

GİRİŞ

Çoğu istatistiksel analiz bir örnekteki bireylerin her birinden alınan bir veya daha fazla gözlem içerir. Burada amaç örneğin seçildiği popülasyonla ilgili yorumlamalarda bulunmaktır. Çalışma sahası genişledikçe bu gözlemler eğriler (curves) veya (surface) yüzeyler halini alır. Bir düzlemin bir parçasında veya bir doğruda her bir noktada gözlenen bir yoğunluk mevcut olduğundan dolayı gözlenen eğriler veya yüzeyler “fonksiyonel veri” olarak adlandırılır. Çoğu durumda gözlemler zamanın veya yakın ilgili bir değişkenin fonksiyonudurlar. Fakat aynı zamanda iki veya üç boyutlu uzayda gözlenen fonksiyonların yüzeyler olduğu uygulamalar da bulunmaktadır. Fonksiyonel verilerin alındığı popülasyonlar hakkında yorumlar yapabilmek amacı ile bilimsel teknolojideki gelişmeler istatistiksel analizin gelişimini gerektirir. Bu tip verileri analiz etmek için metodlar Ramsay ve Dalzell (1991) tarafından adlandırılan “Fonksiyonel Veri Analizi” (Functional Data Analysis) terimi ile adlandırılır. Bu çalışma kapsamında “fonksiyonel veri” olarak tek boyutlu veriler olan eğrisel veriler içermektedir. Fonksiyonel veri analizinin temel felsefesi gözlenen veri fonksiyonlarının birbirini ardına gelen bireysel gözlemlerin bir sırası olarak değil de tekil girişler olarak ele alınmasıdır. Bir diğer deyişle, fonksiyonel veri analizinde girilen veriler bir sayı dizisi olarak görülmeğe asıl ilgilenilen tüm bir fonksiyondur.

Fonksiyonel veri analizinde, veri matrisi ya da p değişkenli şans örneğinde yer alan i .inci gözlem aslında $x_i(t)$ ($i=1,2, \dots, N$) biçiminde reel bir sürekli fonksiyon olmasına rağmen, veriler genellikle birbirlerinden ayrık noktalarda gözlemlenir. N örnek büyüklüğü (hacmi), n_i ise 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, $y_{i1}, y_{i2}, \dots, y_{i(n_i-1)}, y_{in_i}$ 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 özetlenebilir. Klasik çok değişkenli kavramda ise üzerinde çalışılan ölçümler $\{y_{ij}; i=1, 2, \dots, N, j=1, 2, \dots, p\}$ biçiminde belirtilen bir veri matrisidir.

Fonksiyonel veri analizinin 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.

Her bir fonksiyonel gözlem aynı noktalarda gözlendiğinde çok değişkenli veriye benzemekte ise de fonksiyonel veri çok değişkenli veriden en az aşağıdaki noktalarda farklılıklar göstermektedir:

- 1) Bir fonksiyonel veri analizinde, (y_{ij}, t_{ij}) biçiminde gözlenen veri temelde yer alan düzgün fonksiyondan örneklenmiştir. Çok değişkenli veri setinin gözlenen vektörü için böyle bir yapı yoktur.
- 2) Bir fonksiyonel gözlemin boyutu n_i sıklıkla o kadar büyüktür ki bazen fonksiyonel gözlem sürekli bir fonksiyon olarak dikkate alınabilir. Ayrıca boyut genellikle örnek hacminden bir diğer deyişle incelenen birey veya gözlem sayısından büyük olur. Bu durum $n_i > N$ şeklinde ifade edilebilir. Burada ilgili kovaryans matrisi tam ranklı olmadığından dolayı standart çok değişkenli yöntemler başarısızlığa uğrayabilir.
- 3) Gözlem noktaları t_{ij} ($j=1, 2, \dots, n_i$) bir fonksiyonel gözlemden diğerine değişebildiğinden dolayı bu durumda doğrudan çok değişkenli yöntemleri uygulamak mantıklı veya mümkün olmayabilir.

Bir fonksiyonel gözlemin diğerinden bağımsız olduğu varsayılabilir, fakat aynı fonksiyonel veri içerisinde farklı noktalarda gözlenen değerlerin bağımsızlığı gerekli değildir. Bir diğer deyişle, ilgilenilen aralıkta bir zaman noktasında değerlendirilen gözlemin aynı fonksiyonel veri içerisinde diğer noktadan bağımsız olma şartı yoktur.

Fonksiyonel veri analizi sıklıkla zamana ait verilerle ilgilenmesine rağmen, faaliyet alanı ve amaçları zaman serisi analizlerinden farklıdır. Zaman serisi daha çok verilerin modellenmesine ve gelecek gözlemlerin tahminlenmesine odaklanır. Fonksiyonel Veri Analizindeki teknikler ise aslında doğasında açıklayıcıdır, vurgu yörüngelerde ve şekillerdedir(Costanzo, 2005).

Fonksiyonel veri analizinin amaçları aslında istatistiğin diğer branşlarıyla benzerdir ve aşağıdaki biçimde özetlenebilir (Ramsay & Silverman, 2005; 9):

- 1) İleriki analizlere yardımcı olacak biçimde verileri ifade etmek ve dönüştürmek.
- 2) Çeşitli karakteristikleri aydınlatacak biçimde verileri göstermek. Eğriler göz önünde canlandırıldıktan sonra gerekli ise ortalama fonksiyonu ve fonksiyonun türevlerinin gösterimi ile ilgilenmek.
- 3) Veriler arasındaki değişkenliği ve modelin önemli değişkenlik kaynaklarını çalışmak ve bunları açıklamak
- 4) Bazı değişimlerle ilgili olarak birkaç veri setini karşılaştırmak
- 5) Girdi veya bağımsız değişken bilgisini kullanarak çıktı veya bağımlı değişkendeki değişimi açıklamak.

Modern ekipmanlarla, büyük veri setleri kolaylıkla eğriler bir diğer deyişle fonksiyonel veriler formunda taranabilir. Veri setleri olağandışı değildir. Depremlerin sistmik kayıtları, nükleer patlamalar, yürüyüş analizleri, sıcaklık ve yağış verileri, büyüme eğrileri gibi tıpta birçok veriler, finansal veriler, öğrenme eğrileri ve konuşma sırasındaki konuşma üretimi ölçümleri v.b. fonksiyonel verilere örnek oluşturabilir. Fonksiyonel veri analizi için çalışmalar uzun zamandır bulunmakla birlikte, araştırmalardaki dalgalanma ve bir çok uygulama alanındaki ivme Ramsay ve Silverman (1997) çalışmasından sonra başlamıştır. Bu ivmenin bir nedeni de bilgisayar depolama ve bilgisayarın gücündeki teknolojik artıştır. Bu sayede araştırmacılar günden güne otomatik olarak çok daha fazla veri toplayabilir ve bunları büyük veri tabanlarında depolayabilirler. Bununla birlikte, veri üretimi ve veri depolama için artan kapasiteler klasik istatistiksel kavramlara uymayan ve

fonksiyonel veri olarak adlandırılan yeni veri yapılarına neden olmuştur. Bu yeni veri yapıları da taşıdıkları bilgiyi açmak için Fonksiyonel Veri Analizi olarak adlandırılan yeni istatistiksel metodlar gerektirir.

Fonksiyonel perspektiften veri analizini ele almanın pratik nedenleri aşağıdaki biçimde özetlenebilir (Ramsay & Dalzell, 1991):

- 1) Otomatik veri toplama metodları araştırmacılar için daha ulaşılabilir olduğundan beri fonksiyonel gözlemler uygulama çalışmalarında daha çok ortaya çıkmışlardır. Ayrıca düzgünleştirme ve interpolasyon gözlemlerin sonlu setlerinin fonksiyonel gösterimlerini sağlar.
- 2) Gözlemler sonlu sayıda olmasına rağmen bazı modelleme problemlerini fonksiyonel terimlerle düşünmek daha doğaldır.
- 3) Bir analizin amaçları doğal olarak fonksiyonel olabilir. Sonlu veri tüm fonksiyonu tahminlemede kullanılıyorken ilgilenilen diğer fonksiyonlar bu fonksiyonların türevleri v.b. olabilir.
- 4) Fonksiyonel prosesten çıkan düzgünleştirme gibi kavramları ele almanın çok değişkenli veri içi analizlerde gizli bir anlamı olabilir.

Fonksiyonel veriler, bir aralığın $T=[a,b]$ biçiminde sonlu altsetinde değerlendirilen eğrilerden veya gözlenen fonksiyonlardan oluşan verilerdir. Kavramsal anlamda sürekli olarak tanımlandıkları düşünülür. Fonksiyonel verilerin doğasından dolayı, bu tip verileri modellemek Hilbert uzayları gibi fonksiyon uzaylarını ele almayı gerektirir. Bu nedenle x_1, x_2, \dots, x_N biçimindeki fonksiyonel verilerin analizi genel olarak

$$x_i \in H \quad i = 1, 2, \dots, N$$

biçiminde belirtilen bir tam iç çarpım (complete inner product) veya bir diğer deyişle Hilbert uzay yapısını varsayar. Burada H reel bir aralık olan $T=[a,b]$ de tanımlanan fonksiyonlardan oluşan bir vektör uzayıdır.

Genellikle karesinin integrali alınabilir fonksiyonlardan oluşan Hilbert Uzayı ile çalışılır ve bu uzayda tanımlı x fonksiyonunun normu,

$$\|x\| = \langle x, x \rangle^{1/2} = \left(\int_a^b x(t)^2 dt \right)^{1/2}$$

biçimindedir. Eğer fonksiyonlar sonlu normlara sahip olacak biçimde sınırlandırılıyorsa sonuçlanan uzay Hilbert uzayıdır. Bu şekilde karesinin integrali alınabilir fonksiyonlar uzayında ,

$$\langle x, g \rangle = \int_a^b x(t) g(t) dt$$

şeklinde iki fonksiyonun çarpımının integrali bir iç çarpımdır(Ramsay; 1982).

Bu çalışmada örneklenen fonksiyonların, $H^d(t)$ ile belirtilen fonksiyonların vektör uzayında buldukları varsayılır. $H^d(t)$ ile belirtilen, Hilbert uzayı H de tanımlı bir reproducing Kernel Hilbert uzayıdır. Bu fonksiyonlar $T=[a,b]$ aralığında tanımlıdır ve $d-1$ sürekli türeve sahiptirler. d .inci türevin karesi T üzerinde sonlu bir Lebesgue integraline sahiptir. Bu şu demektir, d .inci türev T deki sayılabilir noktalarda süreksiz olabilmesine rağmen, daha düşük dereceli türevler sadece sürekli değil aynı zamanda türevlenebilirdir(Besse & Ramsay; 1986, Ramsay & Dalzell; 1991).

Hilbert uzayları ile ilgili ayrıntılı bilgi için Kreyszig (1989; 127-201), Musayev ve Alp (2000; 98-115) ve He (1999) çalışmalarına başvurulabilir.

Fonksiyonel Veri Analizi için özellikle önemli bir konu boyut indirgemedir. Ana Bileşenler Analizi gibi klasik çok değişkenli analizin ispatlanmış metodları Fonksiyonel Veri Analizinin bazı göze çarpan teknikleri için omurgayı oluşturur. Fonksiyonel Veri Analizi çok değişkenli analizin yanında düzgünleştirmeyi, özellikle Pürüzlü Ceza Yaklaşımı gibi parametrik olmayan regresyonu, Hilbert uzayları ve

ilgilenilen özellikleri gibi fonksiyonel analizi içerir. Bu çalışma kapsamında tüm bu konular üç bölümde ayrıntıları ile açıklanmıştır.

Fonksiyonel veri analizindeki ilk adım her bir fonksiyonun ilgilenilen $T=[a,b]$ aralığında t 'nin tüm değerleri için değerlendirilebilmesi amacıyla, verilerin fonksiyonel forma dönüştürülebilmesidir. Bu şekilde elde edilen fonksiyonlardan genellikle fonksiyonel veri objeleri olarak bahsedilir. Bu dönüşüm verileri farklı yollarla interpolate ederek (interpolation) veya düzgünleştirerek (smoothing) elde edilir.

Verileri interpolate etmede veya düzgünleştirmede en çok kullanılan yaklaşım eldeki probleme uygun, B-Splaynlar gibi bilgisayar dostu bir gösterimi olan baz fonksiyonlar seti seçmektir. Bu yolun çok boyutlu verilerin temel çok değişkenli analizleri üzerinde birçok sayısal avantajı vardır. Sabit bir bazın seçimi öncül bir bilginin girilmesine imkan verir. Örneğin periyodik veriler için Fourier baz girilebilir. Baz fonksiyon yaklaşımı düzensiz örneklenen fonksiyonlarla ve kayıp verilerle uğraşılmasına imkan verir. Bu gösterimin bir diğer etkisi bu gösterim verilerin bireysel veya Fonksiyonel Ana Bileşenler Analizinde ana bileşen fonksiyonlarının elde edilmesinde olduğu gibi topluca düzgünleştirilebilmesine imkan vermesidir. Bir diğer ilginç nokta, bu çalışmada da incelenen En Küçük Kareler ve Pürüzlü Ceza Yaklaşımı gibi, çoğu Fonksiyonel Veri Analizi metodu baz açılımın sayısal katsayıları üzerine çalışır ki bu durum daha az hesaplama sıkıntısı yaratır. Bu fonksiyonlarla çalışmanın bir diğer avantajı da önceden bahsedildiği gibi türev ve integral gibi fonksiyonel preprocessing (önişleme) kullanabilmenin mümkün olmasıdır(Rossi v.d.; 2005). Bu amaçla çalışmanın birinci bölümünde ilk olarak verileri düzgünleştirmede kullanılan baz fonksiyon yaklaşımı ve daha sonra baz açılımın sayısal katsayılarına odaklanan En Küçük Kareler ve Pürüzlü Ceza Yaklaşımı ve bu yaklaşımların ağırlıklandırılmış durumlarına odaklanılmıştır. Bu bölümde ayrıca fonksiyonel veriler için ortalama, kovaryans ve korelasyon fonksiyonlarına ve bunların baz fonksiyon yaklaşımlarına değinilmektedir.

Günümüze kadar yapılan çalışmalarda fonksiyonel verilerin analizi için Fonksiyonel Ana Bileşenler Analizini, Fonksiyonel Doğrusal Modelleri, Fonksiyonel Kanonik Korelasyon ve Diskriminant Analizini içeren birkaç metod ele alınmıştır. Bu çalışmada ikinci bölümde uygulamada sıklıkla kullanılan ve Fonksiyonel Veri Analizi için anahtar teknik olarak adlandırılan Fonksiyonel Ana Bileşenler Analizi, Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi ve bu iki yöntem için baz fonksiyon yaklaşımı incelenmiştir. Son olarak Fonksiyonel Ana Bileşenler ve Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinde baz fonksiyon yaklaşımının anlaşılmasına yardımcı olması açısından gerekli algoritmalar oluşturulmuştur.

Üçüncü bölümde Fonksiyonel Veri Analizinin İstanbul Menkul Kıymetler Borsasında (İMKB) bir uygulamasına değinilmektedir. Bu amaçla öncelikle 2006 yılının dördüncü çeyreğinde İMKB 30 endeksinde yer alan hisse senetleri belirlenmiştir. Daha sonra 61 ayık noktada gözlenen bu 30 hisse senedinin kapanış fiyat değerleri Fonksiyonel Veri Analizi açısından incelenmiştir. Öncelikle bir Splayn Düzgünleştirme Yöntemi olan Pürüzlü Ceza Yöntemi aracılığı ile 61 ayık noktada gözlenen 30 hisse senedinden 30 fonksiyonel gözlem elde edilmiştir. Bir diğer bakış açısıyla (30x61) boyutlu veri matrisi yerini her biri $T=[1,61]$ aralığında tanımlı 30 fonksiyonel gözleme bırakmıştır. İkinci aşamada bu verilerin fonksiyonel gösterimleri oluşturulmuş, bu gösterimler ve birinci türev fonksiyonları incelenmiştir. Bu 30 gözlem için oluşturulan ortalama fonksiyonu ayrı olarak ele alındığında İMKB 30 endeksi ile benzerliği dikkat çekmektedir ve İMKB 30 endeksine bir alternatif olarak düşünülebilir. Fonksiyonel verilere ait kovaryans yüzeyleri oluşturulduktan sonra incelenen gözlem ve değişken sayısı arttıkça daha da artan verilerdeki karmaşıklığı gidermek üzere Düzgünleştirilmiş Ana Bileşenler Analizi uygulanmış ve sonuçta elde edilen ana bileşen fonksiyonu yorumlanmıştır. Burada ana bileşen fonksiyonunun teorik alt yapısı da ortaya konmuştur. Zaman noktalarındaki değişkenlikteki artışın Aralık ayı sonlarından itibaren ortaya çıktığı gözlenmiştir. Bu durum ortalama fonksiyonuna ana bileşen fonksiyonunun eklenip çıkarılmasının etkisiyle de bir kez daha gösterilmiştir. Daha sonra İMKB 30 dan seçilen 11 şirket için bu analizler tekrarlanmış, piyasadaki en yüksek ağırlıklı 11

şirket için oluşturulan ortalama fonksiyonun da İMKB 30 endeksine benzerliği dikkat çekmiştir. Ancak bu 11 şirketin ana bileşen fonksiyonu oluşturulduğunda Kasım ayında değişkenlik bakımından İMKB 30 da yer alan tüm şirketlerin ana bileşen fonksiyonundan farklı olarak bir artışa rastlandığı görülmüştür. Daha sonra belirlenen 2 şirket için aylar bazında fonksiyonları ve birinci türev fonksiyonları incelenmiş ve fonksiyonların türevlerinin alınabilmesinin katkılarına değinilmiştir. Son olarak düzgünleştirme parametresinin değeri değiştirilerek fonksiyonel gösterimlerdeki değişikliklere odaklanılmıştır. Bu inceleme sonucu çalışmada önerilen parametre değerlerinin geçerliliği bir kez daha görsel olarak sınanmıştır.

BİRİNCİ BÖLÜM

KESİKLİ VERİLERİ FONKSİYONEL VERİLERE DÖNÜŞTÜRME

Fonksiyonel veri analizinde, veri matrisi ya da p değişkenli şans örneğinde yer alan i .inci gözlem aslında reel bir sürekli fonksiyon olmasına rağmen, veriler genellikle birbirlerinden ayrık noktalarda gözlemlenir. N örnek büyüklüğü (hacmi), n_i ise 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, $y_{i1}, y_{i2}, \dots, y_{in-1}, y_{in}$ şeklinde verilebilir. N hacimlik şans örneğini oluşturan veri seti ise,

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

şeklinde özetlenebilir.

Genellikle her bir örnek birimi için ölçüm yapılan nokta sayısı eşit alınmak zorunluluğunda olmadığından bu değişebilirliği göstermek üzere n_i notasyonu kullanılmıştır.

Bu durumda fonksiyonel veri analizinin ilk amacı; 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 süreçle dayanıyorsa, kesikli verilerden fonksiyonel verilere yapılan bu dönüşüm süreci, **düzgünleştirme** adı verilen bir süreci daha içerebilir (Ramsay & Silverman, 1997; 9).

