

DEĞİŞEN VARYANS

Hüseyin Taştan¹

¹Yıldız Teknik Üniversitesi
İktisat Bölümü

Ders Kitabı:
Introductory Econometrics: A Modern Approach (2nd ed.)
J. Wooldridge

14 Ekim 2012

Değişen Varyans (Heteroscedasticity)

- ▶ Gauss-Markov varsayımlarından biri hata teriminin varyansının sabit olduğunu söylüyordu (MLR.5: Sabit varyans):

$$\text{Var}(u|x_1, x_2, \dots, x_k) = \sigma^2$$

- ▶ Bu varsayıma göre gözlenemeyen hata terimindeki değişkenlik (varyans) açıklayıcı değişkenlerle ilişkili olamaz.
- ▶ Örnekleme yaptığımız popülasyonun farklı kesimlerinde varyans değişiyorsa bu varsayım sağlanmaz.
- ▶ Yukarıdaki varsayım altında bağımlı değişkenini koşullu varyansı da sabittir:

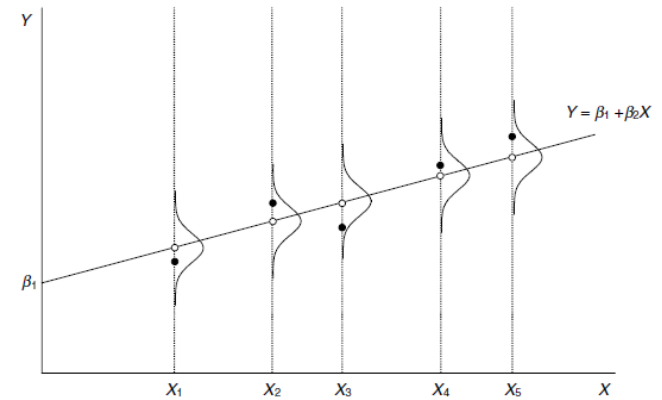
$$\text{Var}(y|x_1, x_2, \dots, x_k) = \sigma^2$$

- ▶ Bu derste MLR.5 Sabit Varyans varsayımının sağlanmadığı durumda OLS tahmincilerinin özelliklerini, değişen varyansın nasıl test edileceğini ve düzeltileceğini öğreneceğiz.

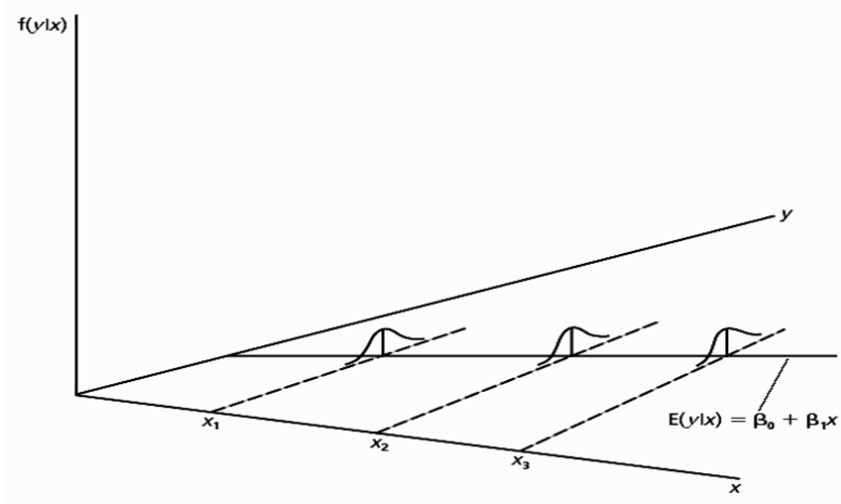
Değişen Varyans

- ▶ Hata teriminin koşullu varyansının değişkenlik göstermesi özellikle yatay-kesit veri analizinde sık rastlanan bir problemdir.
- ▶ Bunun bir sebebi y 'nin koşullu dağılımının popülasyonun farklı kesimlerinde değişkenlik göstermesidir.
- ▶ Örneğin hanehalkı tasarruflarının gelir düzeyi ile açıklandığı bir modelde tasarrufları gelir düzeyine bağlı olarak değişkenlik gösterebilir. Düşük gelir düzeylerinde tasarrufların varyansı daha düşük, yüksek gelir düzeylerinde ise daha yüksek olabilir.
- ▶ Benzer şekilde hanehalkı tüketiminin varyansı düşük ve yüksek gelir düzeylerinde farklı olabilir.
- ▶ Eğitim düzeyine koşullu olarak ücretlerin varyansı değişkenlik gösterebilir.

