# NONLİNEERLİĞİN SAPTANMASI İLE İLGİLİ BAZI TESTLER

Atıf EVREN

**Abstact: There are various methods to detect nonlinearity in regression analysis. Scatter plots; residual plots; some forms of widely used test statistics such as F-tests; polynomial regression models, and piecewise linearization procedures may be useful although not all of the methods work properly since there are a lot of practical restrictions. In such cases a simultaneus application of some of the models may help.**

## 1.1.Uyum Eksikliği Testi

Regresyon modelinin verilere uygun kurulup kurulmadığını test etmek için kullanılan yöntemlerden birisi F testi yardımı ile gerçekleştirilmektedir. Bu test normal dağılım, bağımsızlık ve sabit varyans varsayımlarını şart koşar. Başka bir deyişle test sadece değişkenler arasındaki ilişkinin doğru formüle edilip edilmediği sorusuna cevap aramaktadır. F testinin bu soruya doyurucu yanıt verebilmesi için bağımsız değişkenin en azından bir gözlenen değeri için bağımlı değişkenin birden fazla gözlenen değerinin elde edilmiş olması (replication) gerekmektedir. Bu yöntem kalıntı kareleri toplamının iki bileşene ayrılmasını içermektedir:

$$SS_E = SS_{PE} + SS_{LOF}$$

Burada $SS_{PE}$ pür hata ile ilişkili kareler toplamı; $SS_{LOF}$ ise uyum eksikliği ile ilgili kareler toplamını ifade etmektedir. Bu ayrıştırmayı gösterebilmek için (ij) kalıntı

$$y_{ij} - \hat{y}_i = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i) \qquad (1.1.1)$$

biçiminde yazılsın.

Burada $\bar{y}_i$'nin $x = x_i$ noktasında elde edilen n tane bağımlı değişkenin gözlem değerinin ortalaması olduğunun belirtilmesi gerekiyor. Eşitliğin her iki tarafındaki terimlerin karelerinin alınıp bu ifadelerin de daha sonra toplanmaları halinde

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\hat{y}_i)^2 = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2 + \sum_{i=1}^{m} n_i(\bar{y}_i-\hat{y}_i)^2 \qquad (1.1.2)$$

eşitliği

(çapraz çarpımların toplamı sıfıra eşit olacağı için) elde edilmektedir.

Eşitliğin sol tarafı regresyon analizinin klasik kalıntı kareleri toplamıdır. Sağ taraftaki iki bileşen ise pür hata ve uyum eksikliği bileşenlerinden oluşmaktadır.

Pür kalıntı kareleri toplamının $SS_{PE}=\sum_{i=1}^{m}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$ (1.1.3) olduğunu not etmek gerekiyor. Bu ifade bağımsız değişkenin tekrarlı değerlerinin (replications) elde edildiği noktalarda , y'nin gözlem değerleri ile söz konusu noktalarda elde edilen ortalama değerleri arasındaki farkların karelerinin toplamını vermektedir. Sabit varyans varsayımının yerine gelmesi halinde bu toplam pür hatanın modelden bağımsız bir tahmincisi olarak kullanılabilir. Çünkü bu durumda y'deki değişkenliğin tek nedeni sözkonusu tekrarlı gözlemler olacaktır. Her bir x seviyesinde $n_i-1$ serbestlik derecesi olduğu için pür kalıntı kareleri ile ilgili toplam serbestlik derecesi de Serbestlik derecesi= $\sum_{i=1}^{m}n_i-1=n-m$ (1.1.4) olarak hesaplanmaktadır. Uyum eksikliği kalıntı kareleri toplamı ise

$SS_{LOF}=\sum_{i=1}^{m}n_i(\bar{y}_i-\hat{y}_i)^2$ (1.1.5) formülü ile ifade edilmektedir.

Bu toplam her bir x seviyesindeki y ortalama değerleri ile tahmini değerleri arasındaki farkların karelerinin tartılı bir toplamıdır. Eğer y'nin ortalama değerleri y'nin tahmin değerlerine yakın ise ,bu durumda kareler toplamı göreli olarak küçük olacak ve modelin doğrusal olduğuna hükmedilecektir. Başka bir deyiş ile y'nin tahmin değerleri doğrusal bir model ile gerçekleştirildiği için ortalama değerler ile tahmin değerlerinin birbirine yakın olması modelin doğrusal olduğuna bir delil olarak düşünülmektedir. Eğer y'nin tahmini değerleri ile ortalama değerleri arasındaki farklılık büyük (ve dolayısıyla kareler toplamı büyük ise) modelin doğrusal olmadığına hükmedilecektir. $SS_{LOF}$ ile ilgili serbestlik derecesi; x'in birbirinden farklı (aldığı) m değer bulunduğu için; m-2'dir. (Y'nin tahmin değerlerini elde etmek için iki değişkenli doğrusal regresyon modelinde iki parametre tahminine gidildiği için iki serbestlik derecesi kaybolmaktadır.)

