PARAMETRİK OLMAYAN REGRESYON ANALİZİ

Nuray TEZCAN[1]

Özet: Regresyon analizi, regresyon fonksiyonu hakkında istatistiksel çıkarımda bulunan bir analizdir ve temelde iki değişken arasındaki ilişkinin incelenmesinde kullanılır. Parametrik ve parametrik olmayan regresyon teknikleri, regresyon analizine iki farklı açıdan yaklaşır. Parametrik regresyon güçlü varsayımlara sahipken, parametrik olmayan regresyon ise bu varsayımları gerektirmez. Bununla birlikte parametrik olmayan tahminleyenler, parametrik model geçerli olduğunda, parametrik tahminleyenlere göre daha az etkindirler. Parametrik regresyon ve parametrik olmayan regresyon yöntemleri, regresyon analizi için her ne kadar farklı yaklaşımlar olarak kabul edilseler de, bu durum bir yöntemin diğerini dışlayacağı anlamına gelmez. Parametrik olmayan regresyon parametrik regresyonun önerdiği modelin geçerliliğini doğrulamak için kullanılabilir ya da tam tersi, veriye uygun model parametrik olmayan model tarafından yapılan tahmine göre kurulabilir. Böylece parametrik olmayan regresyon, veri analizinin son aşaması ya da modelleme sürecinde açıklayıcı veya doğrulayıcı bir adım olarak görülebilir. Bu amaçla; İMKB'de işlem gören şirketlerin piyasa değeri/defter değeri oranı çeşitli finansal oranlar ile açıklanmaya çalışılmıştır. Kaldıraç ve karlılık oranları istatistiksel olarak anlamlı bulunmuştur.

Anahtar Kelimeler: Kernel tahmini, düzgünleştirme parametresi, LOWESS yöntemi,

Abstract: Regression analysis refers to methods for statistical inference about the regression function and basically, it is used in analyzing the relationship between two variables. Parametric and nonparametric regression techniques represent two different approaches to the problem of regression analysis. While parametric regression have strong assumptions and particular function form, nonparametric regression doesn't have these requirements. Nonparametric estimator, however, are less efficient than the parametric variety when the parametric model is valid. It should be noted that even though parametric and nonparametric regression models represent different approaches to regression, this does not mean that the use of one approach precludes the use of the other. Indeed, nonparametric regression techniques can be used ases the validity of a proposed parametric model. Thus, nonparametric regression procedures may represent the final stage of data analysis or an explanatory or confirmatory step in the modelling precess. Within this context, market value/book value ratio of companies which is traded on the ISE is explained by various financial ratios. Leverage and profitability ratios are found significant statistically

Key Words: Kernel estimation, smoothing parameter, LOWESS method,

I.Giriş

Regresyon analizi; herhangi bir değişkenin (bağımlı değişken) bir veya birden fazla değişken ile (bağımsız – açıklayıcı değişken) arasındaki

[1]Yrd. Doç. Dr., Haliç Üniversitesi İşletme Fakültesi

ilişkinin matematik bir fonksiyon şeklinde yazılmasıdır. Elde edilen bu fonksiyona ise regresyon denklemi adı verilmektedir (Orhunbilge, 2000:12). Regresyon analizinin temeli; ilk olarak Francis Galton tarafından 19. yüzyılın sonlarında atılmıştır. Galton yaptığı çalışmada; anne-babaların boyu ile çocuklarının boyları arasındaki ilişkiyi incelemiş ve kısa boylu anne-babaların çocuklarının boylarının kısa, uzun boylu anne-babaların çocuklarının boylarının uzun olmasına rağmen, çocuklarının boylarının anakütle boy ortalamasına doğru yaklaşma eğiliminde olduğunu görmüştür. Bu eğilimi "ortaya doğru çekilme = regression to mediocrity" olarak adlandırmıştır. (Galton, 1886: 246-263). Galton'un çalışmaları bugün, değişkenler arasındaki istatistik ilişkileri inceleyen "Regresyon Analizi (Regression Analysis – Relationship Analysis)"nin başlangıcı olmuştur. Günümüzde regresyon analizi için parametrik ve parametrik olmayan olmak üzere iki türlü yaklaşım bulunmaktadır.