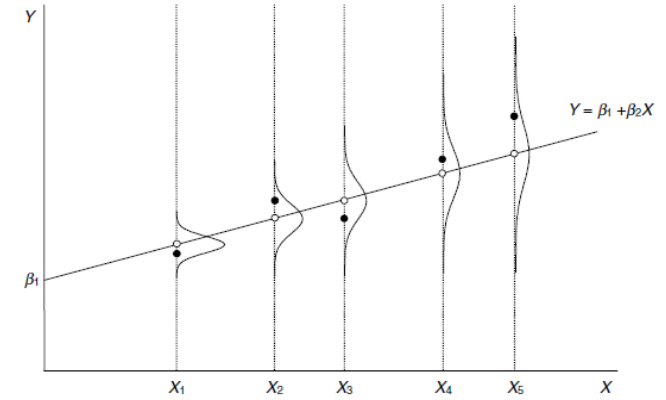
Sabit Varyans (Homoscedasticity) Varsayımı Altında Basit Regresyon Modeli



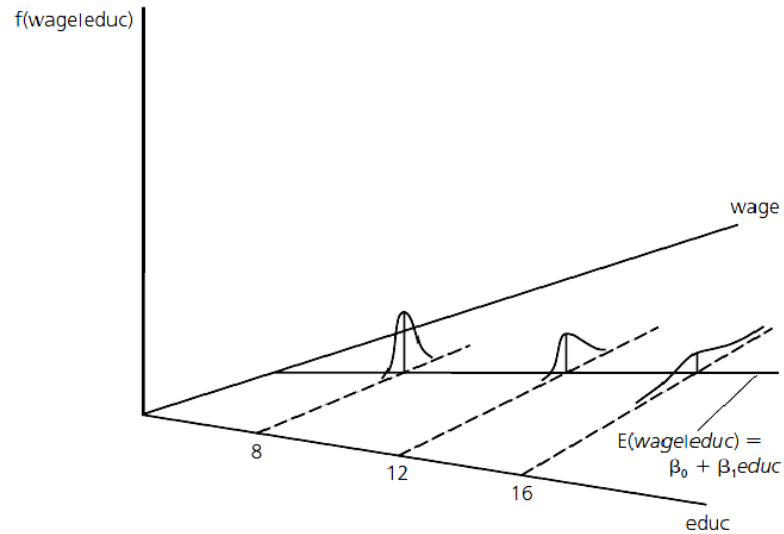
Sabit Varyans (Homoscedasticity) Varsayımı Altında Basit Regresyon Modeli



Değişen Varyans - Heteroskedasticity



Değişen Varyans - Heteroskedasticity



8

Değişen Varyans

- ▶ Basit regresyon modeli çerçevesinde değişen varyans durumunda OLS tahmincilerinin varyans ve standart hatalarını türetelim:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- ▶ Hata varyansının her gözlem için farklı olduğunu düşünelim:

$$\text{Var}(u_i|x_i) = \sigma_i^2, \quad i = 1, 2, \dots, n$$

- ▶ Eğim katsayısının OLS tahmincisinin aşağıdaki gibi yazılabileceğini biliyoruz:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- ▶ Burada $y = \beta_0 + \beta_1 x + u$ yerine yazılıp yeniden düzenlenirse:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x}) u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Değişen Varyans

- Basit regresyonda eğim katsayısının OLS tahmincisi:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

- MLR.1-4 varsayımları sağlandığından OLS sapmasız ve tutarlıdır: $E(\hat{\beta}_1) = \beta_1$. $\text{Var}(\hat{\beta}_1) = E[(\hat{\beta}_1 - \beta_1)^2]$ olduğundan $\hat{\beta}_1$ 'nin varyansı

$$\text{Var}(\hat{\beta}_1) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}$$

olur.

