Ekonometri ve İstatistik Sayı:10 2009 1-19

İSTANBUL ÜNİVERSİTESİ
İKTİSAT FAKÜLTESİ
EKONOMETRİ VE İSTATİSTİK
DERGİSİ

BOX-LJUNG ve NONPARAMETRİK REGRESYON YÖNTEMLERİNİN ETKİNLİKLERİNİN KARŞILAŞTIRILMASI: İMKB-100 ENDEKSİNE YÖNELİK BİR UYGULAMA

Yard.Doç.Dr. Namık Kemal ERDOĞAN *

Doç.Dr.Nevin UZGÖREN**

Abstract

It is a well known fact that time series and data are commonly and frequently used in applied researches and various methods are developed for analysis of these data. The most common method used for analysis of univariate time series is Box- Ljung method which is based on modelling of a time series with its own lagged values and error terms. Another recent method used for analysing the univariate time series is nonparametric regression method which is based on a certain function instead of coefficients and where the estimations are done through this function. The two methods share the same aim which is to model the time series and to forecast making use of this model.

This study aims at realizing a practical comparison of the efficiency of Box-Ljung method and nonparametric regression method, which are used for analysis of univariate time series, on the basis of monthly closing prices of ISE National Index 100. As a result of the analyses performed in line with this aim the comparisons performed as for various performance criteria revealed that the nonparametric regression method gives more effective results than Box-Ljung method.

Keywords: Parametric regression, nonparametric regression , arima , kernel regression
Jel Classification: C140, C530

Özet

Zaman serisi verilerinin, uygulamalı araştırmalarda çok sık ve yoğun bir şekilde kullanıldığı ve bu verilerin analizine yönelik çeşitli yöntemlerin geliştirildiği bilinmektedir. Tek değişkenli zaman serilerinin analizinde kullanılan en bilinen yöntem, bir zaman serisinin kendi gecikmeli değerleri ve hata terimleri ile modellemesine dayalı Box-Ljung (BL) yöntemidir. Tek değişkenli zaman serilerinin analizinde son dönemlerde kullanılan yöntemlerden biri ise, katsayılar yerine belirli bir fonksiyonu temel alan ve kestirimlerin bu fonksiyon üzerinden yapıldığı nonparametrik regresyon yöntemidir. Her iki yöntemin de ortak amacı zaman serilerini modellemek ve bu model yardımı ile öngörüde bulunmaktır.

Bu çalışmanın amacı, tek değişkenli zaman serilerinin analizinde kullanılan Box-Ljung (BL) yöntemi ile nonparametrik regresyon yöntemlerinin etkinliklerini, İMKB-100 Endeksinin aylık kapanış fiyatlarını temel almak suretiyle uygulamalı olarak karşılaştırmaktır. Bu doğrultuda yapılan analizler sonucunda, çeşitli performans kriterlerine göre yapılan karşılaştırmalarda nonparametrik regresyon yönteminin Box-Ljung (BL) yöntemine göre daha etkin sonuçlar verdiği görülmüştür.

Anahtar Kelimeler: regresyon, nonparametrik regresyon, arima, kernel regresyon
Jel Sınıflaması: regresyon, nonparametrik regresyon, arima, kernel regresyon

*** Adres:** *Anadolu Üniversitesi İ.İ.B.F. İşletme Bölümü*
E-Mail: *nkerdoga@anadolu.edu.tr*
**** Adres:** *Dumlupınar Üniversitesi İ.İ.B.F. İşletme Bölümü*
E-Mail: *nkerdoga@anadolu.edu.tr*

1.GİRİŞ

Zaman serileri stokastik bir sürecin gerçekleşmesidir. İstatistiksel veri analizi ise bu süreçlerin sonuçlarını açıklamaya yöneliktir. İstatistiksel veri analizinde yapılan temel işlem gözlenmiş zaman serilerini etkileyen olasılık yasalarını ortaya çıkarmaktır. Ortaya çıkarılan olasılık yasalarıyla zaman serilerini oluşturan dinamikler açıklanmaya çalışılır ve zaman serilerinin analiz edilmesi gelecekteki olaylar hakkında tahminde bulunabilmeyi, gelecekteki olayları kontrol edilebilmeyi ve değerlendirme yapabilme olanağı sağlar.

Zaman serisi verilerinin analizlerinde ARMA, ARIMA, ARCH vb. gibi parametrik modeller kullanılmaktadır. Son yıllarda ise kernel regresyon ve lokal ağırlıklı regresyon gibi nonparametrik regresyon modelleri de kullanılmaya başlamıştır.