II. Parametrik Olmayan Regresyon Analizi

Regresyon analizi; açıklayıcı (bağımsız) değişken(ler) X_i ile bağımlı değişken Y arasındaki genel ilişkiyi belirlerken koşullu beklenti fonksiyonunu (Conditional Expectation Function) $m(x) = E(Y \setminus X = x)$ tahmin etmektedir. Bu amaçla aşağıdaki formül kullanılabilir.

$$Y_i = m(X_i) + \varepsilon_i \qquad i= 1,\ldots,n \qquad (2.1)$$

$m(X)$, matematik anlamda bir fonksiyondur. Birçok durumda teori, $m(X)$'in biçimi hakkında herhangi bir kısıtlama getirmez, diğer bir ifade ile $m(X)$'in doğrusal, kuadratik ya da üstel olup olmadığı hakkında bir varsayımda bulunmaz.

Bağımsız tesadüfi hata terimleri (ε_i) ise, aşağıdaki şartları sağlamaktadır.

$$E(\varepsilon_i \setminus X_i = x) = 0$$

$$Var(\varepsilon_i \setminus X_i = x) = \sigma^2$$

Parametrik olmayan regresyonun güçlü varsayımlarda bulunmaması ve sabit parametrik bir modele bağlı kalmaması dışında; iki değişken arasında var olan ilişkinin açıklanmasında esneklik sağlaması ve ardışık X değerleri arasında interpolasyon ve kayıp gözlemlerin yerine konulması için esnek yöntemler bulması gibi avantajları da vardır (Hardle, 1990: 6).

Parametrik regresyon analizinde, biçimi önceden belirlenmiş modelin parametreleri tahmin edilirken, parametrik olmayan regresyon analizinde ise amaç, regresyon fonksiyonu olan $m(x)$'i doğrudan tahmin etmektir (Hardle v.d., 2004:115). Ayrıca parametrik olmayan regresyon için model belirlenirken bilinmeyen regresyon eğrisini içeren uygun fonksiyon uzayı seçilmekte ve bu seçim yapılırken regresyon fonksiyonunun süreklilik ve türevlenebilme gibi düzgünlük özelliklerine sahip olduğu noktasından hareket edilmektedir (Eubank, 1990: 4).

Parametrik olmayan regresyon aykırı gözlemlerin (outliers) bulunduğu veri setleri için önemli bir analizdir. Literatürde aykırı gözlemlerin etkilerini farklı biçimlerde ele alan güçlü (robust) parametrik yöntemler vardır. Bununla birlikte; aykırı gözlemlerden dolayı parametreler bozulduğu için bu güçlü yöntemler bile uygun çözümler üretemeyebilir ve verinin gerçek yapısı modele yansıtılamaz. Bu durumda parametrik olmayan regresyon, farklı açılımlar sağlamaktadır (Hardle, 1990: 11).

Parametrik olmayan regresyonun açıklanan bu avantajları yanında parametrik regresyona göre bazı dezavantajları da bulunmaktadır. Parametrik regresyona göre artan işlem sayısı ile işlemlerin karmaşıklığı ve bazı durumlarda elde edilen sonuçların açıklanmasındaki zorluk bunlardan bazılarıdır (Fox, 2000a:2) Ayrıca parametrik olmayan regresyonda, yakın gözlemlerin incelenmesi ile değişkenler arasındaki ilişki ortaya çıkarıldığından çok fazla sayıda veriye ihtiyaç duyulmaktadır (Yatchew, 1998: 672).