- Bunun daha önce türettiğimiz sabit varyans varsayımı altındaki varyans formülünden farklı olduğu açıktır. Her gözlem için $\sigma_i^2 = \sigma^2$ ise, yani sabit varyans varsayımı altında:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2 \sum_{i=1}^n (x_i - \bar{x})^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Değişen Varyans

- Sonuç olarak MLR.5 sabit varyans varsayımı altında türettiğimiz OLS tahminci varyansının değişen varyans durumunda geçerli olmadığını gördük.
- Değişen varyans olduğu halde sabit varyans durumunda geçerli formülleri kullanırsak standart hatalar yanlış olur.
- Bu durumda yapacağımız istatistiksel testler, t testleri, F testleri, LM testleri, geçersiz olur.
- Kesit veri regresyon analizinde değişen varyansın varlığı durumunda takip edilecek bir yol değişen varyansa dirençli standart hataların ve test istatistiklerinin hesaplanmasıdır.

Değişen Varyansa Dirençli Standart Hatalar

- White ya da White-Huber-Ecker standart hataları
- Çoklu regresyon modelinde White değişen varyansa dirençli standart hatalar aşağıdaki formülün karekökü olarak tanımlanır:

$$\widehat{\text{Var}}(\hat{\beta}_1) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{u}_i^2}{SST_j^2}$$

\hat{r}_{ij} : x_j 'nin tüme diğer x 'ler üzerine regresyonundan elde edilen inci kalıntı terimi. SST_j^2 : x_j 'deki örneklem değişkenliği

- Bu standart hatalar kullanılarak değişen varyanstan etkilenmeyen t istatistikleri hesaplanabilir.
- GRETL, Eviews, STATA gibi ekonometri paket programlarında bu standart hatalar istendiği takdirde kolayca hesaplanabilmektedir.
- Bu hesapların büyük örneklerde geçerli olduğu unutulmamalıdır.
- Örnek: ücret denklemi

Değişen Varyans Testleri

- Değişen varyansa dirençli standart hatalar (heteroscedasticity-robust standard errors) örneklem yeterince büyük ise geçerlidir.
- Küçük örneklerde ise değişen varyansın yapısına ilişkin belirli varsayımlar altında OLS dışında daha etkin bir tahmin yönteminin kullanılması gerekir. OLS artık BLUE değil.
- Bunun için öncelikle modelimizde değişen varyans olup olmadığını test etmemiz gerekir.
- Çok sayıda değişen varyans testi geliştirilmiştir.
- Bu derste yaygın kullanılan iki testi inceleyeceğiz: Breusch-Pagan testi ve White testi

Değişen Varyans Testleri

- ▶ Değişen varyans testleri MLR.1-4 varsayımlarının geçerli olduğunu dolayısıyla OLS tahmincilerinin sapmasız ve tutarlı olduğunu varsayar.
- ▶ Test etmek istediğimiz hipotez MLR.5 Sabit Varyans varsayımının geçerli olup olmadığıdır:

$$H_0 : \text{Var}(u|x_1, x_2, \dots, x_k) = \sigma^2 \quad \text{sabit varyans}$$

- ▶ Alternatif hipotez ise varyansın değişken olduğunu söyler:

$$H_1 : \text{Var}(u|x_1, x_2, \dots, x_k) \neq \sigma^2 \quad \text{değişen varyans}$$

- ▶ Boş hipotez aşağıdaki gibi de yazılabilir:

$$H_0 : E(u^2|x_1, x_2, \dots, x_k) = E(u^2) = \sigma^2$$

- ▶ Açıklayıcı değişkenler hata teriminin karesiyle ilişkisiz olmalıdır. Değişen varyans testleri bu ilişkinin varlığını araştırır.