Parametrik regresyon modellerinde belirli bir modelin regresyon katsayılarının hesaplanması önemli bir yere sahiptir. Hesaplanan katsayıların aldığı değerleri göre belli bir güven aralığında veya bir hipotez testine göre istatistiksel sonuçlar elde edilir. Parametrik regresyon yaklaşımlarında değişkenler arasındaki ortalama ilişki bir fonksiyon ile ifade edilmesi gerekir. Bu parametrelerin aldığı değerlere göre de istatistiksel çıkarımlarda bulunulur. Parametrik regresyonda eğer fonksiyonel ilişki verilere uygun ise etkili sonuçlar elde edilir.(Takezawa ,2006:20-21) .

Nonparametrik regresyonun ana amacı iki değişken arasındaki ilişkiyi açıklayan bir model sağlamak, belirli bir parametrik modeli referans almadan eldeki gözlemler için bir kestirimde bulunmaktır. Nonparametrik regresyonda regresyon fonksiyonunun süreklilik ve türevlenebilir gibi özellikleri diklate alınır(Eubank,1999:4-5). Ayrıca nonparametrik regresyon modellerinde katsayılar yoktur. İstatistiksel çıkarımlar direkt olarak regresyon fonksiyonu ile ilgilidir (Fox , 2008: 476-477).

Nonparametrik teknikler ilk olarak spectral yoğunluk hesaplarında kullanılmıştır. Daha sonraları bilgisayar teknolojilerindeki gelişmelerle birlikte öncellikle bağımsız gözlemler içeren veriler için regresyon analizinde kullanılmış ve daha sonra bağımlı gözlem içeren zaman serilerine genişletilmiştir (Härdle ve Chen, http://citeseer.ist.psu.edu/228910.html, erişim tarihi:19.03.2009).

Yapılan literatür taramasında, zaman serileri analizinde genellikle Box-Ljung yöntemlerinin kullanıldığı görülmektedir. Ancak son yıllarda nonparametrik regresyon yöntemleri de kullanılmaya başlanmıştır. Bu çalışmanın amacı da bu iki yöntemi İMKB endeks verilerini kullanarak bazı performans kriterlerine göre etkinliklerini karşılaştırmaktır. Aynı doğrultuda yapılan tek bir çalışmaya rastlanmıştır. Rodriguez N. ve Siado P., tarafından 2003 yılında yapılan çalışmada Kolombiya için enflasyon verilerine dayalı olarak arima , star ve nonparametrik regresyon yöntemlerinin karşılaştırması yapılmıştır. Aynı doğrultuda olmamakla beraber farklı yöntemlerin etkinliklerinin karşılaştırılmasına yönelik olarak da bazı çalışmalara ulaşılmıştır. Aydın D. , 2007 yılında yaptığı çalışmada nonparametrik regresyon yöntemlerinden olan kernel ve spline yöntemlerini milli gelir verilerine göre karşılaştırmıştır. Diğer bir çalışma ise, Topal M., Yıldız N. ve Bilgin Ö. C. nin yaptıkları çalışmadır. Bu çalışmada nonparametrik regresyon yöntemi olan Theil yöntemi ile parametrik en küçük kareler yöntemi karşılaştırılmıştır.

Görüldüğü üzere ülkemizde bu iki farklı yöntemin etkinliklerinin karşılaştırılmasına yönelik bir çalışma yapılmamıştır. Bu çalışmanın amacı da zaman serisi verilerinde belirli bir parametrik modeli referans almadan eldeki gözlemler için bir kestirim olanağı sağlayan nonparametrik regresyon yöntemlerden biri olan kernel regresyon ile Box-Lung yöntemiyle bir karşılaştırmasını yapılarak zaman serisi analizleri için bir seçenek verilmiştir. Ayrıca nonparametrik regresyon yöntemler aykırı gözlemlerin olduğu veri kümeleri için önemli bir analiz yöntemi olduğundan İMKB gibi aykırı verilerin olabileceği veri kümelerinde parametrik yöntem kullanılmak istendiğinde bir ön bilgi sağlayacaktır.

2. METODOLOJİ

Zaman serileri analizinin en önemli amacı, ele alınan zaman serisinin bugünkü ve geçmiş değerlerini kullanarak geleceğe dönük öngörülerde bulunmaktır. Bu amaçla geliştirilen çeşitli yöntemler mevcuttur. Bu yöntemlerden birisi Box-Ljung(BL) yöntemidir. Bu çalışmanın amacı, zaman serileri analizinde yaygın olarak kullanılan Box-Ljung yöntemi ile son zamanlarda kullanılmaya başlanılan nonparametrik regresyon yöntemlerinden biri olan kernel regresyon yönteminin etkinliklerini karşılaştırmaktır.

2.1 Box-Ljung Yöntemi

Zaman serilerinin analizinde kullanılan en bilinen yöntem BL yöntemidir. Tek değişkenli zaman serilerinin analizinde kullanılan BL yönteminin esası, zaman serilerinin herhangi bir dönemdeki değerini aynı serinin geçmiş dönemdeki gözlem değerlerinin ve hata terimlerinin doğrusal bir bileşimi ile açıklamaktır. Bu nedenle BL yöntemi literatürde otoregresif entegre hareketli ortalama yöntemi (ARIMA) olarak da bilinmektedir. Bu yöntemin temel varsayımları, ele alınan serinin kesikli ve durağan olmasıdır. Durağan olmayan zaman serilerine BL yönteminin uygulanabilmesi için önce durağanlığı bozan trend ve mevsimsellik gibi unsurların bazı dönüşüm yöntemleriyle ortadan kaldırılarak, serinin durağan hale getirilmesi gerekir (Özmen, 1986: 16-17).

