8. DEĞİŞEN VARYANS SORUNU (devam)

8.3. Değişen Varyans Sorununun Varlığı Saptanabilir mi? (devam)

8.3.2. Biçimsel Yöntemler

1. Goldfeld-Quandt Test'i

Hata terimi varyansındaki değişmeler açıklayıcı değişkenlerden birisi ile ilişkilendirilebiliyorsa bu test uygulanabilir.

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i, \qquad i = 1 \ldots n$

modelinde Var(u) diyelim ki X_2 değişkeni (veya örneğin karesi) ile ilişkilidir. Test birkaç aşamada uygulanır:

1- Önce bu X_2 değişkeni küçükten büyüğe doğru sıralanmalıdır. Sonra, bağımlı değişken ve diğer açıklayıcı değişken verileri de X_2 değişkeninin sıralanmış verilerine karşılık gelecek şekilde yeniden düzenlenmelidir.

2- Yeni sıralanmış verilerin ortasındaki c adet veri atılmalıdır ve $c \approx n/6$ olmalıdır. Kalan n-c gözlem iki eşit sayıda (=(n-c)/2) olmalıdır. Ör. $n = 40$ ise $n/6 \approx 6$ dersek 40-6=34 ikiye bölünebilir: $n_1 = n_2 = 17$.

3- Yukarıda verilen denklem n_1 ve n_2 veri ile iki kez tahmin edilmelidir.

n_1 adet veri ile yapılan tahminde $KKT_1 = \Sigma_1 \hat{u}_i^2$

n_2 adet veri ile yapılan tahminde $KKT_2 = \Sigma_2 \hat{u}_i^2$

bulunmalıdır. Burada KKT_1 düşük varyanslı grup, KKT_2 yüksek varyanslı gruptur.

4- Burada test edilecek hipotez aşağıdaki gibidir:

$H_0: \sigma_1^2 = \sigma_2^2$

$H_1: \sigma_1^2 \neq \sigma_2^2$

Kullanılacak istatistik $F_h = \dfrac{KKT_2/(n_2 - k)}{KKT_1/(n_1 - k)}$

Eğer u_i normal dağılmışsa ve sabit varyans varsayımı geçerli ise $F_{tab} = F(n_2\text{-}k, n_1\text{-}k)$ tablo değeri ile karşılaştırılmalıdır. $F_h > F_{tab}$ ise H_0 reddedilir, varyansın sabit olmadığına karar verilir.

Bu testin gücü c değerine bağlıdır: c büyük olursa yardımcı denklemlerin serbestlik derecesi azalır; küçük olursa gözlemler arasındaki farklılığı belirlemek zordur. Goldfeld-Quandt testinin olumsuz yönü, hata terimi varyansının bir açıklayıcı değişkenle ilişkilendirilmesidir. Özellikle çok açıklayıcı değişken olması durumunda bu ilişkilendirme kolay olmayabilir.

2. White Test'i

White testi bir LM testidir ve diğer LM testlerinde olduğu gibi asıl denkleme ek olarak bir yardımcı denklem tahmini gerektirir. Testin arkasındaki temel düşünce şudur: Eğer sabit varyans varsa $E(u_i^2) = \sigma$ dır ve X'ler veya X'lerin fonksiyonu olan değişkenler u_i^2 yi açıklamaz. Bu nedenle sol taraf değişkeninin u_i^2, sağ taraf değişkenlerinin X'lerin bir fonksiyonu olduğu bir yardımcı denklem tahmin edilir. Değişen varyans sorunundan şüpheleniliyor ama formu hakkında bir fikrimiz yoksa White testi uygun bir testtir.