Değişen Varyans Testleri

- ▶ Eğer boş hipotez yanlışsa, u^2 'nin koşullu beklenen değeri x 'lerin herhangi bir fonksiyonu olabilir.
- ▶ u^2 'nin x 'ler ile doğrusal ilişkili olduğu varsayılırsa

$$u^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + \nu$$

- ▶ Bu model çerçevesinde boş hipotez aşağıdaki gibi yazılabilir:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$$

- ▶ Boş hipotez altında $E(u^2|x_1, \dots, x_k) = \alpha_0$, sabit bir sayı, olmaktadır.
- ▶ u gözlenemediğinden modelden elde edilen kalıntılar \hat{u} kullanılarak F ya da LM testi yapılabilir.

Değişen Varyans Testleri

- ▶ Modelimizi tahmin ettikten sonra kalıntıların karesinin tüm x 'ler üzerine regresyonunu kurarak:

$$\hat{u}^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + \text{hata}$$

$F(k, n - k - 1)$ ya da $LM = nR_u^2 \sim \chi_k^2$ testlerini bu model çerçevesinde hesaplayabiliriz.

- ▶ Yukarıdaki modelin bir bütün olarak anlamlı olup olmadığını test etmekte kullandığımız F istatistiği yeterince büyükse varyansın sabit olduğunu söyleyen boş hipotez reddedilir.
- ▶ Bu testin LM versiyonuna **Breusch-Pagan değişen varyans testi** adı verilir.

Değişen Varyans Testleri

ADIMLAR

1. Modeli tahmin et, kalıntıların karesini \hat{u}^2 kaydet.
2. Kalıntıların karesinin aşağıda verilen tüm x 'ler üzerine regresyonunu kur.

$$\hat{u}^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + \text{hata}$$

ve bu modelden elde edilen determinasyon katsayısını R_u^2 kaydet

3. R_u^2 'yi kullanarak $F \sim F(k, n - k - 1)$ ya da $LM \sim \chi_k^2$ istatistiğini hesaplayarak testi sonuçlandır. (Kritik değerden büyükse H_0 : sabit varyans hipotezi reddedilir, değişen varyansa dair kanıt var denir. Ya da p -değeri hesaplanabilir. 0.05'den küçük bir p -değeri değişen varyansa işaret eder.)

Değişen Varyans Testleri Örnek: Ev fiyatları, hprice1.gdt

1. adımda modelimizi tahmin ediyoruz:

$$\widehat{\text{price}} = -\underset{(29.475)}{21.77} + \underset{(0.0006)}{0.0021} \text{lotsize} + \underset{(0.013)}{0.123} \text{sqrft} + \underset{(9.010)}{13.853} \text{bdrms}$$

$$n = 88 \quad R^2 = 0.672$$

2. adımda bu modelden elde edilen kalıntıların karesinin tüm x 'ler üzerine regresyonunu kuruyoruz:

$$\hat{u}^2 = -\underset{(3259.5)}{5522.79} + \underset{(0.071)}{0.202} \text{lotsize} + \underset{(1.464)}{1.691} \text{sqrft} + \underset{(996.38)}{1041.76} \text{bdrms}$$

$$n = 88 \quad R^2 = 0.1601$$

3. adımda Buradan hareketle $F = 5.34$ p -değeri=0.002, LM istatistiği $LM = 88 \times 0.1601 = 14.09$, p -değeri=0.0028 olarak bulunuyor. Açıkta ki sabit varyans aleyhine güçlü kanıt mevcut. Ev fiyatları modelinde değişen varyans vardır.

Değişen Varyans Testleri Örnek: Ev fiyatları, hprice1.gdt

Ev fiyatları modelinde değişen varyans olduğunu gördük. Logaritma dönüştürmesinin verilerdeki değişkenliği azaltabileceğinden daha önce bahsetmiştik. Örneğimizde oda sayısı haricindeki tüm değişkenlere doğal log dönüştürmesi yaparak yeniden tahmin edelim ve değişen varyans testlerini uygulayalım. Model sonuçları şu şekildedir:

$$\widehat{\ln \text{price}} = -\underset{(0.651)}{1.30} + \underset{(0.038)}{0.17} \ln \text{lotsize} + \underset{(0.093)}{0.70} \ln \text{sqrft} + \underset{(0.028)}{0.04} \ln \text{bdrms}$$

$$n = 88 \quad R^2 = 0.643$$

Bu modelde hesaplanan değişen varyans F ve LM istatistikleri ve p -değerleri şu şekildedir:

$$F = 1.141, \quad p - \text{value} = 0.245, \quad LM = 4.22, \quad p\text{value} = 0.239$$

p -değerleri boş hipotezi reddedecek kadar küçük değil. Dolayısıyla sabit varyans boş hipotezi reddedilemez. Log dönüştürmesini yaptıktan sonra değişen varyans problemi kalmadı.

