



DÜZGÜNLEŞTİRİLMİŞ FONKSİYONEL ANA BİLEŞENLER ANALİZİ İLE İMKB VERİLERİNİN İNCELENMESİ

Doç.Dr. Kadir ERTAŞ *

Araş.Gör.Dr. İstem Köymen KESER **

Abstract

In most situations, modern technological developments give rise to the cases where samples are drawn from a population of real random functions. Functional Data Analysis (FDA) is an appropriate multivariate statistical approximation since the classical multivariate methods can not be used when a random sample consists of such n-real functions. Generally the functions are sampled discretely in time and a certain smoothing technique is used to obtain underlying functions.

In this study we first give a detailed theory of B-Splines and then obtain cubic splines as linear combinations based on the coefficients resulted from an implementation of the Roughness Penalty Method. We then present a comprehensive theoretical background of the functional data analysis with a special attention given to the functional and regularized functional principal components concepts that are very useful to explore and interpret the variability of the functions and also their derivatives especially when one has a large number of functions.

Finally, an application of the regularized functional principal components on the weekly closing share prices data of the thirteen companies belonging to the ISE-100 index is presented. Interpretations of the derivative functions, covariance surface and principal component functions are also given in detail.

Keywords:: Functional Data Analysis, Functional Principal Component Analysis, Regularized Functional Principal Component Analysis, Smoothing, Roughness Penalty Approach, Cubic Spline, Eigenvalue-Eigenfunction Decomposition, Dimension Reduction.

Jel Classification: C130,C190

Özet

Fonksiyonel Veri Analizi Yöntemleri reel bir aralığın sonlu bir alt setinde değerlendirilen eğrilerden veya gözlenen fonksiyonlardan oluşan verileri analiz etmek üzere geliştirilmiştir. Fonksiyonel Veri Analizindeki teknikler, $x_i(t)$ ($i = 1, 2, \dots, N$) şeklinde belirtilen fonksiyonlardan veya onların türevlerinden oluşan rastgele örneklerdeki değişimin (varyasyonun) incelenmesi ve araştırılması amacıyla kullanılabilir. Pratikte bu fonksiyonlar sıklıkla ayrı noktalar da gözlenen verilere uygulanan düzgünleştirme (smoothing) süreçlerinin bir sonucu olarak ortaya çıkarlar. Bu çalışmada da Splayn Düzgünleştirme Yöntemleri bu amaçla ele alınmıştır.

Bu araştırmanın amacı, ayrı noktalar da gözlenen verileri öncelikle B-Splayn Baz Fonksiyonlar ve Pürüzlü Ceza Yaklaşımı kullanarak bir diğer deyişle bu iki yaklaşımın birlikte kullanılması olarak adlandırılan Splayn Düzgünleştirme Yöntemi ile sürekli ve türevlenebilir fonksiyonlar haline dönüştürülmesinin incelenmesidir. Daha sonra da veriler arasındaki yani ilgilenilen zaman aralığında hisse senetlerinin bireysel fonksiyonları arasındaki değişkenlik yapısını ortaya koymak üzere Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinden faydalanılmıştır. Burada ilgilenilen birey sayısı değişken sayısından az olduğundan dolayı klasik yöntemler zaten bu amaç doğrultusunda yetersiz kalmaktadır.

Bu çalışmada, IMKB-100 endeksinde yer alan şirketlerin haftalık hisse senedi kapanış fiyatlarından oluşan bir örnek üzerinde yapılan uygulamaya yer verilmektedir. Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi ile incelenen 13 şirket için özellikle 2000 yılının başlarında ve 360 ıncı günden bir diğer deyişle 2005 yılından itibaren fiyatların değişkenliğinde bir artış olduğu ve zaman noktalarının ardışık olarak birbirleriyle pozitif bir korelasyona sahip olduğu ulaşılan önemli sonuçlardan bir tanesidir.

Geleneksel Ana Bileşenler yönteminin uygulanmasının mümkün olmadığı durumlarda bile uygulanabilen ve sistemdeki gürültü (noise) etkisini de kaldıran bir yöntem olan Düzgünleştirilmiş Ana Bileşenler Analizi sonucu elde edilen harmonikler sayesinde hem verilerin kovaryans yüzeyiyle açıklanamayan değişkenlik yapısı ortaya konulmuş ve hem de genel anlamda Fonksiyonel Veri Analizi ile örneğin türev fonksiyonlarının da incelenebilmesi gibi görsel olarak da kuvvetli bulgular sunulmuştur.

Anahtar Kelimeler: Fonksiyonel Veri Analizi, Fonksiyonel Ana Bileşenler Analizi, Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi, Düzgünleştirme, Pürüzlü Ceza Yöntemi, Kübik Splayn, Özdeğer – Özfonksiyon Ayrışımı, Boyut İndirgeme.

Jel Sınıflaması: C130,C190

* **Adres:** Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi

E-Mail: kadir.ertas@deu.edu.tr

** **Adres:** Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi

E-Mail: istem.koymen@deu.edu.tr

1. Giriş

Çoğu istatistiksel analiz, rastgele bir örnekteki bireylerin her biri üzerinde bir veya daha fazla değişken bakımından yapılan ölçümleri içerir. Burada amaç, örneğin alındığı popülasyonla ilgili yorumlamalarda bulunmaktır. Çalışma sahası genişledikçe bu gözlemler, eğriler (curves) veya yüzeyler (surfaces) halini alır. Bir düzlemin herhangi bir alt setinde veya bir doğru üzerindeki her bir noktada gözlenen bir yoğunluk mevcut olduğundan, sözkonusu eğriler veya yüzeyler “fonksiyonel veri” olarak adlandırılırlar. Çoğu kez gözlemler zamanın, fakat bazı durumlarda da ilişkili başka bir değişkenin fonksiyonudurlar. Fonksiyonel verilerin alındığı popülasyonlar hakkında yorumlar yapabilmek amacı ile bilgisayar teknolojisindeki gelişmeler, fonksiyonel alanda da, istatistiksel analizin gelişimini ortaya çıkarmıştır. Bu tip verileri analiz etmek için geliştirilen yöntemler, ilk olarak, Ramsay ve Dalzell (1991) tarafından kullanılan “Fonksiyonel Veri Analizi” (Functional Data Analysis) terimi ile adlandırılır. Bu çalışma kapsamında “fonksiyonel veri” olarak tek boyutlu durum olan eğrisel veriler dikkate alınmaktadır.

Fonksiyonel veri analizinde temel felsefe olarak, elde edilen veriler bir sayı dizisi olarak dikkate alınmamakta, bunun yerine bu sayıların oluşturduğu fonksiyonun bütünü incelenmektedir.

Fonksiyonel veri analizinin tarihsel gelişimi içerisinde oluşmuş kendisine özgü geleneksel notasyon yapısı şu şekilde özetlenebilir: Veri matrisi ya da p değişkenli şans örneğinde yer alan i .inci gözlem aslında $x_i(t)$ ($i = 1, 2, \dots, N$) biçiminde reel sürekli bir fonksiyon olmasına rağmen, veriler genellikle birbirlerinden ayrı noktalarda gözlemlenir. N örnek büyüklüğü (örnek hacmi), n_i , i .inci örnek bireyi için yapılan ölçüm sayısı, t_{ij} i .inci örnek bireyi için j .inci ölçümlemenin alındığı nokta, y_{ij} , y karakteristiğinin ya da değişkeninin t_{ij} noktasındaki değerini göstermek üzere, çok değişkenli şans örneğindeki i .inci birey için yapılan gözlemler, $y_{i1}, y_{i2}, \dots, y_{in}$ şeklinde verilebilir. N hacimlik şans örneğini oluşturan veri seti ise,

$$y_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n_i$$

şeklinde dir. Klasik çok değişkenli istatistiksel notasyonda ise, üzerinde çalışılan ölçümler $\{y_{ij}; i=1, 2, \dots, n, j=1, 2, \dots, p\}$ biçiminde belirtilen veri matrisidir (Ramsay ve Silverman, 1997: 9).



Fonksiyonel veri analizinin özellikle klasik çok değişkenli analizden farklı olarak zaman noktalarının yerleşimi konusunda geniş bir esnekliği vardır. Genellikle her bir örnek birimi için ölçüm yapılan nokta sayısı eşit alınmak zorunluluğunda olmadığından, bu değişebilirliği göstermek üzere n_i notasyonu kullanılmıştır.

Fonksiyonel veri analizi sıklıkla zamana ait verilerle ilgilenmesine rağmen, faaliyet alanı ve amaçları zaman serilerinin analizinden tamamen farklıdır. Zaman serileri analizi daha çok verilerin modellenmesine ve gelecek gözlemlerin tahminlenmesine odaklanır. Fonksiyonel veri analizindeki teknikler ise aslında doğasında, fonksiyonel verilerin taşıdığı değişkenliğin yapısını ve fonksiyonlar arasındaki karşılıklı ilişkileri açıklayıcıdır, vurgu yörüngelerde ve şekillerdedir (Costanzo, 2005).

Bu çalışmada da ilk adım olarak kesikli veriler Splayn Düzgünleştirme Yöntemleriyle sürekli fonksiyonlar haline dönüştürülmekte ve daha sonra da veriler arasındaki değişkenlik yapısını ortaya koymak üzere Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi kullanılmaktadır. Konunun teorik yapısına ayrı başlıklar altında kısaca değinilmektedir.

2. Kesikli Verileri Fonksiyonel Verilere Dönüştürme

Fonksiyonel veri analizinde ilk adım; gözlenen y_{ij} değerlerini herhangi bir t değeri için hesaplanması mümkün olan bir x_i reel sürekli fonksiyonuna dönüştürmektir. Eğer gözlemlenen değerlerin hatasız oldukları varsayılırsa, bu süreç interpolasyon yöntemi ile yapılır. Fakat verilerde ortadan kaldırılması gereken bazı gözlemsel hatalar varsa, örneğin, verilerin oluşturulması deneysel bir sürece dayanıyorsa, kesikli verilerden fonksiyonel verilere yapılan bu dönüşüm süreci, **düzgünleştirme** adını alır (Ramsay ve Silverman, 1997: 9).

Bu çalışmada da, doğrudan doğruya interpolasyon uygulaması yerine, istatistiksel uygulamalarda çoğu kez tercih edilen splayn düzgünleştirme yöntemi kullanılmaktadır.

2.1. Baz Fonksiyon Yaklaşımı

Verileri interpolate etmede veya düzgünleştirmede en çok kullanılan yaklaşım eldeki probleme uygun, B-Splaynlar gibi bilgisayar dostu bir gösterimi olan baz fonksiyonlar seti seçmektir.

$x(t)$ fonksiyonunu oluştururken esnek yöntemlere ihtiyaç duyulmaktadır. Bunun için K tane baz fonksiyondan (basis function) oluşan bir sistem seçilmektedir. Oluşturulmak



istenilen $x(t)$ fonksiyonu bu baz fonksiyonların ağırlıklandırılmış bir toplamı olarak şu şekilde yazılabilir:

$$x(t) = c_1\theta_1(t) + c_2\theta_2(t) + \dots + c_k\theta_k(t) \quad (2.1)$$

Bu ifadede yer alan $\theta_i(t)$ i.inci baz fonksiyon ve c_i ise bu baz fonksiyona karşılık gelen katsayıdır. Burada c_i , $i=1, 2, \dots, K$ katsayıları, $x(t)$ fonksiyonunun şeklini ve biçimini belirleyen katsayılardır. Bir anlamda parametre olarak yorumlanabilirler. Bu çalışmada ele alınan “Pürüzlü Ceza” (Roughness Penalty) düzgünleştirme yöntemleriyle c_i katsayılarının tahminlenmesi amaçlanmaktadır.

Özetle; K herhangi bir reel pozitif tamsayı ve $c_i \in \mathbb{R}$ olmak üzere $x(t)$ fonksiyonunun,

$$x(t) = \sum_{i=1}^K c_i \theta_i(t) \quad (2.2)$$

şeklinde K tane baz fonksiyonun ağırlıklandırılmış bir toplamı (doğrusal bileşeni) olarak tahminlenmesini amaçlayan yaklaşım “Baz Fonksiyon Yaklaşımı” olarak adlandırılır. Genel olarak en çok kullanılan baz fonksiyonlar, Kuvvetler, Fourier baz ve B-Splayn baz şeklinde sıralanabilir.

Çoğu fonksiyonel veri analizinde fonksiyonel verileri uygun bir bazda depolamak anahtar adımdır. Çoğu kez baz prosedürde açıkça bellidir ve araştırmacının buna dikkat etmesine ihtiyaç olmaz(Ramsay ve Silverman, 2002: 33).