Bağımsız değişken sayısının artması ile yaşanan sorun, parametrik olmayan regresyondaki tahmin yöntemlerinden biri olan lokal polynomial regresyon yönteminde boyut sorunu olarak ortaya çıkmaktadır (curse of dimensionality). Bağımsız değişkenlerin sayısı arttığı zaman, odak noktasının lokal komşuluğundaki noktaların sayısı hızlı bir şekilde azalma eğilimi göstermektedir. Dolayısıyla, yerel tahminler içindeki sabit gözlem sayısını elde edebilmek için komşuluğun oldukça küçük tutulması gerekir. Lokal polynomial regresyonun genel varsayımı gereği; odak noktası etrafındaki komşuluğun büyüklüğünün artması ise tahminin yanlılığını arttırdığı için regresyon fonksiyonuna ait tahminin kalitesini azaltmaktadır. (Fox, 2000b:20)

Birçok parametrik tahminleyen için risk ya da tahminin beklenen hata karesi (expected squared error of estimation) n^{-1} oranında sıfıra doğru yaklaşır. Parametrik olmayan tahminleyenler için genellikle bu oran $\delta \in$ (0,1) olmak üzere $n^{-\delta}$ ‘dir ve m(x)’in düzgünlüğüne bağlıdır. Örnek olarak; eğer regresyon fonksiyonunun m(x) iki kere türevi alınabilirs,e $n^{-\frac{4}{5}}$ oranına sıkça rastlanır. Bundan dolayı, parametrik olmayan regresyon, parametrik regresyon ile karşılaştırıldığında etkinlik bakımından daha zayıftır. Eğer parametrik regresyon ile belirlenen fonksiyonun doğruluğundan şüphe ediliyorsa artık n^{-1} oranı geçerli değildir ve parametrik olmayan regresyon teknikleri daha iyi sonuç verirler (Eubank, 1990: 10).

A.Düzgünleştirme Kavramı

Düzgünleştirici (smoother); bir ya da birden fazla bağımsız değişkenin fonksiyonu olan bağımlı değişkenin sahip olduğu trendi ifade etmek için kullanılan bir araçtır ve bağımlı değişkenin kendisinden daha az değişken bir trend tahmini yapmayı amaçlamaktadır. Bir düzgünleştiricinin en önemli özelliği, değişkenler arasındaki ilişkinin biçimini kesin bir biçimde varsaymamasıdır ve bu özelliğinden dolayı parametrik olmayan regresyonda sık kullanılan bir araçtır (Hastie ve Tibshirani, 1990: 9) Parametrik regresyon ile elde edilen doğru kesin (rigid) parametrik bir biçime sahip olduğu için düzgünleştici değildir Bilinen en basit düzgünleştirici ise hareketli ortalamalardır. Bir düzgünleştirici tarafından tahmin edilen eğriye ise "düzgün" (smooth) adı verilmektedir. Tek bir bağımsız değişkenin olduğu durumlar "serpilme diyagramı düzgünleştirmesi" (scatterplot smoothing) olarak adlandırılmaktadır. Serpilme diyagramındaki noktalar, herhangi bir olasılıksal model olmaksızın düzlem üzerindeki noktaların birikimi olarak kabul edilmektedir (Rupperd v.d., 2003 : 57) .

Serpilme diyagramı düzgünleştiricileri aynı serbestlik derecesinde yaklaşık aynı sonuçlar vermelerine rağmen arasında değerlendirme yaparak karar vermede aşağıdaki kriterler kullanılabilir (Rupperd v.d., 2003 : 87-88).

- Kullanılan bilgisayar programına uygunluk
- İlgilenilen değişkenler arasındaki muhtemel ilişkiyi ortaya çıkarabilme esnekliği
- Kullanılan yöntem sonucunda elde edilen sonuçların anlaşılabilirliği
- Kullanılan yöntemin matematik özelliklerinin analizinin kolaylığı
- Elde edilen cevapların güvenilirliği
- Kullanılan yöntemin, veriyi en etkin biçimde kullanarak sonuç elde etmesi
- Kullanılan yöntemin, kolaylıkla daha karmaşık düzenlemelere uyarlanabilmesi

III. Parametrik Olmayan Regresyon Analizinde Tahmin Yöntemleri

A. Kernel Regresyon

$Y_i = m(x_i) + \varepsilon_,$ $\quad i = 1,...,n \quad$ (x_i, Y_l) gözlemlerine sahip olduğumuzu varsayalım.

m , şekli bilinmeyen "düzgün" bir fonksiyon ve hata terimleri de bağımsız ve 0 ortalamaya sahiptir. $m(x)$ fonksiyonu lokal olarak ağırlıklı ortalama (locally weighted averaging) ile tahmin edilecek olursa Nadaraya – Watson (N-W) tahminleyeni elde edilir.