White Değişen Varyans Testi

- Bölüm 5'de Gauss-Markov varsayımlarının tümünün sağlanması halinde OLS standart hatalarının ve test istatistiklerinin asimptotik olarak geçerli olacaklarını gördük.
- 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.

White Değişen Varyans Testi

- Açıklayıcı değişken sayısı, k , arttıkça yardımcı regresyondaki serbestlik derecesi azalmaktadır.
- Örneğin $k = 3$ için

$$\hat{u}^2 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3$$

$$+ \alpha_4 x_1^2 + \alpha_5 x_2^2 + \alpha_6 x_3^2$$

$$+ \alpha_7 x_1 x_2 + \alpha_8 x_1 x_3 + \alpha_9 x_2 x_3 + \nu$$

- Breusch-Pagan testiyle kıyaslarsak, bu denklemdeki bağımsız değişken sayısının 6 değişken daha fazla olduğunu görürüz.
- White testi LM istatistiğini kullanarak yukarıdaki yardımcı regresyonda sabit haricindeki tüm katsayıların sıfır olup olmadığını test eder:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_9 = 0$$

White Değişen Varyans Testi

- ▶ Bu hipotez F testi ile de sınanabilir. Her ikisi de asimptotik geçerliliğe sahiptir.
- ▶ $k = 6$ olduğunda White testindeki kısıt sayısı 27 olmaktadır. Açık ki bu serbestlik derecesi kaybına yol açar. White testinin yukarıdaki versiyonunun zayıf tarafı budur.
- ▶ Uygulamada genellikle daha az açıklayıcı değişken ile testin yapılmasını sağlayan aşağıdaki versiyonu kullanılır.
- ▶ Yardımcı regresyonda kareler ve çapraz çarpımları açıkça kullanmak yerine, modelin tahmininden elde edilen \hat{y} değerleri ve bunun karesi kullanılabilir:

$$\hat{u}^2 = \alpha_0 + \alpha_1 \hat{y} + \alpha_2 \hat{y}^2 + \nu$$

- ▶ Sabit varyans boş hipotezi aşağıdaki gibi yazılabilir:

$$H_0 : \alpha_1 = \alpha_2 = 0$$

White Değişen Varyans Testi

- ▶ Sabit varyans boş hipotezi aşağıdaki gibi yazılabilir:

$$H_0 : \alpha_1 = \alpha_2 = 0$$

- ▶ Bu hipotez F ya da LM testi ile sınanabilir.
- ▶ Bu testte orijinal modeldeki x sayısı ne olursa olsun, sadece 2 kısıt vardır. Böylece, testin orijinal halindeki serbestlik derecesi (dof) kaybı burada söz konusu değildir.
- ▶ Varyansın y 'nin koşullu beklenen değeriyle birlikte değiştiği durumlarda oldukça yararlı bir testtir.

White Değişen Varyans Testi: Örnek

- ▶ 1. adım: modelin OLS ile tahmini ve kalıntı karelerinin ve fit edilen değerlerin hesaplanması

$$\widehat{lprice} = -1.30_{(0.651)} + 0.17_{(0.038)} llotsize + 0.70_{(0.093)} lsqrft + 0.04_{(0.028)} bdrms$$

$$n = 88 \quad R^2 = 0.643$$

- ▶ 2. adım: \hat{u}^2 'nin \hat{y} ve \hat{y}^2 üzerine regresyonu

$$\widehat{u^2} = 5.047_{(3.345)} - 1.709_{(1.163)} \widehat{lprice} + 0.145_{(0.100)} \widehat{lprice}^2$$

$$n = 88 \quad R^2 = 0.03917$$

- ▶ 3. adım: Test istatistiğinin hesaplanması:
 $LM = nR_u^2 = 88 \times 0.03917 = 3.447$, p -değeri=0.18. Karar: Sabit varyans hipotezi reddedilemez.