Genellikle $SS_{LOF}$ 'un hesaplanması $SS_{PE}$'nin $SS_E$'den çıkartılması ile gerçekleştirilmektedir. Bu durumdaki uyum eksikliğini test etmek kullanılan istatistik;

$$F_0 = \frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)} = \frac{MS_{LOF}}{MS_{PE}} \quad (1.1.6)$$

Bu arada $MS_{PE}$'nin beklenen değeri $\sigma^2$'dir ve $MS_{LOF}$ istatistiğinin beklenen değeri ise

$$E(MS_{LOF}) = \sigma^2 + \frac{\sum_{i=1}^{m} n_i (E(y_i) - \beta_0 - \beta_1 x_i)^2}{m-2} \quad (1.1.7)$$

Ana kütle regresyon fonksiyonunun doğrusal olması halinde $E(y_i) = \beta_0 + \beta_1 x_i$

olacağından

$E(MS_{LOF}) = \sigma^2$ olması beklenir. Ana kütle regresyon fonksiyonunun doğrusal olmaması

halinde

ise $E(MS_{LOF}) > \sigma^2$ olacaktır. Dahası, ana kütle regresyon fonksiyonunun doğrusal

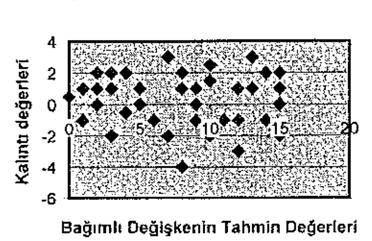
olması halinde $F_0$ m-2 ve n-m serbestlik derecelerinde bir $F$ dağılımına sahip olacaktır.

Dolayısıyla modelin doğrusal olup olmadığının test edilmesi için $F_0$ istatistiği hesaplanmakta ve $F_0 > F_{\alpha, m-2, n-m}$ olması durumunda regresyon fonksiyonunun doğrusal olmadığına hükmedilmektedir.(Montgomery; 1992; s85-89)

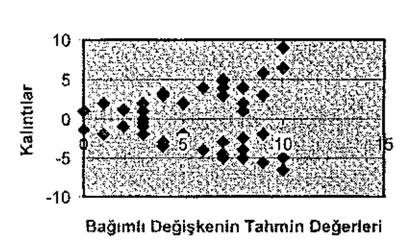
1.2.Kalıntı Analizi ile Nonlineerliğin Belirlenmesi

Modelde varolduğu düşünülen herhangi bir nonlineerliğin saptanması kalıntı analizi yardımı ile de gerçekleştirilebilir. Burada da en pratik yöntem kalıntıların değişkenlere göre seyrinin serpilme diyagramları yardımı ile gözlenmesidir. Kalıntıların bağımlı değişkenin tahmin değerleri karşısında gösterdiği değişim modelin varsayımları ile verilerin uygunluğunu test etmekte kullanılmaktadır. Başka bir ifade ile nonlineerliğin saptanabilmesi için analize modelin doğrusal olduğu, sabit varyanslılığın geçerli olduğu gibi varsayımlarla başlanıp, doğrusal bir model için parametre tahminleri gerçekleştirilebilir. Bu modelden hareketle elde edilen kalıntı değerlerinin bağımlı değişkenin tahmin değerleri karşısında gösterdiği değişim incelenerek başlangıç varsayımları ile ters düşülüp düşülmediği test edilebilir.(Huet;1996; s90-109)

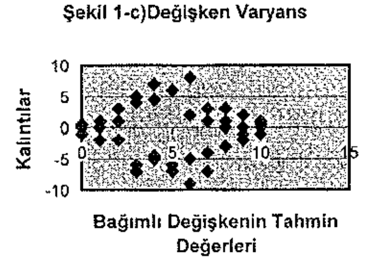
Şekil 1-a)Varsayımlara Uygun



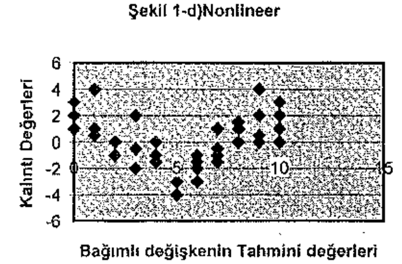

Şekil 1-b)Huni Biçimli Dağılım




Şekil 1-c)Değişken Varyans




Şekil 1-d)Nonlineer




Yukarıdaki dört şekil yardımı ile bu durum ortaya konmaktadır. 1-a'da modelin varsayımları (lineerlik; sabit varyans ) ile veriler arasında herhangi bir uyumsuzluk görülmemektedir. Dolayısıyla şekilden hareketle modelin doğrusal ve sabit varyanslı olduğu düşünülebilir. Şekil 1-b ve Şekil 1-c'de ise değişen varyans durumu gözlenmektedir. Şekil 1-b'de bağımlı değişkenin tahmin değerinin büyümesine paralel olarak kalıntı değerleri de büyümekte ve dolayısıyla sabit varyans varsayımı geçerli olmamaktadır. Şekil 1-c'de ise bağımlı değişkenin tahmin değerinin büyümesine paralel olarak kalıntı değerleri büyümekte, daha sonra da azalmaktadır. Bu durum da değişen varyans durumuna bir başka örnektir. Şekil 1-d ise modeldeki nonlineerliğin gözlenebilmesi bakımından ilginç bir örnektir.(Huet;1996; s74-75)

## 1.3.Polinomiyal Regresyon İle Model Nonlineerliğin Saptanması

Lineer regresyon modeli $y = X\beta + \varepsilon$ (1.3.1) parametrelerinde lineer herhangi bir model için genel formülasyonu sunmaktadır. Bu model hiç şüphesiz bir dizi polinomiyal regresyon modelini de kapsamaktadır.