K , simetrik olasılık yoğunluk fonksiyonu olan Kernel Fonksiyonu, h lokal komşuluğun büyüklüğünü kontrol eden ve negatif olmayan band genişliği ve $K_h(.) = K(./h)/h$ olmak üzere; şekli bilinmeyen $m(x)$ fonksiyonunu tahmin etmek için öncelikle koşullu beklenti fonksiyonunun tanımlanması gerekir (Hardle ve Linton, 1994: 4).

$$m(x) = E(Y/X = x) = \int y \frac{f(x,y)}{f_x(x)} dy = \frac{\int yf(x,y)dy}{f_x(x)} \qquad (3.1)$$

$f(x,y)$'nin Kernel yoğunluk tahmini; $\hat{f}_h(x,y)$ olmak üzere;

$\hat{f}_h(x,y) = n^{-1}\sum_{i=1}^{n} K_h(x - X_i)K_h(y - Y_i)$'dir.

$K(u) = K(-u)$ ve $\int K(u)du = 1$ olduğu gözönüne alınırsa,

$\int \hat{f}_h(x,y)dy = n^{-1}\sum_{i=1}^{n} K_h(x - X_i)$;

$\int y\hat{f}_h(x,y)dy = n^{-1}\sum_{i=1}^{n} K_h(x - X_i)Y_i$

Yukarıda belirtilen ifadeler, (3.1.) nolu eşitlikte yerine konulursa; Nadaraya-Watson kernel tahmini aşağıdaki gibidir (Nadaraya, 1964: 141-142 – Watson, 1964: 359-372)

$$\hat{m}_h(x) = \frac{\sum_{i=1}^{n} K_h(x - X_i)Y_i}{\sum_{i=1}^{n} K_h(x - X_i)} \qquad (3.2.)$$

Band genişliği h, aynı zamanda $\hat{m}_h$ tahmininin düzgünlüğünün derecesini belirlemektedir. Bu durum; h sırasıyla sıfıra ve sonsuza giderken limitleri gözönüne alındığında kolaylıkla görülebilir. Herhangi bir x noktasında, $x = X_i$ olmak üzere $h \to 0$ 'a giderken $\hat{m}_h(X_i) \to Y_i$ ve $h \to \infty$'a giderken $\hat{m}_h(X_i) \to \overline{Y}$. Bu iki limit durumundan açıkça görüldüğü gibi, örnek büyüklüğü n ile ilişkili olan düzgünleştirme parametresi h, sıfıra ne çok hızlı ne de çok yavaş yakınsamalıdır.

B. Lokal Polinomial Regresyon Yöntemi

Lokal polynomial regresyon yöntemi, Kernel yönteminde karşılaşılan bazı eksiklikleri (yanlılığın azaltılması gibi) gidermeye çalışmaktadır. Lokal polynomial regresyon; LOcally WEighted Smoothing Scatterplot ifadesindeki baş harflerin birleştirilmesiyle oluşan LOWESS adı ile bilinmektedir.* LOWESS yöntemi ilk olarak Cleveland tarafından önerilmiştir. Zaman serileri analizinde kullanılan "ağırlıklı bir hareketli ortalama" kavramının uzantısıdır (Sprent ve Smeeton, 2007: 475).

Polynomial regresyon, genel lineer regresyon modelinin özel bir halidir. Regresyon fonksiyonunu eğrisel hale getirmek için, bağımsız değişken(ler)in karesini ve yüksek dereceden değerlerini içerir (Neter v.d., 1996: 223). Bağımsız değişkenlerinde p. dereceden polynomial regresyon modeli aşağıdaki gibi yazılabilir.

$$y = \alpha + \beta_1 x + \beta_2 x^2 + ... + \beta_p x^p + \varepsilon \qquad (3.3)$$

(3.3) ile belirtilen model en küçük kareler yöntemi ile çözülebilir. $p = 1$ ise doğrusal, $p = 2$ ise kuadratik bir fonksiyon elde edilir. Lokal polynomial regresyon, $w_i = K\left[\frac{(x_i - x_0}{h}\right]$ lokal Kernel ağırlıkları kullanılarak, x_0 odak noktasında Kernel tahmininin polynomial tahmin uzantısı olarak kabul edilebilir. Ağırlıklı en küçük kareler modele uygulandığında,

$$y_i = a + b_1(x_i - x_0) + b_2(x_i - x_0)^2 + ... + b_p(x_i - x_0)^p + e_i \quad (3.4)$$

elde edilir (Fox, 2000a:22).