Süreç durağan olduğunda bir otoregresif hareketli ortalama modeli genel olarak aşağıdaki gibi gösterilir ve ARMA(p,q) olarak tanımlanır:

$$Y_t = \xi + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad \textbf{(1)}$$

Burada p otoregresif terim sayısını, q hareketli ortalama terim sayısını, ϕ'ler otoregresif parametreleri ve θ'lar hareketli ortalama parametrelerini göstermektedir. Eğer seri durağan değilse, seriyi durağan hale getirmek için $\nabla^d Y_t = (1-B)^d\, Y_t$ fark alınarak seri durağan hale getirilir. d kaç kez fark alınması gerektiğini belirtmek üzere model ARIMA(p,d,q) olarak ifade edilir.

2.2 Nonparametrik Regresyon Yöntemi

$\{x_i , y_i\}_{i=1}^{n}$ gözlem değerleri olmak üzere nonparametrik regresyonun temel düşüncesi ham verilerin ağırlıklı ortalamasını kullanarak f regresyon fonksiyonunu tahmin etmektir. Söz konusu ağırlıklar x_i noktalarında oluşan $X-uzayındaki$ uzaklıkların azalan bir fonksiyonudur. x_i noktasındaki kestirim için y_j gözlemiyle ilişkili bu tür bir ağırlıklandırma şeması Nadarya (1964) ve Watson (1964) tarafından önerilmiştir:

$$w_{ij} = K\left(\frac{x_i - x_j}{h}\right) / \sum_{j=1}^{n} K\left(\frac{x_i - x_j}{h}\right) = \frac{K(u)}{\sum K(u)} \qquad \textbf{(2)}$$

Burada n gözlemlerin sayısı, K seçilen ve kernel olarak bilinen, sınırlı, sürekli ve integrali 1'e eşit $\left(\int_{-\infty}^{\infty} K(u)\, du = 1\right)$ olan simetrik bir fonksiyon olup, ağırlıkları hesaplamak için kullanır ve h değeri ise bant genişliği veya düzeltme parametresidir.

Uygulamada kullanılan farklı tipte kernel fonksiyonları vardır. Ancak kernel fonksiyonunun seçimi bant genişliğinin seçiminden daha az bir öneme sahiptir. Uygulamada kullanılan bazı kernel fonksiyonları aşağıda verilmiştir.Bu fonksiyonlar sıfıra göre simetrik, negatif olmayan değerler alır ve ikinci mertebeden türevlenebilirdir.(Fox, 2008: 477-478):

- Normal kernel : $K(u) = \left(\frac{1}{\sqrt{2\pi}}\right) e^{(-\frac{1}{2}u^2)}$, $u \in [-\infty, \infty]$,
- Düzgün kernel : $K(u) = \frac{1}{2}$, $u \in [-1,1]$
- Üçgensel kernel : $K(u) = (1-|u|)$, $u \in [-1,1]$
- Epanechnikov kernel: $\frac{3}{4}(1-u^2)$, $u \in [-1,1]$

- Dördüncü dereceden kernel (Quartic): $\frac{3}{4}(1-u^2)^2$, $u \in [-1,1]$

Herhangi bir x_i noktasındaki f regresyon fonksiyonunun *kernel tahmini*,

$$\hat{y}_i = \hat{f}(x_i) = \sum_{j=1}^{n} w_{ij}\, y_j = w_i^{'}\, \boldsymbol{y} \qquad \textbf{(3)}$$

şeklinde tanımlanır. Burada $\mathbf{y} = (y_1, ..., y_n)'$ olarak ifade edilir. (3) denklemi matris formunda yeniden yazılarak aşağıdaki biçimde de ifade edilebilir:

$$\hat{\boldsymbol{f}} = \boldsymbol{W}\ \boldsymbol{y} \qquad \textbf{(4)}$$

i .gözlemin ağırlıkları, $\mathbf{w}_i' = (w_{i1}, ..., w_{in})$ olarak ifade edilir. $\boldsymbol{W}$ matrisine *kernel şapka matrisi* veya *kernel düzeltme matrisi* denir. Buna göre keyfi bir x_1 noktasındaki kernel kestirimleri, i ile 1 yer değiştirilerek (3) denkleminden elde edilebilir:

$$\hat{f}(x_1) = \mathbf{w}_1'\mathbf{y} = (w_{11}, ..., w_{1n})(y_1, ..., y_n)^T \qquad \textbf{(5)}$$