Genel olarak k-mertebeden tek bağımsız değişkenli bir regresyon modeli ;

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_k x^k + \varepsilon$ (1.3.2) denklemi ile ifade edilmektedir.

Bu model çok değişkenli doğrusal regresyon ailesine dahildir. Polinomiyal modeller değişkenlerinde doğrusal olmamalarına karşılık parametrelerinde doğrusal oldukları için çok değişkenli doğrusal regresyon modelinin bir alt kümesi olarak inceleme konusu olmaktadırlar. Polinomiyal modeller ; özellikle fonksiyonel formun bilinmediği durumlarda yararlı olmaktadır. Özellikle karmaşık ve büyük bir olasılıkla nonlineer modellerin yaklaşık olarak tahmin edilmesinde; her fonksiyonun Taylor Serisi açılımı ile ifade edilebilmesi; başka bir deyişle her fonksiyonun bir polinom kestiriminin olması anlamında; yararlı olmaktadırlar. Bunun dışında belirli bir değişim aralığında sürekli olan bir fonksiyonu yüksek dereceden bir polinom ile kestirmek olasıdır. Fakat bu durum özetleyici istatistiklerin yorumlanmasında oldukça büyük karışıklıklara yol açmaktadır. Bunun dışında bağımlı değişkenin alacağı belli değerlerin tahmin edilmesinde de (büyük bir olasılıkla polinomiyal fonksiyonların bir dizi farklı konkavlık özelliği gösteren bölgelere sahip olmasından hareketle) büyük sıkıntılar yaşanmaktadır. (Hocking; 1995; s174-175)

### 1.3.1.Fonksiyonların Seri Açılımı

Bir fonksiyon, bağımsız değişkenin yakınsaklık aralığındaki değerleri için , bir kuvvet serisi ile gösterilebilir. Bağımsız değişkenin verilen bir değeri için, fonksiyonun yaklaşık bir değeri, serinin baştan az sayıda terimlerini kullanarak hesaplanabilir. Bu durumda fonksiyonları daha basit şekillere sokmak sureti ile, fonksiyonların değerlerini küçük yaklaşıklıklarla hesaplamak mümkündür. Fonksiyonların seriye açılımları Maclaurin ve Taylor serileri adı ile anılan iki tür seri yardımı ile yapılır. Maclaurin serisini elde etmek için f(x) fonksiyonunu $f(x)=a_0+a_1x+a_2x^2+a_3x^3+....$ biçiminde ifade etmek isteyelim. Böyle bir açılımın mümkün olabilmesi için a katsayılarının ne olduğuna bakılacak olursa

$$f(x)=a_0+a_1x+a_2x^2+a_3x^3+....$$
$$f'(x)=a_1+2a_2x+3a_3x^2+4a_4x^3+...$$
$$f''(x)=2a_2+3.2.a_3x+4.3.a_4x^2+...+n.(n-1)a_nx^{n-2}$$
..........................................................................................
..........................................................................................
..........................................................................................

x=0 için bu eşitliklerin alacağı şekiller incelenecek olursa

$$f(0)=a_0$$
$$f'(0)=a_1$$
$$f''(0)=2.1.a_2$$
.........................

$f^{(n)}(0)=n!a_n$ denklemleri elde edilecektir.

Bunlardan hareketle a katsayıları çözülürse

$a_k=\frac{1}{k!}f^{(k)}(0)$ k=0,1,2,....,n-1, n, n+1,.... bulunur. Bu katsayılar başlangıçtaki denklemdeki yerine konulacak olursa

$$f(x)=f(0)+\frac{x}{1!}f'(0)+\frac{x^2}{2!}f''(0)+\frac{x^3}{3!}f'''(0)+...+\frac{x^n}{n!}f^{(n)}(0)+...(1.3.1.1)$$

elde edilir.. Bu ifadeye f(x)'in Maclaurin Serisi denilmektedir. Bu ifadeden kolayca görülebilir ki f(x) fonksiyonunun Maclaurin serisine açılabilmesi için f(x) ve türevlerinin x=0 noktasında tanımlı ve sürekli olmaları gerekmektedir.