(3.4) modelinde ağırlıklı hata kareleri toplamı $(\sum_{i=1}^{n} w_i^2 e_i^2)$ minimize edilmektedir.

Lokal olarak ağırlıklı regresyon, düzgünleştirilmiş noktaları bozan olağandışı noktalardan korunma amacı ile güçlü (robust) hale getirilebilir. Bu amaçla sürece, iteratif ağırlıklı en küçük kareler yöntemi adapte edilir. Böylece, güçlü lokal olarak ağırlıklı regresyon, eski bir yöntem olan düzgünleştirme ile yeni bir yöntem olan güçlü (robust) tahminin bir kombinasyonu biçiminde elde edilir (Cleveland, 1979: 829). LOWESS yönteminde; lokal olarak kullanılacak olan polinomun derecesi, ağırlık (kernel) fonksiyonunun ne olacağı, modeli güçlü hale getirmek için kaç iterasyon yapılması gerektiği ve düzgünleştirmede kullanılacak gözlem oranı (span) konuları karar verilmesi gereken kritik noktalardır (Cleveland, 1979: 834).

C. Düzgünleştirme Parametresinin Seçimi

Parametrik olmayan regresyonda, tahminin düzgünlük derecesini kontrol eden düzgünleştirme parametresinin (smoothing parameter) diğer bir ifadeyle bant genişliğinin seçimi, ağırlık fonksiyonunun (Kernel Fonksiyonu) seçiminden daha önemli bir konudur. Uygulamada optimum bant genişliğinin ne olması gerektiği konusu hala tartışılmakla birlikte uygulanan çeşitli yöntemler bulunmaktadır.

$\hat{f}(x)$ fonksiyonunun istatistik özellikleri bant genişliği (h) ile yakından ilişkilidir. Bant genişliği seçimi temelde, tahminin yanlılığı ile varyansı arasında bir denge (trade-off) oluşturulması anlamına gelmektedir. Olması gerekenden daha geniş bir bant genişliği kullanıldığında tahminde daha küçük bir varyans elde edilir fakat gerçek yoğunluk aşırı düzgünleştirilmiş (oversmoothing) olabileceği için yanlı bir tahmin elde edilir (Dinardo ve Tobias, 2001: 11–28). Aynı biçimde daha dar bir bant genişliği kullanılırsa bu kez de yanlılık miktarı azalırken, tahminin varyansı artabilir. Literatürde yanlılık ve varyans arasındaki bu dengeyi kurmak amacı ile çeşitli kriterler kullanılmaktadır. İşte tek bir x noktasında tahminleyenin kuadratik kaybını (quadratic loss) ölçen “Hata Kareleri Ortalaması (Mean Squared Error-MSE)” bu kriterlerden biridir. Uygulamada genellikle tek bir nokta yerine tüm eğrinin tahmini kullanıldığı için x’in çok sayıdaki değeri için global bir ölçü olarak “Bütünleşik Hata Kareleri Ortalaması (BHKO), (Mean Integrated Squared Error =MISE)” kullanılabilir. (Silverman, 1986: 35)

Optimal bant genişliği esasen regresyon fonksiyonunun kendisine ve artıkların varyansına bağlıdır ve veriden elde edilmesi tercih edilen bir değerdir ve Bütünleşik Hata Kareleri Ortalaması (MISE) değerini minimize eden bant genişliği olarak ifade edilmektedir (Brockmann v.d., 1993: 1302-1309).