Böylece bağımlı y değişkeninin tahmini değerler vektörü veya bilinmeyen f fonksiyonun aldığı değerler vektörü, **(5)**'deki formül kullanılarak aşağıdaki şekilde elde edilir (Aydın ,2007:730-731).

$$\hat{\boldsymbol{f}} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}_{\text{nx1}} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}_{\text{nxn}} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{\text{nx1}} \qquad \textbf{(6)}$$

2.2.1 Çapraz geçerlilik (Cross-validation): Kernel fonksiyonlarında h band seçimi oldukça önemlidir.Band genişliği h değerinin nasıl seçileceği ilişkin bir çok araştırma yapılmasına rağmen genel olarak çapraz geçerlilik fonksiyonunu minimum yapan h değeri band genişliği olarak alınır.

$\{x_i \ , y_i\}_{i=1}^{n}$ gözlem değerlerinden herhangi birini atarak kalan (n−1) gözlem değerine dayanarak x_i noktasında düzgün (ikinci mertebeden türevlenebilir) bir fonksiyon için kareli artıkları tahmin etmeye ve kareli artıkların toplamını minimum yapan düzeltme parametresi olarak tanımlanır ve çapraz geçerlilik fonksiyonu ;(Faraway, 2006 : 214-216)

$$\mathbf{CV}(\lambda) = \frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{f}_{\lambda(j)}(x_i))^2 \qquad \textbf{(7)}$$

şeklindedir. Burada (j) değeri j. noktadaki tahmin değerinin çıkarıldığı göstermektedir.

2.2.2 Zaman serilerinde kernel regresyon

Regresyon problemlerinin istatistiksel özelliği bağımsız gözlemler arasındaki ilişkiyi analizi etmektir. $\{x_i \ , y_i\}_{i=1}^{n}$ gözlem değerlerinin i.i.d. varsayımına göre elde edildiğidir. Parametrik ve nonparametrik regresyon tekniklerinde de gözlemler arasındaki bağımlılığın kaldırılması gerekir. Nonparametrik teknikler için literatürde $\alpha - mixing$ ve $\phi - mixing$ koşullarına göre bağımlığın kaldırılması ifade edilmiştir (Heiler, http://129.3.20.41/eps/fin/papers/9904/9904005.pdf).

Nonparanmetrik regresyon kestiricileri bağımlı gözlemler üzerinde kullanıldığında, kestiriciler tüm veri setindeki gözlemlerin bağımlılığı tarafından etkilenir. Bu nedenle gözlemler arasındaki bağımlılığın ortadan kaldırılması gerekir. Nonparametrik regresyon kestiricileri bağımlı gözlemler üzerinde kullanıldığında söz konusu kestiriciler tüm veri setinde değil sadece küçük bir aralıktaki gözlemlerin bağımlılığından etkilenir. Küçük bir aralıktaki gözlemler arasındaki bağımlılık uzun dönemde hatırlanma kabiliyetinden yoksun olacağından bu aralıktaki (periyottaki) gözlemler hemen hemen bağımsız olur. Bu özelliğe

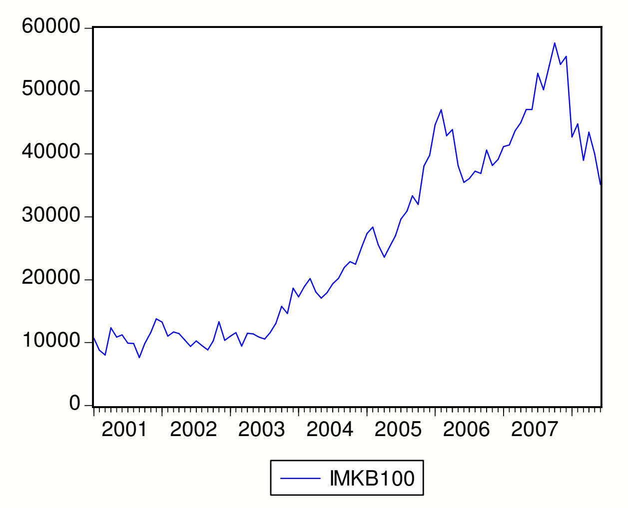
"*Whitening by windowing principle*" denir. Buna göre bağımsız gözlemler için geliştirilen tekniklerin çoğu bağımlı gözlem içeren zaman serilerinde de kullanılabilir (Hart ,1996 :115-117).

3. Uygulama

Bu çalışmanın amacı, İMKB-100 Endeksinin aylık kapanış fiyatlarının Box-Ljung yöntemi ve nonparametrik regresyon yöntemlerinden Kernel Regresyon ile modellemesini yaparak, iki yöntemin etkinliğini belirli performans kriterlerine göre karşılaştırmaktır. Bu amaçla 2001 yılı Ocak ayı ile 2008 yılı Haziran dönemini kapsayan 90 aylık gözlem değerlerinden yararlanılmıştır. İlgili analizlerde **SPSS-13.0, Eviews- 5.0** ve **R** paket programları kullanılmıştır.