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

1.1 FONKSİYONEL VERİLERİ DÜZGÜNLEŞTİRME

1.1.1 Baz Fonksiyon Yaklaşımı

Bu kesimde matematiksel bir açıklama verileceğinden, yukarıda da verilen x_i fonksiyonu için genel olarak $f(t)$ notasyonu kullanılması daha uygun olacaktır. $f(t)$ fonksiyonunu oluştururken esnek yöntemlere ihtiyaç duyulmaktadır. Bunun için K tane baz fonksiyondan (basis function) oluşan bir sistem seçilmektedir. Oluşturulmak istenilen $f(t)$ fonksiyonu bu baz fonksiyonların ağırlıklandırılmış bir toplamı olarak şu şekilde yazılabilir:

$$f(t) = c_1\theta_1(t) + c_2\theta_2(t) + \dots + c_k\theta_k(t) \quad (1.1)$$

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 $f(t)$ fonksiyonunun şeklini ve biçimini belirleyen katsayılardır. Bir anlamda parametre olarak yorumlanabilirler. Zaten “Pürüzlü Ceza” (Roughness Penalty) veya “En Küçük Kareler” düzgünleştirme yöntemleriyle c_i katsayılarının tahminlenmesi amaçlanmaktadır.

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

$$f(t) = \sum_{i=1}^K c_k \theta_k(t) \quad (1.2)$$

şeklinde K tane baz fonksiyonun ağırlıklandırılmış bir toplamı (doğrusal bileşeni) olarak tahminlenmesini amaçlayan yaklaşım “Baz Fonksiyon Yaklaşımı” olarak adlandırılır.

Fonksiyonel verilerin elde edilmesindeki Baz Fonksiyon yaklaşımında B-Splayn veya Fourier Baz Fonksiyonlarının kullanılmasının yararları aşağıdaki şekilde özetlenebilir(Ramsay J.O, 2000):

- $\theta_i(t)$ ile verilen her bir baz fonksiyon hızlı ve kolay bir biçimde hesaplanabilmektedir.
- İhtiyaç duyulduğu yerde gereksinim duyulan eğriliği gösterebilmekte, fakat uygunsa da yaklaşık doğrusal olabilmektedir.
- c_i katsayıları hızlı bir biçimde hesaplanabilmektedir. Değerler matrisi hesaplamayı kolaylaştıracak belirli özelliklerde olabilmektedir.
- Baz fonksiyonların istenildiği kadar türevi alınabilmektedir. Bu da istenilen bir durumdur.
- İstenildiği takdirde üzerine bir çok kısıt konulabilmektedir. Örneğin doğal sınır koşulları veya periyodiklik gibi.

Ayrıca yukarıdaki özelliklere ilaveten baz fonksiyon yaklaşımının kullanılmasının çok önemli bir avantajı da baz fonksiyon yaklaşımının, fonksiyonel verilerin bilgisayar dostu bir gösterimi olmasıdır(Rossi F. v.d., 2005).

Baz fonksiyon yaklaşımı benimsendikten sonra, aşağıda verilen üç konu üzerinde durulması gerekir:

- Hangi durumda hangi baz fonksiyonun uygun olduğu,
- Kaç tane baz fonksiyonun kullanılması gerektiği,
- Son olarak da c_i katsayılarının hesaplanma yöntemleridir.

Bu soruların cevapları akış içerisinde sırayla verilecektir.

1.1.1.1. Baz Fonksiyonlar

Genel olarak en çok kullanılan baz fonksiyonlar,

- Güçler
- Fourier baz
- B-Splayn baz

şeklinde sıralanabilir.

Bu üç baz tipinden B-Splaynlar, uygulamada periyodik olmayan durumlar için sıklıkla kullanıldığından ve bu çalışmanın da kapsamını içerdiğinden ayrı bir bölüm olarak ele alınacaktır.

Çoğu fonksiyonel veri analizinde fonksiyonel verileri uygun bir baz da depolamak anahtar adımdır. Çoğu kez baz prosedür de açıkça bellidir ve araştırmacının buna dikkat etmesine ihtiyaç olmaz(Ramsay & Silverman, 2002; 33).

Ne kadar çok baz fonksiyon içerilirse uyumlaştırılan fonksiyon o kadar daha kompleks olur. y_j gözlem değerlerinin düzgünleştirilme derecesine K baz fonksiyon sayısınınca karar verilir(Ulbricht , 2004).

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 & Silverman, 1997; 46). Ramsay ve Silverman 20-30 baz fonksiyonun genelde göze çarpan görünümleri özetlemek için yeterli olduğunu belirtmişlerdir(Lee, 2004). Ayrıca gözlem sayısının çok fazla olduğu birçok çalışmada baz fonksiyon sayısı gözlem sayısının karekökü olarak kullanılmıştır(Laukaitis & Rackauskas, 2002). 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 & Silverman,1997; 46). Fonksiyonel veri analizinde birçok yerde türev bilgisinden faydalanılır. Örneğin

büyüme eğrilerinin yaşa göre birinci türevi büyüme hızını ve ikinci türevi de ivmeyi göstermektedir(Gasser vd., 1984). Ayrıca bir fonksiyonel gözlemin belirli bir dereceden türevini bir başka dereceden türevine karşı plot ettirerek de değerli bilgi elde edinilebilir. Bu teknik **phase/plane plotting** olarak adlandırılır. Dayanaksız mallar üretim endeksi önemli ekonomik bir göstergedir ve endeksin artış oranına karşı endeksin ivmesi plot edilerek mantıklı çıkarımlar elde edilebilir. Eğer endeks düzgün bir trende ve buna ek olarak sinüsoidal mevsimsel değişime sahip ise uygun ölçek seçimi ile bu plot yaklaşık dairesel olur(Ramsay & Silverman, 2001, Ramsay Silverman, 2005; 13-15). Dayanaksız mallar üretim endeksi ile ilgili benzer bir çalışma 1964 ve 1967 yılları arası için Ramsay ve Ramsey (2002) tarafından da yapılmıştır.

Güçler : En çok bilinen baz fonksiyonlardan ilki 1, t, t^2 ... şeklindeki polinomlar için baz fonksiyonlar olan güçlerdir. Bu tür baz fonksiyonlar periyodik olmayan fonksiyonları modellemek için uygundur.

Polinomlar hesaplanması yönünden kolaydır. Ancak fonksiyonel yaklaşımda polinomların kullanılmasının bazı önemli dezavantajları bulunmaktadır. Eğer veri noktaları çok sayıda ise bu durumda yaklaşılacak polinomun derecesi de yüksek olacaktır, bu nedenle elde edilen polinomun doğruluğu sınanmalıdır. Ayrıca hem türev alınabilmeyi ve hem de eğriselliği arttırmak için polinom t ile çarpılmalıdır. Bir diğer sakınca olarak az sayıda fiziksel kural yüksek dereceli polinom yapısındadır, böylece yaklaşılacak fonksiyon muhtemelen yaklaşım yapılmak istenen fonksiyondan daha az istikrarlı olacaktır. Yüksek dereceli polinomlar fizikte tipik olmayan birçok dönme noktasına sahiptir. Bu gibi nedenlerden polinomlar yeterince esnek değildir ve ancak doğrusal ve küçük eğrisellikler için uygulanması yerinde olur.

Polinomların dezavantajları ile ilgili ayrıntılı bilgilere Schwarz ve Waldvogel(1989; 125), Buchanan ve Turner (1992; 133) gibi çalışmalarda yer verilmektedir. Polinomların nasıl oluşturulduğu ile ilgili detaylı bilgi için ise bu

çalıřmalara ilaveten Burden ve Faires (1997; 103-134) gibi çalıřmalara bařvurulabilir.

Fourier Baz: Bir diđer kullanılan baz fonksiyon Fourier serileridir. Fourier serileri periyodik fonksiyonları modellemek için kullanılır ve sabit bilinen frekanslı periyodik veriler için kullanıřlı olabilir. $w=2\pi/s$ ve s de çalıřılan zaman aralıđının uzunluđunu (periyod) göstermek üzere w nun sabit bilinen bir frekansı için bazlar $1, \sin(wt), \cos(wt), \sin(2wt), \cos(2wt), \dots$ řeklindedir.

Örneđin yıllık periyodik varyasyonla aylık zaman ölçeđinde gözlenen veriler için $s=12$ dir, buna örnek olarak sıcaklık ölçümleri verilebilir. Bu durumda bazlar sırasıyla $1, \sin(\frac{2\pi}{12}t), \cos(\frac{2\pi}{12}t), \dots$ biçiminde seyreder(Lillestol & Ollmar, 2003).

Eđer t_j deđerleri t de eřit ölçeklenmiřse, bu durumda baz ortogonaldir. Bir anlamda katsayıların tahmininde kullanılacak olan apraz-arpım matrisi $\mathbf{B}^T\mathbf{B}$ köřegen olur. Aynı zamanda baz fonksiyonları uygun bir sabite bölerek bu matris birim matrise dönüřtürülebilir. Böylelikle kesim 1.1.2 de bahsedilecek olan gerek en küçük kareler yaklařımı ve gerekse pürüzlü ceza yaklařımında katsayıların tahmininde kullanılan $\mathbf{B}^T\mathbf{B}$ matrisinin tersi rahatlıkla hesaplanabilir.

Fourier serilerinde türev alınabilmesini veya düzgünleřtirmeyi arttırmak için ekstra bir iřlem yapılmasına gerek yoktur, bunlar zaten sonsuz türev alınabilen fonksiyonlardır.

Fourier serileri özellikle oldukça durađan (stable) fonksiyonlar için uygundur. Fonksiyonun durađan olması burada güçlü bir lokal deđiřimin olmadıđı ve eđriselliđin hemen her yerde aynı dereceden olduđu anlamına gelir(Ramsay & Silverman, 1997; 47).

Fourier serilerini baz olarak kullanan Splayn'larla ilgili ayrıntılı bilgilere Anselone ve Laurent (1968) ve Ramsay ve Silverman (2005; 45) çalışmasında yer verilmektedir.

1.1.1.2 B Splayn

Splayn adı eski bir çizim aracından gelmektedir. Bu araç belirlenmiş noktalardan geçen düzgün bir eğri oluşturmak için kullanılan ince esnek bir tahta parçası olarak kullanılmaktaydı. Aynı zamanda bir interpolasyon aletiydi. Varsaydığı şekil minimum depolanmış enerjiye sahip şekildi ve neredeyse doğal bir kübik splayndi. Sınır (end) koşullarını sağlaması doğal bir özelliktir, çünkü şekil sınır (extreme) düğümlerin (knot) ötesinde düz bir çizgi halinde devam etmektedir (Buchanan & Turner, 1992; 180).

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 reel 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,

$$f(t) = \sum_{i=1}^K c_i B_{i,d} \quad (1.3)$$

şeklinde ifade edilebilirler.

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

$$B_{j,d}(t) = \frac{t - t_j}{t_{j+d} - t_j} B_{j,d-1}(t) + \frac{t_{j+1+d} - t}{t_{j+1+d} - t_{j+1}} B_{j+1,d-1}(t) \quad (1.4)$$

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.

Ayrıca sıfırcı dereceden birinci mertebeden bir B-Splayn daima aşağıdaki denklemi sağlar:

$$B_{j,0}(t) = \begin{cases} 1 & t_j \leq t \leq t_{j+1} \\ 0 & \text{d.d} \end{cases} \quad (1.5)$$

Yukarıdaki biçimde tanımlanan B-Splaynlardan aynı zamanda

$$\sum_{i=1}^K B_{i,d}(t_j) = 1, \quad i = 1, 2, \dots, K, \quad (1.6)$$

$$j = 1, 2, \dots, n$$

sağladığından dolayı normalize B-Splayn olarak bahsedilmektedir. Ancak kullanım kolaylığı bakımından literatürde normalize B-Splayndan kısaca B-Splayn olarak söz edilmektedir.

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:

$$\text{Baz fonksiyon sayısı} = \text{Düğüm sayısı} + \text{mertebe} - 2 \quad (1.7)$$

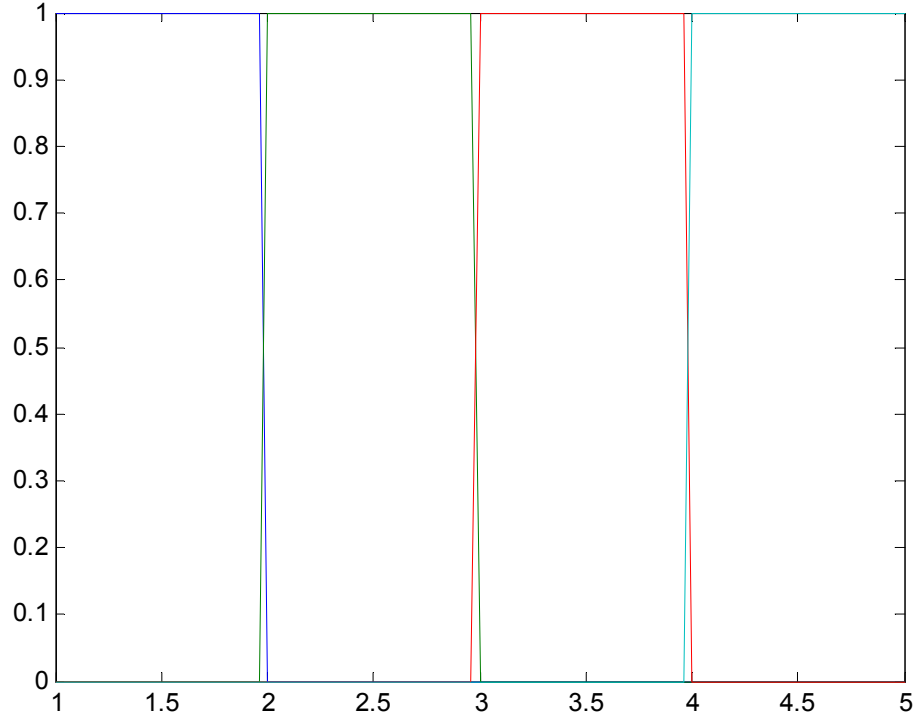
Splaynlarla ilgili literatürde düğümden bahsedildiğinde genel olarak kesme noktası (break point) ile aynı kavram kastedilmektedir. Ancak daha açık bir biçimde ifade edildiğinde düğüm denildiğinde birden fazla çakışan noktadan bahsedilebiliyor iken, kesme noktasından bahsedildiğinde düğümlerin yerleştiği t_j değerlerinin sadece artan sırası dikkate alınmaktadır(Ramsay, [2000](#)).

B-Splaynlar oluşturulurken düğümlerin ve dolayısıyla da kesme noktalarının nasıl seçileceği önemli bir konudur. Bununla birlikte fonksiyonel veri analizinde gözlemler olmadan düğümleri yerleştirmek anlamsızdır. Çoğu durumda her bir alt aralıkta en az bir gözlenen t_j değerinin bulunması gereklidir. Düğümlerin nasıl yerleştirileceği genellikle fonksiyonel veri analizinde iki yöntemle belirlenir. Birincisi düğümleri eşit aralıklarla yerleştirmektir, fakat bu durumda her bir alt aralıkta en az bir gözlemin bulunmasına dikkat etmek gerekir. Noktaların bu şekilde eşit aralıklarla seçimi numerik analiz çalışmalarında bile her zaman çok tercih edilen bir yöntem değildir. Bu çalışmada ve fonksiyonel veri analizinde de sıklıkla kullanılmakta olan bir diğer ikinci yöntem de düğüm olarak her bir gözlem değerini yerleştirmektir(Ramsay, 2000).

B-Splayn baz fonksiyonları oluştururken alt ve üst sınırlar uzatılabilir, ancak burada baz fonksiyonların belirlenen aralığın dışında nasıl bir davranış gösterdikleri göz ardı edilebilir, bunun nedeni aralığın dışındaki şekillerin aralığın içini etkilememesidir. Alt ve üst sınırları genişletirken bazı ilave düğümlerin içerilmesi gerekir. Bu durumda $m-1$ ilave ön (initial) ve arka (final) düğüme ihtiyaç duyulur. Ön sınırlar $\alpha_1 \leq \alpha_2 \leq \dots \leq \alpha_{m-1} \leq t_1$ ve arka sınırlar $t_n \leq \alpha_{n+1} \leq \alpha_{n+2} \dots \leq \alpha_{m-1}$ biçiminde ifade edilebilir. İlave düğümlerin gerçek değerleri keyfidir, ancak fonksiyonel veri analizinde ilave ön sınırları birbirine t_1 'e, ilave arka sınırları da birbirine ve t_n 'e eşit yapmak gelenekseldir.

Aşağıda çeşitli dereceden B-Splaynlara örnekler verilmiştir:

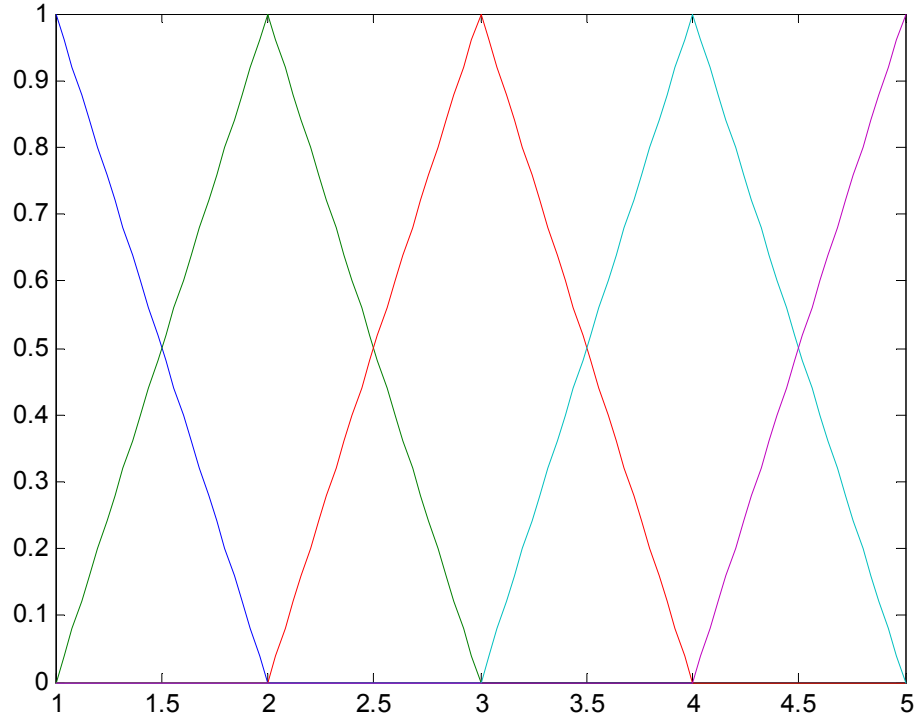
$\underline{t} = [1 \ 2 \ 3 \ 4 \ 5]$ için $B_{j,0}(t)$ aşağıdaki biçimdedir:



Şekil (1.1): Sıfırıncı Dereceden B-Splayn

$B_{j,0}(t)$ birim kutu fonksiyondur ve splayn fonksiyonun ilkel bir tipidir.