Band genişliği h sabittir, fakat bununla birlikte bant genişliği kolaylıkla en yakın komşuluk tipi band genişliğine uyarlanabilir. Uyarlama kernel fonksiyonları için basittir. Pencere içine sabit sayıda gözlemin (m) düşmesi için h değeri ayarlanabilir. n toplam gözlem sayısı olmak üzere; m/n oranı, Kernel düzgünleştiricisinin “Sabit Örnek Büyüklüğü (Span)” olarak adlandırılır (Fox, 2000a: 7).

Parametrik olmayan regresyon analizinde düzgünleştirme parametresinin belirlenmesinde “Deneme-Yanılma” ya da “Çapraz Geçerlilik (Cross-Validation-CV)” yöntemi en çok kullanılan yöntemlerdir. Çapraz geçerlilik yönteminin temelinde, x_i odak noktasında hesaplanan lokal regresyondan i. gözlemin ihmal edilmesi düşüncesi yatmaktadır. Aynı mantık span için de geçerlidir. Çapraz geçerlilik fonksiyonu

$$CV(s) = \frac{\sum_{i=1}^{n}\left[\hat{y}_{-i}(s) - y_i\right]^2}{n}$$ olmak üzere

Burada amaç; span değeri s için $\hat{y}_{-i} \mid x_i$ 'in tahmini $\hat{y}_{-i}(s)$ olmak üzere, CV fonksiyonunu minimize edecek s değerini bulmaktır. (Fox, 2000a:22)

IV.Uygulama

A. Uygulamanın Amacı ve Kapsamı

İşletmelerin borsa performanslarının bir göstergesi olan Piyasa Değeri/ Defter Değeri (PD/DD) oranı ile çeşitli finansal oranlar arasındaki ilişki parametrik olmayan regresyon analizi ile incelenmiştir. Bu amaçla İstanbul Menkul Kıymetler Borsası (İMKB) İmalat Sanayiinde işlem gören firmaların 2007 yılına ait 12 aylık finansal tabloları kullanılmıştır.** Analizdeki en önemli kısıt, işletme sayısının azlığıdır. Kayıp değerlerden dolayı bazı işletmeler analiz dışı bırakıldığı için gözlem sayısı 150'nin altına düşmüştür.

B. Uygulamada Kullanılan Değişkenler

İşletme performansının belirlenmesinde çeşitli karlılık, kaldıraç, verimlilik, likidite ve büyüme oranları kullanılabilir. Parametrik olmayan regresyon analizinde bağımlı değişken olarak oranı PD/DD oranı, bağımsız değişken olarak borsa performansının belirlenmesinde etkili olacağı düşünülen çeşitli finansal oranlar seçilmiştir. Seçilen oranlar aşağıda belirtilmiştir.

Bağımsız Değişkenler:

1) Cari Oran
2) Toplam Yabancı Kaynaklar / Özsermaye
3) Net Kar / Satışlar
4) Net Kar / Özsermaye
5) (Faiz ve Vergiden Önceki Kar + Ödenen Faiz) / Toplam Aktif
6) Aktif Büyüme Hızı
7) Kar Payı Dağıtım Oranı

C. Parametrik Olmayan Regresyonda Kullanılan Yöntem ve Uyum İyiliğinin (Goodness of Fit) Ölçülmesi

İşletmelerin borsa performansını finansal oranlar ile açıklamak amacı ile S-PLUS ve R programından yararlanılmıştır. Tahmin yöntemi olarak LOESS kullanılmıştır. LOESS yöntemi en yakın komşuluk tipi band genişliği diğer bir ifade ile sabit örnek büyüklüğü olan span'ı kullandığı için, bu oranın ne olması gerektiği önemli bir konudur. (0,1) aralığında değer alabilen span değerinin belirlenmesinde deneme-yanılma yöntemi kullanılmıştır. LOESS yönteminde span değeri olarak 0,5 civarında oranlar önerilmektedir (Cleveland, 1979: 834 - Fox, 2000a: 23 - Neter v.d.,1996:426). Bu nedenden dolayı analizler 0.4-0.5 aralığındaki span değerleri denenerek yapılmıştır.