Asıl denklem aşağıdaki gibi olsun

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} + u_i, \qquad i = 1 \ldots n$$

1- Asıl denklem tahmin edilerek hata tahmin kareleri bulunur: $\hat{u}_i^2$

2- Aşağıdaki yardımcı denklem tahmin edilir:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{i2} + \ldots + \alpha_k X_{ik} + \alpha_{k+1} X^2_{i2} + \ldots + \alpha_{2k} X^2_{ik} + \alpha_{2k+1} X_{i2} X_{i3} + \ldots + v_i$$

Yani asıl denklemin hata tahmin karelerinin bağımlı, açıklayıcı değişkenlerin kendileri, kareleri ve çarpımlarının açıklayıcı değişken olduğu denklem tahmin edilir. Açıklayıcı değişkenlerin daha yüksek dereceleri de kullanılabilir.

Asıl denklemde sabit terim olsa da olmasa da yardımcı denklemde vardır.

Bu yardımcı denklem için R^2 hesaplanır. Buna R_Y^2 diyelim.

3- Boş hipotez değişen varyans olmadığı şeklindedir:
H_0: $\alpha_2 = \alpha_3 = \ldots = 0$
H_1: $\alpha_2, \alpha_3, \ \ldots \neq 0$

R_Y^2 nin n ile çarpımı asimptotik olarak ki-kare dağılımına sahiptir ve serbestlik derecesi yardımcı denklemde yer alan sabit dışındaki açıklayıcı değişken sayısıdır.

$nR_Y^2 \sim \chi^2(k+f-1)$

(burada f asıl denklemde bulunmayıp yardımcı denklemde bulunan değişken sayısıdır)

4- Eğer hesaplanan χ^2 değeri tablo değerinden büyükse H_0 reddedilir. Yani değişen varyans sorunu var demektir. Eğer büyük değilse değişen varyans sorunu yoktur.

Test edilen hipotez bakımından düşünülürse White testi Goldfeld-Quandt testinden daha geneldir.

Olumsuz tarafı: çok sayıda açıklayıcı değişken olduğunda yardımcı denklemde serbestlik derecesi hızla düşer. Test istatistiğinin anlamlı bulunduğu durumlarda bunun nedeni değişen varyans olmak zorunda değildir, tanımlama hataları da olabilir veya ikisi birden olabilir. Hangisinin olduğunu bilmek ise zordur.

3. ARCH-LM Test'i

Engle ARCH sorununun varlığını test etmek için bir LM testi önermiştir. Diğer LM testlerinde olduğu gibi ARCH için yapılan LM testinde de asıl denkleme ek olarak bir yardımcı denklem tahmin edilir.

Asıl denklem ve yardımcı denklem sırasıyla şöyledir:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + u_t, \qquad t = 1 \ldots n$$

$$\hat{u}_t^2 = c_0 + c_1 \hat{u}_{t-1}^2 + c_2 \hat{u}_{t-2}^2 + \ldots + c_p \hat{u}_{t-p}^2 + e_t$$

Yardımcı denklemdeki gecikme sayısı p araştırmacıya kalmıştır ancak üç aylık veriler kullanılıyorsa gecikme sayısı 4 e çıkabilir.

White testinde olduğu gibi test istatistiği olarak χ^2 dağılımlı değişkenler kullanılabilir:

$nR_Y^2 \sim \chi^2(p)$

H_0: $c_1, c_2, \ldots, c_p = 0$

H_1: $c_1, c_2, \ldots, c_p \neq 0$

Hesaplanan değer tablo değerinden büyükse H_0 reddedilir ve modelde ARCH vardır sonucuna ulaşılır.

8.4. Değişen Varyans Sorununun Çözümü Var mıdır?

Değişen varyans sorununu nasıl çözeceğimiz, değişen varyansın formunu belirleyip belirleyemediğimize bağlıdır. Değişen varyans sorununun nasıl çözüleceğine geçmeden önce daha önce belirttiğimiz bir şeyi tekrarlayalım: değişen varyans model spesifikasyonunun yanlış olmasından kaynaklanabilir. Verinin logaritmasının alınması değişen varyans sorununun azalmasını veya ortadan kalkmasını sağlayabilir. Bu nedenle ele alacağımız metodları uygulamadan önce spesifikasyonun doğru olduğundan emin olmak gerekir.

Değişen varyansın formu tam olarak biliniyorsa GEKK yöntemi kullanılır. Değişen varyansın formunu bilmiyorsak veya tahmin edemiyorsak White standart hatalar kullanılır.