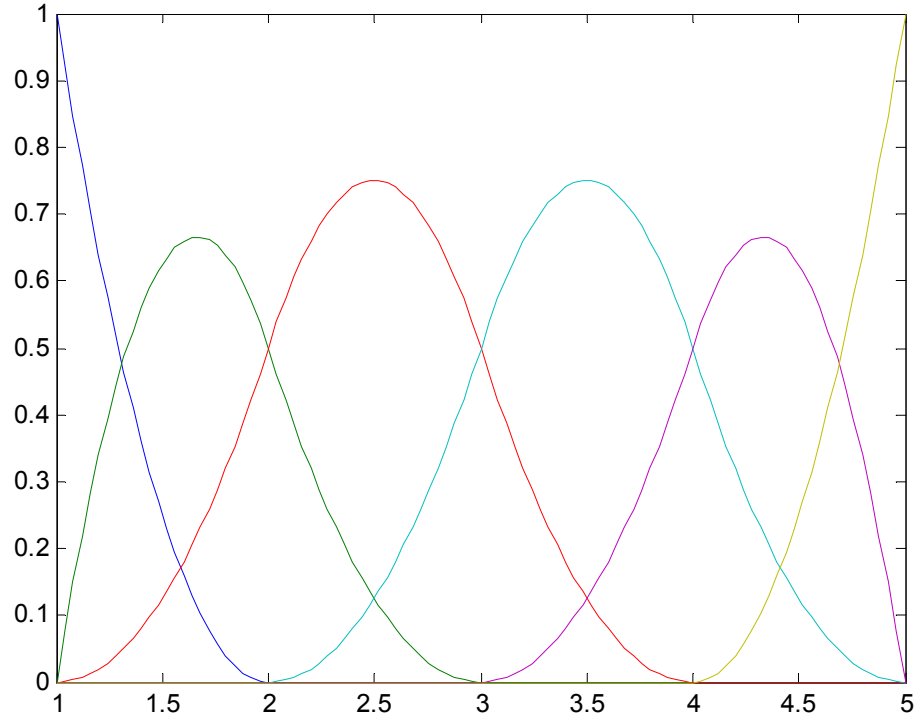
$\underline{t} = [1 \ 2 \ 3 \ 4 \ 5]$ için $B_{j,1}(t)$ aşağıdaki biçimdedir:



Şekil (1.2): Birinci Dereceden B-Splayn

$B_{j,1}(t)$ şeklinde birinci derece ve ikinci mertebeye sahip B-Splaynlar şekil (1.2)'den de görüldüğü gibi çadır fonksiyonlardır. Birinci dereceden bir B-Splayn iki doğrusal parçadan oluşmaktadır. B-Splaynların sınır değerlerinin dışında B-Splaynlar sıfırdır. Buradaki her bir B-Splayn üç düğüme dayanmaktadır.

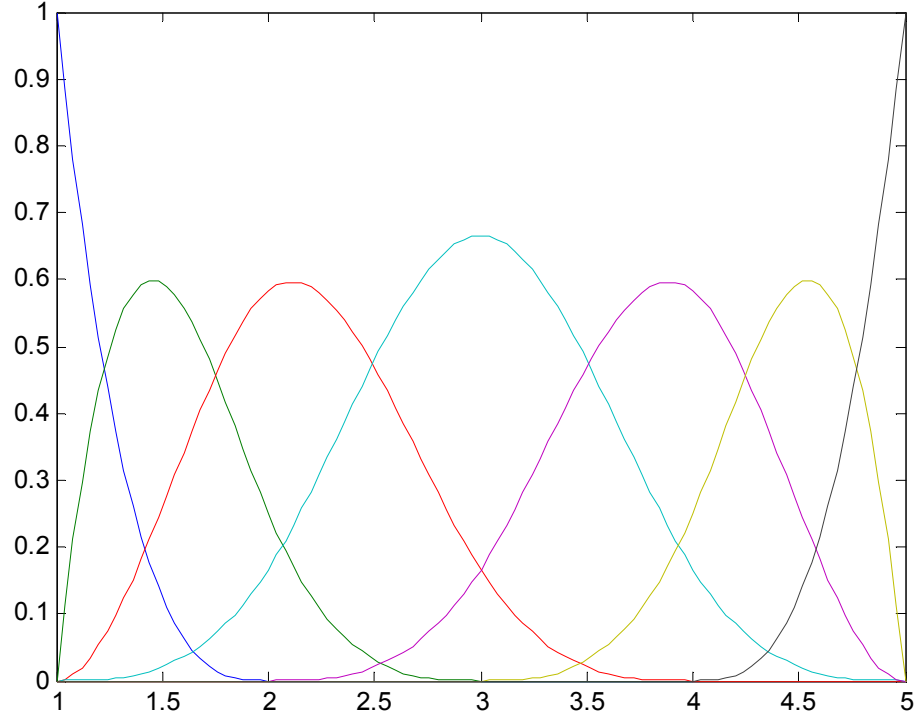
$\underline{t} = [1 \ 2 \ 3 \ 4 \ 5]$ için $B_{j,2}(t)$ aşağıdaki biçimdedir:



Şekil(1.3): İkinci Dereceden B-Splayn

Yukarıda ikinci dereceden B-Splaynlara örnekler verilmektedir. Burada her bir B-Splayn iki düğümde birleşen üç kuadratik parça içerir. Birleşme noktalarında sadece polinom parçalarının ordinatları değil birinci türevleri de eşittir. İkinci dereceden bir B-Splayn dört ardışık düğümüne bağlıdır.

$\underline{t} = [1 \quad 2 \quad 3 \quad 4 \quad 5]$ için $B_{j,3}(t)$ ařađıdaki biçimdedir:



řekil(1.4): Üçüncü Dereceden B-Splayn

Artan düđüm deđerleri için ikinci dereceden veya daha yüksek dereceden B-Splaynlar sınır deđerler civarı dıřında řekil (1.3) ve (1.4)'den de görüldüđü gibi çan eđrilerine benzerdir.

B-Splaynlar “Düzgünleřtirme Özellikleri” denilen ařađıdaki ilave özelliklere de sahiptir:

- $B_{j,0}(t)$ ve $B_{j+1,0}(t)$ t_{j+1} kesim noktasında (düđümünde) aynı deđerlere sahiptir. Topamlar süreklidir.

- $B_{j,1}(t)$ ve $B_{j+1,1}(t)$ sırasıyla t_{j+1} ve t_{j+2} kesim noktalarında aynı değerlere ve aynı türev değerlerine sahiptir. Bu iki B-Splaynın toplamı da sürekli birinci dereceden türeve sahiptir.

Genellersek $B_{j,d-1}(t)$ ve $B_{j+1,d-1}(t)$ sırasıyla t_{j+1} ve t_{j+2} kesim noktalarında aynı değerlere ve $d-1$ 'e kadar aynı türev değerlerine sahiptirler. Toplamları da sürekli $(d-1)$ 'inci türeve sahiptir.

Bir diğer deyişle tüm düğümler farklı ise, bu durumda doğrusal Splayn sürekli, kuadratik splayn sürekli birinci türeve sahiptir, kübik splaynın ikinci türevi de sürekli(Lyche & Morken, 2002; 39).

Düzgünleştirmenin kontrolü splayn teorisinin temel amacı olup, düzgünleştirme (1.4) numaralı denklemlerle verilmiş olan tekrarlı ilişki ile tanımlanır. Tekrarlı ilişkiler yoluyla, yukarıda özetlendiği gibi, verilen bir düzgünleştirme seviyesindeki B-Splayn baz fonksiyonları kullanılarak bir derece veya bir merteye daha yüksek, bir ilave aralığa daha yayılmış, daha yüksek düzgünleştirme seviyesindeki yeni bir B-Splayn elde edilebilir. Benzer tekrarlı ilişki B-Splaynların türevleri için de geçerlidir. d 'inci dereceden bir B-Splaynın m 'inci türevi, burada D bir türev operatörünü göstermek üzere, aşağıdaki tekrarlı ilişki ile tanımlanabilir:

$$D^m B_{j,d}(t) = d \left\{ \frac{D^{m-1} B_{j,d-1}(t)}{t_{j+d} - t_j} - \frac{D^{m-1} B_{j+1,d-1}(t)}{t_{j+1+d} - t_{j+1}} \right\} \quad (1.8)$$

Düğümünlerin genellikle artan bir sırada oldukları varsayılır, ancak durum daha önce de bahsedildiği gibi her zaman böyle olmayabilir. Belirli bir t değerinde birden fazla sayıda düğüm bulunabilir. Bu gibi bir durumda t ye eşit olan her bir ilave değer için düzgünleştirmeden bir derece daha kaybedilir, bir diğer deyişle bir derece az türeve sahip olunur.

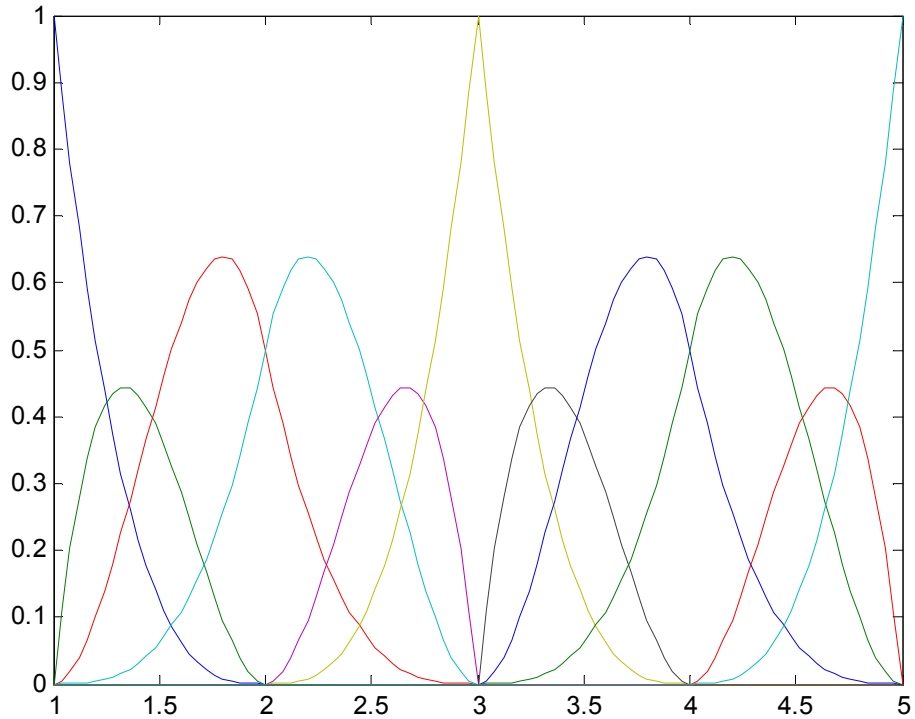
(1.5) ifadesi dikkate alındığında,

$$B_{j,0}(t) = 0 \quad t_j = t_{j+1} \quad (1.9)$$

olacağından, belirli bir t değerinde birden fazla düğüm bulunduğu, çoklu düğüm durumunda dikkat edilmesi gereken bir diğer nokta, karşılaşılabilecek sıfıra bölüm sonucunda belirsizlik durumundan kaçınılmış olur. Bununla birlikte (1.4) de tekrarlı düğüm değerleri için ilgilenilen terim, katsayısının ne olacağı önemli olmaksızın sıfır olur(Ulbricht, 2004).

Tekrarlı düğüm değerleri için B-Splaynlara aşağıdaki gibi bir örnek verilebilir.:

$\underline{t} = [1 \ 2 \ 2 \ 3 \ 3 \ 3 \ 4 \ 4 \ 5]$ için $B_{j,3}(t)$ aşağıdaki biçimdedir:



Şekil (1.5): Tekrarlı Düğüm Değerli Üçüncü Dereceden B-Splayn

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'dur (1978; 109-153). Konu ile ilgili ayrıntılı bilgi için ayrıca Nürnberger (1989; 95-99), Schumaker (1993; 108-139) ve Dierckx (1993; 1-22) çalışmalarına başvurulabilir.

B-Splaynlar ve dolayısıyla da Splaynlar bilgisayar destekli tasarımlarda, görüntü işlemede (image processing), tıpta ve istatistikte geniş bir uygulama alanına sahiptirler. Tıp biliminde genlerle ilgili araştırmalarda B-Splaynları kullanarak özellikle kümeleme analizinde birçok çalışma yapılmıştır. Ayrıntılı bilgi için Joseph vd. (2003), Luan ve Li (2003), Ando v.d. (2004) ve Leng ve Müler (2006) incelenebilir. Başka bir çalışmada ise, bir hokkabazın toparlarla yaptığı hokkabazlıklar modellenirken, kesikli veriler fonksiyonlara 51 eşit ölçekli nokta için 6 mertebeli 55 B-Splayn aracılığı ile döndürülmüştür(Ramsay, 2001). Finans sektöründe de kesikli verileri fonksiyonel verilere dönüştürürken Splaynların kullanılması ile ilgili örnekler bulunmaktadır(Dabos M. & Bugallo F., 2000, Laukaitis & Rackauskas, 2002). Ayrıca büyüme eğrilerinde de çoğu çalışmada odak, ağırlık ve boy olmak üzere, Splaynlardan geniş bir ölçüde yararlanılmıştır(Gasser T. vd., 1984, Ramsay, 1998, Ramsay & Silverman, 2001, Gasser & Kneip, 1995). Bunlara ilaveten insan hareketlerini modellemede de Splaynlardan geniş ölçüde yararlanılmıştır (Faraway, 2000, Ramsay & Silverman, 1997; 6)

Splaynları oluştururken doğal (natural) splaynlar da baz olarak kullanılabilir. Doğal splaynlar baz olarak kullanılıyorsa Reinsch'ın "Smoothing By Spline Functions" (1967) isimli makalesinde ayrıntılı bir biçimde incelenmiş olan Reinsch algoritması kullanılır ancak bu algoritma B-Splaynların hesaplanması için gereken algoritmadan daha kısa olsa bile B-Splaynların oluşturulması numerik olarak daha istikrarlı, dolayısıyla da daha doğrudur. Sonuç olarak elde edilen Splaynlar düğüm noktaları aynı olmak kaydı ile aynı sonucu verir(Green & Silverman, 1994; 45-46)

Reinsh'in algoritmasının geliştirilmiş hali "Smoothing By Spline Functions II" isimli makalesinde yeniden ele alınmıştır(Reinsh, 1971). Splaynlarla ilgili daha ayrıntılı bilgiler için Hutchinson ve Hoog (1985), Craven ve Wahba (1979) çalışmalarına başvurulabilir.

1.1.2. Pürüzlü Ceza Yaklaşımı

$t_j, j=1,2, \dots, n$ düğümlerinde y_j gözlemlerine sahip olunan regresyon problemi ele alınsın ve gözlemlerin,

$$y_j = x(t_j) + \varepsilon_j \quad (1.10)$$

modelini sağladığı varsayılınsın. Bu çalışmada düğümlerin $t_1 \leq t_2 \leq \dots \leq t_n$ sıralamasını sağladığı ve ε_j şeklinde gösterilen hataların ortalamalarının sıfır olduğu ve ilişkisiz oldukları varsayılmıştır. İlk olarak ε_j 'lerin varyanslarının eşit oldukları (homoskedasticity) varsayılır, fakat daha sonra bu yaklaşım esnetilmiş ve farklı varyanslılık (heterokedasticity) durumu da ele alınmıştır(Silverman,1985).

Baz fonksiyon yaklaşımında temel olarak y_j 'lerin az sayıda K baz fonksiyonun doğrusal bir kombinasyonu olarak ifade edilmesine odaklanılmıştır ve amaç bu kısıt altında artık kareler toplamını (residual sum of squares) minimize etmektir. Bu amaçla bir eğrinin uyumunu sağlamak için, en bilinen yaklaşım artık kareler toplamını minimize etmeyi amaçlayan En Küçük Kareler Metodudur. $B_j(t)$ baz fonksiyonlarından oluşan bir sete sahip olunduğu varsayılınsın. $B_i(t_j), i=1,2, \dots, K; j=1,2, \dots, n$ elemanlarına sahip $(n \times K)$ boyutlu bir \mathbf{B} matrisi tanımlansın. Eğer katsayı vektörü \underline{c} , $[\underline{c} = (c_1 \ c_2 \ \dots \ c_K)^T]$ ise, bu durumda düğüm noktalarındaki değerlerin vektörü $\mathbf{B}\underline{c}$ olur.

y_j ve $\sum_{i=1}^K c_i B_i(t_j)$ arasındaki sapmaların karelerinin toplamını minimize etmek için En Küçük Kareler Metodu kullanılabilir. \underline{y} (nx1) boyutlu gözlem vektörü, $\mathbf{B} = [B_i(t_j)]$, $i = 1, 2, \dots, K$, $j = 1, 2, \dots, n$ değerlerinden oluşan (nxK) boyutlu B-Splayn değerleri matrisi, \underline{c} (Kx1) boyutlu katsayılar vektörü olmak üzere, standart istatistiksel en küçük kareler teorisine göre, Artık Kareler Toplamı (AKT),

$$AKT = \sum_{j=1}^n \left(y_j - \sum_{i=1}^K c_i B_i(t_j) \right)^2 \quad (1.11)$$

veya vektör notasyonunda olmak üzere,

$$\begin{aligned} AKT &= [\underline{y} - \mathbf{B}\underline{c}]^T [\underline{y} - \mathbf{B}\underline{c}] \\ &= \underline{y}^T \underline{y} - \underline{y}^T \mathbf{B}\underline{c} - \underline{c}^T \mathbf{B}^T \underline{y} + \underline{c}^T \mathbf{B}^T \mathbf{B} \underline{c} \\ &= \underline{y}^T \underline{y} - 2 \underline{c}^T \mathbf{B}^T \underline{y} + \underline{c}^T \mathbf{B}^T \mathbf{B} \underline{c} \end{aligned} \quad (1.12)$$

şeklinde yazılabilir. Bu Artık Kareler Toplamını minimize eden katsayılar vektörü $\hat{\underline{c}}$,

$$\begin{aligned} \frac{\partial AKT}{\partial \underline{c}} &= -2\mathbf{B}^T \underline{y} + 2\mathbf{B}^T \mathbf{B} \underline{c} \\ &= -2\mathbf{B}^T \underline{y} + 2\mathbf{B}^T \mathbf{B} \hat{\underline{c}} = 0 \\ 2\mathbf{B}^T \underline{y} &= 2\mathbf{B}^T \mathbf{B} \hat{\underline{c}} \\ \mathbf{B}^T \underline{y} &= (\mathbf{B}^T \mathbf{B}) \hat{\underline{c}} \\ (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \underline{y} &= (\mathbf{B}^T \mathbf{B})^{-1} (\mathbf{B}^T \mathbf{B}) \hat{\underline{c}} \\ \hat{\underline{c}} &= (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \underline{y} \end{aligned} \quad (1.13)$$

şeklinde tahminlenir.