Parametrik regresyon analizinde; "Belirlilik katsayısı (Coefficient of Determination)" bağımlı değişkendeki değişkenliğin azaltılmasında , diğer bir ifade ile bağımlı değişkenin tahmin edilmesi için belirsizliğin azaltılmasında, bağımsız değişken ya da değişkenlerin ne ölçüde etkili olduğunun bir göstergesidir (Neter v.d.,1996:81). Parametrik olmayan regresyon analizi için de belirlilik katsayısı hesaplamak mümkündür. Birimden bağımsız olarak hesaplanan bu katsayının formülü aşağıda verilmiştir (Hayfield ve Racine, 2008: 10).

y_i ; Bağımlı değişkenin gerçek değerini

$\hat{y}_i$; Bağımlı değişkenin parametrik olmayan regresyon ile elde edilen tahmini değerini

$\bar{y}$; Bağımlı değişkenin ortalamasını

$i = 1,......n$; Gözlem sayısını göstermek üzere

$$R^2 = \frac{\left[\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{y})\right]^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}$$

Hesaplanan R^2 her zaman $[0,1]$ aralığında değer almaktadır. 1 değeri fonksiyonunun örnek verisine mükemmel bir uyum sağladığını, 0 değeri ise; herhangi bir tahmin gücü olmadığını göstermektedir. Elde edilen R^2, parametrik regresyonda; fonksiyon doğrusal olduğu zaman en küçük kareler yöntemi ile elde edilen R^2 ile eştir (Racine, 2008:47).

Parametrik olmayan regresyonda bağımsız değişkenlerin modele olan katkısı artan F testi (incremental F test) ile araştırılabilir. Bu testte alternatif modellerin hata kareleri toplamı ile serbestlik dereceleri karşılaştırılmaktadır (Fox, 2000b:19).

$$F = \frac{(RSS_0 - RSS_1)/(df_1 - df_0)}{RSS_1 / df_{res}}$$

RSS_1 =Tam modelin hata kareleri toplamı

df_1 = Tam modelin serbestlik derecesi

RSS_0=İhmal edilen j. bağımsız değişken dışındaki modelin hata kareleri toplamı

df_0= İhmal edilen j. bağımsız değişken dışındaki modelin serbestlik derecesi

$df_{res} = n - df_1$

D. Parametrik Olmayan Regresyon Analizi Sonuçları

İMKB İmalat Sanayinde 157 adet hisse senedi işlem görmesine rağmen, bu sayı parametrik olmayan regresyonun çok sayıda veriye ihtiyaç duymasından dolayı ikiden fazla bağımsız değişkenin modelde yer almasına imkan vermemiştir. PD/DD ile bağımsız değişkenler arasında Toplam Borç/Özsermaye (Kaldıraç) ile Aktif Karlılığı(Karlılık) değişkenleri arasında istatistiksel olarak ilişki bulunmuştur.

Parametrik olmayan regresyon analizine göre elde edilen R^2 değeri %21'dir. Parametrik regresyon analizinde elde edilen R^2 değeri ise %16,75'tir.

Tablo 1: *F Testi Sonuçları*

	Hata Kareleri Toplamı	Serbestlik Derecesi	F-Değeri	p
Kaldıraç Değişkeninin Yer Aldığı Model	300,291	5,090	1,845	0,09047*
Tam Model	272,374	8,950		
Karlılık Değişkeninin Yer Aldığı Model	329,32		3,97	0,001084**
Tam Model	272,374	8,950		

*0,10 **0,01 düzeyinde istatistiksel olarak anlamlıdır.

V. Sonuç

Parametrik olmayan regresyon analizinin uygulanması amacı ile İstanbul Menkul Kıymetler Borsası (İMKB) İmalat sanayinde işlem gören işletmelerin borsa performansı (piyasa değeri/defter değeri) ile uygun finansal oranlar arasındaki ilişki araştırılmıştır. Kurulan modele göre; kaldıraç ile karlılık oranlarının modelde anlamlı ve en yüksek açıklayıcılığı sahip olduğu görülmüştür. Elde edilen belirlilik katsayısının (%21) yüksek olmaması modelde yer alması gereken başka değişkenlerin de olduğunu göstermektedir. Aynı veri ve değişkenlerle parametrik regresyon analizi uygulandığında ise belirlilik katsayısı %16 olarak bulunmuştur. Buna göre;

parametrik olmayan regresyon ile elde sonuç, parametrik regresyon ile elde edilen sonuçtan daha iyidir.

