) $\sigma_i^2 = (\alpha_0 + \alpha_1 X_{i1})^2$ olur
- Kısacası DV sonucuna varılması için α_0-ın istatistiki olarak anlamlı olup olmamasına gerek yoktur. Bu nedenle sadece modelin geneline F-testi uygulayarak DV olup olmadığına karar verilebilir. Tabi bu durumda σ_i^2-nin formu hakkında kesin bir bilgi sahibi olamayız.
- Bu durumda neden DV-nin formu $\alpha_1^2 X_1^2$ şeklinde oldu? ⟶ Çünkü DV-yi ortadan kaldırmak için ya artıkların kareleri $\sum \hat{u}_i^2$ $\alpha_1^2 X_1^2$ ile ağırlıklandırılmalı ya da tüm değişkenler $\alpha_1 X_1$ ile ağırlıklandırılır.
  ⟶ yani $Y^* = \dfrac{Y}{\alpha_1 X_1}$; $X_1^* = \dfrac{X_1}{\alpha_1 X_1}$; $X_2^* = \dfrac{X_2}{\alpha_1 X_1}$
- Bir sonraki konuda Ağırlıklandırılmış En Küçük Kareler yönteminde bu durumu ayrıntılarıyla inceleyeceğiz.

DV Testleri Hakkında Önemli Notlar

★ DV testlerini yaparken GDR1 - GDR6 varsayımlarının sağlandığını varsayıyoruz.
★ Bu varsayımlar sağlanmadan (ör: ihmal edilmiş değişken varsa ya da log-log yerine level-level fonksiyonel form kullanıldıysa), DV testi uygulanırsa modelde sabit varyans varken bile hipotez testi sonucunda Ho red edilip DV sonucuna varılabilir. Bir başka ifadeyle, 1. Tip hata (α) olasılığı artar.
★ Bu nedenle ekonometrisler değişen varyans testlerini "yanlış biçim seçimi" (misspecification) testleri olarak değerlendirilir. Ancak fonksiyon formu seçimi doğrudan başka testlerle test edilmelidir. Yanlış fonksiyonel form seçimi DV-den daha ciddi bir sorundur.