Burada öncelikle ele alınacak iki durum sözkonusudur. Eğer değerlendirme noktalarına eşit veya daha az sayıda baz fonksiyon var ise, $K \leq n$ durumunda, y_j

ve $\sum_{i=1}^K c_i B_i(t_j)$ arasındaki sapmaların karelerinin toplamını minimize etmek amacı ile

baz fonksiyonları en küçük kareler ile uyumlaştırılabilir. $K < n$ iken açılım orijinal verilerin düzgünleştirilmiş bir versiyonu halini alır. $K = n$ ise, bu durumda

$x(t) = \sum_{i=1}^K c_i B_i(t_j)$ y_j değerlerini tamamen interpolate eder. Ancak böyle bir

interpolant istatistikçilerce büyük bir olasılıkla hızlı iniş çıkışlar inandırıcı olmadığından dolayı reddedilecektir(Ramsay & Silverman, 2002; 33).

Düğüm noktalarının gözlem noktaları ile aynı olduğu varsayıldığında, düzgünleştirme matrisi,

$$\mathbf{S} = \mathbf{B} (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \quad (1.14)$$

şeklinde verilebilir.

Düzgünleştirme matrisi \mathbf{S} bu durumda ortogonal bir izdüşüm matrisidir, çünkü simetriktir ve $\mathbf{S}^2 = \mathbf{S}$ özelliğini sağlamaktadır, yani idempotent bir matristir. Böylece \mathbf{S} idempotent matrisi, \mathbf{B} matrisinin kolonlarından oluşan baz fonksiyonların tanımladığı uzaya izdüşümün yapılmasını sağlayan bir en küçük kareler baz fonksiyon düzgünleştiricisidir(Ramsay & Silverman, 1997; 44).

Genel olarak ε_j 'lerin sıfır ortalama ile dağıldığı fakat varyanslarının eşit olmadığı durumlarda ağırlıklandırma uygulanır. Bu durumda ağırlıklar genellikle artıkların varyanslarının tersleri olarak alınır(Green & Silverman, 1994; 40). Böylece \mathbf{W} matrisi, her bir artığın karesine farklı ağırlık verilebilmesine izin veren pozitif belirli simetrik bir matris olmak üzere, Artık Kareler Toplamı, Ağırlıklandırılmış Artık Kareler Toplamı (AAKT) olarak aşağıdaki şekilde genişletilebilir:

$$AAKT = [\underline{y} - \underline{B}\underline{c}]^T \underline{W} [\underline{y} - \underline{B}\underline{c}] \quad (1.15)$$

$$AAKT = \underline{y}^T \underline{W} \underline{y} - \underline{y}^T \underline{W} \underline{B}\underline{c} - \underline{c}^T \underline{B}^T \underline{W} \underline{y} + \underline{c}^T \underline{B}^T \underline{W} \underline{B}\underline{c}$$

$$AAKT = \underline{y}^T \underline{W} \underline{y} - 2\underline{c}^T \underline{B}^T \underline{W} \underline{y} + \underline{c}^T \underline{B}^T \underline{W} \underline{B}\underline{c}$$

Ağırlıklandırılmış Artık Kareler Toplamını minimize eden katsayılar vektörü $\hat{\underline{c}}$,

$$\frac{\partial AAKT}{\partial \underline{c}} = -2\underline{B}^T \underline{W} \underline{y} + 2\underline{B}^T \underline{W} \underline{B}\underline{c}$$

$$-2\underline{B}^T \underline{W} \underline{y} + 2\underline{B}^T \underline{W} \underline{B}\hat{\underline{c}} = 0$$

$$2\underline{B}^T \underline{W} \underline{y} = 2\underline{B}^T \underline{W} \underline{B}\hat{\underline{c}}$$

$$\underline{B}^T \underline{W} \underline{y} = (\underline{B}^T \underline{W} \underline{B})\hat{\underline{c}}$$

$$(\underline{B}^T \underline{W} \underline{B})^{-1} \underline{B}^T \underline{W} \underline{y} = (\underline{B}^T \underline{W} \underline{B})^{-1} (\underline{B}^T \underline{W} \underline{B})\hat{\underline{c}}$$

$$\hat{\underline{c}} = (\underline{B}^T \underline{W} \underline{B})^{-1} \underline{B}^T \underline{W} \underline{y} \quad (1.16)$$

şeklinde tahminlenir.

Düğüm noktalarının gözlem noktaları ile aynı olduğu varsayıldığında, düzgünleştirme matrisi \underline{S} , $\underline{B} (\underline{B}^T \underline{W} \underline{B})^{-1} \underline{B}^T \underline{W}$ biçimindedir ve yine bir ortogonal izdüşüm operatörüdür, çünkü $\underline{W}\underline{S}$ simetriktir. \underline{S} matrisi ile tanımlanan doğrusal mapping (linear mapping) işlemi, genellikle \underline{W} metriğinde bir izdüşüm olarak adlandırılır (Ramsay & Silverman, 1997; 45).

Bir diğer yandan, eğer değerlendirme noktalarından daha fazla sayıda baz fonksiyon var ise, ki üçüncü dereceden B-Splaynlarda bu durum geçerlidir, bu durumda en küçük kareler yaklaşımında verilen değerleri $x(t) = \sum_{i=1}^K c_i B_i(t_j)$ şeklinde interpolate eden birden fazla \underline{c} seçimi bulunmaktadır. Bunların arasından seçim yapabilmek için, eğrinin iniş çıkışını minimize eden parametreleri seçmek gerekir (Ramsay & Silverman, 2002; 36). Aynı zamanda katsayıların tahmininde kullanılan ters matris işlemi matris tekile yaklaştığı için zorluk çıkarmaktadır. Bu

durumda matrisi tekillikten kurtarmak (Besse v.d. 2005), eğrinin iniş çıkışını minimize eden tek bir parametre değeri elde etmek için B-Splaynlar kullanılıyorsa aşağıda detaylı bir biçimde açıklanacak olan **Pürüzlü Ceza Yöntemi** bu amaçlar doğrultusunda kullanılabilir.

$$\text{Ayrıca, } x(t) = \sum_{i=1}^K c_i B_{i,d} \text{ şeklinde B-Splaynların doğrusal bir kombinasyonu}$$

olarak oluşturulan splaynlar üzerinde baz fonksiyonların sayısını sınırlandırarak düzgünleştirme sürekli olarak kontrol altında tutulamaz, bu ve yukarıda bahsedilen sebepler doğrultusunda sıklıkla düzgünleştirme üzerinde sürekli kontrol sağlayacak bir yöntem olan ve en küçük karelerin bir uzantısı olan Pürüzlü Ceza Yöntemi kullanılmaktadır.

Genel olarak ele alındığında, herhangi bir x_i eğrisinin bir parametrik regresyon tekniği ile tahmin edilmemesinin nedenlerinden biri regresyonun temelde iki amacının bulunmasıdır. Birincisi dizayn değişkeni ve yanıt değişkeni arasındaki ilişkiyi sunmanın ve açıklamanın bir yolunu sağlamak, ikincisi ise gözlemlerin tahminlerini elde etmektir. Burada bu iki amaca ve özellikle birinci amaca ilaveten parametrik olmayan regresyon metodunun kullanılmasının temel nedenlerinden biri üçüncü bir amaç içindir. Elde edilen splayn, x eğrisinin ilgilenilen özelliklerinin tahminlerini verir. Örneğin, büyüme eğrileri ile ilgili bir çalışmada büyümenin maksimum miktarı önemli bir kavramdır ve böylece x eğrisinin maksimum türevi ile ilgilenilebilir. Özellikle dizayn değişkeni ve yanıt değişkeni arasındaki ilişkiyi sunmanın ve açıklamanın bir yolunu sağlaması ve elde edilen x eğrisinin ilgilenilen özelliklerinin tahminlerinin elde edilmesi parametrik olmayan bir metodu arzu edilir kılar(Silverman,1985). Ayrıca yaşa karşı uzunluk verilerini modelleme gibi uygulamanın klasik alanlarında mevcut parametrik modeller ciddi bir biçimde yetersiz olabilir(Kneip & Gasser, 1992). Bir diğer taraftan kimi durumlarda veriler birkaç parametrik fonksiyonu birbirinden ayırt etmeye izin vermez(Walker, 2001). Bu görüşlere ilaveten, bir diğer görüşe göre de splayn düzgünleştirme yöntemleri parametrik ve parametrik olmayan yöntemler arasında bir köprü olmakta, bunlar “overparametric” veya “anonymous” modeller olarak da adlandırılmaktadır. Burada

parametreler mevcuttur, ancak bunların bir bilimsel yorumlaması yoktur(Eiler & Marx , 1996).

Çalışma kapsamında Kübik Splaynları kullanan parametrik olmayan regresyon, eğri tahmini için çekici, esnek ve geniş uygulanabilir bir yaklaşımdır.

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 Örnekleme Varyansının toplamına eşittir. Örnekleme 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,

$$PEN_2(x) = \int_a^b (D^2 x(t))^2 dt = \|D^2 x\|^2 \quad (1.17)$$

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 & Silverman, 1997; 59). Ayrıca, $\|D^2x\|^2$, 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),

$$CKT_\lambda = \sum_j (y_j - x(t_j))^2 + \lambda \|D^2x\|^2 \quad (1.18)$$

şeklinde tanımlanır.

Pürüzlü ceza terimi $\lambda \|D^2x\|^2$ belirli bir eğrinin cezalı en küçük karelerinin sadece $\sum_j (y_j - x(t_j))^2$ şeklinde Artık Kareler Toplamı ile ölçülen verilere uyum iyiliği ile değil, aynı zamanda $\|D^2x\|^2$ şeklinde pürüzlülüğüne de bakarak karar verilmesini garanti altına alır(Green & Silverman, 1994; 5). Arzu edilen en yüksek mertebeden türev m ise, bu durumda fonksiyon iki türev daha yüksekten cezalandırılır. Örneğin büyüme verileri ile ilgili olarak ivme ile ilgileniyorsak bu durumda $\|D^4x\|^2$ belirli integrali ikinci türevin eğriselliğini de kontrol eder(Timini, 2003).

Düzleştirme parametresi λ Artık Kareler Toplamı ile ölçülen ‘verilerin eğriye uyumu’ ve $\|D^2x\|^2$ 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 değişken olur, 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. Uyumlaştırılan eğri verileri daha fazla değişkenlik olsa bile daha yakın izler. Burada pürüzlülük üzerine daha az ceza konduğundan dolayı eğri daha değişken 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 düzgünleştirme parametresinin elde edilmesi ile ilgili Kesim 1.1.2.1 de daha ayrıntılı bilgi verilecektir.

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) :

- 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).

S, Silverman'ın belirttiği, doğal sınır koşulları da dahil olmak üzere tüm özellikleri sağlayan Kübik Splaynlardan oluşan bir uzay olsun. Her bir $i = 1, 2, \dots, K, j = 1, 2, \dots, n$ için, $B_i(t_j)$ Kesim 1.1.1.2 de ayrıntılı olarak bahsedildiği gibi

B-Splayn olarak adlandırılır ve $B_i \in S$ olmak üzere Kübik Splaynların tüm özelliklerini taşır.

B-Splaynlar numerik analizde oldukça geniş bir kullanım alanına sahiptir. Kübik Splaynlar oluşturulurken B-Splaynların ilgilenilen önemli bir özelliği, Kübik Splaynların, $x(t) = \sum_{i=1}^K c_i B_{i,d}$ şeklinde B-Splaynların doğrusal bir kombinasyonu olarak ifade edilebilmesidir.

\underline{c} ve $\underline{\mathbf{B}}(t)$ ($K \times 1$) boyutlu sırasıyla katsayılar ve B-Splaynlar vektörü olmak üzere, d.inci dereceden B-Splaynların doğrusal bir kombinasyonu olarak $x(t)$ şeklinde ifade edilmiş bir Splayn, vektör bazında,

$$x(t) = \sum_{i=1}^K c_i B_{i,d} = \underline{c}^T \underline{\mathbf{B}}(t) \quad (1.19)$$

şeklindedir.

\underline{y} vektörü düzgülendirilecek y_j değerlerinden oluşan ($n \times 1$) boyutlu bir vektörü, $\underline{\mathbf{B}}$ $B_i(t_j)$ baz fonksiyon değerlerinden oluşan ($n \times K$) boyutlu matrisi, \underline{t} t_j düğüm noktalarından oluşan ($n \times 1$) boyutlu bir vektörü ve son olarak da $x(\underline{t})$ uyumlaştırılan değerlerin ($n \times 1$) boyutlu vektörünü belirtmek üzere Cezalı Kareler Toplamı (CKT_λ) vektör bazında,

$$\begin{aligned} CKT_\lambda &= [\underline{y} - \underline{\mathbf{B}}\underline{c}]^T [\underline{y} - \underline{\mathbf{B}}\underline{c}] + \lambda \underline{c}^T \underline{\mathbf{R}} \underline{c} \\ &= \underline{y}^T \underline{y} - \underline{y}^T \underline{\mathbf{B}}\underline{c} - \underline{c}^T \underline{\mathbf{B}}^T \underline{y} + \underline{c}^T \underline{\mathbf{B}}^T \underline{\mathbf{B}}\underline{c} + \lambda \underline{c}^T \underline{\mathbf{R}} \underline{c} \\ &= \underline{y}^T \underline{y} - 2 \underline{c}^T \underline{\mathbf{B}}^T \underline{y} + \underline{c}^T \underline{\mathbf{B}}^T \underline{\mathbf{B}}\underline{c} + \lambda \underline{c}^T \underline{\mathbf{R}} \underline{c} \end{aligned} \quad (1.20)$$

şeklindedir.

Ayrıca klasik ceza yaklaşımı için bir x eğrisinin ne kadar pürüzlü olduğunu ölçmenin global olarak kabul edilen bir yolu $\|D^2 x\|^2$ şeklindeydi. Bu değer \mathbf{R} ($K \times K$) boyutlu simetrik bir pürüzlü ceza matrisi olmak üzere, vektör notasyonunda gösterilecek olursa,

$$\begin{aligned}
 \text{PEN}_2(x) &= \int_a^b (D^2 x(t))^2 dt = \|D^2 x\|^2 = \int (D^2 \underline{c}^T \underline{\mathbf{B}}(t))^2 dt \\
 &= \int (D^2 \underline{c}^T \underline{\mathbf{B}}(t)) (D^2 \underline{\mathbf{B}}^T(t) \underline{c}) dt = \underline{c}^T \int (D^2 \underline{\mathbf{B}}(t)) (D^2 \underline{\mathbf{B}}^T(t)) dt \underline{c} \\
 &= \underline{c}^T \mathbf{R} \underline{c} \tag{1.21}
 \end{aligned}$$

şeklindedir.

Cezalı Kareler Toplamını minimize eden pürüzlü ceza tahminleri vektörü $\hat{\underline{c}}$,

$$\begin{aligned}
 \frac{\partial \text{CKT}_\lambda}{\partial \underline{c}} &= -2 \mathbf{B}^T \underline{y} + 2 \mathbf{B}^T \mathbf{B} \underline{c} + 2 \lambda \mathbf{R} \underline{c} \\
 &= -2 \mathbf{B}^T \underline{y} + 2 \mathbf{B}^T \mathbf{B} \hat{\underline{c}} + 2 \lambda \mathbf{R} \hat{\underline{c}} = 0 \\
 2 \mathbf{B}^T \underline{y} &= 2 \mathbf{B}^T \mathbf{B} \hat{\underline{c}} + 2 \lambda \mathbf{R} \hat{\underline{c}} \\
 \mathbf{B}^T \underline{y} &= (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R}) \hat{\underline{c}} \\
 (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R})^{-1} \mathbf{B}^T \underline{y} &= (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R})^{-1} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R}) \hat{\underline{c}} \\
 \hat{\underline{c}} &= (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R})^{-1} \mathbf{B}^T \underline{y} \tag{1.22}
 \end{aligned}$$

şeklinde elde edilir.

Veri uyum vektörü ise $\hat{\underline{y}} = \mathbf{B} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R})^{-1} \mathbf{B}^T \underline{y} = \mathbf{S}_{\mathbf{B},\lambda} \underline{y}$ şeklindedir.

Burada düzgünleştirme matrisi $\mathbf{S}_{\mathbf{B},\lambda}$ ile belirtilir.

Buraya kadar splaynları düzgünleştirirken öncelikle kısaca Ağırlıklandırılmamış ve Ağırlıklandırılmış En Küçük Kareler yaklaşımı ve daha sonra Ağırlıklandırılmamış Pürüzlü Ceza Yaklaşımı açıklanmıştır. Son olarak da Pürüzlü Ceza Yaklaşımında artıkların ağırlıklandırıldığı durum ele alınacaktır. w_1, w_2, \dots, w_n şeklinde pozitif ağırlıklar olduğu varsayıldığında, Cezalı Ağırlıklandırılmış Kareler Toplamı ($CAKT_\lambda$) aşağıdaki biçimde ifade edilebilir:

$$CAKT_\lambda = \sum_j w_j (y_j - x(t_j))^2 + \lambda \|D^2x\|^2 \quad (1.23)$$

Ağırlıklandırma en küçük kareler yaklaşımında bahsedildiği gibi artıkların varyanslarının tersi olacak biçimde seçilir.

Cezalı Ağırlıklandırılmış En Küçük Kareler Kriteri vektör bazında ifade edilirse,

$$CAKT_\lambda = [\underline{y} - \mathbf{B}\underline{c}]^T \mathbf{W} [\underline{y} - \mathbf{B}\underline{c}] + \lambda \underline{c}^T \mathbf{R}\underline{c} \quad (1.24)$$

şeklinindedir. Cezalı Ağırlıklandırılmış En Küçük Kareler değerini minimize eden $\hat{\underline{c}}$ 'nin doğrusal kombinasyonu ile oluşturulan Splayn tüm iki kez türev alınabilen eğriler içinde bu değeri minimize eden en düzgün eğridir. Cezalı Ağırlıklandırılmış Kareler Toplamını minimize eden katsayılar vektörü $\hat{\underline{c}}$, (1.22) numaralı formüle benzer biçimde

$$\hat{\underline{c}} = (\mathbf{B}^T \mathbf{W} \mathbf{B} + \lambda \mathbf{R})^{-1} \mathbf{B}^T \mathbf{W} \underline{y} \quad (1.25)$$

şeklinde tahminlenir.

Veri uyum vektörü ise $\hat{\underline{y}} = \mathbf{B} (\mathbf{B}^T \mathbf{W} \mathbf{B} + \lambda \mathbf{R})^{-1} \mathbf{B}^T \mathbf{W} \underline{y} = \mathbf{S}_{B,\lambda} \underline{y}$ biçimindedir. Dikkat edilecek olunursa, ağırlıklandırılmamış durumda \mathbf{W} şeklinde belirtilen ağırlık matrisinin yerini birim matris alır.

Ağırlıklandırılmış ve Ağırlıklandırılmamış 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).

1.1.2.1 Düzgünleştirme parametresinin belirlenmesi

Düzgünleştirme parametresi λ 'nın seçimi eğri tahmininde birçok yerde karşılaşılan bir problemdir. Örneğin verilere polinomlar aracılığı ile eğri uyumlaştırılıyorsa, bu durumda düzgünleştirme parametresinin seçimi uyumlaştırılan polinomun derece seçimine eşittir. Splayn düzgünleştirmede ise düzgünleştirme parametresi açık bir biçimde belirlenmektedir.