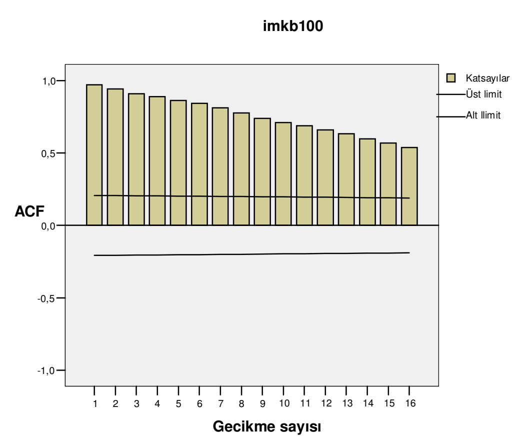
3.1. Box-Ljung yöntemi ile model tahmini

Verilerin zamana göre grafiği çizildiğinde (Şekil 1) İMKB-100 serisinin artan bir eğilime sahip olduğu ve aynı zamanda birbirini izleyen yılların aynı aylarında benzer davranışların var olduğu görülmektedir. Bu durum seride hem trendin hem de mevsimsel dalgalanmaların olabileceğini göstermektedir.



Şekil 1. İMKB-100 serisinin grafiği

Ancak serinin durağan olup olmadığı otokorelasyon fonksiyonu yardımıyla daha güvenilir bir şekilde belirlenebilir. İMKB-100 serisinin orijinal değerlerine göre ACF grafiği Şekil 2'de verilmiştir.



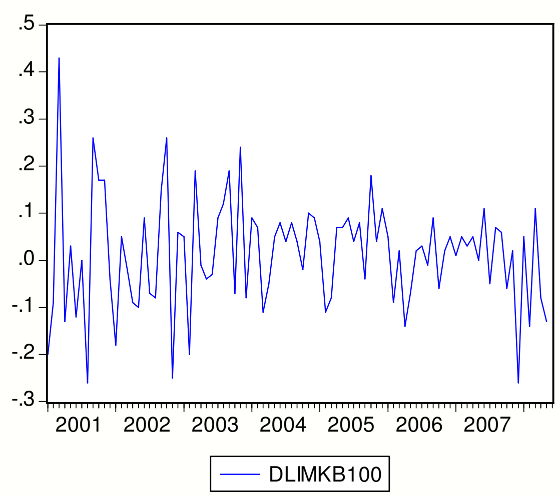
Şekil 2. İMKB-100 serisinin otokorelasyon fonksiyon grafiği (ACF)

Şekil 2 incelendiğinde 16 gecikme için tüm otokorelasyon katsayılarının anlamlı olduğu ve serinin durağan olmadığı görülmektedir. Dolayısıyla ilk olarak seriyi durağan hale getirmek ve daha sonra uygun ARIMA modeline karar vermek gerekir. Bu amaçla logaritmik zaman serisinin birinci dereceden diferansiyeli alınarak seri durağan hale getirilmiştir. Şekil 3 incelendiğinde serinin artık trend unsuru içermediği görülmektedir.

$Y_{t=}$İMKB100

LİMKB100= $\ln(Y_t)$

DLİMKB100= $\ln(Y_t) - \ln(Y_{t-1})$



Şekil 3. DLİMKB100 serisinin grafiği

Çalışmanın bu aşamasında birçok model denemesi yapılmış olup, en uygun modelin belirlenmesinde parametrelerin anlamlılık düzeylerinin yanı sıra uyum iyiliği testlerinden Akaike bilgi kriteri (AIC) ve Schwartz Bayesian bilgi(BIC) kriterinden yararlanılmıştır**.** Bu analizler sonucunda geçici uygun model olarak mevsimsel ARIMA $(1,1,0)(1,1,0)_{12}$ modeli benimsenmiştir. Serin alınan birinci diferansiyeli trendi, alınan on ikinci diferansiyel ise aylık mevsimselliği elimine etmiştir. Tahmin sonuçları Tablo 1'de ve uyum iyiliği göstergeleri ise Tablo 2'de gösterilmiştir.