Ne kadar çok baz fonksiyon içerilirse uyumu sağlanan fonksiyon o kadar daha kompleks olur. y_j gözlem değerlerinin düzgünleştirilme derecesine K baz fonksiyon sayısına dayalı olarak karar verilir(Ulbricht, 2004: 10).

Baz fonksiyonların arzu edilen karakteristiği çalışılmak istenilen fonksiyon tipleriyle uyuşan özelliklere sahip olmasıdır. İdeal olan, nisbeten küçük bir K değeri kullanarak mükemmel bir yaklaşım başaracak biçimde bir baz seçimidir(Ramsay ve Silverman, 1997: 46). Ramsay ve Silverman 20-30 baz fonksiyonun genelde göze çarpan görünümleri özetlemek için yeterli olduğunu belirtmişlerdir(Lee, 2004: 9). Ayrıca gözlem sayısının çok fazla olduğu birçok çalışmada baz fonksiyon sayısı gözlem sayısının karekökü olarak kullanılmıştır(Laukaitis ve Rackauskas, 2002: 59). Fonksiyon tahmini için iyi çalışan bir baz nispeten zayıf türev tahminleri verebilir. Bu nedenle bir baz için seçim kriterlerinden biri de yaklaşımın bir veya daha fazla türevinin mantıklı hareket edip etmediğidir(Ramsay ve Silverman,1997: 46).

2.2. B-Splayn Baz Fonksiyonlar

Yukarıda bahsedilen üç baz tipinden B-Splaynlar, uygulamada periyodik olmayan durumlar için sıklıkla kullanıldığından ve bu çalışmanın da kapsamını içerdiğinden ayrı bir kesim olarak ele alınmıştır. B-Splaynlar periyodik olmayan durumlar için polinomlara çoğu durumda tercih edilmektedir. B-Splaynların esnek ve numerik olarak sağlam olması ve değerlerin bant matris yapısı B-Splaynların yaygın kullanımını arttırır.

d herhangi bir negatif olmayan tam sayı olmak üzere polinomun derecesini, $\underline{t} = (t_j)$, $j=1,2, \dots, n$ B-Splaynları oluşturmada kullanılan düğüm vektörünü, $\{B_{i,d}\}_{i=1}^K$, B-Splaynları göstermek üzere, Splaynlar, B-Splaynların doğrusal bir kombinasyonu olarak,

$$x(t) = \sum_{i=1}^K c_i B_{i,d} \quad (2.3)$$

şeklinde ifade edilebilirler.

Yukarıdaki tanımlamalara dayalı olarak, d 'inci dereceden m 'inci mertebeye sahip bir j . inci B-Splayn aşağıdaki tekrarlı ilişki ile tanımlanabilir:

$$B_{j,d}(t) = \frac{t - t_j}{t_{j+d} - t_j} B_{j,d-1}(t) + \frac{t_{j+1+d} - t}{t_{j+1+d} - t_{j+1}} B_{j+1,d-1}(t) \quad (2.4)$$

B-Splaynın mertebesi genellikle m ile gösterilir ve derecenin bir fazlasıdır. Burada t bağımsız değişken olarak ele alınmıştır ve tüm reel sayılar için tanımlıdır. $B_{j,d}(t)$ şeklindeki bir B-Splayn sadece $(t_k)_{k=j}^{j+d+1}$ düğümlerine bağlıdır.

Düzgünleştirmenin kontrolü splayn teorisinin temel amacı olup, düzgünleştirme (2.4) numaralı denklemlerle verilmiş olan tekrarlı ilişki ile tanımlanır. Tekrarlı ilişkiler yoluyla, verilen bir düzgünleştirme seviyesindeki B-Splayn baz fonksiyonları kullanılarak, bir derece veya bir mertebeye daha yüksek, bir ilave aralığa daha yayılmış, daha yüksek düzgünleştirme seviyesindeki yeni bir B-Splayn elde edilebilir.

Bir B-Splayn oluşturabilmek için düğümlerin sayısının en az $d+2$ tane olması gerekmektedir. Eğer düğüm sayısı tekrarlı değil ise bu durumda oluşturulacak baz fonksiyon sayısı fonksiyonel veri analizinde geleneksel olarak yaklaşımın doğruluğunu garantiye alacak büyüklükte olması bakımından aşağıdaki eşitlik yardımı ile verilmektedir:

$$\text{Baz fonksiyon sayısı} = \text{Düğüm sayısı} + \text{mertebeye} - 2 \quad (2.5)$$

Tüm düğümler farklı ise, bu durumda doğrusal Splayn süreklidir, kuadratik splaynın birinci türevi, kübik splaynın ikinci türevi de süreklidir(Lyche ve Morken, 2002: 39).

Splaynlar ve bunların interpolasyonu ve düzgünleştirilmesi fikri ilk olarak Wittaker (1923)'a dayanmaktadır, ancak 1960'lardan itibaren bu konuda çalışmalar yoğun bir biçimde yapılmaya başlanmıştır. Splaynlarla ilgili ilk çalışmaların büyük bir çoğunluğu Fransız bilim adamları tarafından yapılmıştır.

B-Splayn ve Splayn teorisi ile ilgili klasik referans de Boor (1978: 109-153) dur. Konu ile ilgili ayrıntılı bilgi için ayrıca Nürnbergger (1989: 95-99), Schumaker (1993: 108-139) ve Dierckx (1993: 1-22) çalışmalarına başvurulabilir.

B-Splayn baz fonksiyonları elde ettikten sonra $x_i(t)$ fonksiyonlarını elde etmek için ikinci adım c_i katsayılarının Pürüzlü Ceza Yaklaşımı ile tahminlenmesidir.

2.3. Pürüzlü Ceza Yaklaşımı

Fonksiyonel veri analizinde verilere bir eğrinin uyumunu sağlarken tek amaç yalnızca iyi bir uyum yapmak değil, aynı zamanda bu amaçla aslında çatışan diğer bir amaç da çok fazla iniş çıkış göstermeyen bir eğri tahmini elde etmektir.

Fonksiyonel veri analizinde splaynları düzgünleştirirken yaygın olarak kullanılan Pürüzlü Ceza Yaklaşımının temel amacı eğrinin pürüzlülüğünü ölçmek ve verilerin eğriye uyumu ve eğrinin pürüzlülüğü arasında bir uzlaşma sağlamaktır.

Bu iki çatışan amaç bir anlamda istatistiğin temel prensibinin iki elemanına karşılık gelebilir. Bilindiği gibi, Ortalama Karesel Hata, Sapmanın karesi ile Örneklem Varyansının toplamına eşittir. Örneklem varyansını azaltmak için sapmadan biraz taviz verilebilir, bu da tahminlenen eğriye düzgünleştirme yüklenmesinin temel nedenidir.

$[a,b]$ kapalı aralığında tanımlı bir x eğrisi verilsin. Bu x eğrisinin ne kadar pürüzlü, bir diğer deyişle inişli çıkışlı olduğunu ölçmenin birkaç farklı yolu vardır. İki kez türevi alınabilen bir x eğrisinin pürüzlülüğünü ölçmenin global olarak kabul edilen bir yolu,

$$PEN_2(x) = \int_a^b (D^2 x(t))^2 dt = \|D^2 x\|^2 \quad (2.6)$$

şeklinde ikinci türevin karesinin integralini almaktır (Eubank, 1999: 227). Bu değer x deki toplam eğriselliği değerlendirir, veya bir diğer deyişle, x eğrisinin bir doğrudan sapma derecesidir. Bu yüzden, yüksek değişkenli fonksiyonların yüksek $PEN_2(x)$ değerleri sağlayacağı beklenebilir, çünkü ikinci türevleri ilgilenilen aralığın çoğunda büyüktür (Ramsay ve Silverman, 1997: 59). Ayrıca, $\|D^2 x\|^2$, x fonksiyonunun değişkenliğinin hızının bir ölçüsü olarak da değerlendirilebilir (Leurgans vd., 1993).

x fonksiyonu $[a,b]$ aralığında tanımlı iki kez türevi alınabilen bir fonksiyon ve $\lambda > 0$ düzleştirme parametresi olsun. Cezalı kareler toplamı (CKT_λ),

$$CKT_\lambda = \sum_j (y_j - x(t_j))^2 + \lambda \|D^2 x\|^2 \quad (2.7)$$

şeklinde tanımlanır. Pürüzlü Ceza Yaklaşımında (2.7)den de görüldüğü gibi amaç En Küçük Kareler Yönteminde olduğu gibi yalnız Artık Kareler Toplamını değil eğrilerin pürüzlülüğünü de dikkate alan Cezalı Kareler Toplamını minimize etmektir.

Düzleştirme parametresi λ Artık Kareler Toplamı ile ölçülen ‘verilerin eğriye uyumu’ ve $\|D^2 x\|^2$ ile ölçülen ‘ x fonksiyonunun pürüzlülüğü’ arasındaki değişim oranını ölçer. Eğer λ çok büyük ise bu durumda doğrusal olmayan fonksiyonlar CKT_λ da büyük bir pürüzlülük cezası içerir. Bu nedenle $\lambda \rightarrow \infty$, uyumu sağlanan x eğrisi, gözlenen verilerin standart doğrusal regresyonuna yaklaşır. Bunun tam tersine λ azalırken, eğri gitgide daha inişli çıkışlı hale gelir, pürüzlülüğüne daha az ceza konmuştur ve $\lambda \rightarrow 0$ x eğrisi $f(t_j) = y_j$ ($j = 1, 2, \dots, n$) şeklinde verileri interpolate etmeye yaklaşır. Bir diğer deyişle, Cezalı Kareler Toplamındaki temel katkı Artık Kareler Toplamı ile olur. Bu limit durumunda bile interpolate edilen eğri keyfi değişken değildir, bunun yerine, bu eğri tüm iki kez türev alınabilen eğriler içinde verilere uyum gösteren en düzgün eğridir.

λ düzleştirme parametresinin seçimi ile ilgili Green ve Silverman (1994: 29) iki farklı yaklaşımdan bahsetmişlerdir. Bunlardan bir tanesi subjektif bir değeri ise otomatik seçimdir. İlk yaklaşım, düzleştirme parametresinin serbest seçimini prosedürün avantajlı bir özelliği olarak ele almaktır. Düzleştirme parametresini değiştirerek farklı ölçeklerde ortaya çıkan veri özellikleri incelenebilir ve eğer tek bir tahmine ihtiyaç var ise subjektif bir seçimle en iyi görünümü veren parametre değeri seçilebilir. Bu şekildeki subjektif bir yaklaşım gerçekte en kullanışlı olanı olabilir ve uygulamada sıklıkla kullanılmaktadır.

Ramsay ve Silverman’a (2002: 56) göre düzleştirme uygulayan çoğu araştırmacının düzleştirme seviyesinin seçimi ile ilgili bakış açısı, ilgilenilen kararlı ve yorumlanabilir tahmin elde edebilme ile verilere uyum arasında bir denge kurmaktır. λ ’nın seçimi araştırmacının kendi ihtiyaçlarından çıkar ve aynı verilere farklı değerler uygulanan analizleri görmekten, bir diğer deyişle parametrenin subjektif seçiminden memnun olurlar.

Buna ilave olarak Ramsay ve Li’nin “Curve Registration (1998)” isimli çalışmasında uygulamalarla ilgili kapsamlı bir araştırma sonucu düzleştirme parametresi λ için 10^{-4} , 10^{-3} ve 10^{-2} değerlerinin iyi çalıştığı gözlenmiştir.



Bu alanda Ulbricht (2004: 60) tarafından yapılan bir tez çalışmasında ise λ düzgünleştirme parametresi, baz fonksiyon sayısı gözlem sayısından daha büyük olduğu durumda; ki bizim yaklaşımımızda bu durum geçerlidir,

$$\lambda=10^{-4} [\text{tr} (\mathbf{B}^T \mathbf{B}) / \text{tr} \mathbf{R}] \quad (2.8)$$

şeklinde tanımlanmıştır.

Cezalı Kareler Toplamını minimize eden eğri tahmini **düzgünleştirme ve uyum iyiliği arasındaki en iyi uzlaşmadır**. Buna göre Cezalı Kareler Toplamını minimize eden eğri tahmini \hat{x} aşağıdaki özelliklere sahiptir(Silverman, 1985: 3) :

- Her bir (t_j, t_{j+1}) aralığında bir kübik polinomdur.
- t_j şeklinde belirtilen düğümlerde eğri ve ilk iki türev süreklidir, fakat üçüncü türevde süreksizlik olabilir.
- $(-\infty, t_1)$ ve $(t_n, +\infty)$ aralıklarının herbirinde ikinci türev sıfırdır ve böylece veri aralığının dışında \hat{x} doğrusaldır.