Öte yandan Taylor Teoremi herhangi bir f(x) fonksiyonunun a ve b noktalarında tanımlı ve n. mertebeden türevlenebilir olması durumda fonksiyonun f(b) değerini f(a) değeri cinsinden hesaplamanın olası olduğunu söylemektedir. (Karadeniz;1995; s35-49)

$$f(b)=f(a)+\frac{b-a}{1!}f'(a)+\frac{(b-a)^2}{2!}f''(a)+\ldots+\frac{(b-a)^{n-1}}{(n-1)!}f^{(n-1)}(a)+\frac{(b-a)^n}{n!}f^{(n)}(x_1)$$

( $a<x_1<b$ olmak üzere) (1.3.1.2)

Yukarıdaki formülasyondan hareket edilirse bazı çok kullanılan fonksiyonların Mac Laurin serisi açılımlarını şu şekilde elde etmek mümkündür:

$$e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots+\frac{x^n}{n!}+\ldots \quad (1.3.1.3)$$

$$\sin x=x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\ldots+\frac{(-1)^{n-1}x^{2n-1}}{(2n-1)!}\mp\ldots(1.3.1.4)$$

$$\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\ldots+\frac{(-1)^n x^{2n}}{(2n)!}\mp\ldots(1.3.1.5)$$

$$\arcsin x=x+\frac{1}{2}\frac{x^3}{3}+\frac{1.3}{2.4.}\frac{x^5}{5}+\frac{1.3.5}{2.4.6.}\frac{x^7}{7}+\ldots(1.3.1.6)$$

### 1.3.2. Polinomiyal Modellerin Kullanılmasındaki Kimi Yetersizlikler

Polinomiyal modeller her ne kadar model belirlenmesinde yararlı da olsa, belli yetersizlikler sergilemektedir. Bu yetersizliklere ve kimi önerileri şu şekilde özetlemek mümkündür:

1)Polinomiyal bir modelin formülasyonunda olabildiğince düşük mertebeli modellerin tercih edilmesi gerekmektedir. Genel bir kural olarak- aksi yönde güçlü bir sezgi delil vb. olmadığı takdirde- mertebesi yüksek (k>2) modeller tercih edilmemelidir. Düşük mertebeden ve dönüştürülmüş bir model her zaman yüksek mertebeden ama orijinal verilerle çalışan bir modele tercih edilmelidir. Bu arada n gözlem değerinden oluşan bir veri kümesi için her zaman n-1 mertebesinden bir polinomiyal model yaratmak olasıdır. Fakat böyle bir modelin iyi bir çözüm olacağı düşünülmemelidir.

2)Polinomun mertebesinin vb. belirlenebilmesi için çeşitli modeller denenmelidir. Bunun için sınamaya en düşük mertebeli model ile başlamak; istatistiksel olarak anlamlı olduğu ölçüde (t-değerleri anlamlı olduğu ölçüde) mertebe arttırımına gitmek düşünülebilir.

3)Ekstrapolasyon işleminin de tehlikeli olduğunun belirtilmesi ve veri setinin değişim aralığı dışında kalan bölgeler için genellemeler yapılmaması gerektiğinin altının çizilmesi gerekiyor. Genel olarak, hem enterpolasyon ve hem de ekstrapolasyon işlemleri için aynı tehlikenin geçerli olduğunu ve polinomiyal modellerin farklı bölgelerde farklı konkavlıklara sahip olmalarından hareketle bu gibi yöntemlerle yapılan çıkarsama ya da tahminlerin yanıltıcı olacağının altının çizilmesi gerekiyor.

4)Polinomun mertebesi arttıkça genel bir kural olarak $X'X$ matrisinin (ill-conditioned olduğu için) tersinin alınamadığının belirtilmesi gerekiyor. Başka bir ifade ile $X'X$ matrisinde sütunlar ya da satırlar arasındaki doğrusal bağımlılık nedeni ile sözkonusu matrisin determinantının değerinin sıfır olması polinomiyal modellerde oldukça sık gözlenen bir durumdur. Matrisin determinantının sıfıra yakın çıkması halinde de parametrelerin standard hatalarının yüksek çıkması modelin geçersizliğine istatistiksel bir delil olacağı için benzer problemlerle karşılaşılacaktır. (Montgomery; 1992; s202-226)

1.4. 1.Piecewise Regresyon ile Nonlineerliğin Saptanması

Bazen de bağımsız değişkenin tanım aralığını parçalara bölerek ve her bir bölge için ayrı bir eğri denklemi elde etmek yoluna gidilmektedir. Spline fonksiyonlar yardımı ile bu türden parçalı polinomiyal eğri tahminleri gerçekleştirilmektedir.

Spline fonksiyonlar k mertebeden polinomlardır. Farklı eğri parçalarının birbirleri ile birleştiği noktalar (knots) düğüm noktaları olarak adlandırılmaktadır. Sözkonusu spline'ın k-1 tane türev noktasında sürekli bir fonksiyon olduğunun düşünülebilmesi için k-1 tane türev değerinin sözkonusu düğüm noktalarında birbirleri ile eşit olması gerekmektedir.

Yine de literatürde pek çok pratik sorunun çözümü için kübik spline fonksiyonları (k=3) önerilmektedir. h tane düğüm noktası bulunan kübik bir spline fonksiyonu; $(t_1 < t_2 < t_3 < \dots < t_h$

olmak üzere) sürekli birinci ve ikinci türev fonksiyonuna sahip olması koşuluyla; şu şekilde ifade edilebilir:

$$E(y) = S(x) = \sum_{j=0}^{3} \beta_{0j} x^j + \sum_{i=1}^{h} \beta_i (x - t_i)_+^3 \quad (1.4.1.1.)$$

$$(x - t_i)_+ = \begin{cases} (x - t_i); & x - t_i > 0 \quad ise \\ 0 & x - t_i \le 0 \quad ise \end{cases}$$ biçiminde tanımlanmaktadır. Burada düğüm

noktalarının bilindiği varsayılmaktadır. Bununla birlikte düğüm noktalarının bilinmemesi halinde ise doğrusal olmayan bir regresyon problemi ile düğüm noktalarının tahmin edilebileceği söylenmelidir.Bununla birlikte düğüm noktalarının bilinmesi halinde yukarıdaki tipte bir denklemin parametrelerinin tahmin edilmesinin doğrusal regresyon yardımı ile kolaylıkla gerçekleştirilebileceğinin de altının çizilmesi gerekir. (s210)