Tablo 1. Mevsimsel ARIMA (1,1,0)(1,1,0)$_{12}$ modeli tahmin sonuçları

Parametre Tahmini

	Tahminler	Standart Hata	t	Anlamlılık
Mevsimsel olmayan gecikmeler AR1	-0,276	0,111	-2,497	0,015
Mevsimsel gecikmeler Mevsimsel AR1	-0,524	0,100	-5,250	0,000

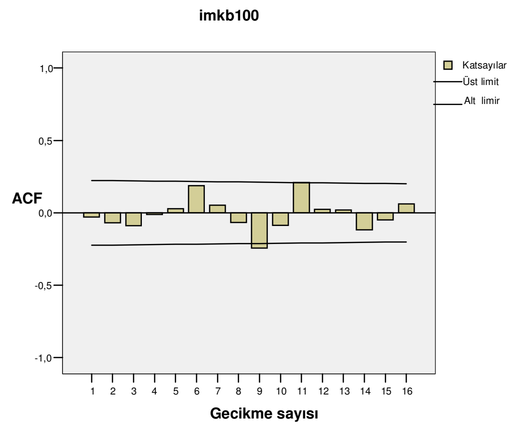
Tablo 2. Mevsimsel ARIMA (1,1,0)(1,1,0)$_{12}$ modeli uyum iyiliği göstergeleri

Artıklar Göstergeleri

Artık sayısı	77
Parametre sayısı	2
Artıklar serbestlik derecesi	75
Düzeltilmiş artık kareler Toplamı	1,331
Artık kareler toplamı	1,348
Artık varyansı	0,017
Modelin standart hatası	0,130
Log-Likelihood	46,962
AIC bilgi kriteri	-89,923
BIC bilgi kriteri	-85,235

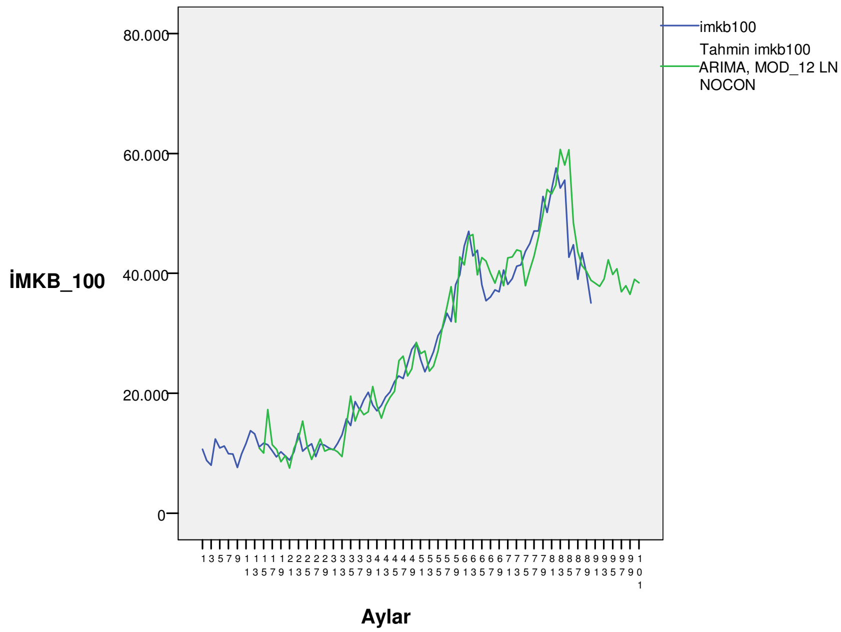
Sabit terim anlamlı bulunmadığından, modelde yer verilmemiştir. Tahmin sonuçları incelendiğinde mevsimsel ve mevsimsel olmayan birinci dereceden otoregresif parametrelerin sırasıyla 0,05 düzeyinde istatistiksel açıdan anlamlı olduğu görülmektedir. Elde edilen bu sonuçlar, uyum iyiliği göstergeleri ile birlikte değerlendirildiğinde benimsenen geçici modelin uygun model olabileceğini göstermektedir. Ancak kesin bir sonuca ulaşabilmek için benimsenen modelin tahmin hatalarının incelenmesi gerekir.

Bu aşamada tahmin edilen geçici mevsimsel ARIMA (1,1,0)(1,1,0)$_{12}$ modelinin verilere uygunluğunun testi yapılmıştır. Bir ARIMA modelinin uygunluğu, tahmin hataları serisinin otokorelasyon katsayılarının analiz edilmesiyle yapılır. Bu amaçla ilk olarak modele ilişkin tahmin hataları bulunmuş ve daha sonra tahmin hatalarının ACF (otokorelasyon fonksiyonu) grafiği (Şekil4) çizdirilmiştir.



Şekil 4. Tahmin hatalarının otokorelasyon fonksiyonu

Şekil 4 incelendiğinde , otokorelasyon katsayılarının 9. gecikme haricinde standart hata limitleri içinde ($\pm 2/\sqrt{77} = \pm 0{,}23$, n=90-13=77) kaldığı ve dolayısıyla istatistiksel açıdan anlamlı olmadıkları görülmektedir. Diğer bir ifadeyle, hatalar rassaldır ve herhangi bir zaman unsuru göstermemektedir. Bu nedenle seçilen mevsimsel ARIMA $(1,1,0)(1,1,0)_{12}$ modelinin İMKB-100 serisi için uygun bir model olduğu kabul edilir. Ayrıca seçilen modelin uygunluğu Box-Ljung istatistiğine göre de incelenmiştir. 16. gecikmeye kadar hesaplanan BL istatistiklerinin herbirinin istatistiklerinin istatistiksel açıdan anlamlı olmadığı görülmüştür. Dolayısıyla otokorelasyon katsayılarının eşanlı olarak sıfır olduğunu ileri süren sıfır hipotezini reddetmek istatistiksel açıdan mümkün değildir. Böylece bir kez daha tahmin hatalarının rassal olarak dağıldığı ve modelin verilere uygun olduğuna karar verilmiştir. Son olarak gerçekleşen ve tahmin edilen değerlerin aynı grafik üzerinde çizimine yer verilmiştir (Şekil 5).