İlk iki koşulu sağlayan herhangi bir eğri t_j düğümlerine sahip bir düzgünleştirilmiş Kübik Splayn olarak adlandırılır. Bu özellikler tahmine empoze edilmemiştir, pürüzlü cezanın $\|D^2x\|^2$ olarak seçiminden otomatik olarak ortaya çıkmaktadırlar. Bu düzgünleştirilmiş Kübik Splayn üçüncü özellik olan doğal sınır koşullarından dolayı, bir diğer deyişle sınırlardaki ikinci ve üçüncü türevlerinin sıfır olması nedeniyle Doğal Kübik Splayn (Natural Cubic Splayn) olarak adlandırılabilir(Simonoff,1996: 170).

Pürüzlü Ceza Yönteminde derece seçiminde genellikle üçüncü dereceden splaynlar bir diğer deyişle Kübik Splaynlar kullanılır. B-Splaynların doğrusal bir kombinasyonu olarak oluşturulan Kübik Splaynlarda her bir B-Splayn bir kübik polinomdur ve her bir polinom parçası ikinci türevlerine kadar eşlenerek birleştirilebildiğinden bu eğrinin görsel olarak düzgün olması demektir. İkinci türev eğrinin eğriselliğini gösterdiğinden ve eğrisellikler kesme noktalarında eşleştiğinden, böylece eğrisellik düzgün bir şekilde değişir(Ramsay, 2000). Ayrıca kübik splaynlar kolaylıkla değerlendirilir ve genel olarak tatmin edici sonuçlar verir(Reinsh, 1967).

3. Fonksiyonel Ana Bileşenler Analizi

Fonksiyonel veri analizinde bireylerin sayısı arttıkça fonksiyonel gözlemlerden oluşan büyük bir grubu üst üste yerleştirmek çok fazla aydınlatıcı ve görünüm olarak da çok cazip

olmayabilir. Ayrıca birçok birbiriyle çakışan ve karışan eğri olduğundan dolayı bir tek eğrinin bireysel davranışı diğer birçok eğrinin davranışları arasında kaybolur (Jones ve Rice, 1992: 140). Ana Bileşenler Analizi bu tip kompleks populasyonların yapısının anlaşılmasıyla ilgili oldukça genel bir probleme etkili bir çözüm sağlar. Veri karmaşıklığını gidermek üzere eğriler arasındaki değişimin önemli modlarını tanımlamak için Ana Bileşenler Analizinin (ABA) kullanımı önerilebilir. Ana Bileşenler Analizi sistemde olması beklenen ve aynı zamanda da önceden fark edilmeyen ilişkileri ortaya çıkarır. Bu nedenlerden dolayı Ana Bileşenler Analizi fonksiyonel veri analizinde ele alınan anahtar tekniktir.

Klasik çok değişkenli analizde varyans–kovaryans ve korelasyon matrislerinin olduğu gibi, fonksiyonel veri analizinde de varyans, kovaryans ve korelasyon fonksiyonlarının yorumlanması zor olabilir ve gözlenen verilerdeki değişkenlik yapısı ile ilgili tamamen anlaşılır bir gösterim vermeyebilir. Ana Bileşenler Analizi varyans–kovaryans yapısına daha aydınlatıcı bir biçimde bakmayı sağlar (Ramsay & Silverman, 1997: 85).

Fonksiyonel veriler için uygulanan Fonksiyonel Ana Bileşenler Analizinde asıl amaç çok değişkenli veriler için uygulanan Ana Bileşenler Analizi ile aynı olup verilerdeki değişimi etkili bir biçimde tanımlayan birkaç ortogonal fonksiyon elde etmektir. Aralarındaki temel fark γ_j biçiminde belirtilen ortogonal fonksiyonlar olan ana bileşen ağırlıkları (bu ağırlıklar genelde harmonik olarak da adlandırılır) şimdi zamanın veya ilgili başka bir değişkenin fonksiyonlarıdır.

Fonksiyonel kavramda her bir ana bileşen bir fonksiyonel veri ile aynı t aralığında tanımlı, verilerin temel “Değişim Modlarını” tanımlayan bir ana bileşen ağırlık fonksiyonu ($\gamma(t)$) ile belirtilir ve doğrusal kombinasyon aşağıdaki biçimde tanımlanır:

$$Y_j = \langle \gamma_j, x - E(x) \rangle = \int \gamma_j(t) \{ x(t) - E x(t) \} dt \quad (3.1)$$

Bu çalışmada verilen tüm ifadelerde yer alan integraller [a,b] sonlu reel aralığında belirli integrali temsil etmektedirler.

Bundan sonra artık γ_j ile belirtilen ağırlıklar $\gamma(t)$ değerlerine sahip bir ağırlık fonksiyonu halini alır. Burada Y_j , her bir $x(t)$ için γ_j üzerine $\{ x(t) - E x(t) \}$ nin izdüşüm (projection) miktarıdır (Castro v.d., 1986: 332).

Fonksiyonel Ana Bileşenlerin ilk adımında ağırlık fonksiyonu (ana bileşen fonksiyonu veya harmonik fonksiyonu) γ_1 ,

$$\|\gamma_1\|^2 = \int \gamma_1(t)^2 dt = 1 \quad (3.2)$$



kısıtı altında, doğrusal bileşenin varyansı olan,

$$\text{Var}(Y_j) = \text{Var} \langle \gamma_j, x - E(x) \rangle = \iint \gamma_j(s) \text{Cov}(s,t) \gamma_j(t) ds dt \quad (3.3)$$

ifadesini **maksimum** yapacak biçimde belirlenir. İkinci ağırlık fonksiyonunun hesaplanması için Klasik Ana Bileşenler Analizinde olduğu gibi, Fonksiyonel Ana Bileşenler Analizinde de ağırlık fonksiyonunun,

$$\langle \gamma_j, \gamma_m \rangle = \int \gamma_j(t) \gamma_m(t) dt = 0 \quad (j \neq m) \quad (3.4)$$

şeklindeki ilave kısıt olan ortogonalite koşullarını da sağlaması gerekir. Her bir ağırlık fonksiyonunun eğrilerdeki değişimin en önemli modunu tanımlama görevi vardır ve burada her bir modun önceki adımlarda tanımlanan modlara ortogonal olması gerekir (Ramsay ve Silverman, 1997: 88). Ağırlık fonksiyonları her aşamada maksimum değişimi açıklayabilecek biçimde oluşturulan ortogonal baz fonksiyonlar setidir.

(3.3) numaralı formülle belirtilmiş olan doğrusal bileşenin varyansının maksimum yapılması problemi Klasik Ana Bileşenler Analizinde olduğu gibi fonksiyonel veriler için de özdeğer- özfonksiyon ayrışımı ile çözümlenebilir.

Fonksiyonel özdenklemler aşağıdaki biçimdedir:

$$\int \text{Cov}(s,t) \gamma(t) dt = \psi \gamma(s) \quad (3.5)$$

Bu ifadede,

$$\text{Cov}(s,t) = N^{-1} \left\{ \sum_{i=1}^N x_i(s) x_i(t) \right\} \quad (3.6)$$

şeklinde. Bu kovaryans ifadesinde $x_i(t)$ her bir birimden ortalama fonksiyon değerinin çıkarılmış halidir. Burada amaç kovaryans fonksiyonunun ortogonal ayrışımını sağlayarak fonksiyonel değişimin dominant bileşenlerini ayırmaktır (Costanzo, 2005). Artık yukarıdaki özdenklemleri sağlayan Klasik Ana Bileşenler Analizinde olduğu gibi farklı **özdeğer- özvektör** değil **özdeğer- özfonksiyon** çiftleri vardır. $\gamma_1, \gamma_2, \dots, \gamma_{N-1}$ biçimindeki özfonksiyonlar karşılık geldikleri $\psi_1 \geq \psi_2 \geq \dots \geq \psi_{N-1}$ özdeğerlerine göre sıralanırlar. Klasik çok değişkenli analizdeki felsefeye paralel bir biçimde j.inci bileşenin varyansı j.inci özdeğere eşit olmaktadır:

$$\text{Var}(Y_j) = \iint \gamma_j(s) \text{Cov}(s,t) \gamma_j(t) ds dt = \psi_j \quad (3.7)$$



yani ψ_j biçimindeki özdeğerler, (3.7) numaralı formülden de görüldüğü üzere her bir bileşene yüklenebilir varyans miktarını göstermektedir.

x_i ($i = 1, 2, \dots, N$) şeklinde belirtilen N eğriden oluşan bir veri setini ve $\text{Cov}(s,t)$ kovaryans fonksiyonunu belirtmek üzere,

$$\int \text{Cov}(s,t) \gamma(t) dt = \psi \gamma(s) \quad (3.8)$$

şeklinde belirtilen fonksiyonel özdeğer - özfonksiyon probleminin çözümü için bir yaklaşım, bu denklemi yaklaşık olarak buna denk olan bir matris formuna dönüştürmektir.

Özdenklemini kesikli veya matris formuna döndürmenin bilinen popüler bir yolu, her bir fonksiyon x_i 'yi bölüm 2de de bahsedildiği gibi,

$$x_i(t) = \sum_{k=1}^K c_{ik} \theta_{ik}(t) \quad i=1, 2, \dots, N; \quad k=1, 2, \dots, K \quad (3.9)$$

şeklinde bilinen baz fonksiyonların doğrusal bir kombinasyonu olarak ifade etmektir. Bu gösterimi genellersek,

$$\underline{\mathbf{x}} = [x_1, x_2, \dots, x_N]^T \quad (3.10)$$

fonksiyon değerlerine sahip bir vektörü ve

$$\underline{\boldsymbol{\theta}} = [\theta_1, \theta_2, \dots, \theta_K]^T \quad (3.11)$$

baz fonksiyonlardan oluşan bir vektörü belirtmek üzere, \mathbf{C} ($N \times K$) katsayılar matrisi iken, tüm N eğrilerinin eş zamanlı açılımı,

$$\underline{\mathbf{x}} = \mathbf{C} \underline{\boldsymbol{\theta}} \quad (3.12)$$

şeklinindedir. Baz fonksiyon yaklaşımında kovaryans fonksiyonu da, baz fonksiyon yaklaşımına göre

$$\text{Cov}(s,t) = \underline{\boldsymbol{\theta}}^T(s) \text{Cov}(\mathbf{C}) \underline{\boldsymbol{\theta}}(t) \quad (3.13)$$

şeklinde ifade edilebilir.

Ana bileşen ağırlık fonksiyonu (özfonksiyon) ,

$$\gamma(t) = \sum_{k=1}^K b_k \theta_k(t) = \underline{\boldsymbol{\theta}}^T(t) \underline{\mathbf{b}} \quad \text{iken,} \quad (3.14)$$

$$\int \text{Cov}(s,t) \gamma(t) dt = \psi \gamma(s)$$

şeklindeki sürekli fonksiyonel özdenklemin, matris notasyonunda aşağıdaki biçimde ifade edilebilir:

$$\int \underline{\boldsymbol{\theta}}^T(s) \text{Cov}(\mathbf{C}) \underline{\boldsymbol{\theta}}(t) \underline{\boldsymbol{\theta}}^T(t) dt \underline{\mathbf{b}} = \psi \underline{\boldsymbol{\theta}}^T(s) \underline{\mathbf{b}} \quad (3.15)$$



Burada

$$\mathbf{J} = \int \underline{\boldsymbol{\theta}}(t) \underline{\boldsymbol{\theta}}^T(t) dt \quad (3.16)$$

şeklinde (KxK) boyutlu bir matris tanımlandığında, (3.15) numaralı eşitlik aşağıdaki hali alır:

$$\underline{\boldsymbol{\theta}}^T(s) \text{Cov}(\mathbf{C}) \mathbf{J} \underline{\mathbf{b}} = \psi \underline{\boldsymbol{\theta}}^T(s) \underline{\mathbf{b}} \quad (3.17)$$

veya

$$\text{Cov}(\mathbf{C}) \mathbf{J} \underline{\mathbf{b}} = \psi \underline{\mathbf{b}} \quad (3.18)$$

$\|\gamma\|^2 = \int \gamma(t)^2 dt = 1$ olduğundan $\underline{\mathbf{b}}^T \mathbf{J} \underline{\mathbf{b}} = 1$ olur. Benzer şekilde sadece ve sadece fonksiyonel skaler çarpım $\langle \gamma_j, \gamma_m \rangle = \int \gamma_j(t) \gamma_m(t) dt = \underline{\mathbf{b}}_j^T \mathbf{J} \underline{\mathbf{b}}_m = 0$ ise γ_j ve γ_m ortogonal olur.