Değişen Varyans Testleri

- ▶ Bu değişen varyans testlerini yaparken MLR.1-MLR.4 varsayımlarımızın sağlandığını varsayıyoruz.
- ▶ Sağlanmazsa, örneğin, regresyonun fonksiyonel biçimi yanlış belirlenmiş ise (ihmal edilmiş değişken varsa ya da log-log yerine level model seçilmişse vs.), heteroscedasticity testi varyans sabitken bile boş hipotezi reddedebilir.
- ▶ Başka bir ifadeyle, 1. Tip Hata olasılığı nominal düzeyden daha fazla olabilir.
- ▶ Bu yüzden, ekonometristler değişen varyans testlerini “yanlış biçim seçimi” (misspecification) testleri olarak değerlendirirler.
- ▶ Ancak, fonksiyon biçimi (form) seçimi doğrudan başka testler kullanılarak test edilmeli. Yanlış biçim seçimi değişen varyanstan daha ciddi bir sorundur.

Ağırlıklandırılmış En Küçük Kareler (Weighted Least Squares - WLS)

- ▶ Yaptığımız testler sonucunda modelde değişen varyans olduğu sonucuna ulaştığımızı düşünelim.
- ▶ Bu durumda takip edebileceğimiz opsiyonlardan biri, daha önce bahsettiğimiz, değişen varyansa dirençli standart hatalar ve test istatistiklerinin kullanılmasıdır (heteroscedasticity-robust standard errors)
- ▶ Ancak bu standart hatalar büyük örneklerde geçerlidir. Ayrıca modeli değişen varyanstan arındırarak yeniden tahmin etmek isteyebiliriz.
- ▶ Bunun için OLS yöntemini değil WLS yöntemini tercih ediyoruz.
- ▶ Değişen varyans durumunda WLS, OLS'den daha etkin tahmin ediciler sunar.

Ağırlıklandırılmış En Küçük Kareler (Weighted Least Squares - WLS)

- ▶ WLS yöntemi değişen varyansın formuna ilişkin bilgiyi gerektirir.
- ▶ Çoklu regresyon modelinde

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i$$

MLR.5 varsayımının sağlanmadığını, ve değişen varyansın formunun aşağıdaki gibi olduğunu düşünelim:

$$\text{Var}(u_i | x_{i1}, x_{i2}, \dots, x_{ik}) = \sigma^2 h(x)$$

- ▶ Burada $h(x) > 0$ x 'lerin herhangi bir fonksiyonudur. $h(x)$ 'in bilindiğini düşünelim.
- ▶ Orijinal modeli $h(x)$ fonksiyonunu kullanarak hata terimi sabit varyanslı olacak şekilde dönüştürebiliriz. Bu dönüştürülmüş modeli OLS ile tahmin edebiliriz. Buna WLS tahmini denir.

Ağırlıklandırılmış En Küçük Kareler (WLS)

- ▶ Modelimizi $1/\sqrt{h(x)}$ ile çarparak dönüştürüyoruz:

$$\frac{y_i}{\sqrt{h(x)}} = \beta_0 \frac{1}{\sqrt{h(x)}} + \beta_1 \frac{x_{i1}}{\sqrt{h(x)}} + \beta_2 \frac{x_{i2}}{\sqrt{h(x)}} + \dots + \beta_k \frac{x_{ik}}{\sqrt{h(x)}} + \frac{u_i}{\sqrt{h(x)}}$$

- ▶ Dönüştürülmüş modeli aşağıdaki gibi yazabiliriz:

$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + \dots + \beta_k x_{ik}^* + u_i^*, \quad i = 1, 2, \dots, n$$

$$x_{i0}^* = \frac{1}{\sqrt{h(x)}}$$

- ▶ Bu modelde hata varyansını hesaplırsak:

$$\begin{aligned} \text{Var}(u_i^* | x) &= E(u_i^{*2} | x) = E\left(\left(\frac{u_i}{\sqrt{h(x)}}\right)^2 | x\right) = \frac{1}{h(x)} E(u_i^2 | x) \\ &= \frac{1}{h(x)} \sigma^2 h(x) = \sigma^2 \end{aligned}$$