1.4.2. Parçalı (Piecewise) Doğrusal Regresyon

Yukarıdaki modelin önemli bir özel hali parçalı doğrusal regresyon modelidir. Bu model, doğrusal spline'lar kullanılarak oluşturulabilir. Örneğin t gibi tek bir düğüm noktasının varlığı ve bu noktada hem eğim değişikliği ve hem de düğüm noktasında bir süreksizlik sözkonusu olsun. Bu durumdaki lineer spline modeli

$$E(y) = S(x) = \beta_{00} + \beta_{01} x + \beta_{10} (x - t)_+^0 + \beta_{11} (x - t)_+^1 \quad (1.4.2.1)$$

denklemi ile ifade edilecektir. Bu arada x<=t olması durumunda modelin

$E(y) = S(x) = \beta_{00} + \beta_{01} x$ biçimde olacağı , x>t olması halinde ise

$$E(y) = S(x) = \beta_{00} + \beta_{01} x + \beta_{10} (1) + \beta_{11} (x - t) \quad (1.4.2.2)$$

$$E(y) = S(x) = (\beta_{00} + \beta_{10} - \beta_{11} t) + (\beta_{01} + \beta_{11}) x \quad (1.4.2.3)$$

Başka bir ifade ile x<=t olması durumunda modelin $\beta_{00}$ kesim noktasına ve $\beta_{01}$ eğimine eşit olduğunun söylenmesi gerekir. Eğer x>t olursa bu durumda da kesim noktasının $(\beta_{00}+\beta_{10}-\beta_{11}t)$ ve eğiminin de $(\beta_{01}+\beta_{11})$'e eşit olduğunun da not edilmesi gerekiyor.

Son olarak $\beta_{10}$'un t düğüm noktasında bağımlı değişkenin ortalama değerleri arasındaki fark olduğunun da vurgulanması gerekiyor. Eğer regresyon fonksiyonunun düğüm noktasında sürekli olması arzu ediliyorsa, bu durumda elde edilecek regresyon fonksiyonunun da daha düzgün (smoother) bir fonksiyonun olacağının belirtilmesi gerekiyor. Bu durum $\beta_{10}(x-t)^0$ teriminin

orijinal modelden çıkarılması ile sağlanacaktır. Bu durumda model

$$E(y)=S(x)=\beta_{00}+\beta_{01}x+\beta_{11}(x-t)_+^1 \quad (1.4.2.4.)$$

biçimini alacaktır. x<=t, olması halinde model

$$E(y)=S(x)=\beta_{00}+\beta_{01}x$$ biçiminde,

x>t olması halinde ise

$$E(y)=S(x)=\beta_{00}+\beta_{01}x+\beta_{11}(x-t) \quad (1.4.2.5)$$ biçimini almaktadır.

## 1.5 Gösterge Değişkenlerden Hareketle Nonlineerlik Testi

Hatırlanacağı gibi Chow testi bir doğrusal regresyon modelinde yapısal bir değişikliğin olup olmadığını test eden bir tür F testidir. Chow testinin yapısal değişiklik olup olmadığı sorusuna olumlu yanıt vermesi hakinde ise sonuçları yorumlamada bazı karışıklıklar yaşanmaktadır. Bilindiği gibi Chow testi yapısal değişikliği saptaması halinde sözkonusu değişikliğe hangi parametre ya da parametrelerin neden olduğunu bize söyleyememektedir. Bunu saptamanın yolu ise kukla değişkenler kullanarak basit bir regresyon analizi gerçekleştirmekten geçmektedir. Bu yöntem Chow testine göre daha üstündür. Çünkü eğer sözkonusu ise yapısal değişikliğin hangi parametre ya da parametrelerden kaynaklandığının bilgisini de vermektedir.

Deminki soruya yanıt bulabilmek için gölge değişken yaklaşımı şu şekilde özetlenebilir:

Toplam n bağımlı değişken gözlem değerinden oluşan örneklem iki parçaya bölünsün.