Cov (C) ve J matrisleri ayrı ayrı simetriktir, ancak (Cov (C) J) matrisi simetrik değildir. Bu nedenle J matrisi simetrik olduğundan, $\underline{\mathbf{u}} = \mathbf{J}^{1/2} \underline{\mathbf{b}}$ olarak tanımlanır ve daha sonra,

$$\mathbf{J}^{1/2} \text{Cov}(\mathbf{C}) \mathbf{J}^{1/2} \underline{\mathbf{u}} = \psi \underline{\mathbf{u}} \quad (3.19)$$

şeklindeki (Cov (C) J) matrisi ile aynı özdeğerlere sahip olan, simetrik özdeğer probleminin çözülmesine ve ters dönüşüm olan $\underline{\mathbf{b}} = \mathbf{J}^{-1/2} \underline{\mathbf{u}}$ ifadesinin hesaplanmasına ihtiyaç duyulur. Böylece eğriler arasındaki değişimi etkili biçimde açıklayan ana bileşen ağırlık fonksiyonu elde edilir.

Fourier baz şeklindeki ortonormal fonksiyonel baz için J matrisi birim matrise eşit olur. Bu durumda Fonksiyonel Ana Bileşenler Analizi katsayıların kovaryans matrisine uygulanan Klasik Ana Bileşenler Analizi problemine indirgenir.

Ana bileşen ağırlık fonksiyonlarının açık bir biçimde yorumlanması her zaman mümkün olmayabilir. Bu durumda sonuçların yorumlanmasına yardımcı olacak bir yaklaşım aşağıdaki biçimde belirtilmiş olan ana bileşen skorlarının işaretlenmesidir.

$$\gamma_{ij} = \langle \gamma_j, x_i - E(x) \rangle = \underline{\mathbf{b}}_j^T \mathbf{J} (\underline{\mathbf{c}}_i - \underline{\mathbf{c}}) \quad (3.20)$$

Burada γ_{ij} ana bileşen skoru, i.inci gözlemin j.inci ana bileşen fonksiyonuna göre koordinatı anlamına gelir ve bileşenlere bazı yorumlamalar eklerken yardımcı olabilir. Ana bileşen ağırlık fonksiyonlarını ve aynı zamanda ana bileşen skorlarının dağılımını analiz ederek ele alınan fonksiyonların yapısı ve dinamikleri değerlendirilebilir(Benko v.d., 2006:



5). Ana bileşen skorlarının dağılımı sapan gözlemlerin (outlier) saptanmasında da kullanılabilir.

Ana bileşen ağırlık fonksiyonlarının yorumlanmasına yardımcı olacak bir diğer ikinci yöntem de ağırlık fonksiyonunun bir çarpanıyla ortalama fonksiyonunun karşılaştırılması olabilir. Burada ortalama fonksiyonuna, ilgilenilen uygun bir çarpanla çarpılmış ana bileşen fonksiyonu eklenerek ve çıkarılarak elde edilen fonksiyonlarla birlikte ortalama fonksiyonu aynı düzlemde çizdirilir ve karşılaştırmalar yapılarak çeşitli yorumlamalarda bulunulabilir.

Bu çizimi oluşturmada hangi çarpanın kullanılacağını seçmek önemlidir. Ramsay (2003) çarpan olarak ana bileşen fonksiyonunun oluşturulmasında her bir özfonksiyona karşılık gelen özdeğerlerin kareköklerini almayı önermiştir. Böylece her bir bileşen için $\hat{\mu} \pm \hat{\gamma}_j \sqrt{\Psi_j}$ plot edilir. Ancak bu konu araştırmaya açıktır ve konuyla ilgili farklı önerilerde de bulunulabilmektedir. Tüm bileşenler için aynı sabitin kullanımı da önerilebilir (Ramsay ve Silverman, 1997: 93).

Klasik Ana Bileşenler Analizinde hesaplanabilecek ana bileşen sayısı için üst limit değişken sayısıdır. Fonksiyonel Ana Bileşenler Analizinin baz fonksiyon yaklaşımında ise, ana bileşen sayısı için üst limit ($K < N$ olması durumunda) bazın boyutu olan K 'dır. Bu durumda baz fonksiyon yaklaşımında, $\sum_{i=1}^K \psi_i$ değeri K doğrusal bileşence açıklanan toplam varyansı verir.

Örnekleme problemlerinde bilinmeyen kovaryans fonksiyonu $Cov(s,t)$, örnek kovaryans fonksiyonu $Co\hat{v}(s,t)$ ile tahminlenir. Dauxois v.d. (1982), örnek kovaryans fonksiyonu $Co\hat{v}(s,t)$ 'nin özdeğerleri ve özfonksiyonlarının sırasıyla ψ_j ve γ_j 'nin tutarlı tahminleyicileri olduğunu göstermişlerdir. Ayrıca Leurgans v.d. (1993) düzgünleştirme uygulandığında bile bu tutarlılığın devam ettiğini belirtmişlerdir.

Fonksiyonel Ana Bileşenler Analizinin baz fonksiyon yaklaşımı ile ilgili ayrıntılı bilgi için Ramsay ve Silverman (1997; 101), Ramsay ve Silverman (2005; 161-163), James v.d. (2000), Barra (2004), Benko (2004), Lee (2004), Lober ve Villa (2004), Yamanishi ve Tanaka (2005) ve Hall ve Nasab M. (2006) çalışmalarına başvurulabilir.

3.1. Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi

Fonksiyonel Ana Bileşenler Analizi ile elde edilen özfonksiyonlar bir diğer deyişle ana bileşen ağırlıkları pürüzlü olabilir. Bu pürüzlülük örnekleme varyansından veya gözlem

gürültüsünden (observation noise) ve kullanılan fonksiyonel bazın esnekliğinden kaynaklanabilir. Bu kesimde daha durağan ve daha yorumlanabilir sonuçlara sahip olmak için özfonksiyonların düzgünleştirilme metodları ele alınacaktır.

Fonksiyonel Ana Bileşenler Analizinin değeri düzgünleştirmenin Ana Bileşenler Analizine dahil edilmesi ile biraz daha artar. Fonksiyonel Ana Bileşenler Analizini düzgünleştirme, sadece Klasik Ana Bileşenler Analizi ile elde edilen bileşenleri düzgünleştirmek değildir. Düzgünleştirme Ana Bileşenlerin orijinal tanımının içine dahil edilir. Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinde klasik ortonormallik kısıtları fonksiyonların pürüzlülüğünü de hesaba katan bir ortonormallikle yer değiştirir.

Düzgünleştirilmiş bir fonksiyonel ana bileşen elde etmek için, sadece γ 'nın hacminin (size) değil aynı zamanda pürüzlülüğünün (iniş çıkışlarının) de kontrol edilmesi gerekir.

$$PEN_2(\gamma) = \int (D^2 \gamma(t))^2 dt = \|D^2 \gamma\|^2 \quad (3.21)$$

olmak üzere, Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinde Cezalı Ana Bileşen Varyansı (CABV),

$$CABV = \text{Var}(Y_j) = \frac{\int \int \gamma_j(s) \text{Cov}(s, t) \gamma_j(t) ds dt}{\int \gamma_j(t)^2 dt + \lambda \int (D^2 \gamma_j(t))^2 dt} \quad (3.22)$$

şeklinde tanımlanır (Silverman, 1996 :5). λ parametresi bölüm 2 de bahsedilen düzgünleştirme parametresidir. λ sabit olarak alındığında, ana bileşen ağırlık fonksiyonu γ 'nin pürüzlülüğü arttıkça $PEN_2(\gamma)$ 'nın değeri artacağından Cezalı Ana Bileşen Varyansı azalır. Ayrıca λ sıfıra yaklaşırken Cezalı Ana Bileşen Varyansı düzgünleştirilmemiş durumun varyansına döner. Bir diğer bakış açısıyla ise, λ 'nın değeri arttıkça, Cezalı Ana Bileşen Varyansı γ 'in pürüzlülüğünden daha çok etkilenir.

$$\|\gamma_j\|^2 = \int \gamma_j(t)^2 dt = 1$$

şeklindeki Fonksiyonel Ana Bileşenler Analizi için daha önceden verilmiş olan kısıt, γ 'nin pürüzlülüğünü de dikkate alan,

$$\int \gamma_j(t)^2 dt + \lambda \int (D^2 \gamma_j(t))^2 dt = 1 \quad (3.23)$$

kısıtı ile yer değiştirir (Ramsay ve Silverman, 2002: 26). Bu aslında standart Sobolev iç çarpımlarının genelleştirilmiş halidir (Musayev ve Alp, 2000: 114).



Pürüzlü ceza ikinci, üçüncü ve daha yüksek dereceli düzgülendirilmiş ana bileşenlere ilave kısıtlar ekler. j.inci bileşen fonksiyonu, (3.23) kısıtına bağılı olarak $\text{Var}(Y_j)$ 'yi ,

$$\int \gamma_j(t) \gamma_m(t) dt + \lambda \int (D^2 \gamma_j(t)) (D^2 \gamma_m(t)) dt = 0 \quad j \neq m \quad (3.24)$$

ilave kısıtı altında maksimum yapar.

Cezalı Ana Bileşen Varyansını maksimize eden özfonksiyon γ ,

$$\int \text{Cov}(s, t) \gamma(t) dt = \psi (I + \lambda D^4) \gamma(s) \quad (3.25)$$

şeklinde ifade edilmiş olan özdenklemin en büyük özdeğerine karşılık gelen özfonksiyondur. Bir diğere deyişle, harmonik fonksiyon veya ana bileşen ağırlık fonksiyonudur.

Burada yukarıdaki özdenklemini çözerek gereksinim duyulan tüm özfonksiyonların tahminleri elde edilebilir. Bu eşitlikten elde edilen özfonksiyon tahminleri çok düzgün fonksiyonlardır.

Özetle, ana bileşen ağırlık fonksiyonu γ_j ,

$$\text{CABV} = \text{Var}(Y_j) = \frac{\langle \gamma, \text{Cov} \gamma \rangle}{\|\gamma\|^2 + \lambda \|D^2 \gamma\|^2}$$

şeklinde belirtilen Cezalı Ana Bileşen Varyansını,

$$\int \gamma(t)^2 dt + \lambda \int (D^2 \gamma(t))^2 dt = 1$$

kısıtı altında maksimize eder ve önceki tahminlenen bileşenlerle ortogonalliğin modifiye edilmiş şekli

$$\int \gamma_j(t) \gamma_m(t) dt + \lambda \int (D^2 \gamma_j(t)) (D^2 \gamma_m(t)) dt = 0, \quad j \neq m$$

ifadesiyle temsil edilir.

Pratikte genelleştirilmiş özdeğer – özfonksiyon problemi Klasik Fonksiyonel Ana Bileşenler Analizinde olduğu gibi uygun bir baz fonksiyon yaklaşımı ile çalışılarak daha kolaylıkla çözümlenebilir.

J klasik Fonksiyonel Ana Bileşenler Analizinde olduğu gibi (KxK) boyutlu bir matris ve **K**, elemanları $\langle D^2 \gamma_j, D^2 \gamma_m \rangle$ olan (KxK) boyutlu bir matris olmak üzere, Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinde amaç, ana bileşen ağırlık fonksiyonu (özfonksiyon),

$$\gamma(t) = \sum_{k=1}^K b_k \theta_k(t) = \underline{\theta}^T(t) \underline{b} \quad \text{iken,}$$



$$\int \gamma(t)^2 dt + \lambda \int (D^2 \gamma(t))^2 dt = \underline{\mathbf{b}}^T (\mathbf{J} + \lambda \mathbf{K}) \underline{\mathbf{b}} = 1 \quad (3.26)$$

kısıtı altında,

$$\iint \gamma(s) \text{Cov}(s,t) \gamma(t) ds dt = \underline{\mathbf{b}}^T \mathbf{J} \text{Cov}(\mathbf{C}) \mathbf{J} \underline{\mathbf{b}} \quad (3.27)$$

şeklindeki kuadratik formun maksimize edilmesidir. Bu maksimizasyon süreci aşağıdaki algoritma ile özetlenebilir:

Algoritma

1) Yukarıdaki denklemi çözmek için birinci adım olarak $\mathbf{L}^T \mathbf{L} = (\mathbf{J} + \lambda \mathbf{K})$ eşitliğini sağlayacak bir \mathbf{L} matrisi bulmak amacıyla Cholesky ayrışımı kullanılır. Burada \mathbf{L} üst üçgen matristir.