Regresyon analizine farklı bir bakış açısı getiren parametrik olmayan regresyon, bazı dezavantajlara sahip olmasına rağmen, özellikle parametrik regresyonun uygulanamadığı veya uygulandığı fakat varsayımların sağlanamadığı durumlarda iyi bir alternatiftir. Ayrıca, parametrik regresyon ile kurulan modelin geçerliliği doğrulanmak isteniyorsa yine parametrik olmayan regresyon analizine başvurulabilir.

Notlar

*LOWESS adı basit regresyon (tek bağımsız değişken) uygulamalarında kullanılmaktadır. Bağımsız değişken sayısı arttığında aynı yöntemin adı LOESS olarak kullanılmaktadır.

**Hesaplanan tüm finansal oranlar İMKB'nin Çevrimiçi; http:// www.imkb.gov.tr adresli sayfasından alınmış veya hesaplanmıştır.

Kaynaklar

Cleveland, W. S. (1979); "Robust Locally Weighted Regression and Smoothing Scatterplots", Journal of the American Statistical Association , 74, pp 829

Eubank Randall L. (1990); Nonparametric Regression And Spline Smoothing, Second Edition, s.4

Dinardo John, Justin Tobias (2001); "Nonparametric Density and Regression Estimation", Journal of Economic Perspectives, Volume 15, Number 4, 11–28.

Fox John (2000a) Nonparametric Simple Regression: Smoothing Scatterplots, Sage University Paper, Series No: 07–130, s.22

Fox, John (2000b); Multiple and Generalized Nonparametric Regression, Sage University Paper, 131, s. 20

Galton Francis (1886); "Regression Towards Mediocrity in Hereditary Stature", Journal of Anthropological Institute of Great Britain and Ireland, Vol. 15, 246–263

Hastie T and R.Tibshirani (1990); Generalized Additive Models, Chapman-Hall, s.9

Hardle, Wolfgang (1990); Applied Nonparametric Regression, Cambridge University Press, s. 4-5

Hardle W, Müller M, Sperlich S, Werwatz A (2004); Nonparametric and Semiparametric Models: An Introduction, Springer Series in Statistics, s.115

Hardle, W, Linton Oliver (1994); Applied Nonparametric Methods, Cowles Foundation Discussion Paper 1069, s.4

Hayfield Tristen, Jeffrey Racine (2008); "Nonparametric Econometrics: The np Package"; Journal of Statistical Software, Volume 27, Issue 5, s.10

Michael Brockmann, Theo Gasser, Eva Hermann (1993); "Locally Adaptive Bandwidth Choice for Kernel Regression Estimators", Journal of the American Statistical Association, Vol 88, No.424, pp.1302–1309

Nadaraya E.A.(1964); "On estimating regression", Theory of Probability and Its Applications, 9, 141–142

Neter , Kutner, Nachtsheim,Wasserman, (1996); Applied Linear Statistical Models, Fourth Edition s.223

Orhunbilge Neyran (2000); Uygulamalı Regresyon ve Korelasyon Analizi, 2.Baskı, İstanbul, s.12

Peter Sprent, Nigel Smeeton (2007); Applied Nonparametric Statistical Methods, Fourth Edition, Chapman&Hall, s.475

Ruppert, Wand, Carroll (2003); Semiparametric Regression, Cambridge University Press, s.57

Racine Jeffrey (2008); "Nonparametric Econometrics: A Primer", Foundations and Trends in Econometrics, Vol. 3, No.1, s.47

Silverman B.W (1986); Density Estimation for Statistics and Data Analysis, Chapman and Hall, s.35

Yatchew Adonis (1998); "Nonparametric Regression Techniques in Economics", Journal of Economic Literature, 36, 2, s. 672

Watson G.S.(1964); "Smooth Regression Analysis", Sankhya Ser. A 26, 359–372.