WLS

- ▶ Örneğin, y : tasarruflar, x : gelir düzeyi, olsun:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

$$\text{Var}(u_i | x_i) = \sigma^2 x_i$$

- ▶ Burada, $h(x) = x_i$ 'dir. Tasarrufların varyansı gelir düzeyi ile birlikte değişmektedir.
- ▶ Gelir pozitif olduğundan varyans her zaman pozitif olacaktır.
- ▶ Modeli $\sqrt{x_i}$ 'ye bölerek WLS dönüştürmesi yapabiliriz.
- ▶ Dönüştürülmüş model sabit varyanslı olacağından OLS yöntemi ile etkin tahmin mümkün olur.

Genelleştirilmiş En Küçük Kareler (Generalized Least Squares - GLS)

- Dönüştürülmüş modele OLS uygulanırsa, GLS tahmincileri elde edilir.
- GLS tahmincileri OLS tahmincilerinden farklı olacaktır. Ancak yorumlar orijinal model çerçevesinde yapılır.
- Değişen varyansı düzelteren GLS dönüştürmesi sonucunda elde edilen modele OLS uygulanırsa WLS tahmincileri elde edilir.
- Zaman serileriyle kurulan modellerde otokorelasyon durumunda da GLS dönüştürmesi kullanılabilir.
- GLS tahmincileri, β_j^* , BLUE özelliklerini sağlarlar.
- Dönüştürülmüş modelin R^2 'si uyum iyiliği ölçütü olarak kullanılmaz. Ancak test istatistiklerinin hesaplanmasında kullanılabilir.

GLS, WLS, OLS

- Dönüştürülmüş modele OLS uygulanırsa amaç fonksiyonu aşağıdaki gibi yazılabilir:

$$\sum_{i=1}^n \hat{u}_i^2 = \sum_{i=1}^n \frac{1}{h(x)} (\hat{y}_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_k x_{ik})^2$$

- OLS, WLS'in özel bir halidir. Her gözleme eşit ağırlık verir.
- WLS yönteminde ise her bir gözlem varyansın tersi ile ağırlıklandırılır. Böylece regresyon düzleminden uzak gözlemler cezalandırılır.
- Büyük varyanslı gözlemlere daha düşük, düşük varyanslı gözlemlere ise daha büyük ağırlık verilir.

WLS Örnek: Tasarruf-Gelir ilişkisi

Independent Variables	(1) OLS	(2) WLS	(3) OLS	(4) WLS
<i>inc</i>	.147 (.058)	.172 (.057)	.109 (.071)	.101 (.077)
<i>size</i>	—	—	67.66 (222.96)	-6.87 (168.43)
<i>educ</i>	—	—	151.82 (117.25)	139.48 (100.54)
<i>age</i>	—	—	.286 (50.031)	21.75 (41.31)
<i>black</i>	—	—	518.39 (1,308.06)	137.28 (844.59)
<i>intercept</i>	124.84 (655.39)	-124.95 (480.86)	-1,605.42 (2,830.71)	-1,854.81 (2,351.80)
Observations	100	100	100	100
R-Squared	.0621	.0853	.0828	.1042

Tahmin Edilebilir Genelleştirilmiş En Küçük Kareler - FGLS

- FGLS: Feasible Generalized Least Squares
- GLS-WLS yönteminin uygulanabilmesi için $h(x)$ fonksiyonunun yapısının ve hangi x 'lere bağlı olduğunun bilinmesi gerekir.
- Uygulamada çoğu zaman bu bilinmez.
- Bu durumda değişen varyansın yapısı verilerden hareketle tahmin edilebilir.
- Oldukça genel bir çerçeve sunan varyans formu aşağıdaki gibidir:

$$\text{Var}(u|x) = \sigma^2 \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k)$$