2) İkinci adımda $\underline{\mathbf{b}}^T (\mathbf{J} + \lambda \mathbf{K}) \underline{\mathbf{b}} = \underline{\mathbf{u}}^T \underline{\mathbf{u}}$ olacak biçimde $\underline{\mathbf{u}} = \mathbf{L} \underline{\mathbf{b}}$ tanımlaması yapılır. Böylece $\underline{\mathbf{u}}$, $(\mathbf{L}^{-1})^T \mathbf{J} \text{Cov}(\mathbf{C}) \mathbf{J} \mathbf{L}^{-1}$ matrisinin özvektörüdür. $\underline{\mathbf{u}}$ birim uzunluğa sahip olacak biçimde normalize edilir, böylece $\underline{\mathbf{u}}$ özvektörü $\underline{\mathbf{u}}^T \underline{\mathbf{u}} = 1$ kısıtı altında, $(\mathbf{L}^{-1} \underline{\mathbf{u}})^T \mathbf{J} \text{Cov}(\mathbf{C}) \mathbf{J} \mathbf{L}^{-1} \underline{\mathbf{u}}$ yi maksimize eder.

3) $\underline{\mathbf{b}} = \mathbf{L}^{-1} \underline{\mathbf{u}}$ ters dönüşümü hesaplanır. Burada $\underline{\mathbf{b}}$ düzgünleştirilmiş fonksiyonel ana bileşen ağırlık fonksiyonunun baz katsayı vektörüdür.

4) Son olarak dördüncü adımda $\gamma(t) = \sum_{k=1}^K b_k \theta_k(t) = \underline{\boldsymbol{\theta}}^T(t) \underline{\mathbf{b}}$ doğrusal

kombinasyonu aracılığı ile ana bileşen ağırlık fonksiyonları elde edilir.

Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinin baz fonksiyon yaklaşımı ile ilgili ayrıntılı bilgi için Ramsay ve Silverman (1997; 116-118), Ramsay ve Silverman (2002; 37-38), Ramsay ve Silverman (2005; 181-182) Benko (2004), Yamanishi (2004), Yao ve Lee (2006) çalışmalarına başvurulabilir.

Fonksiyonel Ana Bileşenler Analizinin uygulama alanları bilgisayar algoritmalarının gelişimi ile büyük bir artış göstermiştir. Rao (1958) büyüme eğrileri ile ilgili uygulamalarda Fonksiyonel Ana Bileşenler ile ilgili öncül fikirleri geliştirmiştir. Tıpta DNA genlerinden kaynaklanan eğriler setinin analizi bu setteki değişimin temel modlarını gösteren Fonksiyonel Ana Bileşenler Analizi kullanılarak uygulanabilir. Benzer değişimler gösteren genler bir araya toplanır ve gen profillerinin karakteristik parametreleri özetlenir (Barra, 2004). Tıp biliminde Ana Bileşenler Analizinin uygulandığı daha birçok benzer çalışma bulunmaktadır (Zhang, 1999; Müller, 2005). Kaliforniya da ozon seviyeleri ile ilgili bir çalışma yapılmış, 47 hafta için günlük maksimum ozon seviyeleri ile ilgili veriler



toplanmıştır. Burada 47 hafta için tüm veriler tamdır, bu çalışmada kayıp verileri interpolate etmeye çalışmak yerine sadece tüm verileri eksiksiz olan haftalar dikkate alınmıştır ve uygulanan Fonksiyonel Ana Bileşenler Analizi sonucunda ozon seviyeleri arasındaki değişimin temel modlarının büyük oranda iki temel bileşence açıklanabildiği görülmüştür. Bu değişim modlarından ilki yaz aylarındaki mevsimsel değişim, ikincisi ise bazı belirli haftalardan kaynaklanan değişim olarak açıklanabilir(Jones ve Rice, 1992). Dünyada internetin hızla yayılmasıyla internetten toplanan verilere Fonksiyonel Ana Bileşenler Analizi uygulamaları da hızla yayılmaya başlamıştır. Hyde vd. (2006)'nın yaptığı bir araştırmada internette yapılan açık arttırmalar için ana bileşenleri kullanarak zamana karşı fiyat artırma faaliyetlerinin nasıl değiştiğine karar vermeye çalışılmıştır. Bu çalışma iki tanınmış saat markası için yapılmıştır. Ayrıca fonksiyonel veri analizinin özelliklerinden biri olarak açık artırma süresince hız ve ivme de araştırılmıştır. Bu çalışmadan sonra benzer bir çalışma Jank ve Shmueli (2006) tarafından da gerçekleştirilmiştir. Kriminolojide yapılan bir araştırmada da poliste kaydı bulunan 11 ile 35 yaş arasındaki 413 suçlunun zamana karşı işledikleri suç faaliyetleri ele alınarak Fonksiyonel Ana Bileşenler Analizi ile suç faaliyetlerinin bir sınıflandırılmasına gidilmiştir(Ramsay & Silverman, 2002). Bir başka uygulama olarak da 35 Kanada hava istasyonundan alınan sıcaklık verilerine Fonksiyonel Ana Bileşenler Analizi uygulanarak sıcaklıklardaki değişimlerin temel nedenleri açıklanmaya çalışılmıştır (Silverman, 1995; Ramsay ve Silverman, 1997: 85-109). Bu çalışmanın benzeri ve biraz daha gelişmiş 50 Japonya hava istasyonundan alınan veriler için uygulanmıştır(Yamanishi, 2004). Bir başka çalışmada ise Nordic elektrik pazarındaki finansal sözleşmelere Fonksiyonel Ana Bileşenler Analizi uygulanması ile ilgili örnekler verilmiştir(Lillestol & Ollmar, 2003).

Fonksiyonel Ana Bileşenler Analizi için klasik referans Besse ve Ramsay (1986), Castro v.d.(1986), Ramsay ve Dalzell (1991) , Ramsay ve Silverman (1997) ve Ramsay ve Silverman(2005) dir.

4.Uygulama

4.1. Amaç: Bu araştırmanın amacı, İstanbul Menkul Kıymetler Borsasında işlem gören 13 hisse senedinin 408 ayrık noktada gözlenen kapanış fiyatlarının öncelikle B-Splayn Baz Fonksiyonlar ve Pürüzlü Ceza Yaklaşımını kullanarak bir diğer deyişle bu iki yaklaşımın birlikte kullanılması olarak adlandırılan Splayn Düzgünleştirme Yöntemi ile sürekli ve türevlenebilir fonksiyonlar haline dönüştürülmesinin incelenmesidir. Daha sonra da veriler

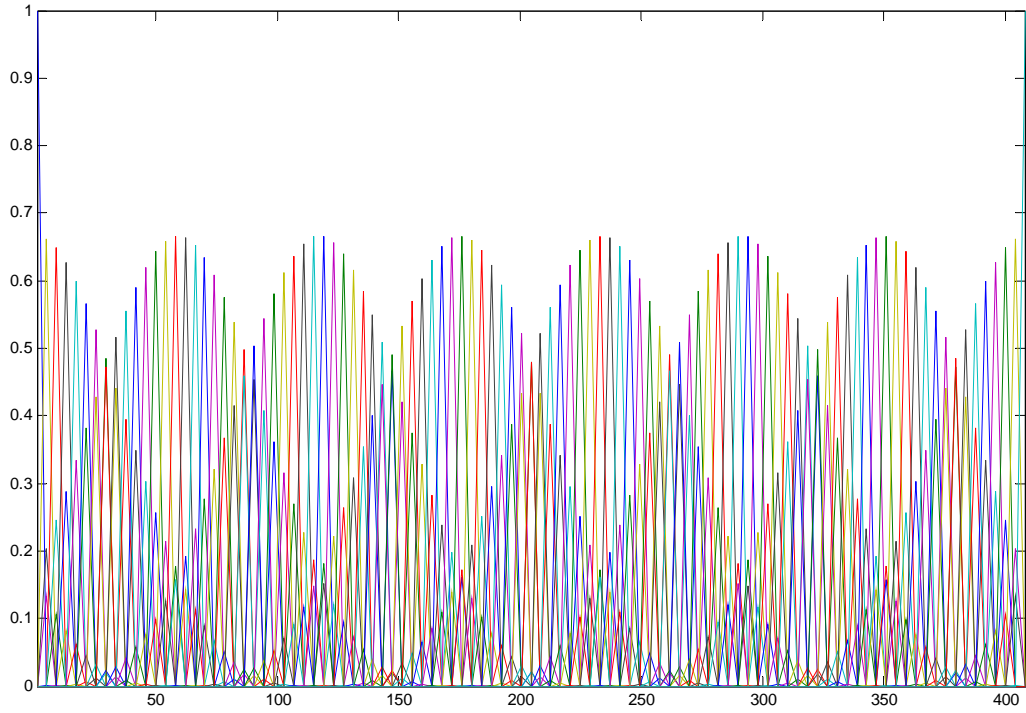


arasındaki yani ilgilenilen zaman aralığında hisse senetlerinin bireysel fonksiyonları arasındaki değişkenlik yapısını ortaya koymak üzere Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinden faydalanılmıştır.

Bu amaçla 97 yılının Temmuz ayından başlayarak 2005 yılının Aralık ayının sonuna kadar Cuma günü kapanış fiyatları, şirketlerin bileşik endeksi (İMKB 100 endeksi) oluşturmadaki ağırlıkları göz önünde bulundurularak belirlenen 13 şirket için ele alınmıştır. Bu şirketler, Akbank (AKBNK), Garanti Bankası (GARAN), İş Bankası C (ISCTR), Finansbank (FINBN), Sabancı Holding (SAHOL), Koç Holding (KCHOL), Doğan Holding (DOHOL), Tüpraş (TUPRS), Ereğli Demir Çelik (EREGL), Türk Hava Yolları (THYAO), Aksigorta (AKGRT), Anadolu Sigorta (ANSGR) ve son olarak da Akçansa (AKCNS) şeklindedir.

4.2. Yöntem: Öncelikle 11.07.1997 ile 30.12.2005 tarihleri arasında yukarıda belirtilen 13 hisse senedinin 408 ayrıık noktada gözlenen kapanış fiyatları Splayn Düzgünleştirme Yöntemi ile sürekli bir fonksiyon haline dönüştürülmüş ve öncelikle oluşturulan bu 13 farklı fonksiyon ve ortalama fonksiyonu elde edilerek bunların değişim yapıları incelenmiştir. Bu çalışmada baz fonksiyon sayısı değerlendirme noktalarından daha fazladır, bu durumda eğrinin iniş çıkışını minimize eden ve pürüzlülük üzerinde sürekli bir kontrol sağlayan Pürüzlü Ceza Yöntemi her bir fonksiyonu tahminlemede kullanılmıştır. Daha sonra bu 408 tane değişkene (burada günler olarak ele alınmıştır) ait kovaryans yüzeyleri oluşturulmuş ve Pürüzlü Ceza Yöntemi ile tahminlenen katsayılarla Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi uygulanarak tüm fonksiyonlar birlikte ele alındığında gözlemlenebilmesi ve algılanabilmesi güç olan fonksiyonlar arasındaki değişim ortaya konulmaya çalışılmıştır. Bu çalışmada incelenen verilere üzerinde durulan yöntemlerin uygulanabilmesi amacıyla MATLAB 7.0 paket programı kullanılmıştır.

4.3. Analiz ve Bulgular: Uygulamada birinci adım B-Splayn baz fonksiyonları oluşturmaktır. Oluşturulan 410 tane B-Splayn baz fonksiyon Şekil (4.1) de verilmektedir.