Şekil 5. Gerçekleşen ve tahmin serilerinin grafiği

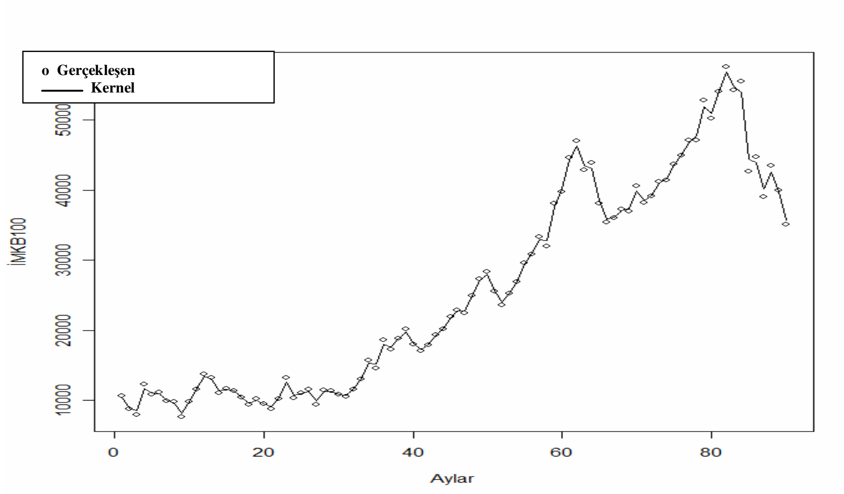
3.2. Nonparametrik regresyon yöntemi ile tahmin

Bu çalışmadaki veri seti için nonparametrik regresyon modellerinden biri olan kernel regresyon uygulanmıştır. Kernel regresyon fonksiyonlarından normal kernel adı verilen fonksiyon tercih edilmiştir. Band genişliği h değeri çapraz geçerlilik fonksiyonu yardımıyla çapraz geçerlilik fonksiyonunu minimum yapan h band genişliği değeri 1,38 olarak tespit edilmiştir. Buna göre şekil 6 da tahmin sonuçlarının grafiği verilmiştir. Grafikten görüleceği üzere band genişliğinin 1,38 olması durumunda gerçekleşen değerler ile kernel tahmin değerleri birbirine oldukça yakındır.

Verileri temsil eden değişkenler aşağıdaki gibi tanımlanmıştır.

İMKB100: Ocak 2001-Haziran 2008 arasında aylık kapanış değerleri

Aylar: Söz konusu yıllara ilişkin aylar.



Şekil 6 : Kernel regresyona göre gerçekleşen ve kernel tahmin değerlerinin grafiği

4. Yöntemlerin Karşılaştırılması

Modellerinin performanslarını karşılaştırmak için uygulamada hata kareler ortalaması (MSE) , onun karekökü (RMSE) , ortalama mutlak hata (MAE) , ortalama mutlak hata yüzdesi (MAPE) gibi kriterler kullanılmaktadır. Bu kriterlere göre örneğin hata kareler ortalaması (MSE) değeri ne kadar küçükse gözlenen değer ile beklenen değer arasındaki sapma da küçük olur. Böylece gözlem değerleri ve beklenen değer arasındaki sapma küçüldükçe modelin gerçek uyumunun daha iyi olduğu belirtilir. İlgili performans kriterlerine ilişkin formüller aşağıdaki gibidir:

$$MSE = \frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2 \quad ,$$

$$RMSE = \sqrt{MSE} \ ,$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | \ y_t - \hat{y}_t | \ \ ,$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{| \ y_t - \hat{y}_t |}{|y_t|}$$

gibi kriterler kullanılır. Burada y_t, gerçekleşen gözlem değeri , $\hat{y}_t$ ise öngörülen değerdir (Fan & Yao, 2005:244-245).

Kernel regresyonda analizler orijinal gözlem değerleri (Yt=İMKB100) üzerinden yapılırken, BL yönteminde ise, logaritmik gözlem değerlerine [$\ln Y_t$=ln(İMKB100)] dayalı olarak yapılmıştır. İki yöntemin etkinliklerinin karşılaştırılabilmesi için bağımlı değişkenin aynı olması gerekir. Bu nedenle BL yöntemine ilişkin Tablo 2'de verilen uyum iyiliği göstergeleri kullanılmayıp, performans kriterleri uygun dönüşümler ile yeniden hesaplanmıştır.