Düzgünleştirme parametresinin seçimi ile ilgili Green ve Silverman (1994; 29) iki farklı yaklaşımdan bahsetmişlerdir. Bunlardan bir tanesi subjektif bir diğeri ise otomatik seçimdir. İlk yaklaşım, düzgünleştirme parametresinin serbest seçimini prosedürün avantajlı bir özelliği olarak ele almaktır. Düzgünleştirme 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 olan olabilir ve uygulamada sıklıkla kullanılmaktadır.

Düzgünleştirme parametresinin seçimi için, subjektif yaklaşımın karşıtı gibi görünen diğer bir yaklaşım otomatik seçim metodudur. Bu tip bir metod için otomatik kelimesi objektiften daha uygundur. Bunun nedeni de metodun kendisinin seçiminde içerilen keyfi kararlar bulunmasıdır(Green & Silverman, 1994; 29).

Deneyimsiz bir kullanıcı metod tamamen otomatik ise kendisini güvende hisseder. Otomatik seçim daha sonraki uygulamalar için bir başlangıç noktası olabilir. Sonuçlarını sunan ve karşılaştırmak isteyen bilim adamları standart bir metodu referans vermek isteyebilirler. Eğer düzgünleştirme metodu geniş ve tekrarlı değerler içeren büyük bir prosedürün bir parçası olarak uygulanacak ise bu durumda otomatik metod gerekli olabilir. Otomatik metodlarla ilgili olarak kullanıcı uyarılmalı ve bu metodlar tamamen körü körüne uygulanmamalıdır. Bu nedenle de Silverman (1985) kontrol parametrelerinin açıkça belirlenmediği bu metodlara objektif demektense otomatik demeyi tercih etmektedir. Genelde otomatik metodlara sıkı sıkıya uymaktansa onların bir yönlendirici olarak kullanılması tavsiye edilmektedir(Ramsay & Silverman, 2002; 39).

Düzgünleştirme parametresinin belirlenmesinde en yaygın olarak 6 tane otomatik metod kullanılmaktadır. Bunlardan dördü klasik metod diğer ikisi risk tahmin yöntemi olarak bilinmektedir. Düzgünleştirme parametresinin seçimi için kullanılan klasik metodlar Çapraz Geçerlilik (Cross-Validation (CV)), Genelleştirilmiş Çapraz Geçerlilik (Generalized Cross-Validation (GCV)), Geliştirilmiş Akaike Bilgi Kriteri (Improved Akaike Information Criterion (AIC_c)) ve Mallows'un C_p kriteri (Mallows C_p Criterion) , kalan iki risk tahmin metodu Klasik Pilotları Kullanan Risk Tahmini (Risk Estimation Using Classical Pilots (RECP)) ve Lokal Risk Tahmin (Local Risk Estimation (LRS)) kriteridir. Ayrıntılı bilgi için Craven ve Wahba (1979), Eubank (1985), Hutchinson ve Hoog (1985), Raz vd. (1989), Hardle (1997;147-187), Hurvich vd. (1997), Wei (2005), Aydın (2005) çalışmalarına başvurulabilir.

Yukarıda belirtilen metodlardan en çok kullanılanları Çapraz Geçerlilik ve Genelleştirilmiş Çapraz Geçerlilik'tir. Bu iki metod eşit ölçeklenmiş veriler için çok yakın λ değerleri verirler, ancak eşit olmayan ölçekli veriler için Genelleştirilmiş Çapraz Geçerliliğin kullanılması önerilmektedir. Ayrıca çapraz geçerliliğin hesaplanması daha çok vakit kaybettirmektedir(Simonoff, 1996; 172).

Ramsay ve Silverman 'a (2002; 56) göre düzgünleştirme uygulayan çoğu araştırmacının düzgünleştirme 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" isimli çalışmasında uygulamalarla ilgili kapsamlı bir araştırma sonucu düzgünleştirme parametresi λ için 10^{-4} , 10^{-3} ve 10^{-2} değerlerinin iyi çalıştığı gözlenmiştir(Ramsay & Li;1998).

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

$$\lambda=10^{-4} [\text{tr} (\mathbf{B}^T \mathbf{B}) / \text{tr} \mathbf{R}] \quad (1.26)$$

şeklinde tanımlanmıştır. Diğer taraftan, W ağırlıklar matrisi dikkate alınırsa, benzer bir tanımlama

$$\lambda = 10^{-4} [\text{tr} (\mathbf{B}^T \mathbf{W} \mathbf{B}) / \text{tr} \mathbf{R}] \quad (1.27)$$

şeklinde yapılabilir.

Bu çalışmada da uygulamada sıklıkla kullanılan düzgünleştirme parametresi ilk önce (1.26) numaralı denklemlerle verilmiş olan şekilde belirlenecek ve buradan yola çıkarak çeşitli subjektif belirlenen λ değerleri için karşılaştırmalar yapılacaktır.

1.2. FONKSİYONEL VERİ İÇİN TANIMLAYICI İSTATİSTİKLER

1.2.1. Ortalama Fonksiyonu

Tek değişkenli veri için klasik özet istatistikler benzer biçimde fonksiyonel verilere de uygulanır.

Fonksiyonlar setinin yerleşimi (location) bireyler üzerinden nokta ortalamaları alınarak özetlenebilir. Bir diğer deyişle, ortalama fonksiyonu, bireylere karşılık gelen nokta değerli fonksiyonların ortalamasıdır. İstatistiksel analiz için, N gözlem eğrisinin t_1, t_2, \dots, t_n noktalarında gözleendiği varsayılmaktadır. i .inci eğrideki j .inci nokta $x_i(t_j)$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, n$ ile belirtilmektedir .

Özetlemek gerekirse ortalama fonksiyonun elde edilmesi için her bir düğüm değerine karşılık gelen ortalamalara uyan eğri,

$$\bar{x}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t_j) \quad (1.28)$$

şeklinde elde edilir.

Alternatif olarak eğer fonksiyonlar büyük bir gürültü (noise) bileşenine veya diğer arzu edilmeyen düzgün olmayan bileşenlere sahip ise ilk olarak verilere uygun bir splayn ile yaklaşılması ve daha sonra bu fonksiyonların ortalamasının alınması önerilebilir(Ramsay, 1982). Bir diğer deyişle, tahminin doğruluğu düzgünleştirmeyi de ilave ederek geliştirilebilir(Rice & Silverman, 1991).

Bir doğal düzgünleştirme yaklaşımı Kesim 1.1.2 de görüldüğü gibi Cezalı En Küçük Kareleri kullanmaktır. $\underline{y}_i = (y_{i1}, y_{i2}, \dots, y_{in})^T$ olsun. Ortalama fonksiyonu μ ,

$$\frac{1}{N} \sum_{i=1}^N (\underline{y}_i - \mu)^2 + \lambda \| D^2 \mu \|^2 \quad (1.29)$$

değerini minimum yapan $\hat{\mu}$ eğrisi ile tahminlenir.

Verilen bu ifadeyi minimum yapmak,

$$\sum_j \left\{ \bar{y}_j - \mu(t_j) \right\}^2 + \lambda \left\| D^2 \mu \right\|^2 \quad (1.30)$$

ifadesini minimum yapmak ile aynı anlama gelir .

Bu, nokta ortalamalara uygulanan splayn düzgünleştirme yaklaşımı olarak bilinir. Tahminleyici $\bar{x}(t)$, t_j düğümlerine sahip bir kübik splayndır. Bir diğer deyişle düzgünleştirilmiş bir ortalama fonksiyonu tahminlemek için düzgünleştirme matrisini **ortalama koordinat vektörü** ile çarpmak yeterlidir. Bunun nedeni (1.30) denklemini minimize eden katsayı vektörü baz fonksiyon yaklaşımına göre,

$$\underline{\hat{c}} = (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R})^{-1} \mathbf{B}^T \underline{\bar{y}} \quad (1.31)$$

biçiminde tahminlenir ve dolayısıyla

$$\underline{\hat{\mu}} = \mathbf{B} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{R})^{-1} \mathbf{B}^T \underline{\bar{y}} = \mathbf{S}_{B,\lambda} \underline{\bar{y}} \quad (1.32)$$

şeklindedir.

1.2.2. Varyans ve Kovaryans Fonksiyonları

Fonksiyonel veri analizinde artık varyans ve kovaryanslardan değil, varyans ve kovaryans fonksiyonlarından bahsedilmektedir. Varyans fonksiyonu şu şekilde tanımlanabilir:

$$\text{Var}_X(t) = \frac{1}{N-1} \sum_{i=1}^N (x_i(t) - \bar{x}(t))^2 \quad t \in [a, b] \quad (1.33)$$

Kovaryans fonksiyonu farklı s ve t deęerleri karřısında gözlemlerin baęımlılıęını özetler ve tüm s ve t deęerleri için,

$$\text{Cov}_x(s, t) = \frac{1}{N-1} \sum_{i=1}^N \{x_i(s) - \bar{x}(s)\} \{x_i(t) - \bar{x}(t)\} \quad s, t \in [a, b] \quad (1.34)$$

řeklindedir. $\text{Cov}_x(s, t)$ fonksiyonu için literatürde aynı anlamda olmak üzere $v(s, t)$ ifadesi de kullanılabilir.

$\text{Cov}_x(s, t)$ bir varyans-kovaryans yüzeyi tanımlar. Bu yüzeyin yükseklięi, zamanın (veya ilgili deęişkenin) her bir noktasında eğrilerin deęişkenlięini gösterir ve birlikte deęişiminin ölçüsünü verir.

Gözlem deęerlerinin,

1. birey için $\underline{y}_1 = [2 \quad 4 \quad 5 \quad 3 \quad 1]^T$

2. birey için $\underline{y}_2 = [5 \quad 1 \quad 1 \quad 7 \quad 6]^T$

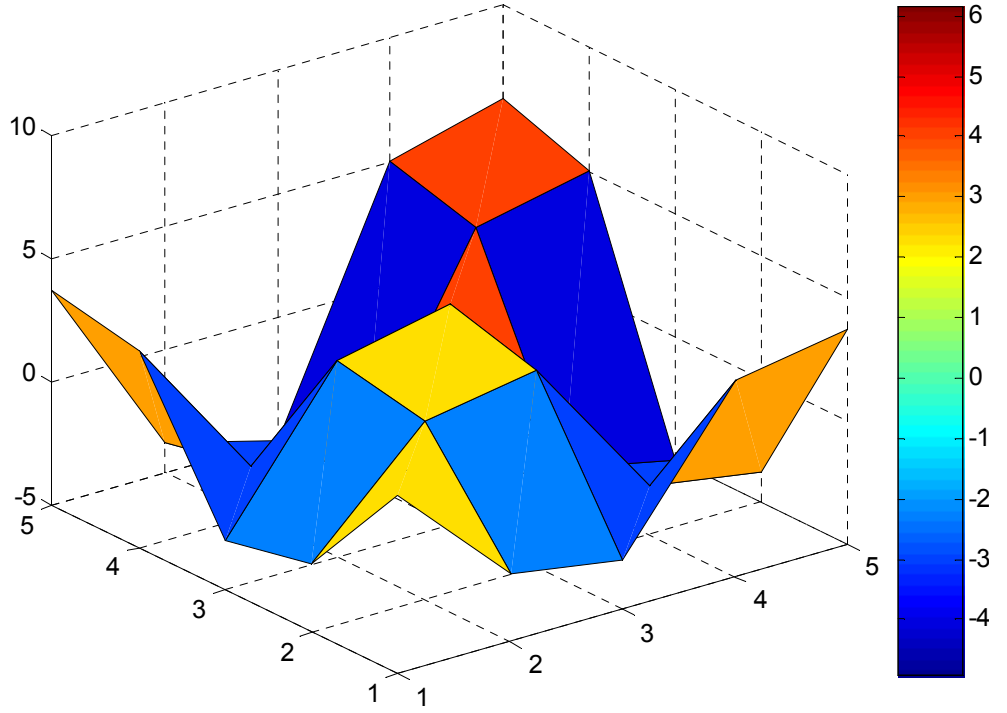
düğüm deęerlerinin de iki birey için de aynı olup,

$$\underline{t} = [1 \quad 2 \quad 3 \quad 4 \quad 5]$$

olduęu varsayıldığında ilgili kovaryans yüzeyi, s ve t noktalarında ilgili kovaryans fonksiyonunun aldığı deęerler ařaęıdaki biçimde olmak üzere,

s	t	Cov(s,t)	s	t	Cov(s,t)
1	1	2.2495	2	4	-2.9975
1	2	-2.2504	2	5	-3.7523
1	3	-2.9964	3	3	3.9913
1	4	2.9963	3	4	-3.9911
1	5	3.7507	3	5	-4.9960
2	2	2.2514	4	4	3.9909
2	3	2.9977	4	5	4.9958
			5	5	5

Tablo (1.1): Varyans Kovaryans Değerleri



Şekil (1.6): Kovaryans Yüzeyi

şeklindedir.

Korelasyon fonksiyonu ise

$$\text{Corr}_x(s,t) = \frac{\text{Cov}_x(s,t)}{\sqrt{\text{Var}_x(s)\text{Var}_x(t)}}$$

(1.35)

şeklindedir.

Daha önceden de bahsedildiği gibi fonksiyonel veri analizinin popüler bir yolu baz fonksiyon yaklaşımını kullanmaktır. Pratikte, fonksiyonel objelerin analizi,

$$C = \{c_{ik}, i=1,2, \dots, N, j=1, 2, \dots, K\}$$

şeklinde belirtilen, arzu edilen yönteme göre tahminlenen katsayılar matrisi yardımıyla yapılmaktadır. Bu anlamda ortalama, varyans, kovaryans ve korelasyon fonksiyonları B- Splayn baz fonksiyon yaklaşımına göre sırasıyla aşağıdaki biçimde özetlenebilir:

$$1\text{-Ortalama fonksiyonu} = \bar{x}(t) = \underline{\bar{c}}^T \underline{\mathbf{B}}(t) \quad (1.36)$$

$$2\text{-Varyans Fonksiyonu} = \text{Var}_x(t) = \underline{\mathbf{B}}(t)^T \text{Cov} (C) \underline{\mathbf{B}}(t) \quad (1.37)$$

$$3\text{- Kovaryans Fonksiyonu} = \text{Cov}_x(s,t) = \underline{\mathbf{B}}(s)^T \text{Cov} (C) \underline{\mathbf{B}}(t) \quad s,t \in [a, b] \quad (1.38)$$

$$4\text{- Korelasyon Fonksiyonu} = \text{Corr}_x(s,t) = \frac{\text{Cov}_x(s,t)}{\{\text{Var}_x(t)\text{Var}_x(s)\}^{1/2}} \quad (1.39)$$

Burada $\text{Cov}(C)$ katsayıların varyans kovaryans matrisidir ve aşağıdaki biçimde tahminlenir:

$$\text{Cov}(C) = \frac{1}{N-1} \sum_{i=1}^N (\underline{c}_i - \bar{c})(\underline{c}_i - \bar{c})^T \quad (1.40)$$

İKİNCİ BÖLÜM

FONKSİYONEL ANA BİLEŞENLER ANALİZİ

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 & Rice, 1992). 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) kullanımını 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).

2.1. ÇOK DEĞİŞKENLİ ANA BİLEŞENLER ANALİZİ

Ana Bileşenler Analizi bazı cebirsel ve geometrik kriterlere göre verilerin optimal bir gösterimidir. Ana Bileşenler Analizinin temel fikri, varyans–kovaryans yapısını orijinal değişkenlerin maksimum varyansa sahip doğrusal kombinasyonu aracılığı ile açıklamaya çalışarak veri setinin boyutunu indirgemektir. Toplam değişkenliği açıklamak için, p değişken sayısını göstermek üzere, p bileşen gerekli olsa bile, bu değişkenliğin büyük bir kısmı genellikle k sayıda ana bileşen tarafından

açıklanabilir. Amaç toplam değişkenliğin mantıklı bir oranını açıklayacak bir k değerinin seçimidir.

\sum varyans – kovaryans matrisinin $\psi_1 \geq \psi_2 \geq \dots \psi_p \geq 0$ şeklinde sıralanmış özdeğerleri olsun.

$\Psi = \text{diag}(\psi_1, \psi_2, \dots, \psi_p)$ tanımlı iken,

Γ matrisi,

$$\Gamma^T \Sigma \Gamma = \Psi \quad (2.1)$$

eşitliğini sağlayan ortogonal bir matris olmak üzere, ana bileşen transformasyonu,

$$\underline{x} \rightarrow \underline{y} = \Gamma^T (\underline{x} - \underline{\mu}) \quad (2.2)$$

şeklinde yazılabilir.

\underline{x} şans vektörünün i.inci ana bileşeni,

$$Y_j = \underline{\gamma}_j^T (\underline{x} - \underline{\mu}) \quad (j=1,2, \dots, p) \quad (2.3)$$

olarak tanımlanabilir. $\underline{\gamma}_j$, sütunları ψ_j şeklindeki özdeğerlerine karşılık gelen özvektörlerden oluşan Γ ortogonal matrisinin j.inci sütunudur ve ana bileşen yükleri (principal component loadings) j.inci vektörü olarak adlandırılabilir. $k=1, 2, \dots, p$ olmak üzere γ_{jk} 'nin büyüklüğü k.inci ana bileşen için j.inci değişkenin önemini ölçer.

Ana Bileşenler Analizinde Y_j şeklindeki doğrusal bileşen,

$$\begin{aligned} 1) \underline{\gamma}_j^T \underline{\gamma}_m &= \langle \underline{\gamma}_j, \underline{\gamma}_m \rangle = 1 & (j=m \text{ ise ; } j,m= 1,2, \dots, p) \\ 2) \underline{\gamma}_j^T \underline{\gamma}_m &= \langle \underline{\gamma}_j, \underline{\gamma}_m \rangle = 0 & (j \neq m \text{ ise ; } j,m= 1,2, \dots, p) \end{aligned} \quad (2.4)$$

kısıtları altında $V(Y_j) = \psi_j$ şeklinde maksimum varyansa sahiptir ve $Cov(Y_j, Y_m) = 0$ olur. İlk adımda hesaplanan maksimize edilen $V(Y_j)$ değişkenlerdeki değişimin en güçlü ve en önemli modudur. İkinci ana bileşen geriye kalan toplam varyansa maksimum katkıda bulunmaktadır. Bir diğer deyişle, birinci ana bileşen en çok diğer bileşenler ise gittikçe azalan oranda toplam varyansa katkıda bulunmaktadır. Bu nedenle, az sayıda bileşenle toplam varyansın büyük bir kısmı açıklanabilmektedir. Ana Bileşenler Analizi ile birbiriyle ilişkisiz daha az sayıda yeni değişkenler türetildiğinden dolayı bu yeni değişkenler regresyon analizi varsayımlarından çoklu doğrusal bağlantıyı ortadan kaldırması ve boyutunun indirgenmiş olması nedeni ile diğer istatistiksel analizler için girdi oluşturabilir. Ancak burada yeni değişkenlerin mantıksal olarak da anlamlı olmaları gerekmektedir. Ana Bileşenler Analizi temel olarak orijinal değişkenlerin birbirleriyle ilişkili olduklarında kullanılması uygun bir teknik olarak görülmektedir. Bunun nedeni değişkenler arasındaki korelasyon arttıkça boyut indirgeme düzeyi artmakta bir diğer deyişle ilk temel bileşenin açıkladığı varyans oranı artmaktadır. Bunun tersi durumda da aksi gerçekleşmektedir.