Eğrinin bir bölgesi için ($n_1$ gözlem değeri için) bir doğrusal denklem, diğer bölgesi için de ($n_2$ gözlem değeri için) farklı bir doğrusal denklem sözkonusu olsun. ($n = n_1 + n_2$ olmak üzere)

$n_1$ ve $n_2$ gözlemden oluşan iki küme birleştirilerek şu regresyon denklemi oluşturulsun:

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta_1 X_i + \beta_2 (D_i X_i) + \varepsilon_i$$

Bu modelden hareketle

$$E(Y_i / D_i = 0, X_i) = \alpha_1 + \beta_1 X_i$$

$$E(Y_i / D_i = 1, X_i) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_i$$

denklemleri elde edilecektir. Yukarıdaki denklemlerde $\alpha_2$ kesim noktasındaki farkı; $\beta_2$ de eğimdeki farkı temsil etmektedir. Bu katsayıların istatistiksel olarak yorumlanması ise şüphesiz modelde yapısal bir değişiklik olup olmadığını, var ise bunun hangi parametre ya da parametrelerden kaynaklandığını göstermeye yetecektir. Böyle bir model için elde edilecek olan parametre tahminlerinin standard hataları ile kıyaslanması hiç şüphesiz hangi parametrenin/ parametrelerin modele dahil edilmesinin istatistiksel olarak anlamlı olduğunu söyleyecektir.(Gujarati; 1995;s 509-522)

**Uygulama**

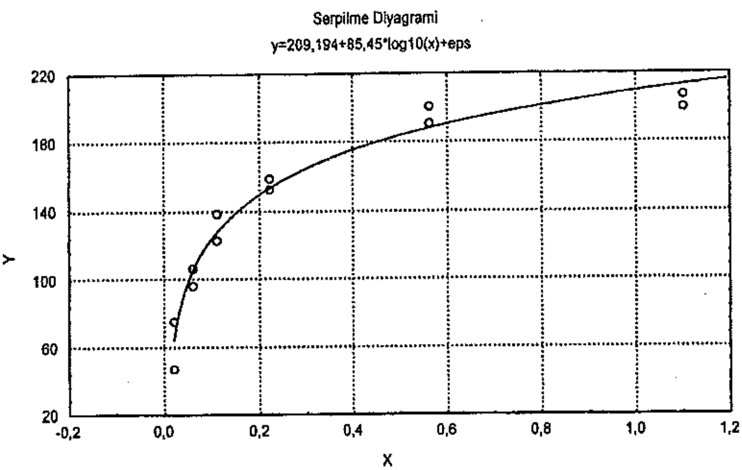
Nonlineerlik testlerin eşzamanlı bir uygulamasını tartışabilmek için Bates ve Watts (1988) ın verileri kullanılacaktır. Bu çalışmada aşağıdaki veriler için önerilen nonlineer model

$Y_i = \dfrac{\theta_1 X_i}{X_i + \theta_2} + \varepsilon_i$ denklemi ile ifade edilen Michaelis-Menten modelidir. Kimyasal bir maddenin enzimatik reaksiyon hızının (Y) maddenin yoğunluğu (X) ile olan ilişkisini ortaya çalışmaktadır. Adı geçen kaynakta yukarıdaki model doğrusal olmayan yöntemlerle çözülmektedir. Veri tablosu ise şöyledir:

| X(ppm) | Y(counts/min)/mi |
|---|---|
| 0,02 | 76 |
| 0,02 | 47 |
| 0,06 | 97 |
| 0,06 | 107 |
| 0,11 | 123 |
| 0,11 | 139 |
| 0,22 | 159 |
| 0,22 | 152 |
| 0,56 | 191 |
| 0,56 | 201 |
| 1,1 | 207 |
| 1,1 | 200 |

Verilerle ilgili olarak elde edilen

serpilme diyagramı da aşağıdadır.




Yukarıdaki serpilme diyagramından değişkenler arasındaki ilişkinin doğrusal olmadığı görülecektir.

Yine de modelin doğrusal olduğu gibi bir varsayımdan hareket edilecek olursa şu STATİSTİCA sonuçları elde edilecektir.

Regresyon Sonuçları
R= ,83103618 $R^2$= ,69062113 Ayarlı $R^2$= ,65968324
F(1,10)=22,323 p<,00081 Tahminin Standard Hatası: 30,898

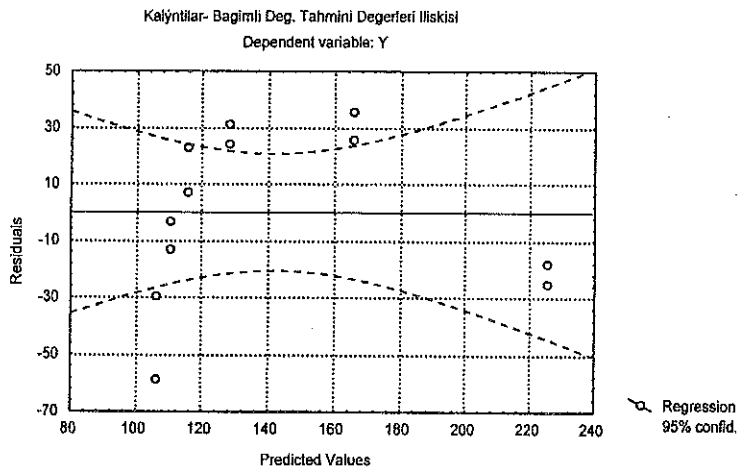
| | BETA | Std.Err. BETA | B | St. Err. of B | t(10) | p-olasılığı |
|---|---|---|---|---|---|---|
| Kesim | 103,4881 | 12,02377 | 8,606956 | ,000006 | | |
| X | ,831036 | ,175892 | 110,4211 | 23,37100 | 4,724704 | ,000811 |

Durbin-Watson d (new.sta)
ve Kalıntıların Serisel Korelasyonu

| | Durbin-Watson d | Serisel Korr. |
|---|---|---|
| Tahmin, | 683174 | ,620056 |

Durbin-Watson ve Serisel korelasyon istatistiklerine bakıldığında modelin yukarıda kurgulanan doğrusal hali ile verilere iyi uymadığı görülecektir.