Performans kriterleri incelendiğinde Kernel regresyon ile elde edilen sonuçların **ARIMA (1,1,0** $(1,1,0)_{12}$ modeline göre daha küçük değer aldığı görülmektedir. Dolayısıyla aynı verilerin modellenmesinde kernel regresyonun daha etkin bir tahmin yöntemi olduğu görülmektedir.

Tablo 3: İMKB_100 için modellerin performans değerleri

Performans kriterleri	ARIMA (1,1,0 $(1,1,0)_{12}$	Kernel regresyon
MSE	13034605,6	228782,5
RMSE	3610,3470	478,3121
MAE	2712,64	347,9461
MAPE	0,1071 (%10,71)	0,01559 (%1,559)

5. Sonuç

Bu çalışmada zaman serisi analizlerinde sıkça kullanılan parametrik BL yöntemi ile nonparametrik kernel regresyon yöntemi İMKB aylık endeks verileri kullanılarak karşılaştırılmıştır. MSE, RMSE, MAE ve MAPE performans kriterlerine göre yapılan karşılaştırma da kernel regresyonun BL parametrik yöntemine göre daha iyi sonuç verdiği görülmüştür. Bu alanda yapılan tek çalışma olan Rodrigez N. ve Siado P.' de benzer bulgulara ulaşmıştır. Bu çalışmada Kolombiya için enflasyon verileri kullanılarak arima, star ve kernel regresyon yöntemleri karşılaştırılmıştır ve kernel regresyonun diğer iki yönteme göre daha etkin olduğu sonucu elde edilmiştir.

Ulaşılan bulgular çerçevesinde İMKB verilerini kullanarak Box-Ljung yöntemine göre analiz yapan bilgi kullanıcıları, aynı zamanda kernel regresyon yöntemini de kullanarak önerdikleri modelin geçerliliğini test edebilirler. Böylece İMKB verileri kullanılarak yapılacak öngörülerin daha doğru ve güvenilir olması sağlanabilir.

Kaynaklar

Aydın,D. “A comparision of the nonparametric regression models using smoothing and kernel regression “, Proceeding of world academy of science, engineering and technology ,Volume 26 December 2007.

Eubank, R. L., (1999). Nonparametric Regression and Smoothing Spline, Marcel Dekker Inc., NewYork.

Faraway J.J.(2006) , Extending the linear model with R, Chapman and Hall /CRC

Fazekas M.,“Application of ARIMA Models”,

http://www.telmae.karlov.mff.cuni.cz/.../articles.nsf/89B69C1700CD9453C1256C7F0056D248/$FILE/Fazekas.PDF?OpenElement**, (**09.03.2007).

Fox J.(2008) , Applied Regression Analysis and Generalized Linear Models , Sage Publications Inc.

Härdle , W. (1990), Applied Nonparametric Regression,Cambridge University Press.

Hart J.D.(1996) , Some automated methods of smoothing time-dependent data , Journal of nonparametric statistics, 6 , 115-142

Härdle , W. ve Chen , R. Nonparametric Time Series Analysis, a selevtive review

with examples , http://citeseer.ist.psu.edu/228910.html, (19.03.2009).

Heiler S, A survey on nonparametric time series analysis , http://129.3.20.41/eps/fin/papers/9904/9904005.pdf **,** (11.02.2008).

Nadarya, E.A. (1964), *On Estimating Regression,* Theory Pb. Appl., Vol.10.

Özmen, A. (1986), Zaman Serisi Analizinde Box-Jenkins Yöntemi ve Banka Mevduat Tahmininde Uygulama Denemesi, Anadolu üniversitesi Yayınları No. 201, Eskişehir.

Rodriguez N.N. and Siado C.P. (2003),”Un pronostico no parametrico de la inflacion Colombiana”,Revista Colombiana de estadistia , diciembre vol 26 , 89-128

Sevüktekin M. ve Nargeleçekenler M. (2007) , Ekonometrik Zaman Serileri Analizi ,Nobel Yayınevi, 2007.

Shumway ,R.H., Stoffer D. S.(2006) , *Time Series Analysis and Its applications with R Examples ,* Springer Texts in Statistics.

Teknomo , K, *Numerical Microsoft Excel Tutorials* ,

http://people.revoledu.com/kardi/tutorial/Regression/WhatIsRegression.html **,** (11.02.2008)

Takezawa K.(2006) , Introduction to Nonparametric Regression, John Wiley and Sons Inc.

Topal M., Yıldız N. ve Bilgin Ö.C. ,Farklı Dağılış Gösteren Verilerde Parametrik ve Nonparametrik Regresyon Metotlarının İncelenmesi, 4uzbk.sdu.edu.tr/4UZBK/BGB/4UZBK_086.pdf , (09.12.2009)

Watson,G.S.(1964), Smooth regression analysis,Sankhya,Series A,26 .