Ana Bileşenler Analizi araştırmanın amacına bağlı olarak genellikle her bir gözlemden ortalamayı çıkardıktan sonra verilerdeki değişimin dominant modlarını bulmak üzere kullanılır, Fonksiyonel Ana Bileşenler Analizinde de kendi ortalamasından ayrılan fonksiyonel veri setlerini açıklar, bir diğer deyişle her bir fonksiyonel verinin ortalamadan sapmasını ölçer. Bu sebeple, uygulamadan önce her bir değişken için karşılık gelen ortalama çıkartılır. Bu yapıldığında ana bileşen skorunun ortalama karesini maksimize etmek kendi varyansını maksimize etmeye karşılık gelir.

Ana bileşenler, veri setinde yer alan değişkenlerin ölçüm birimlerinin farklı olduğu ve değişkenliklerinin farklı olduğu durumlarda standardize veri matrisinden ya da korelasyon matrisinden hesaplanmalıdır(Johnson & Wichern, 1998; 468).

Çok değişkenli Ana Bileşenler Analizini içeren bir çok önemli eser bulunmaktadır. Konu ile ilgili ayrıntılı bilgi için Anderson (2003; 459-486), Mardia v.d.(1989; 213-146), Johnson ve Wichern (1998; 458-512) ve Jolliffe (2002) incelenebilir.

2.2. FONKSİYONEL ANA BİLEŞENLER ANALİZİ

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 (bunlar genelde harmonik olarak da adlandırılır) şimdi zamanın veya ilgili başka bir değişkenin fonksiyonlarıdır.

Fonksiyonel Ana Bileşenler Analizinde çok değişkenli analizdeki değişken değerleri fonksiyon değerleri $x_i(t)$ ile, kesikli indeks j sürekli indeks t ile, γ_{jk} biçimindeki ağırlıklar $\gamma_j(t)$ fonksiyon değerleri ile ve son olarak da vektör uzayındaki skaler çarpım fonksiyon uzayındaki skaler çarpımlarla yer değiştirir. Aynı zamanda toplam indisi integrale dönüşür.

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:

$$Y_j = \langle \gamma_j, x - E(x) \rangle = \int \gamma_j(t) \{ x(t) - E x(t) \} dt \quad (2.5)$$

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).

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

$$\|\gamma_1\|^2 = \int \gamma_1(t)^2 dt = 1 \quad (2.6)$$

kısıtı altında, doğrusal bileşenin varyansı olan,

$$\text{Var}(Y_j) = \text{Var} \langle \gamma_j, x - E(x) \rangle = \iint \gamma_j(s) \text{Cov}(s,t) \gamma_j(t) ds dt \quad (2.7)$$

ifadesini maksimum yapacak biçimde belirlenir. (2.7) numaralı denklemin sağ tarafı klasik çok değişkenli analizde şans vektörlerinin doğrusal kombinasyonlarının varyansını gösteren kuadratik forma karşılık gelir (Lobster & Villa, 2004). Burada $\text{Cov}(s,t)$ biçiminde ifade edilen kovaryans fonksiyonu yerine korelasyon fonksiyonunun kullanılması için daha az neden vardır. Bunun nedeni fonksiyon değerlerinin hepsinin aynı birim veya ölçekte olmasıdır.

İkinci ağırlık fonksiyonunun hesaplanması için Klasik Ana Bileşenler Analizinde olduğu gibi Fonksiyonel Ana Bileşenler Analizinde de ağırlık fonksiyonunun

$$\langle \gamma_j, \gamma_m \rangle = \int \gamma_j(t) \gamma_m(t) dt = 0 \quad (j \neq m) \quad (2.8)$$

şeklindeki ilave kısıt olan ortogonallik 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 & Silverman, 1997; 88). Ağırlık fonksiyonları her aşamada

maksimum deęişimi açıklayabilecek biçimde oluşturulan ortogonal baz fonksiyonlar setidir.

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

Fonksiyonel özdenklemler aşağıdaki biçimdedir:

$$\int \text{Cov}(s,t) \gamma(t) dt = \psi \gamma(s) \quad (2.9)$$

Bu ifadede,

$$\text{Cov}(s,t) = N^{-1} \left\{ \sum_{i=1}^N x_i(s) x_i(t) \right\} \quad (2.10)$$

şeklinde. Bu kovaryans ifadesinde $x_i(t)$ her bir birimden ortalama fonksiyon değerinin çıkarılmış halidir. Payı ortalamalar tahminlendiğinden dolayı N yerine N-1'e bölmek tavsiye edilebilir, ancak Fonksiyonel Ana Bileşenler Analizi için aralarında önemli bir fark yoktur(Ramsay & Silverman, 1997; 91). Burada amaç kovaryans fonksiyonunun ortogonal ayrışımını sağlayarak fonksiyonel deęişimin dominant bileşenlerini ayırmaktır(Costanzo, 2005). Artık yukarıdaki özdenklemleri sağlayan farklı **özdeğer- özvektör** değil **özdeğer- özfonksiyon** çiftleri vardır. $\gamma_1, \gamma_2, \dots$ biçimindeki özfonksiyonlar karşılık geldikleri $\psi_1 \geq \psi_2 \geq \dots$ özdeğerlerine göre sıralanırlar. Özdeğerler ve özfonksiyonlar arasında aşağıdaki biçimde bir ilişki bulunmaktadır:

$$\text{Var}(Y_j) = \iint \gamma_j(s) \text{Cov}(s,t) \gamma_j(t) ds dt = \psi_j \quad (2.11)$$

Burada ψ_j biçimindeki özdeğerler, (2.11) numaralı formülden de görüldüğü üzere her bir bileşene yüklenebilir varyans miktarını göstermektedir. Orijinal verilere

tatmin edici bir yaklaşımı başarmak için bu değişim modlarından kaç tanesine gereksinim duyulduğu bilinmek istenebilir.

$k < p$ iken, örneğin ilk k tane ana bileşenin belirlediği bir sistemin gücü,

$$\frac{\sum_{j=1}^k \psi_j}{\sum_{j=1}^p \psi_j} \quad (2.12)$$

şeklindedir. Sistemin gücünün ne olması gerektiği veya bir başka ifadeyle kaç tane ana bileşenin dikkate alınması gerektiği konusunda literatürde en yaygın kullanılan testlerden birisi aynı zamanda görsel bir yöntem olan scree plot (yamaç eğimi) yöntemidir. Burada en büyükten en küçüğe doğru sıralanmış özdeğerlerin büyüklükleri özdeğerlerin sıralarına karşı bir çizgi grafiği ile gösterilir. Bileşenlerin uygun sayısına karar vermek için çizgi grafiğinin dirsek noktasına bakılır. Ana bileşenlerin sayısı geri kalan özdeğerlerin nisbeten küçük ve hemen hemen aynı olduğu noktaya kadar alınır (Johnson & Wichern, 1998; 475). Literatürde diğer bazı farklı yöntemlere de rastlanmaktadır.

2.2.1 Fonksiyonel Ana Bileşenlerin Elde Edilmesinde Baz Fonksiyon Yaklaşımı

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

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

şeklinde belirtilen sürekli 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 1de de bahsedildiği gibi,

$$x_i(t) = \sum_{k=1}^K c_{ik} \theta_{ik}(t) \quad i=1, 2, \dots, N; k=1, 2, \dots, K \quad (2.13)$$

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

$$\underline{\mathbf{x}} = [x_1, x_2, \dots, x_N]^T \quad (2.14)$$

şeklinde fonksiyon değerlerine sahip bir vektörü ve

$$\underline{\boldsymbol{\theta}} = [\theta_1, \theta_2, \dots, \theta_K]^T \quad (2.15)$$

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

$$\underline{\mathbf{x}} = \mathbf{C} \underline{\boldsymbol{\theta}} \quad (2.16)$$

şeklindedir. Baz fonksiyon yaklaşımında kovaryans fonksiyonu da

$$\text{Cov}(s,t) = \underline{\boldsymbol{\theta}}^T(s) \text{Cov}(\mathbf{C}) \underline{\boldsymbol{\theta}}(t) \quad (2.17)$$

şeklinde ifade edilebilir. Kovaryans fonksiyonunun B-Splayn baz fonksiyon açılımına Kesim 1.2.2 de değinilmiştir.

Ana bileşen ağırlık fonksiyonu (özfonksiyon) ,

$$\gamma(t) = \sum_{k=1}^K b_k \theta_k(t) = \underline{\boldsymbol{\theta}}^T(t) \underline{\mathbf{b}} \quad \text{iken,} \quad (2.18)$$

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

şeklindeki sürekli fonksiyonel özdenkleme matris notasyonunda aşağıdaki biçimde ifade edilebilir:

$$\int \underline{\theta}^T(s) \text{Cov}(\mathbf{C}) \underline{\theta}(t) \underline{\theta}^T(t) dt \underline{\mathbf{b}} = \psi \underline{\theta}^T(s) \underline{\mathbf{b}} \quad (2.19)$$

Burada

$$\mathbf{J} = \int \underline{\theta}(t) \underline{\theta}^T(t) dt \quad (2.20)$$

şeklinde (KxK) boyutlu bir matris tanımlandığında, (2.19) numaralı eşitlik aşağıdaki hali alır:

$$\underline{\theta}^T(s) \text{Cov}(\mathbf{C}) \mathbf{J} \underline{\mathbf{b}} = \psi \underline{\theta}^T(s) \underline{\mathbf{b}} \quad (2.21)$$

veya

$$\text{Cov}(\mathbf{C}) \mathbf{J} \underline{\mathbf{b}} = \psi \underline{\mathbf{b}} \quad (2.22)$$

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

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

$$\mathbf{J}^{1/2} \text{Cov}(\mathbf{C}) \mathbf{J}^{1/2} \underline{\mathbf{u}} = \psi \underline{\mathbf{u}} \quad (2.23)$$

şeklindeki $\text{Cov}(\mathbf{C}) \mathbf{J}$ 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.

$\mathbf{J}^{1/2} \text{Cov}(\mathbf{C}) \mathbf{J}^{1/2}$ simetrik matrisinin özdeğerlerinin ve özvektörlerinin numerik çözümü herhangi bir yüksek kaliteli özdeğer - özfonksiyon yazılım paketi ile başarılabilir ve simetrik olmayan $\text{Cov}(\mathbf{C}) \mathbf{J}$ matrisinin özdeğer ve özfonksiyonlarının doğrudan doğruya bulunmasından daha çok tercih edilmektedir(Castro v.d., 1986).

Yukarıdaki belirtilen süreç aşağıda bir algoritma ile özetlenecektir:

1) $\text{Cov}(\mathbf{C})$ ve \mathbf{J} matrisleri hesaplanır.

2) $\mathbf{J}^{1/2}$ matrisini hesaplamak için Cholesky ayrışımı kullanılabilir.

Bir simetrik ve pozitif tanımlı \mathbf{A} matrisi verildiğinde Cholesky ayrışımı $\mathbf{A} = \mathbf{B}\mathbf{B}^T$ özelliğini sağlayan bir \mathbf{B} alt üçgen matrisi oluşturur. \mathbf{B} matrisi bazen \mathbf{A} matrisinin karekökü olarak da adlandırılabilir. Cholesky ayrışımı sıklıkla \mathbf{A} matrisinin tersinin ve determinantının hesaplanmasında kullanılır. \mathbf{A} matrisinin determinantı \mathbf{B} matrisinin köşegen elemanlarının çarpımlarının kareköküne eşittir. Cholesky ayrışımının farklı bir formu $\mathbf{A} = \mathbf{L}^T\mathbf{L}$ biçimindedir. Burada \mathbf{L} üst üçgen matristir. Matlab programı bu formu kullanmaktadır.

3) Simetrik matris özdeğer - özvektör ayrışımından $\mathbf{J}^{1/2} \text{Cov}(\mathbf{C}) \mathbf{J}^{1/2}$ matrisinin özdeğerleri (ψ) ve özvektörleri ($\underline{\mathbf{u}}$) hesaplanır.

4) $\underline{\mathbf{b}} = \mathbf{J}^{-1/2} \underline{\mathbf{u}}$ ters dönüşümü hesaplanır.

5) $\gamma(t) = \sum_{k=1}^K b_k \theta_k(t) = \underline{\boldsymbol{\theta}}^T(t) \underline{\mathbf{b}}$ doğrusal kombinasyonu aracılığı ile ana

bileşen ağırlık fonksiyonları elde edilir.

Fourier baz şeklindeki ortonormal fonksiyonel baz için \mathbf{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 plot edilmesidir.

$$\gamma_{ij} = \langle \gamma_j, x_i - E(x) \rangle = \mathbf{b}_j^T \mathbf{J} (\mathbf{c}_i - \bar{\mathbf{c}}) \quad (2.24)$$

Burada γ_{ij} ana bileşen skoru, i.inci gözlemin j.inci ana bileşenin 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). Ana bileşen skorlarının dağılımı sapan gözlemlerin (outlier) saptanmasında da kullanılabilir.

Ana bileşen ağırlık fonksiyonlarının yorumlanmasına yardımcı olacak bir diğer ikinci yöntemde ağırlık fonksiyonunun bir çarpanıyla ortalama fonksiyonun karşılaştırılması olabilir. Burada ortalama fonksiyona ilgilenilen, uygun bir çarpanla çarpılmış, ana bileşen fonksiyonu eklenerek ve çıkarılarak elde edilen fonksiyonlarla ortalama fonksiyon plot edilir ve karşılaştırmalar yapılarak çeşitli yorumlamalarda bulunulabilir.

Bu plotu 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 $\hat{\mu} \pm \sqrt{\lambda_j} \hat{\gamma}_j$ plot edilir. Ancak bu konu araştırmaya açıktır ve konuyla ilgili farklı önerilerde bulunulabilir. Tüm bileşenler için aynı sabitin kullanımı da önerilebilir (Ramsay & 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 bazın boyutu 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 ettiği belirtmişlerdir.

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

2.3. DÜZGÜNLEŞTİRİLMİŞ FONKSİYONEL ANA BİLEŞENLER ANALİZİ

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 'nin hacminin (size) değil aynı zamanda pürüzlülüğünün (iniş çıkışlarını) de kontrol edilmesi gerekir. Bu durumda , $PEN_2(\gamma) = \int (D^2 \gamma(t))^2 dt = \|D^2 \gamma\|^2$ olmak üzere, $\|\gamma_j\|^2 = \int \gamma_j(t)^2 dt = 1$ şeklindeki kısıt γ 'nın pürüzlülüğünü de dikkate alan,

$$\int \gamma_j(t)^2 dt + \lambda \int (D^2 \gamma_j(t))^2 dt = 1 \quad (2.25)$$

kısıtı ile yer değiştirir(Ramsay J, Silverman ,2002; 26). Bu aslında standart Sobolev iç çarpımlarının genelleştirilmiş halidir(Musayev & Alp, 2000; 114). Bu durumda $Var(Y_j)$, $\int \gamma_j(t)^2 dt + \lambda \int (D^2 \gamma_j(t))^2 dt$ kısıtına bölünerek Cezalı Ana Bileşen Varyansı (CABV),

$$CABV = Var(Y_j) = \frac{\int \int \gamma_j(s) Cov(s, t) \gamma_j(t) ds dt}{\int \gamma_j(t)^2 dt + \lambda \int (D^2 \gamma_j(t))^2 dt} \quad (2.26)$$

şeklinde elde edilir(Silverman; 1996). λ parametresi bölüm 1 de ayrıntılı olarak bahsedildiği gibi düzgünleştirme parametresidir. Düzgünleştirme parametresi λ sabit olarak alındığında, ana bileşen ağırlık fonksiyonu γ 'nin pürüzlülüğü arttıkça $PEN_2(\gamma)$ 'nın değeri artacağından dolayı Cezalı Ana Bileşen Varyansı azalır. Ayrıca λ sıfıra yaklaşırken Cezalı Ana Bileşen Varyansı düzgünleştirilmemiş varyansına

döner. Bir diğer bakış açısıyla ise, λ 'nın değeri arttıkça Cezalı Ana Bileşen Varyansı γ 'in pürüzlülüğünden daha çok etkilenir.

Pürüzlü ceza ikinci, üçüncü ve daha yüksek dereceli düzgünleştirilmiş ana bileşenlere ilave kısıtlar ekler. j.inci bileşen fonksiyonu (2.25) kısıtına bağlı olarak $\text{Var}(Y_j)$ 'yi,

$$\int \gamma_j(t) \gamma_m(t) dt + \lambda \int (D^2 \gamma_j(t)) (D^2 \gamma_m(t)) dt = 0 \quad j \neq m \quad (2.27)$$

ilave kısıtı ile maksimize eder.

Pürüzlü cezayı alternatif bir formda açıklamak mümkündür. Bunun için ele alınacak fonksiyonlara bazı ilave düzen koşulları empoze edilmesi gerekebilir. γ 'nin dördüncü dereceye kadar türevlerinin karelerinin integralinin alınabildiği varsayıldığında γ aşağıdaki iki koşuldan birini sağlar(Ramsay&Silverman, 1997;113):

Fonksiyonların $[a,b]$ aralığında tanımlı olduğu varsayıldığında,

- 1) $D^2 \gamma$ ve $D^3 \gamma$, $[a,b]$ aralığının sonlarında sıfırdır.
- 2) $D^2 \gamma$ ve $D^3 \gamma$ $[a,b]$ de periyodik sınır koşullarını sağlar, çünkü bu fonksiyonların $\|D^2 \gamma\|^2$ 'nin anlamlı olduğu fonksiyonlardan oluşan uzayda periyodik olması gerekmektedir.

Yukarıdaki iki koşuldan birini sağlayan fonksiyonların $[a,b]$ kapalı aralığında çok düzgün oldukları varsayılır. Bu koşullardan birincisi doğal ve ikincisi de periyodik sınır koşulları olarak adlandırılır. Bu çalışmada periyodik olmayan durumlar ile ilgilenildiğinden birinci koşul dikkate alınacaktır.

x eğrisinin, sürekli türe ve sahip olduğu ve aynı zamanda ikinci türevlerinin karelerinin integralinin alınabildiği fonksiyonlardan oluşan uzayda herhangi bir fonksiyon olduğu ve γ 'nin çok düzgün olduğu varsayıldığında,

$$\boxed{\begin{array}{l} D^2\gamma(t) = u \rightarrow D^3\gamma(t)dt = du \\ D^2x(t)dt = dv \rightarrow D x(t) = v \end{array}} \text{ olduğundan,}$$

iç çarpımlar,

$$\langle D^2x, D^2\gamma \rangle = \int_a^b D^2x(t) D^2\gamma(t) dt \quad (2.28)$$

$$= D^2\gamma(t) Dx(t) \Big|_a^b - \int_a^b D x(t) D^3\gamma(t) dt \quad (2.29)$$

şeklinde ifade edilebilir.