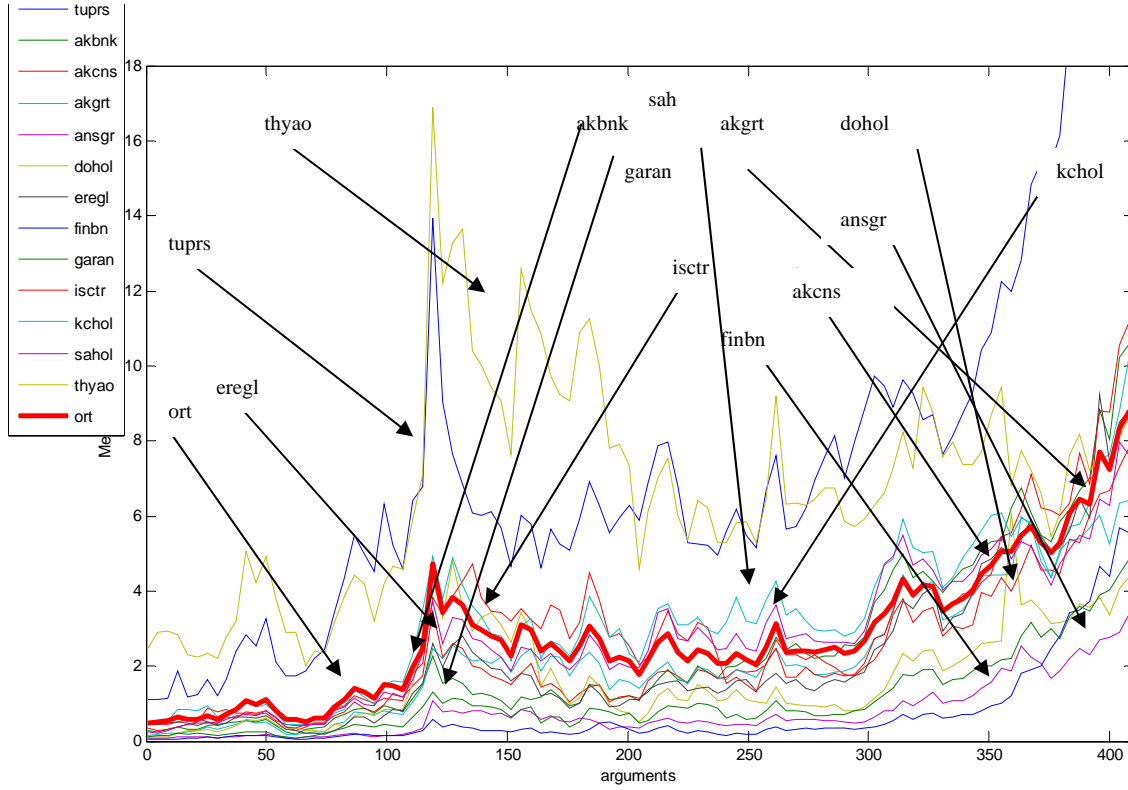


Şekil(4.1): 410 elemanlı B-Splayn Baz Fonksiyonlar Seti

Bu çalışmada,

$$\lambda=10^{-4} [\text{tr} (\mathbf{B}^T \mathbf{B}) / \text{tr} \mathbf{R}]$$

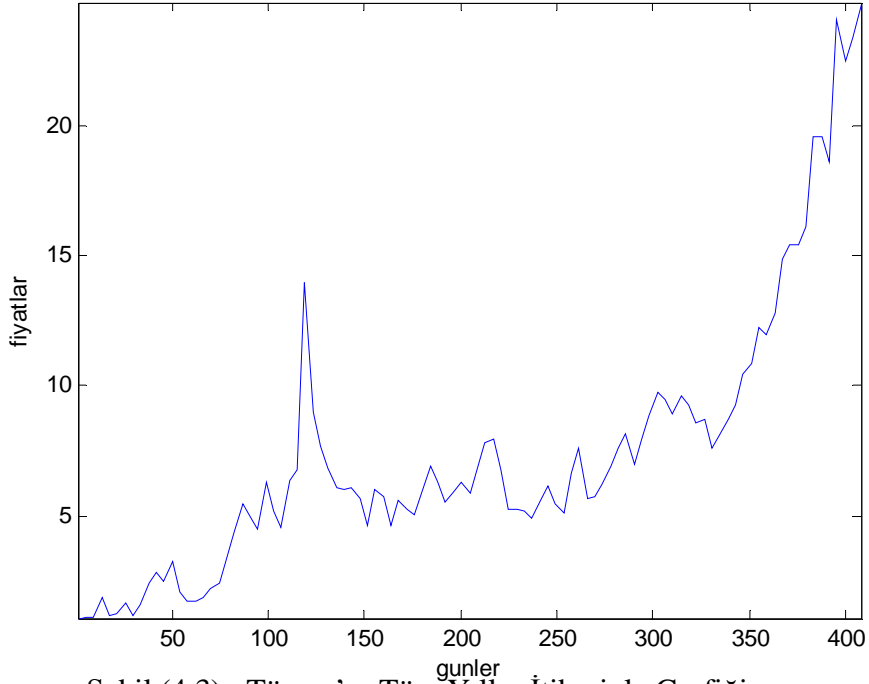
formülü ile düzenleme parametresi belirlenmiştir. Buradan düzenleme parametresinin değeri, $\lambda=10^{-4} \times 0.1769$ olarak elde edilmiştir. İkinci adımda her bir şirketin genel seyrini nokta işaretlemelere nazaran (kesikli verilere nazaran) çok daha belirgin bir şekilde gösteren ve aynı zamanda türevlerinin de incelenebilmesini mümkün kılan 13 tane bireysel fonksiyon elde edilerek bunlar Şekil (4.2) de verilmiştir.



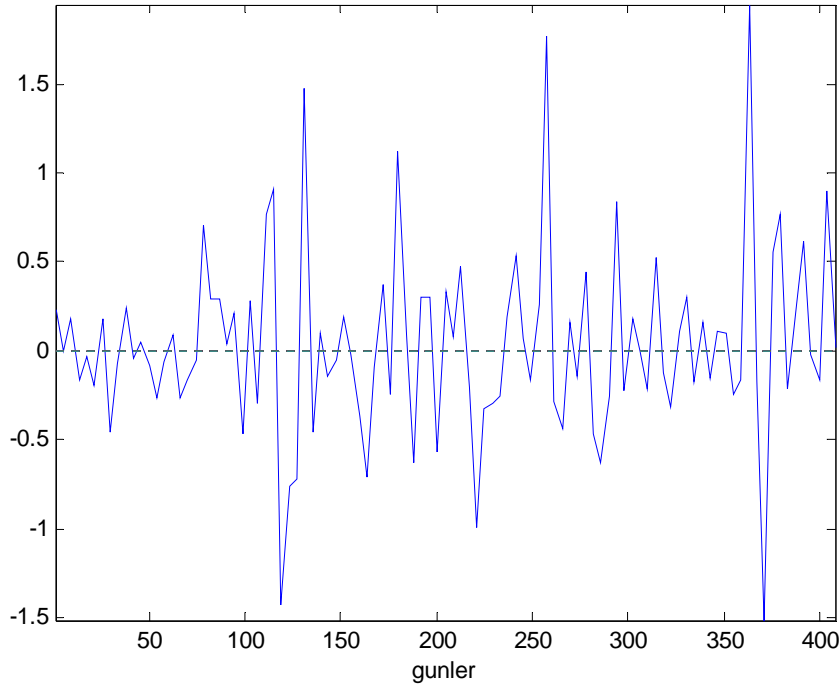
Şekil(4.2): 13 Bireysel Fonksiyon

Yukarıdaki çizimden de anlaşılacağı gibi birçok fonksiyonun bireysel davranışı diğer fonksiyonların arasında kaybolmuştur. Şekilde en rahat Tüpraş ve Türk Hava Yolları şirketlerinin seyri belirgin bir şekilde görülmektedir ki bunlar diğer hisse senetlerine göre kapamış fiyatları açısından daha yüksek seviyede seyretmektedirler. Bu şirketlerin işlem gördüğü fiyatlar da diğer şirketlerin fiyatlarına yakın olsaydı, bunların fonksiyonları da diğer fonksiyonların arasında yer alacaktı ve her zaman karşılaşılan genel bir zorluk olan yorumlama güçlükleri bu şirketler için de geçerli olacaktı. Burada ortalama fonksiyonu belirgin olması açısından diğerlerine göre daha kalın olarak çizilmiştir. Şekilden, incelenen şirket sayısı arttıkça görselliğin iyice karmaşık bir hale geleceği, fonksiyonların bireysel davranışlarının diğerlerinin arasında kaybolacağı ve bu nedenle de fonksiyonların bireysel ve genel seyirleri hakkında bilgi edinilmesinin zorlaşacağı açıkça görülmektedir. Verilen şekilde, 1-25 günleri arası 1997 yılına, 26-72 günleri arası 1998 yılına, 73-117 günleri arası 1999 yılına, 118-165 günleri arası 2000 yılına, 166-216 günleri arası 2001 yılına, 217-264 günleri arası 2002 yılına, 265-313 günleri arası 2003 yılına, 314 - 361 günleri arası 2004 yılına ve son olarak da 362- 408 günleri arası 2005 yılına karşılık gelmektedir.

Ayrıca elde ettiğimiz bu fonksiyonlar, bir diğer deyişle Kübik Splaynlar sayesinde birinci ve ikinci türev fonksiyonları da incelenebilir hale gelmektedir. İncelenen şirketler içerisinde örneğin Tüpraş ele alındığında fonksiyonun tek başına grafiği Şekil (4.3) ile verildiği biçimdedir.



Şekil (4.3) : Tüpraş'ın Tüm Yıllar İtibariyle Grafiği



Şekil (4.4) : Tüpraş'ın Birinci Türev Fonksiyonu

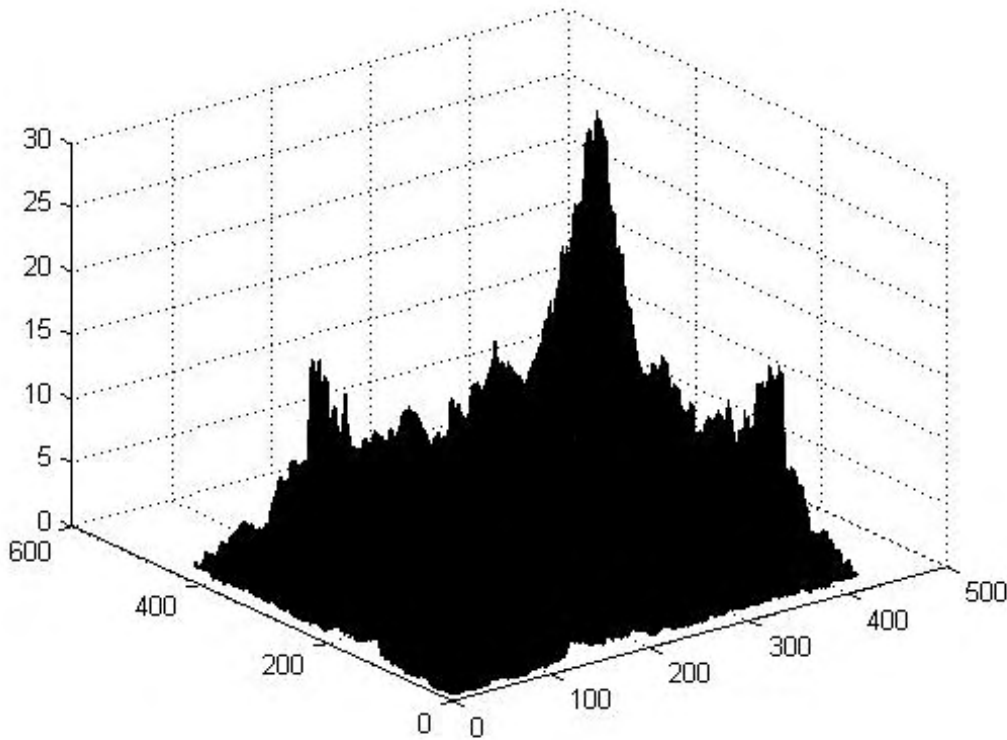
Tüpraş'ın birinci türev fonksiyonu da Şekil (4.4) de verilmektedir. Zamana göre birinci türev kapanış fiyatlarının artış ya da azalış hızını vermektedir. Bazı zaman aralıkları için asıl fonksiyonda fark edilmesi oldukça güç olan iniş ve çıkışlar, türevin özelliğinden dolayı, türev fonksiyonları incelendiğinde daha rahatlıkla görülebilmektedir. Şekil (4.4) ele alındığında türevlerin negatif ve pozitif olduğu bölgeler, bir diğer deyişle fonksiyonun artan ve azalan olduğu bölgeler ile türev fonksiyonunun sıfır değerini aldığı maksimum, minimum noktalar rahatlıkla gözlemlenebilmektedir. Ani iniş ve çıkışların gücü de türev fonksiyonunda daha rahatlıkla yakalanabilmektedir. Örneğin 26 Kasım 1999 da 113. gün civarındaki iniş, fonksiyondan da görülebilmesine karşın 8 Şubat 2002 ye karşılık gelen 221. gün civarındaki ani iniş türev fonksiyonunda daha net görülebilmekte ve bu iki inişin boyutu orijine göre daha rahatlıkla kıyaslanabilmektedir. Arzu edildiği takdirde Kübik Splaynın ikinci türev fonksiyonu da incelenebilir.