8.4.1. σ_{ui}^2 değerleri biliniyorsa: Genelleştirilmiş EKK (GEKK) Yöntemi (Generalized Least Squares - GLS)

Değişen varyans sorununda hata terimi varyans-kovaryans matrisinin

Var, Cov(**u**) = E(**uu'**) = $\sigma^2 \mathbf{I}_n$ yerine

$$\text{Var, Cov}(\mathbf{u}) = E(\mathbf{uu'}) = \mathbf{\Omega} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}$$

geçmektedir.

GEKK yöntemi asıl denklemden bir dönüştürülmüş denklem elde edip bu dönüştürülmüş denklemi EKK ile tahmin etmek anlamına gelmektedir. Bu durumda dönüştürülmüş denklem aşağıdaki gibidir.

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_{2i}}{\sigma_i} + \dots + \beta_k \frac{X_{ki}}{\sigma_i} + e_i$$

Dönüştürülmüş denklemin hata terimi $e_i = u_i/\sigma_i$ dir ve e'nin varyansı :
$\text{Var}(e_i) = E(e_i)^2 = E(u_i/\sigma_i)^2 = (1/\sigma_i^2)E(u_i)^2 = (1/\sigma_i^2)\sigma_i^2 = 1$ dir ve sabittir.

Dönüştürülmüş denklemin EKK ile tahmin edilmesi, asıl denklemin GEKK ile tahmin edilmesi anlamına gelmektedir. Buradaki dönüştürülmüş denklemin EKK uygulanması Ağırlıklı EKK (weighted least squares) olarak adlandırılır. Çünkü her bir gözlem hata terimi varyansının tersi ile ağırlıklandırılmıştır. Ağırlıklı EKK, daha genel bir yöntem olan GEKK'in özel bir durumudur. Daha sonraki bölümlerde GEKK'in farklı özel durumlarını da ele alacağız.

Ağırlıklı EKK'da ağırlıkların kullanılması şu anlama gelir: yüksek varyansa sahip gözlemler tahminde daha düşük ağırlığa sahiptir. Daha genel olarak şu söylenebilir: en yüksek kalitedeki gözlemlere en yüksek ağırlık verilir, en düşük kalitedeki gözlemlere en düşük ağırlık verilir.

Bazı durumlarda

$$\text{Var, Cov}(\mathbf{u}) = E(\mathbf{uu}') = \sigma^2\boldsymbol{\Omega} = \sigma^2\begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \omega_n \end{bmatrix} = \begin{bmatrix} \sigma^2\omega_1 & 0 & \dots & 0 \\ 0 & \sigma^2\omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2\omega_n \end{bmatrix}$$

Olduğu varsayılır. Yani her bir hata terimi için $E(u_i^2) = \sigma^2\omega_i$ dir. Yani her bir varyansın σ^2 gibi sabit bir kısmı vardır fakat ω_i nedeniyle her bir varyans birbirinden farklıdır:

$\text{Var}(u_1) = \sigma^2\omega_1$, $\text{Var}(u_2) = \sigma^2\omega_2$, … $\text{Var}(u_n) = \sigma^2\omega_n$

Bu durumda dönüştürülmüş denklemi şöyle ifade edebiliriz:

$$\frac{Y_i}{\sigma\sqrt{\omega_i}} = \beta_1\frac{1}{\sigma\sqrt{\omega_i}} + \beta_2\frac{X_{2i}}{\sigma\sqrt{\omega_i}} + \cdots + \beta_k\frac{X_{ki}}{\sigma\sqrt{\omega_i}} + \frac{u_i}{\sigma\sqrt{\omega_i}}$$

veya

$$\frac{Y_i}{\sqrt{\omega_i}} = \beta_1\frac{1}{\sqrt{\omega_i}} + \beta_2\frac{X_{2i}}{\sqrt{\omega_i}} + \cdots + \beta_k\frac{X_{ki}}{\sqrt{\omega_i}} + e_i$$