Doğal koşullar altında $D^2\gamma(b) = D^2\gamma(a) = 0$ olur. Bu durumda (2.29)'daki sınır değerlerinin ikisi de sıfırdır. Böylece,

$$\langle D^2x, D^2\gamma \rangle = D^2\gamma(b) Dx(b) - D^2\gamma(a) Dx(a) - \int_a^b D x(t) D^3\gamma(t) dt \quad (2.30)$$

$$= - \int_a^b D x(t) D^3\gamma(t) dt \quad (2.31)$$

şeklindedir. Diğer taraftan,

$$\boxed{\begin{array}{l} D^3\gamma(t) = u \rightarrow D^4\gamma(t)dt = du \\ D x(t)dt = dv \rightarrow x(t) = v \end{array}} \text{ olduğundan,}$$

iç çarpımlar,

$$\langle D^2 x, D^2 \gamma \rangle = - \int_a^b D x(t) D^3 \gamma(t) dt \quad (2.32)$$

$$= - (D^3 \gamma(t) x(t) \Big|_a^b - \int_a^b x(t) D^4 \gamma(t) dt) \quad (2.33)$$

şeklinde ifade edilebilir. Bu durumda doğal koşullar altında $D^3 \gamma(b) = D^3 \gamma(a) = 0$ olur. Bu durumda (2.33)'daki sınır değerlerinin ikisi de sıfır olur. Böylece,

$$\langle D^2 x, D^2 \gamma \rangle = - (D^3 \gamma(b) x(b) - D^3 \gamma(a) x(a) - \int_a^b x(t) D^4 \gamma(t) dt) \quad (2.34)$$

$$= \int_a^b x(t) D^4 \gamma(t) dt = \langle x, D^4 \gamma \rangle \quad (2.35)$$

şeklinde ifade edilebilir.

(2.35) numaralı formüle benzer biçimde

$$\|D^2 \gamma\|^2 = \langle \gamma, D^4 \gamma \rangle \quad (2.36)$$

yazılabilir.

Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizinde amaç, yalnızca varyansı maksimize etmek değil aynı zamanda varyans maksimizasyonu ile ana bileşen ağırlık fonksiyonu γ 'nın pürüzlülüğünün çok büyük olmasının engellenmesi arasında bir denge kurmaktır. Bu amaçla $\int \gamma_j^2(t) dt + \lambda \int (D^2 \gamma_j(t))^2 dt = 1$ kısıtı altında yukarıda bahsedildiği gibi (2.26) numaralı formülle daha önce verilmiş olan,

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

Cezalı Ana Bileşen Varyansının (CABV) maksimize edilmesi amaçlanmaktadır.

Cezalı Ana Bileşen Varyansı (CABV),

$$CABV = \text{Var} (Y_j) = \frac{\langle \gamma, \text{Cov} \gamma \rangle}{\|\gamma\|^2 + \lambda \|D^2 \gamma\|^2} \quad (2.37)$$

şeklinde yeniden ifade edilebilir.

(2.36) numaralı formüle göre $\|D^2 \gamma\|^2 = \langle \gamma, D^4 \gamma \rangle$ eşitliği yazılabilir. Bu durumda (2.37) eşitliği tekrar,

$$CABV = \text{Var} (Y_j) = \frac{\langle \gamma, \text{Cov} \gamma \rangle}{\|\gamma\|^2 + \lambda \langle \gamma, D^4 \gamma \rangle} \quad (2.38)$$

şeklinde ifade edilebilir. Burada,

$$\|\gamma\|^2 + \lambda \langle \gamma, D^4 \gamma \rangle = \langle \gamma, (I + \lambda D^4) \gamma \rangle \quad (2.39)$$

eşitliği sağlanmaktadır. Bu durumda Cezalı Ana Bileşen Varyansı (CABV) son olarak,

$$CABV = \text{Var} (Y_j) = \frac{\int \int \gamma_j(s) \text{Cov}(s, t) \gamma_j(t) ds dt}{\int \gamma_j(t) (I + \lambda D^4) \gamma_j(t) dt} \quad (2.40)$$

şeklinde düzenlenmiştir.

Cezalı Ana Bileşen Varyansını maksimize eden özfonksiyon γ ,

$$\int \text{Cov}(s, t) \gamma(t) dt = \psi (I + \lambda D^4) \gamma(s) \quad (2.41)$$

şeklinde ifade edilmiş olan özdenklemin en büyük özdeğerine karşılık gelen özfonksiyondur. Bir diğer 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 formu

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

şeklinde dir. Modifiye ortogonallik koşulunun kullanımı,

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

şeklindeki tek bir λ değeri için genelleştirilmiş özdeğer – özfonksiyon problemini çözerek tüm gereksinim duyulan özfonksiyonların tahminlerinin elde edilebilmesi demektir(Ramsay & Silverman, 1997; 115).

2.3.1 Düzgünleştirilmiş Fonksiyonel Ana Bileşenlerin Elde Edilmesinde Baz Fonksiyon Yaklaşımı

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

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

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

$$\int \gamma(t)^2 dt + \lambda \int (D^2 \gamma(t))^2 dt = \underline{\mathbf{b}}^T (\mathbf{J} + \lambda \mathbf{K}) \underline{\mathbf{b}} = 1 \quad (2.42)$$

kısıtı altında,

$$\iint \gamma(s) \text{Cov}(s, t) \gamma(t) ds dt = \underline{\mathbf{b}}^T \mathbf{J} \text{Cov}(\mathbf{C}) \mathbf{J} \underline{\mathbf{b}} \quad (2.43)$$

şeklindeki kuadratik formun maksimize edilmesidir. Bu amacı gerçekleştirmenin yollarından biri aşağıdaki adımları uygulamaktır.

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 için Cholesky ayrışımı kullanılır. Burada \mathbf{L} üst üçgen matristir.

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

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

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

kombinasyonu aracılığı ile ana bileşen ağırlık fonksiyonları elde edilir.

Son olarak özetlemek gerekirse, \mathbf{u}_j vektörü, $(\mathbf{L}^{-1})^T \mathbf{J} \text{Cov}(\mathbf{C}) \mathbf{J} \mathbf{L}^{-1}$ matrisinin normalize özvektörü ve $\mathbf{L}^{-1} \mathbf{u}_j$ vektörü de j . inci düzgünleştirilmiş ana bileşen ağırlık fonksiyonunun baz katsayısı vektörüdür.

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

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

2.4 FONKSİYONEL ANA BİLEŞENLER ANALİZİNE LİTERATÜRDEN BAZI ÖRNEKLER

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 ortalama 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 & Rice, 1992). Dünyada internetin hızla yayılmasıyla internetten toplanan verilere Fonksiyonel Ana Bileşenler Analizi uygulamaları da hızla yayılmaya başlamıştır. Hyde vd.'nin (2006) yaptığı bir araştırmaya göre 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 deęişimlerin temel nedenleri açıklanmaya çalışılmıştır (Silverman, 1995, Ramsay& 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).

ÜÇÜNCÜ BÖLÜM

UYGULAMA

Bu bölümde Fonksiyonel Veri Analizinin İstanbul Menkul Kıymetler Borsasında (İMKB) bir uygulamasına değinilmektedir. Bu amaçla öncelikle 2006 yılının dördüncü çeyreğinde İMKB 30 endeksinde yer alan hisse senetleri belirlenmiştir. Bu hisse senetleri sırasıyla Arçelik (ARCLK), Doğan Yayın Holding (DYHOL), Ereğli Demir Çelik (EREGL), Garanti Bankası (GARAN), İş Bankası C (ISCTR), Migros (MIGRS), Sabancı Holding (SAHOL), Türk Hava Yolları (THYAO), Tüpraş (TUPRS), Vakıflar Bankası (VAKBN), Denizbank (DENIZ), Doğu Otomotiv (DOAS), Finansbank (FINBN), Gsd Holding (GSDHO), İş Gayri Menkul Yatırım Ortaklığı (ISGYO), Pektim (PETKM), Şekerbank (SKBNK), Tofaş Oto Fabrika (TOASO), Turkcell (TCELL), Vestel (VESTL), Akbank (AKBNK), Doğan Holding (DOHOL), Hürriyet Gazetecilik (HURGZ), Koç Holding (KCHOL), Petrol Ofisi (PTOFS), Şişe Cam (SİSE), Türkiye Sınai Kalkınma Bankası (TSKB), Ülker Gıda (ULKER), Yapı ve Kredi Bankası (YKBNK) ve son olarak da Fortisbank (FORTIS) şeklindedir.

Bu uygulamada öncelikle 01.10.2006 ile 31.12.2006 tarihleri arasında yukarıda belirtilen 30 hisse senedinin 61 ayrıık noktada gözlenen kapanış fiyatları Splayn Düzgünleştirme ile sürekli bir fonksiyon haline dönüştürülmüş ve öncelikle oluşturulan bu 30 farklı fonksiyon ve ortalama fonksiyonu elde edilerek bunların değişim yapıları incelenmiştir. Daha sonra bu 61 tane değışkene ait kovaryans yüzeyleri oluşturulmuş ve Pürüzlü Ceza Yöntemi ile tahminlenen katsayılara Düzgünleştirilmiş Ana Bileşenler Analizi uygulanıp tüm fonksiyonlar birlikte ele alındığında gözlenmesi güç olan fonksiyonlar arasındaki değışim ortaya konulmaya çalışılmıştır. Bu adımlardan sonra İMKB 30 endeksinin oluşumunda en yüksek ağırlığa sahip ilk 11 şirket belirlenerek, bu hisse senetleriyle İMKB 30 endeksine bir yaklaşım önerilmiştir. Bu seçilen 11 şirketin içerisinde farklı sektörlerden belirlenen 2 tanesinin aylar bazında fonksiyonları ve bunların türev fonksiyonları verilmiştir. Son olarak da farklı düzgünleştirme parametresi değerleri için fonksiyonlardaki ve ana bileşen fonksiyonlarındaki değışimler incelenmiştir.

3.1. KULLANILAN PROGRAM

Bu çalışmada Matlab matematiksel programlama dili kullanılarak programlama yapılmıştır. Bu program aşağıda verilmektedir. Programın temel adımları kısaca yanlarına açıklanmıştır.

```
function [coef2, coef3, coef4, coef5, fdobj, fdobj2, fdobj3, fdobj4, fdobj5, e, w, covC, basismat] = istem1b(knots, y)
```

```
r1=knots(1,1);  
r2=knots(end,1);  
rng = [r1,r2];  
knots = knots;  
norder = 4;  
nbasis = length(knots) + norder - 2;  
hgtbasis = create_bspline_basis(rng, nbasis, norder, knots);  
plot(hgtbasis);  
pause
```

Yandaki kod grubu B-Splaynları oluşturmakta ve B-Splaynların çizimlerini yapmaktadır. Ramsay (2003) tarafından B-Splaynları tanımlayabilmek ve işlem yapabilmek için gerekli alt yapı oluşturulmuştur.

```
Lfd = int2Lfd(2);  
lambda = lambda;  
hgtfdPar = fdPar(hgtbasis, Lfd, lambda);  
[fdobj, coef, SSE, penmat, y2cMap, basismat, yhat, basisobj, fdnames] =  
smooth_basis_ist(knots, y, hgtfdPar);
```

Yukarıdaki kod grubu hgtfdPar, knots ve y de depolanan verilere dayanarak arzu edilen düzgünleştirme parametresine ve Splaynın derecesine dayalı olarak verileri düzgünleştirmek için gerekli alt yapıyı hazırlar.

```
b= ( basismat'*basismat );  
lambda=10^(-4)* [ sum(diag(b)) / sum(diag(penmat)) ]
```

Yandaki kod grubu önerilen düzgünleştirme parametresini hesaplar.

```
penmat=full(penmat);  
basismat=full(basismat);  
[sat,sn]=size(yhat);  
for i=1:sn  
    ye=y(:,i);  
    yhate=yhat(:,i);  
    ee=ye-yhate;  
    ee2=cov(ee);  
    eeet=ee*ee';  
    ee_diag=diag((diag(eeet)));  
    we=pinv(ee_diag);  
    e{1,i}=ee;  
    w{1,i}=we;  
end
```

Yandaki döngü farklı varyanslılık durumunda Pürüzlü Ceza Yöntemini ağırlıklandırmak üzere çalışır. Burada **W** ağırlık matrisi oluşturulur.

```
for i=1:sn  
    ye=y(:,i);  
    coef2e=pinv(basismat'*basismat)*basismat'*ye;  
    coef3e=pinv(basismat'*w{1,i}*basismat+lambda*penmat)*basismat'*w{1,i}*ye;  
    coef4e=pinv(basismat'*basismat+lambda*penmat)*basismat'*ye;  
    coef2(:,i)=coef2e;  
    coef3(:,i)=coef3e;  
    coef4(:,i)=coef4e;  
end
```

Yukarıdaki döngü sırasıyla En Küçük Kareler Yöntemi, Pürüzlü Ceza ve Ağırlıklandırılmamış Pürüzlü Ceza için gerekli c_i katsayılarını hesaplar.

```

for i=1:sn
    ye=y(:,i);
    yhatea=basismat*coef2;
    eea=ye-yhatea(:,i);
    ee2a=cov(eea);
    eeeta=eea*eea';
    eea_diag=diag((diag(eeeta)));
    wea=pinv(eea_diag);
    ea{1,i}=eea;
    wa{1,i}=wea;
end

```

Yandaki döngü farklı varyanslılık durumunda En Küçük Kareler Yöntemini ağırlıklandırmak üzere çalışır. Burada W ağırlık matrisi oluşturulur.

```

for i=1:sn
    ye=y(:,i);
    coef5e=pinv(basismat'*wa{1,i}*basismat)*basismat'*wa{1,i}*ye;
coef5(:,i)=coef5e;
end

```

Yukarıdaki döngü Ağırlıklandırılmış En Küçük Kareler Yönteminin katsayılarını tahminlemek üzere çalışır.

```
fdobj2 = fd(coef2, basisobj, fdnames);  
fdobj3 = fd(coef3, basisobj, fdnames);  
fdobj4 = fd(coef4, basisobj, fdnames);  
fdobj5 = fd(coef5, basisobj, fdnames);
```

```
pause  
plot(fdobj)  
hold on  
plot_ist(mean(fdobj))  
hold off  
pause  
plot(fdobj2)  
hold on  
plot_ist(mean(fdobj2))  
hold off  
pause  
plot(fdobj3)  
hold on  
plot_ist(mean(fdobj3))  
hold off  
pause  
plot(fdobj4)  
hold on  
plot_ist(mean(fdobj4))  
hold off  
pause
```

Yandaki kod grubu sırasıyla En Küçük Kareler Yöntemi, Ağırlıklandırılmış Pürüzlü Ceza Yöntemi, Pürüzlü Ceza Yöntemi ve son olarak da Ağırlıklandırılmış En Küçük Kareler Yöntemi için sırasıyla katsayı tahminleri olan coef'leri, baz fonksiyonların özelliklerini belirten basisobj'leri ve etiketleri belirten fdnames'leri fdobj de depolar, böylece fonksiyonel veri objeleri oluşturulur.

```
plot(fdobj5)
hold on
plot_ist(mean(fdobj5))
hold off
pause
```

Yukarıdaki kod grubu sırasıyla bütün yöntemler için fonksiyonları ve ortalama fonksiyonlarını çizmektedir.

```
legend('arclk','dyhol','eregl','garan','isctr','migrs','sahol','thyao','tuprs','vakbn','deniz','d
oas','finbn','gsdho','isgyo','petkm','skbnk','toaso','tcell','vestl','akbnk','dohol','hurgz','kc
hol','ptofs','sise','tskb','ulker','ykbk','fortis','ort');
```

Yukarıdaki kod fonksiyonları etiketlendirir.

```
[harmcoef,harmscr, eigvecs, Lmat, Jmat, Cmat,pcastr,covC] =istem_pca(fdobj,
2,lambda)
```

Yukarıdaki kod grubu önce fdobj’yi hangi yönteme göre girerseniz o yönteme göre tahminlenmiş katsayılara ana bileşenler analizi uygular ve ana bileşen fonksiyonlarının oluşturulması için gerekli alt yapıyı hazırlar. Burada 2 değeri dikkate alınacak temel bileşen sayısını vermektedir.

```
t=2
for i=1:t
harmonikler=basismat*harmcoef(:,i)
plot(harmonikler)
pause
hold off
end
istem_plotpca_mean(pcastr,61);
```

Yandaki döngü
temel bileşen
fonksiyonlarını
oluşturur ve çizer.

Yukarıdaki kod ortalama fonksiyonu ve ortalama fonksiyona belirli bir katsayıyla çarpılan ana bileşen fonksiyonunun eklenip çıkarılmasının etkilerini gösterir.

```
covar=basismat*covC*basismat';  
x=knots';  
y=x;  
[X,Y] = meshgrid(x,y);  
surf(X,Y,covar);  
pause
```

Yandaki kod grubu,
kovaryans yüzeyini
oluşturur.

```
scatter(harmscr(:,1),harmscr(:,2));  
pause
```

Yandaki kod grubu,
temel bileşen skorlarını
plot eder.

```
turev=fdobj;  
t=2;  
d=t+1;  
for i=0:t  
subplot(1,d,1)  
deriva= deriv(turev,i);  
subplot(1,d,i+1)  
plot(deriva)  
title([num2str(i),' türev']);  
end  
hold off  
pause
```

Yandaki kod grubu öncelikle fonksiyonların kaçınıcı dereceden arzu ediliyorsa o dereceye göre türevlerini alır ve hangi yöntemle göre arzu ediliyorsa ki burada pürüzlü ceza yöntemi ele alınmıştır, o yöntemle göre, fonksiyonun kendisini, birinci ve ikinci türevini v.b. yan yana çizer. Burada t=2 kaçınıcı dereceye kadar türev alınması arzu ediliyorsa onu belirten sayıdır.

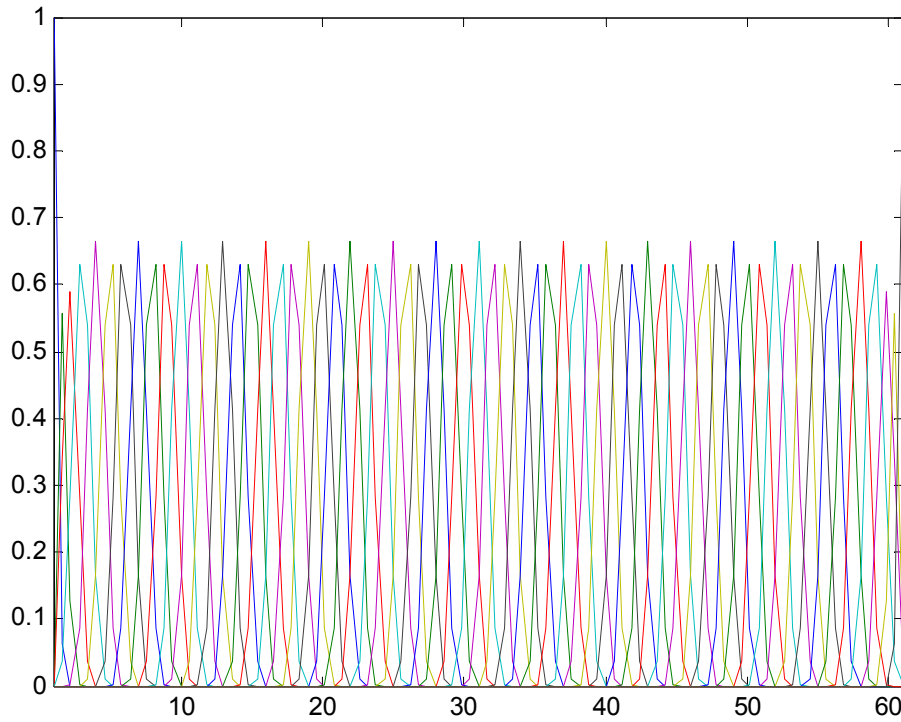
Yukarıda oluşturulan program içinde kullanılan ve Ramsay (2003) tarafından hazırlanan alt programlar EK 1 de verilmektedir.