Öte yandan kalıntıların Y'nin tahmin değerleri karşısındaki seyrine bakıldığında da modelin Doğrusallık vb. gibi varsayımları ile veriler arasındaki uyumsuzluk görülecektir. Kalıntılara Ilişkin grafik şu şekildedir:




Diğer taraftan modelin doğrusal olup olmadığı ile ilgili F-testi sonuçları da bir EXCEL tablosu yardımı şu şekilde özetlenmektedir:

## F-testi Sonuçları

| Gözlem Sırası | X | Y | Y-Yort | (Y-Yort)^2 | *Ytahminleri* | *Kalıntılar* | *Kalıntı Kareleri* |
|---|---|---|---|---|---|---|---|
| 1 | 0,02 | 76 | 14,5 | 210,25 | 105,6964834 | -29,6964834 | 881,8811263 |
| 2 | 0,02 | 47 | -14,5 | 210,25 | 105,6964834 | -58,6964834 | 3445,277163 |
| 3 | 0,06 | 97 | -5 | 25 | 110,1133265 | -13,11332647 | 171,9593311 |
| 4 | 0,06 | 107 | 5 | 25 | 110,1133265 | -3,113326468 | 9,692801696 |
| 5 | 0,11 | 123 | -8 | 64 | 115,6343803 | 7,365619696 | 54,25235351 |
| 6 | 0,11 | 139 | 8 | 64 | 115,6343803 | 23,3656197 | 545,9521838 |
| 7 | 0,22 | 159 | 3,5 | 12,25 | 127,7806987 | 31,21930126 | 974,644771 |
| 8 | 0,22 | 152 | -3,5 | 12,25 | 127,7806987 | 24,21930126 | 586,5745534 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 8 | 0,22 | 152 | -3,5 | 12,25 | 127,7806987 | 24,21930126 | 586,5745534 |
| 9 | 0,56 | 191 | -5 | 25 | 165,3238648 | 25,67613517 | 659,2639174 |
| 10 | 0,56 | 201 | 5 | 25 | 165,3238648 | 35,67613517 | 1272,786621 |
| 11 | 1,1 | 207 | 3,5 | 12,25 | 224,9512463 | -17,95124626 | 322,2472422 |
| 12 | 1,1 | 200 | -3,5 | 12,25 | 224,9512463 | -24,95124626 | 622,5646898 |

| | | | |
|---|---|---|---|
| SSPE= | 697,5 | SSE= | 9547,096754 |
| MSPE= | 116,25 | | |
| SSLOF= | 8849,597 | | |
| MSLOF | 2212,399 | | |
| Fo= | 19,03139 | | |
| Ftablo= | 5,143249 | | |
| Sonuç= | Nonlineerlik var! | | |

Yukarıdaki sonuçlara göre de modelin doğrusal olmadığına hükmedilebilir.

**Parçalı-Doğrusallaştırma (Piecewise Linearization) Sonuçları**

Öte yandan nonlineerliğin saptanmasında parçalı doğrusallaştırmadan da yararlanılabilir.Bununla ilgili olarak STATİSTİCA kullanılmış ve şu sonuçlar elde edilmiştir:

Model is: Piecewise linear regression with breakpoint
Dependent variable: Y Loss: Least squares
Loss Fonksiyonu: 1451,9513053 R= -- Açıklanan Değişim: -- %

| | Sabit B0 | X | Sabit.B0 | X | Kesişim Nokt. |
|---|---|---|---|---|---|
| Tahmin | 49,75820 | 764,3443 | 153,1469 | 50,82938 | 141,5833 |

İlk olarak parçalı doğrusallaştırmanın da doğrusal olmayan algoritmalar ile gerçekleştirildiğini belirtmek gerekiyor. Bu durumda incelenen eğri iki doğrunun toplamı olarak ele alınmakta ve bu iki doğrusun kesim noktası tahmini program tarafından gerçekleştirilmektedir. Fakat parçalı doğrusallaştırmanın doğrusal olmayan optimizasyonla çözülmesi doğrusal olmayan regresyonun da muzdarip olduğu (yakınsamanın sağlanamaması gibi) bazı sorunlarla karşılaşılmasına neden olmaktadır. Buna ek olarak parçalı doğrusallaştırmada bir fonksiyon yerine (örneğimizde iki) birden fazla fonksiyonun kullanılması hiç şüphesiz serbestlik derecesinin azalmasına yol açmaktadır. Dolayısıyla gözlem sayısının azlığına ya da çokluğuna bakılarak parçalı doğrusallaştırma yapılıp yapılmayacağına karar verilmelidir.

Örneğimize dönülecek olursa bu yöntem ile anlamlı bir sonuç bulunamadığını söylemek gerekiyor.