Kovaryans fonksiyonu farklı s ve t değerleri için gözlemlerin bağımlılık durumunu özetler ve tüm s ve t değerleri için,

$$\text{Cov}_x(s, t) = \frac{1}{N-1} \sum_{i=1}^N \{x_i(s) - \bar{x}(s)\} \{x_i(t) - \bar{x}(t)\} \quad s, t \in [a, b] \quad (4.1)$$

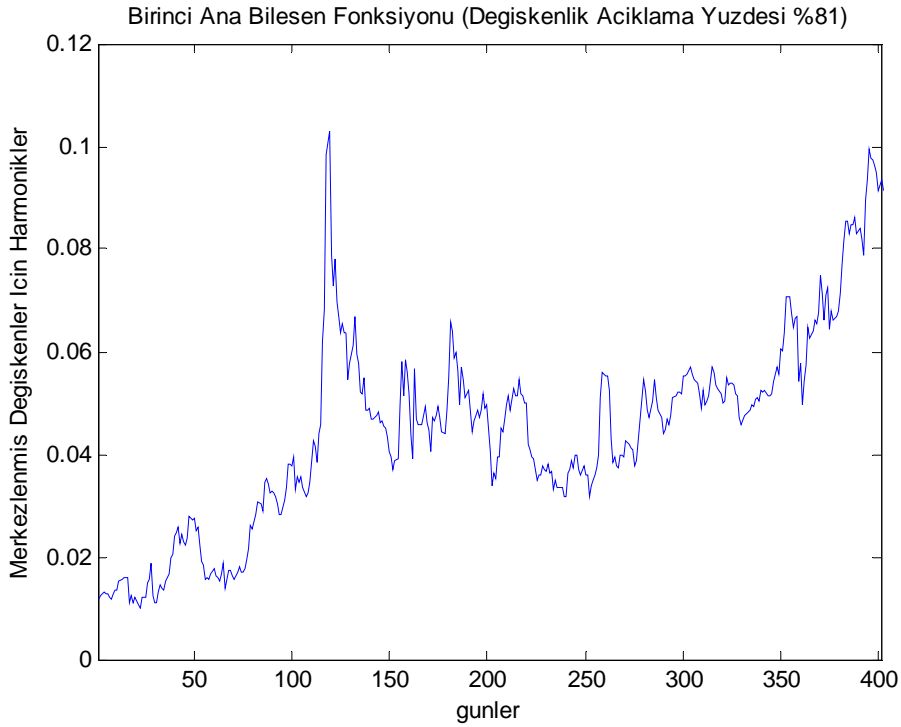
şeklindedir. $Cov_x(s,t)$ fonksiyonu için literatürde aynı anlamda olmak üzere $v(s,t)$ ifadesi de kullanılabilir.

$Cov_x(s,t)$ bir varyans-kovaryans yüzeyi tanımlar. Bu yüzeyin yüksekliği, zamanın (veya ilgili değişkenin) her bir noktasında eğrilerin değişkenliğini gösterir ve birlikte değişiminin ölçüsünü verir. Şekil (4.5) ile Pürüzlü Ceza Yöntemine göre düzgünleştirilmiş kovaryans yüzeyi verilmektedir.



Şekil(4.5): Kovaryans Yüzeyi

Bu kovaryans yüzeyi 408 ayrı nokta da 13 birey için inceleme yapıldığından dolayı oldukça karmaşıklaşmış ve gözlemlerin birlikte değişimleri gözlemlenemez ve açıklanamaz hale gelmiştir. Yüzey üzerinde hiçbir yükseklik neredeyse tanımlanamaz haldedir. Kovaryans yüzeyi incelendiğinde fonksiyonlar arasındaki değişimin belirlenmesinin oldukça güç hatta bu uygulama verileri için neredeyse imkansız olduğu görülmüş ve bu karmaşık yapının yorumlanmasını mümkün kılabilmek için ana bileşen fonksiyonları elde edilerek bunların yorumlanması yoluna gidilmiştir. Elde edilen birinci ana bileşen fonksiyonu Şekil(4.6) ile verilmektedir.



Şekil(4.6) : Birinci Ana Bileşen Fonksiyonu

En yüksek değişkenliği veren bu ana bileşen fonksiyonu incelendiğinde, ardışık inişler ve çıkışlar birbirini izleyen iki şans değişkeni (burada gün) arasındaki kovaryansın yönü tarafından belirlenir. Bu artış yada azalışların uzunluğu da kovaryansın mutlak değeri tarafından belirlenir, ardışık olarak ters ilişki varsa fonksiyon sifıra göre pozisyon değiştirir. Ayrıca sıfırdan y eksenini boyutundaki mutlak uzaklıklar değişkenin varyansı ile doğrudan ilişkilidir. Cauchy – Schwarz eşitsizliği gereği, ardışık değişkenlerin varyanslarının yüksekliği bunlar arasındaki kovaryansın yüksek bir değer alabilmesini mümkün kılmaktadır. Burada şekilden de anlaşıldığı gibi fonksiyon sıfırın üzerinde seyrettiği için tüm ardışık kovaryans değerleri pozitifdir. Fonksiyonda sıfırdan mutlak uzaklıklar varyanslar ile doğrudan orantılı olduğuna göre, bu şekilden özellikle 2000 yılının Ocak ayının ilk yarısında (07.01.2000 tarihi 118.güne karşılık gelmektedir) ve 360ıncı günden bir diğer deyişle 2005 yılından itibaren şirketlerin fiyatlarının değişkenliğinde bir artış olduğu ve zaman noktalarının ardışık olarak birbirleriyle pozitif bir korelasyona sahip olduğu gözlemlenebilmektedir. Görüldüğü gibi verilen kovaryans yüzeyi ve ana bileşen fonksiyonu kıyaslandığında ana

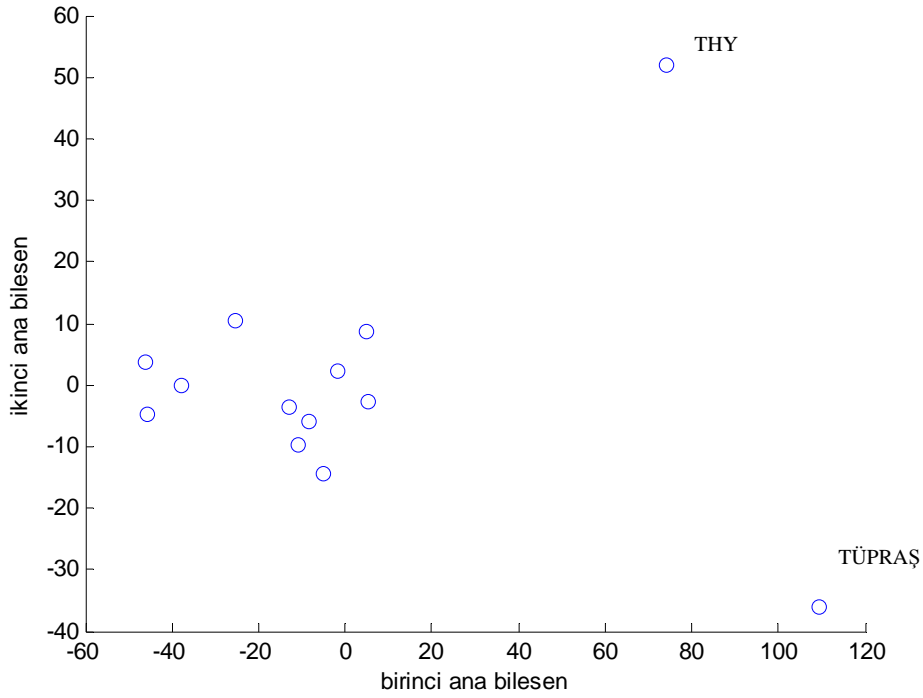


bileşen fonksiyonu çok daha rahatlıkla yorumlanabildiği açıktır. Tüm bunlara ilaveten birinci ana bileşen fonksiyonu %81 gibi yüksek bir değişkenlik açıklama yüzdesine sahiptir.

Ana bileşen fonksiyonu incelendikten sonra özellikle bu fonksiyonun yorumlanmasının güç olduğu durumlarda alternatif yardımcı yöntem olarak kullanılan ana bileşen skorları Tablo (4.1) de ve bunların dağılımı Şekil (4.7) de verilmektedir.

Tablo (4.1): Ana Bileşen Skorları

	Ana Bileşen 1	Ana Bileşen 2
Tüpraş	109.2795	-36.0090
Akbank	-4.8971	-14.5288
Akcansa	-13.0615	-3.6060
AkSigorta	-8.2390	-6.0059
Anadolu Sigorta	-45.9984	3.8401
Doğan Holding	-25.5378	10.4437
Ereğli Demir Ç.	-10.9511	-9.6925
Finansbank	-45.8305	-4.6457
Garanti Bankası	-37.6974	-0.0672
İş Bankası C	5.2674	-2.5818
Koç Holding	5.0913	8.6581
Sabancı Hol.	-1.7828	2.2484
Türk Hava Yol.	74.3574	51.9466

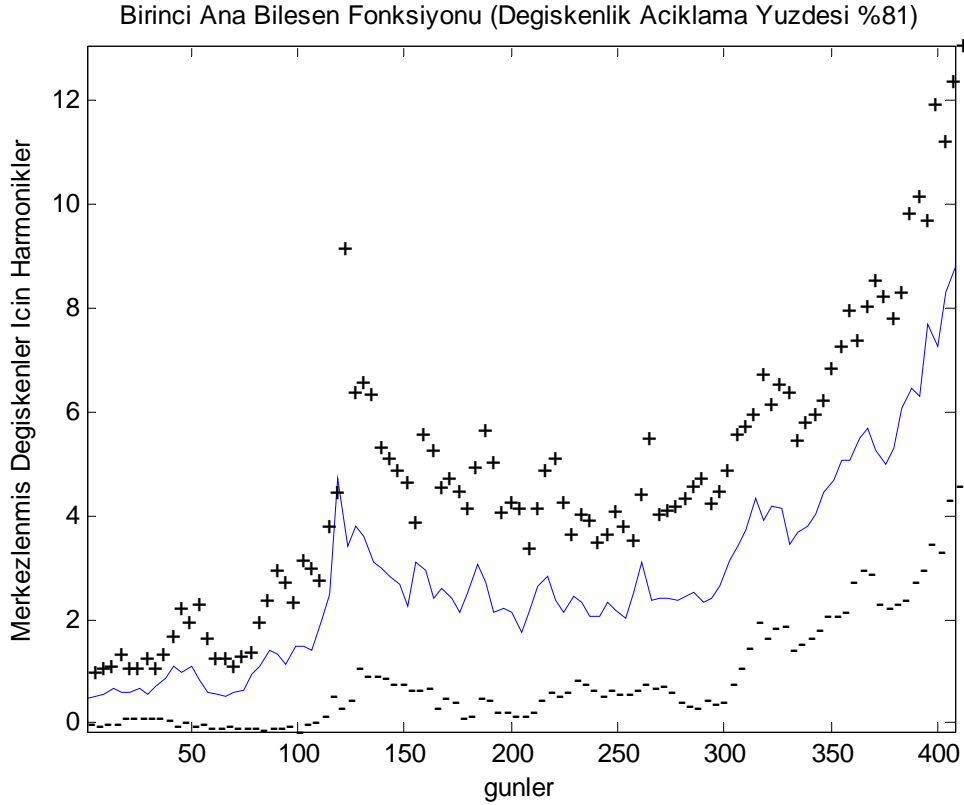


Şekil(4.7): Ana Bileşen Skorlarının Dağılımı

Ana Bileşen Skorları özellikle %81 varyans açıklama gücüne sahip birinci ana bileşen açısından incelendiğinde Türk Hava Yolları ve Tüpraş şirketlerinin diğerlerine oranla farklı bir davranış sergiledikleri görülmüştür. Bunlar sapan gözlem olarak da adlandırılabilir. Diğerlerine oranla ana bileşen skorları açısından daha yüksek değer taşımaktadırlar. Bunun nedeni fiyatlarının diğer şirketlere oranla tüm yıllar boyunca ayrıca değişkenliklerin yüksek olduğu zaman aralıkları da dahil olmak üzere daha yüksek seviyede seyretmesidir.

Son olarak da yine ana bileşen fonksiyonunun yorumlama güçlüğü olduğu durumlarda ana bileşen fonksiyonunun yorumlanmasına yardımcı olması açısından ortalama fonksiyonu ile ana bileşen fonksiyonunun bir çarpanının karşılaştırılması ele alınmıştır. Ortalama fonksiyonu incelendiğinde bu çalışmada dikkate alınan İMKB-30 endeksine dahil yüksek ağırlıklı 13 tane şirketin fiyatlarındaki artış yada azalışların genel olarak büyük ölçüde bu fonksiyon tarafından yansıtıldığı gözlemlenmektedir. + ve - noktalar ortalama fonksiyonuna sırasıyla ana bileşen fonksiyonunun bir çarpanının eklenmesinin ve çıkarılmasının etkilerini göstermektedir. Şekil (4.8) ile verilen bu karşılaştırma ile ana bileşen fonksiyonuna benzer şekilde başlarda ortalama fonksiyonundan sapmaların daha düşük olduğu ve ana bileşen fonksiyonunda olduğu gibi özellikle 2000 yılının Ocak ayının başlarında ve 2005 yılından

itibaren ortalamalardan sapmalarda bir diğer deyişle şirketlerin fiyatlarının değişkenliğinde bir artış olduğu görülmüştür.