$\frac{Y_i}{\sqrt{\omega_i}} = \beta_1\frac{1}{\sqrt{\omega_i}} + \beta_2\frac{X_{2,i}}{\sqrt{\omega_i}} + \ldots + \beta_k\frac{X_{k,i}}{\sqrt{\omega_i}} + e_i$ dönüştürülmüş denklemin hata terimi $e_i = u_i/\sqrt{\omega_i}$

e'nin varyansı : $\text{Var}(e_i) = E(e_i)^2 = E(u_i/\sqrt{\omega_i})^2 = (1/\omega_i)E(u_i)^2 = \sigma^2\omega_i/\omega_i = \sigma^2$ dir ve sabittir.

8.4.2. σ_{ui}^2 değerleri bilinmiyorsa:

GEKK yönteminin bu şekilde uygulanabilmesi için Ω değerlerinin bilinmesi gerekir. Bu değerlerin bilinmemesi durumunda üç yöntem uygulanabilir.

i- Değişen varyans ile ilgili varsayım

Ω genellikle bilinmediği için hata terimi varyansının açıklayıcı değişkenlerden birisi ile ilişkili olarak değiştiği varsayımı yapılır: Örneğin

a- Hata terimleri varyansı X_2^2 ile orantılıdır: $Var(u_i) = \sigma^2 X_{2i}^2$ $(\omega_i = X_{2i}^2)$

Bu durumda **Ω** matrisi şu şekilde tanımlanmış olmaktadır:

$$\mathbf{\Omega}= \begin{bmatrix} X_{2,1}^2 & 0 & \dots & 0 \\ 0 & X_{2,2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{2,n}^2 \end{bmatrix}$$

Sonuç olarak dönüştürülmüş model şu şekildedir:

$$\frac{Y_i}{X_{2i}} = \beta_1 \frac{1}{X_{2i}} + \beta_2 \frac{X_{2i}}{X_{2i}} + \dots + \beta_k \frac{X_{ki}}{X_{2i}} + e_i \qquad e_i = u_i / X_{2i}$$

e'nin varyansı : $Var(e_i) = E(e_i)^2 = E(u_i / X_{2i})^2 = (1/X_{2i}^{\ 2})E(u_i)^2 = \sigma^2 X_{2i}^2 / X_{2i}^{\ 2} = \sigma^2$ dir. (sabit)

b- Hata terimleri varyansı X_2 ile orantılıdır: $Var(u_i) = \sigma^2 X_{2i}$ $(\omega_i = X_{2i})$

dönüştürülmüş model:

$$\frac{Y_i}{\sqrt{X_{2i}}} = \beta_1 \frac{1}{\sqrt{X_{2i}}} + \beta_2 \frac{X_{2i}}{\sqrt{X_{2i}}} + \dots + \beta_k \frac{X_{ki}}{\sqrt{X_{ki}}} + e_i \qquad e_i = u_i / \sqrt{X_{2i}}$$

$Var(e_i) = E(e_i)^2 = E(\frac{u_i}{\sqrt{X_{2i}}})^2 = \frac{1}{X_{2i}} \sigma^2 X_{2i} = \sigma^2$ dir. (sabit)

ii- White değişen varyansla tutarlı varyansları

(White’s heteroscedasticity-consistent variances and Standard errors)

White göstermiştir ki büyük örneklemlerde gerçek parametre değerleri ile ilgili istatistiki çıkarımlarda bulunulabilir. Bu yöntemde katsayı varyans kovaryans matrisinin hesaplanmasında, σ_i^2 yerine onun tahmini olarak $\hat{u}_i^2$ kullanılır.

(White cov matrisi: $(X'X)^{-1}[\Sigma_{i=1}^{n} u_i^2 x_i x_i'](X'X)^{-1}$)

Pek çok ekonometri paketi White değişen varyansla tutarlı varyansları ve standart hataları vermektedir. Bu değerler EKK değerlerinden büyük veya küçük olabilir.

Bu yöntemin uygulanabilmesi için büyük örneklem olması gerekir.