- Burada

$$h(x) = \exp(\delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_k x_k)$$

FGLS: Adımlar

1. Modeli OLS ile tahmin et. Kalıntıları sakla, \hat{u}
2. Kalıntıların karesinin doğal logaritmasını hesapla $\Rightarrow \log(\hat{u}^2)$
3. $\log(\hat{u}^2)$ 'nin x_1, x_2, \dots, x_k üzerine regresyonunu kur. Fit edilen değerleri sakla: \hat{g}
4. \hat{g} 'nin üstelini hesapla $\Rightarrow \hat{h} = \exp(\hat{g})$
5. $1/\hat{h}$ ağırlıklarını kullanarak modeli WLS ile tahmin et.

FGLS Örnek: Smoke.gdt

$$\widehat{\text{cigs}} = - \underset{(24.079)}{3.64} + \underset{(0.728)}{0.88} \text{ lincome} - \underset{(5.773)}{0.751} \text{ lcigpric} - \underset{(0.167)}{0.50} \text{ educ} \\ + \underset{(0.160)}{0.771} \text{ age} - \underset{(0.001)}{0.009} \text{ agesq} - \underset{(1.111)}{2.826} \text{ restaurn} \\ n = 807 \quad R^2 = 0.053 \quad F(6, 800) = 7.4231 \quad \hat{\sigma} = 13.405 \\ \text{(standard errors in parentheses)}$$

White test istatistiği: LM = 36.15, p-değeri = 0.000079
Değişen varyansa dair güçlü kanıt.

FGLS Örnek

$\log(\hat{u}^2)$ 'nin x 'ler üzerine regresyonu:

$$\widehat{\log(\hat{u}^2)} = - \underset{(2.563)}{1.92} + \underset{(0.077)}{0.29} \text{ lincome} + \underset{(0.614)}{0.195} \text{ lcigpric} - \underset{(0.017)}{0.079} \text{ educ} \\ + \underset{(0.017)}{0.20} \text{ age} - \underset{(0.0001)}{0.002} \text{ agesq} - \underset{(0.118)}{0.627} \text{ restaurn} \\ n = 807 \quad \bar{R}^2 = 0.2417$$

Bu modelden edilen fit edilen değerler: \hat{g}
Ağırlıklar: $1/\exp(\hat{g})$

FGLS Örnek

$1/\exp(\hat{g})$ ağırlıkları kullanılarak WLS tahmini yapılırsa:

$$\widehat{\text{cigs}} = \underset{(17.803)}{5.64} + \underset{(0.437)}{1.295} \text{ lincome} - \underset{(4.46)}{2.94} \text{ lcigpric} - \underset{(0.120)}{0.46} \text{ educ} \\ + \underset{(0.097)}{0.482} \text{ age} - \underset{(0.0009)}{0.006} \text{ agesq} - \underset{(0.795)}{3.46} \text{ restaurn} \\ n = 807 \quad R^2 = 0.113$$

GRETl: Model \Rightarrow Other Linear Models \Rightarrow Weighted Least Squares
weights: $1/\exp(\hat{g})$

Doğrusal Olasılık Modeli ve FGLS

- ▶ Doğrusal olasılık modelinin (LPM) değişen varyanslı olduğunu görmüştük.
- ▶ OLS tahmincileri bu durumda BLUE olmaz. LPM modeli için uygun ağırlıklar nasıl bulunabilir?
- ▶ Doğrusal olasılık modelinde hata varyanslarının aşağıdaki gibi yazılabildiğini görmüştük:

$$\text{Var}(y|x) = p(x)(1 - p(x))$$

- ▶ Burada $p(x)$ başarı olasılıklarını gösteriyordu ve x 'lerin lineer bir kombinasyonu olarak yazılmıştı.
- ▶ Açıktır ki varyans fit edilen değere göre değişecektir. Bu nedenle $h(x)$ fonksiyonu kolayca hesaplanabilir:

$$\hat{h}_i = \hat{y}_i(1 - \hat{y}_i)$$

- ▶ 0 – 1 aralığı dışındak kalan \hat{y}_i değerlerini 0.01 ve 0.99 olarak alıp FGLS yöntemini uygulayabiliriz.