Şekil (4.8) : Ana Bileşen ve Ortalama Fonksiyonu Karşılaştırması

5. Tartışma ve Sonuç

Günümüzde hızla ilerleyen teknolojik gelişmeler sayesinde artık çoğu durumda klasik çok değişkenli istatistiksel yöntemler elde edilen verileri analiz etmek için yetersiz kalmaya başlamıştır. Bu çalışmada da incelenen gözlem noktası sayısı araştırmaya alınan şirket sayısından büyük olduğu, bir diğer deyişle, örnek hacmi değişken sayısından küçük olduğu için standart Çok Değişkenli Ana Bileşenler Analizi uygulanamaz. Gözlemler ilgilenilen veri aralığı genişledikçe bir sayı dizisi olarak görülmektense, zamanın veya yakın ilgili değişkenin bir fonksiyonu olarak görülmeye başlanmıştır. Bu amaçla bu çalışmada, özellikle Ramsay ve Silverman (1997) den sonra bir ivme kazanan ve fonksiyonel veriler için geleneksel veri analizlerinin bir uzantısı olan Fonksiyonel Veri Analizi tanıtılmaya çalışılmıştır. Fonksiyonel Veri Analizinde özellikle bilgisayar uygulamaları ve sayısal istikrar sağlaması açısından uygulamalarda sıkça tercih edilen bir yaklaşım olan Baz Fonksiyon Yaklaşımı ve bu yaklaşımdan B-Splaynlar araştırmacılara yardımcı olması açısından detayları ile açıklanmıştır. Fonksiyonel Veri Analizi Yöntemleri olarak günümüze kadar olan çalışmaların



çoğu Fonksiyonel Ana Bileşenler Analizi, Fonksiyonel Doğrusal Modeller, Fonksiyonel Kanonik Korelasyon ve Diskriminant Analizini içermektedir. Bu çalışmada Fonksiyonel Veri Analizi için anahtar teknik olan ve incelenen olayların sayısı arttıkça sistemde bulunması beklenen veya araştırmacı açısından sürpriz olarak nitelendirilebilecek durumları ortaya çıkarmadaki yeteneği ile çoğu uygulayıcının can kurtaranı olarak kullanılan Fonksiyonel Ana Bileşenler Analizi, Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi ve bu analizlerde baz fonksiyon yaklaşımı ele alınmıştır. Baz fonksiyon yaklaşımlarının ve dolayısıyla da analizlerin anlaşılmasına ve bilgisayar algoritmalarının oluşturulmasında araştırmacılara yardımcı olması amacıyla gerekli algoritmalar oluşturulmuştur.

Fonksiyonel Veri Analizinin gittikçe öneminin artmasının nedenleri, önceden de bahsedildiği üzere gittikçe ilerleyen teknolojiyle birlikte elde edilen verilerin analizi için gerekli yöntemlerin yetersiz kalması, Fonksiyonel Veri Analizinin interpolasyon ile düzensiz örneklenen fonksiyonlarla ve kayıp verilerle uğraşılmasına imkan vermesi ve düzgünleştirme ile oluşturulan fonksiyonların türevlerinin de incelenebilmesi gibi görsel olarak da çoğu konuda araştırmacılara yardımcı olması ve bu açıdan veri analizine yeni bir bakış açısı getirmesidir. Fonksiyonel Veri Analizinde özellikle Ana Bileşenler Analizinde esas olan görselliktir. Fonksiyonel Ana Bileşenler Analizi ve Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi ile Klasik Ana Bileşenler Analizi karşılaştırıldığında, Klasik Ana Bileşenler Analizinin veri yapısı nedeniyle uygulanmasının uygun olmadığı durumlar haricinde bu iki yöntem arasındaki temel fark, bireylere ait fonksiyonlar, ortalama fonksiyonu, kovaryans yüzeyleri, ana bileşen fonksiyonları ve elde edilen fonksiyonların türevlerinin görsel olarak da incelenebilmesi ve bir çok görünmeyeni ortaya çıkarmadaki yeteneğidir.

Bu çalışmada uygulama olarak İMKB verileri ele alınmıştır. İMKB de gerçekleşen veriler anlık olarak elektronik ortamda saklandığından ve güvenilirlik sorunu olmadığından, ayrıca araştırma için uygun görülen zaman aralığındaki verilere ulaşılması mümkün olduğundan tercih edilmiştir. Bilindiği gibi İMKB de şirketlere veya endekslere ait fiyat zaman grafikleri ele alınan zaman periyodu için [saat, gün, hafta, ay] kesikli verilere dayalı olarak grafikler oluşturularak araştırmacıya veya ilgiliye sunulmaktadır. Doğal olarak bu fonksiyonların herhangi bir şekilde türev yapıları matematiksel olarak incelenemez, çünkü elimizde gerçek bir matematiksel fonksiyon yoktur. Bu yaklaşım ile istenilen dereceden türevlenebilir matematiksel fonksiyonların (Splaynlar) oluşturulması mümkün olmaktadır.



Uygulamada, İMKB 100 endeksine dahil onüç tane şirketin Cuma günleri için haftalık kapanış fiyatları ele alınmıştır. Bu şirketlerden Tüpraş'ın türev fonksiyonu da elde edilerek incelenmiştir. Türev fonksiyonunun incelenmesi ile bireysel fonksiyonlarda gözlenmesi oldukça güç olan çok küçük iniş çıkışların bile rahatlıkla yakalanabildiği, türevlerin negatif ve pozitif olduğu bölgeler ile türev fonksiyonunun sıfır değerini aldığı asıl fonksiyonun maksimum ve minimum noktaları görsel olarak rahatlıkla gözlemlenebilmektedir. Ayrıca kapanış fiyatlarının ana bileşen fonksiyonlarının incelenmesi ile fonksiyonların görünümünden ve kovaryans yüzeyinden ortaya çıkarılması oldukça güç olan zaman noktaları arasındaki değişim ortaya konulmuş ve bunun altındaki teorik yapı da belirtilmiştir. İncelenen 13 şirket için özellikle 2000 yılının başlarında ve 360 ıncı günden bir diğer deyişle 2005 yılından itibaren fiyatların değişkenliğinde bir artış olduğu ve zaman noktalarının ardışık olarak birbirleriyle pozitif bir korelasyona sahip olduğu sonucuna ulaşılmıştır.



Kaynaklar

- Barra V. (2004) Analysis of Gene Expression Data Using Functional Principal Components, *Computer methods and programs in biomedicine*, 75(11).
- Benko M. (2004). *Functional Principal Components analysis, Implementation and Applications*. A Master Thesis. Humboldt University Center of Applied Statistics and Economics, Berlin.
- Benko M., Hardle W., Kneip A.(2006). *Common Functional Principal Components*, SFB 649, Discussion Paper, Erişim:10.11.2006, <http://ideas.repec.org/p/hum/wpaper/sfb649dp2006-010.html>
- Besse P., Ramsay J.O. (1986). Principal Components Analysis Of Sampled Functions, *Psychometrica*, 51(2)
- Boor C . (1978). *A Practical Guide to Splines*. Springer-Verlag: New-York
- Castro P. E, Lawton W. H., Sylvestre E. A. (1986). Principal Modes Of Variation for Processes with Continuous Sample Curves, *Technometrics*, 28(4).
- Costanzo G.D. (2005). *Functional Principal Component Analysis of Financial Time Series*, Cnam-Paris.
- Dauxois J., Pousse A., Romain Y. (1982). Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference, *J. Multivariate Analysis*, 12
- Dierckx P. (1993). *Curve and Surface Fitting with Splines*. Oxford University Press:New York
- Eubank R.L. (1999). *Nonparametric Regression and Spline Smoothing*. Marcel Dekker: USA.
- Green.P.J., & Silverman B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman & Hall:London.
- Hall P. , Nasab H. M. (2006). On Properties Of Functional Principal Components Analysis. *Journal of the Royal Statistical Society: Series B*, 68(1).
- Hyde V., Moore E., & Hodge A. (2006). *Functional Pca For Exploring Bidding Activity Times for Online Auctions*. Erişim:06.07.2006, <http://www.rhsmith.umd.edu/ceme/statistics/functionalpca.pdf>
- James G. M., Hastie T.J., Sugar C.A. (2000) Principal Components Models For Sparse Functional Data, *Biometrika*, 87(3).
- Jank W., & Shmueli G. (2006). Functional Data Analysis in Electronic Commerce Research. *Statistical Science*, 21(2).
- Jones M. C., Rice J. A. (1992). Displaying The Important Features Of Large Collections Of Similar Curves. *The American Statistician*, 46(2).
- Laukaitis A., Rackauskas A. (2002). Functional Data Analysis of Payment Systems. *Nonlinear Analysis: Modelling and Control*, 7(2).
- Lee H.J. (2004). *Functional Data Analysis: Classification and Regression*. Doctor of Philosophy, Texas A&M University.
- Leurgans S.E., Moyeed R.A, Silverman B.W. (1993). Canonical Correlation Analysis when the Data are Curves. *Journal of the Royal Statistical Society: Series B.*, 55(3).
- Lillestol J., & Ollmar F. (2003). *Functional Data Analysis: Introduction and Applications to Financial Electricity Contracts*. Erişim: 01.03.2005, <http://www.nhh.no/for/dp/2003/0603.pdf>
- Lober E.M., & Villa C. (2004). *Functional Principal Component Analysis of the Yield Curve*, Erişim: 05.05.2005 , http://www.u-cergy.fr/AFFI_2004/IMG/pdf/MATZNER.pdf



- Lyche, T., Morken, K. (2002). *Spline Methods Draft*, Erişim: 10.07.2005, <http://www.ifi.uio.no/in329/nchap1.pdf>
- Musayev B., & Alp M. (2000). *Fonksiyonel Analiz*. Balcı Yayınları: Kütahya
- Müler, H. G. (2005). Functional Modelling and Classification of Longitudinal Data, *Scandavian Journal of Statistics*, 32(2).
- Nürnberg G. (1989). *Approximation by Spline Functions*. Springer – Verlag: Berlin.
- Ramsay, J. O., Dalzell C. (1991). Some Tools For Functonal Data Analysis, *Journal of the Royal Statistical Society: Series B.*,53 (3)
- Ramsay J.O., Silverman B.W. (1997). *Functional Data Analysis*. Springer – Verlag: New York.
- Ramsay J. O. , Li X. (1998). Curve Registration, *Journal of the Royal Statistical Society: Series B*, 60(2).
- Ramsay J.O., (2000). *Basis Functions*, Erişim:11.04.2005, <ftp://ego.psych.mcgill.ca>
- Ramsay J.O, Silverman B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer – Verlag: New York.
- Ramsay J.O., Silverman B.W. (2005). *Functional Data Analysis. Second Edition*. Springer : USA
- Rao, C. R. (1958). Some Statistical Methods for Comparison of Growth Curves. *Biometrics*, 14(1).
- Reinsch C.H. (1967). Smoothing By Spline Functions. *Numerische Mathematik*, 10.
- Schumaker L.L. (1993). *Spline Functions: Basic Theory*. Krieger Publishing Company: Florida
- Silverman B.W. (1985). Some Aspects Of The Spline Smoothing Approach To Non-Parametric Regression Curve Fitting. *Journal of the Royal Statistical Society: Series B.*, 47(1).
- Silverman B. W. (1995). Incorporating Parametric Effects into Functional Principal Components Analysis *Journal of the Royal Statistical Society: Series B.*,57(4).
- Silverman B.W.(1996). Smoothed Functional Principal Component Analysis By Choice Of Norm. *The Annals of Statistics*, 24(1)
- Simonoff J.S. (1996). *Smoothing Methods in Statistics*, Springer – Verlag: New- York.
- Ulbricht J.(2004). *Representing Functional Data as Smooth Functions*. A Master Thesis, Humboldt University Institute of Statistics and Econometrics, Berlin.
- Wittaker E.T. (1923). On a New Method of Graduation, *Proc. Edinburg Mathematics Society*, 41.
- Yamanishi Y., & Tanaka Y. (2005). Sensitivity Analysis in Functional Principal Component Analysis, *Computational Statistics*, 20(2).
- Yamanishi Y., (2004). *Statistical Case Studies: Biostatistics and Geostatistics*, Erişim: 01.07.2005, <http://www.quantlet.com/mdstat/scripts/xcs/pdf/xcspdf.pdf>
- Yao F., Lee T.C.M. (2006). Penalized Spline Models For Functional Principal Component Analysis, *Journal of the Royal Statistical Society: Series B.*, 68(1).
- Zhang, J. T. (1999). *Smoothed Functional Data Analysis*, Doctor of Philosophy, University of North Carolina, 1999.