**Denenen Bazı Modeller**

**1)Y=A+B*Karekök(x) Modeli**

Regresyon Sonuçları
R= ,92205062 $R^2$= ,85017735 Ayarlı $R^2$= ,83519509
F(1,10)=56,746 p<,00002 Tahminin Standard Hatası: 21,502

| | BETA | St. Err. of BETA | B | St. Err. of B | t(10) | p-olasılığı |
|---|---|---|---|---|---|---|
| Kesim | 67,1521 | 11,66863 | 5,754929 | ,000184 | | |
| SQRV1 | ,922051 | ,122402 | 149,6498 | 19,86599 | 7,532966 | ,000020 |

Durbin-Watson d (new.sta)
Ve Kalıntıların Serisel Korelasyonu

| | Durbin-Watson d | Serisel Korr. |
|---|---|---|
| Tahmin | ,891426 | ,543329 |

**2)Y=A+B*Ln(x) Modeli**

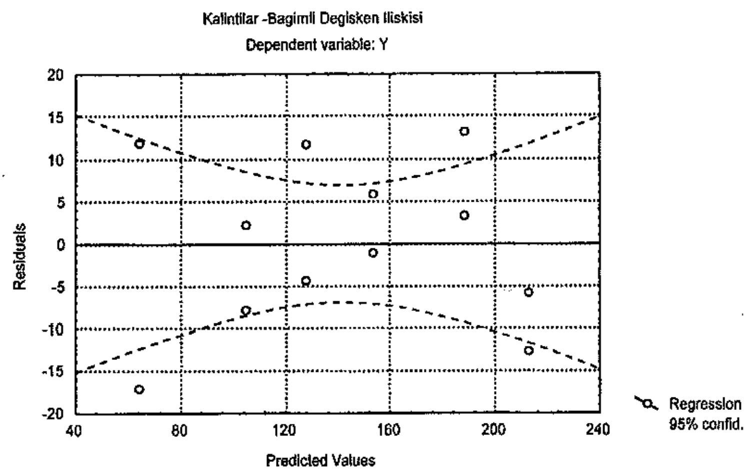
| | St. Err. of BETA | B | St. Err. of B | t(10) | p-level |
|---|---|---|---|---|---|
| Kesim | 209,1945 | 5,045082 | 41,46503 | ,000000 | |
| LN-V1 | ,059017 | 37,1104 | 2,229301 | 16,64667 | ,000000 |

Durbin-Watson d (new.sta)
Ve Kalıntıların Serisel Korelasyonu

| | Durbin-Watson d | Serisel Korr. |
|---|---|---|
| Tahmin | 2,197898 | -,050929 |

Bu iki modelden başka modeller de (örneğin polinomiyal modeller) denenmiştir. Fakat araştırma bulguları istatistiksel açıdan anlamlı olmadığı için buraya aktarılmamıştır.

Yukarıdaki iki modelden birincisinde Durbin-Watson ve serisel korelasyon istatistikleri model belirlenmesinde (specification) hata yapıldığını göstermektedir. İkinci modelde ise bu değerler kabul edilebilir nitelikte bulunmuş ve bu model kabul edilmiştir. Öte yandan bu model ile ilgili kalıntı diyagramı aşağıdaki gibidir.




**Sonuç:** Nonlineerliğin saptanmasında her yöntemin kendine göre yetersizlikleri vardır. Modelin doğrusal olup olmadığının test edilmesinde F-testinin uygulanabilmesi için bağımsız değişkenin en azından bir noktasında bağımlı değişkenin tekrarlı gözlem değerlerinin (replication) elde edilmesi gerekir. Parçalı doğrusallaştırma ise doğrusal olmayan algoritmalar ile çözüldüğü için doğrusal olmayan regresyon yöntemlerinin sorun yaşadığı noktalarda aynı sorunlardan muzdarip olmaktadır. Polinomiyal modellerde; X'X matrisinin determinantının sıfıra yakın olması (çoklu doğrusal bağıntı) problemi sıklıkla yaşandığı için; doğrusal olmama durumunun saptanmasında tıkanıklıklar sıklıkla yaşanmaktadır. Bu durumlarda parametrelerin standard hataları çok büyük olmaktadır. Bunun dışında doğrusal olmayan modellerde genellikle birden fazla çözüm sözkonusu olduğu için bu durum da sıkıntı yaratmaktadır. Yine de bütün bu modeller eşanlı olarak denenebilir. Bunlara ek olarak örneğin serpilme diyagramları; kalıntı diyagramları gibi her zamanki görsel analiz araçları yardıma çağrılabir.

**Kaynaklar:**

1)Bates;D.M.; Watts; D.G.;Nonlinear Regression;John Wiley;1988

2)Gujarati; D.N.;Basic Econometrics; Mac Graw Hill; Third Edition; 1995

3) Hocking; R.R.; Methods and Applications of Linear Models: Regression and the Analysis of Variance; John Wiley; 1996

4)Huet; S. vd.;Statistical Tools for Nonlinear Regression; Springer Series in Statistics;1996

5)Karadeniz;A.;Yüksek Matematik; Çağlayan Yayınevi; 7. baskı; 1995

6)Montgomery; D.C.; Peck; E.A.; Introduction to Linear Regression Analysis; John Wiley; Second Edition; 1992

---

* Dr. Araştırma Görevlisi İstanbul Üniversitesi İktisat Fakültesi