3.2. PROGRAMIN SONUÇLARI VE YORUMLAR

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 fonksiyonu tahminlemede kullanılmıştır. Uygulamanın verileri EK 2 de verilmektedir. Buradan 10., 30. ve 50. inci v.b. günlerin hangi tarihlere karşılık geldiği açıkça görülmektedir.

3.2.1. İMKB 30 Endeksi ve İMKB 30 da Yer Alan Şirketlerin Fonksiyonel Analizi

Öncelikle İMKB 30 da bulunan şirketler analize dahil edilmişlerdir. Çalışmada birinci adım B-Splayn baz fonksiyonları elde etmektir.



Şekil (3.1): B-Splayn Baz Fonksiyonlar

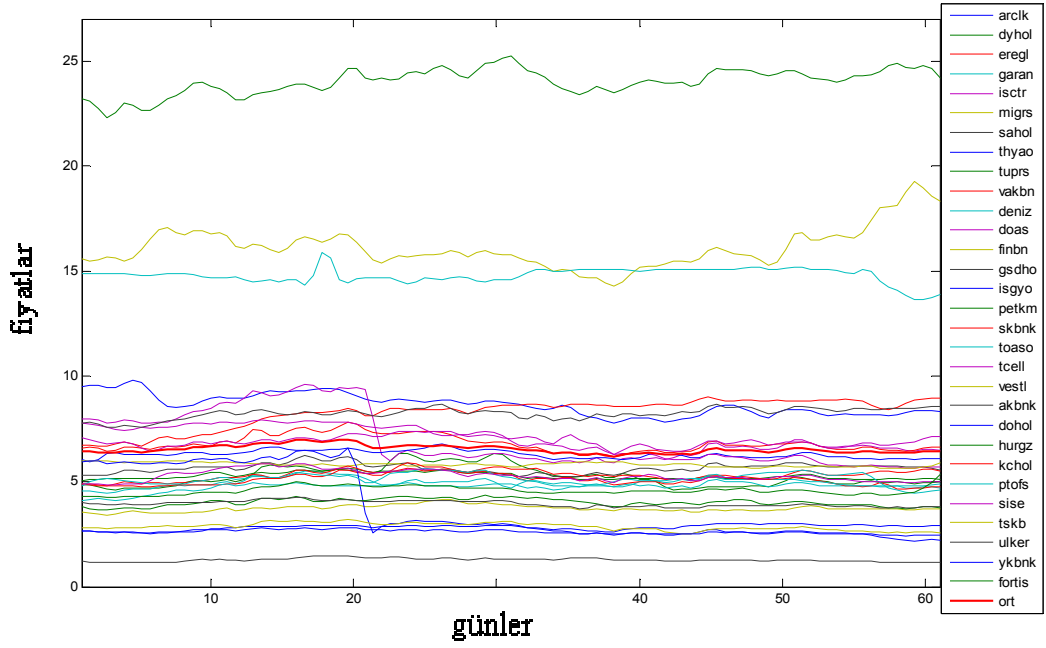
Bu çalışmada,

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

formülü ile düzgünleştirme parametresi belirlenmiştir. Bu çalışmada düzgünleştirme parametresinin değeri

$$\lambda=10^{-4} \times 0.1346$$

olarak elde edilmiştir.

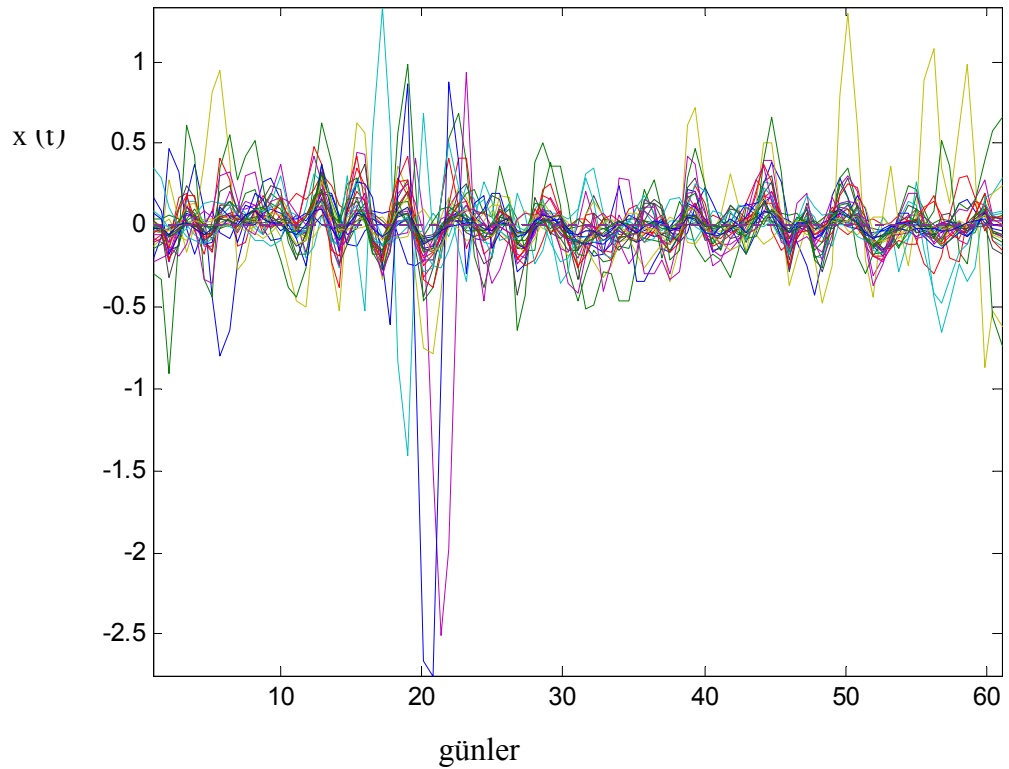


Şekil (3.2): Pürüzlü Ceza Yöntemine göre Oluşturulan Fonksiyonlar

Yukarıdaki çizimden de anlaşılacağı gibi birçok fonksiyonun bireysel davranışı diğer fonksiyonların arasında kaybolmuştur. Yukarıdaki şekilden en rahat olarak Migros, Tüpraş ve Denizbank şirketlerinin seyri belirgin bir şekilde görülmektedir ki bunlar diğer hisse senetlerine göre daha yüksek seviyede seyretilmektedirler. 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 eğrisi 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.

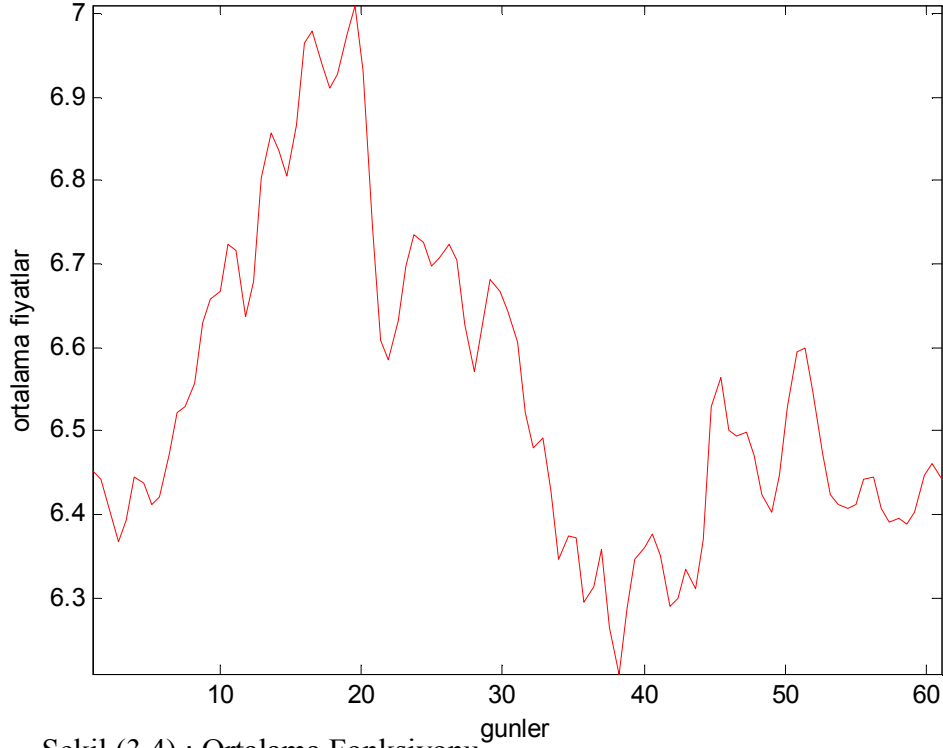
Şekil (3.3) de fonksiyonların zamana göre birinci türevleri incelenmiş ve aşağıdaki biçimde verilmiştir. Zamana göre birinci türev kapanış fiyatlarının artış ya da azalış hızını vermektedir.



Şekil(3.3):İMKB 30 Endeksi Birinci Türev Fonksiyonları

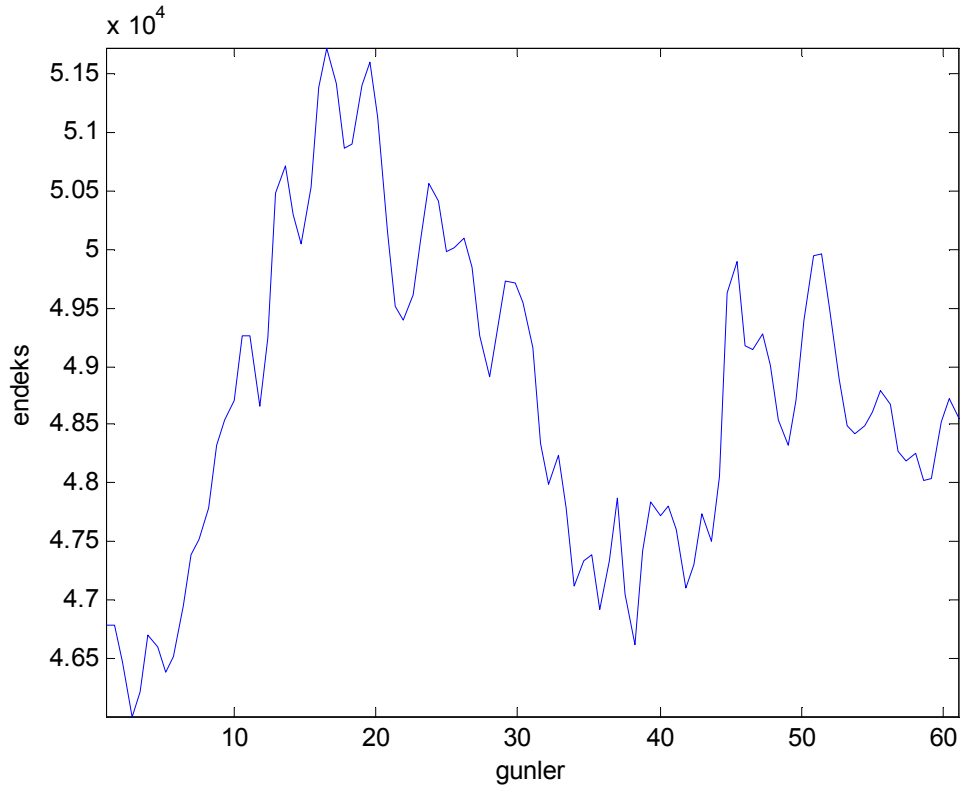
Şekil (3.2) de bazı fonksiyonlar için fark edilmesi oldukça güç olan iniş ve çıkışlar, türevin özelliğinden dolayı, türev fonksiyonları incelendiğinde daha rahatlıkla görülebilmektedir. Bu durumda fonksiyonların özellikle tek tek türevleri incelendiğinde de bu değişimler, iniş ve çıkışlar, orijinal fonksiyonlara oranla daha rahatlıkla yakalanabilir. Seçilen bazı şirketler için bu incelemeler kesim 3.2.2 de aylar bazında yapılacaktır.

B-Splaynlar, düzgünleştirilmiş fonksiyonlar ve birinci türev fonksiyonları oluşturulduktan sonra fonksiyonel verilerin ortalama fonksiyonu ayrı olarak çizdirilmiş ve Şekil (3.4) de verilmektedir.



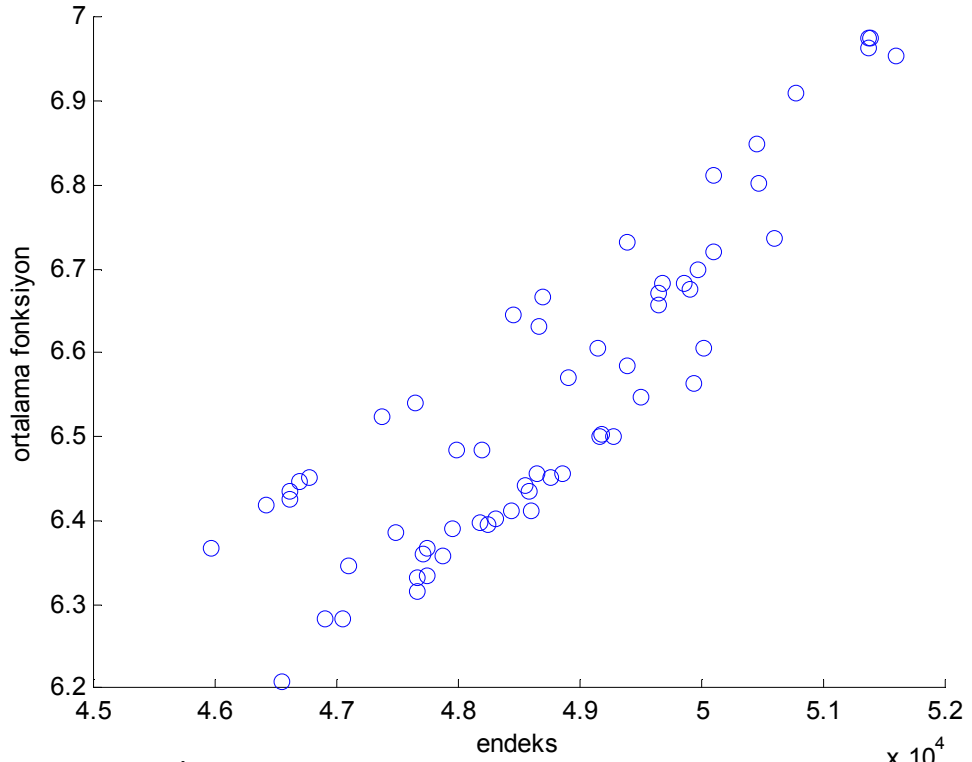
Şekil (3.4) : Ortalama Fonksiyonu

Ortalama fonksiyonunun fiyatlar açısından piyasanın genel seyrini göstermedeki yeterliliği İMKB de piyasanın genel seyrini gösteren Şekil (3.5) de verilmekte olan İMKB 30 endeksi ile kıyaslama yapılarak ortaya konulmuştur



Şekil (3.5): İMKB 30 Endeksi

Şekil (3.4) ve Şekil (3.5) birlikte ele alındığında iki şeklin birbirine çok büyük benzerlikler gösterdiği, aynı dönemlerde artışlara ve azalışlara rastlandığı yine İMKB 30 endeks fonksiyonunun da ortalama fonksiyonu gibi 3 temel bölgeye ayrıldığı görülebilmektedir. Ortalama fonksiyonu ile İMKB 30 endeksi arasındaki korelasyon Şekil (3.6) de verildiği gibidir.

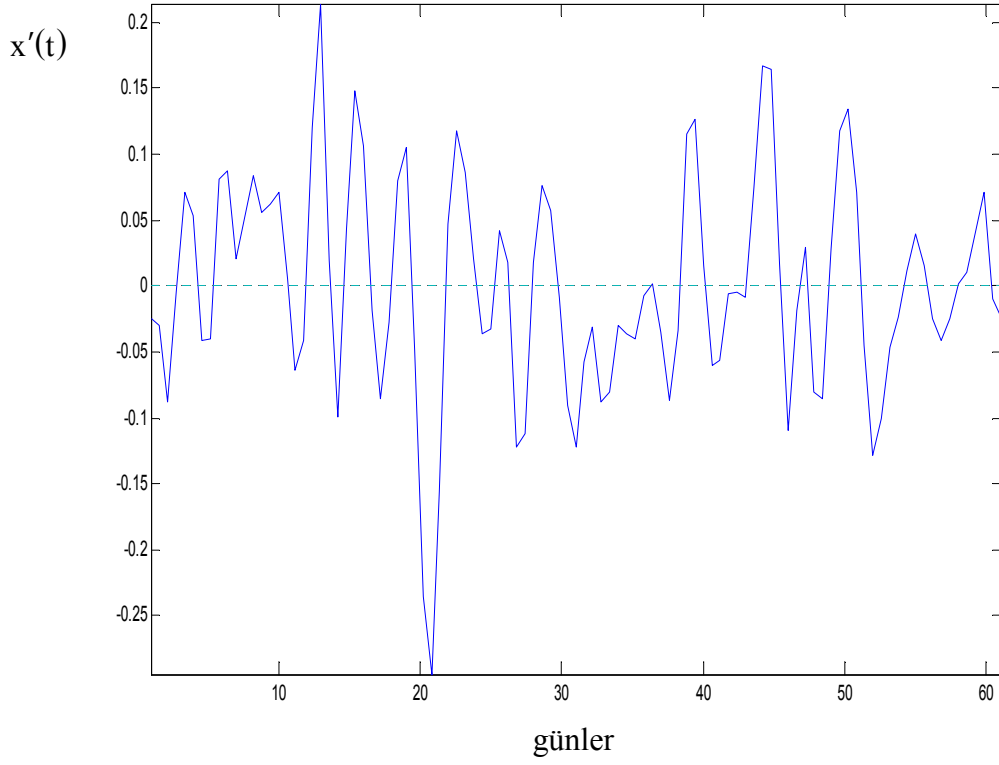


Şekil (3.6) : İMKB 30 Endeksi ile Ortalama Fonksiyonu Arasındaki

Korelasyon

Bu güçlü korelasyona ilaveten ortalama fonksiyonunun hesaplama kolaylığı ve arzu edilen dereceden türevlerinin incelenebilmesi özelliğinden dolayı ortalama fonksiyonunun İMKB 30 endeksine alternatif bir yöntem olarak kullanılması önerilebilir.

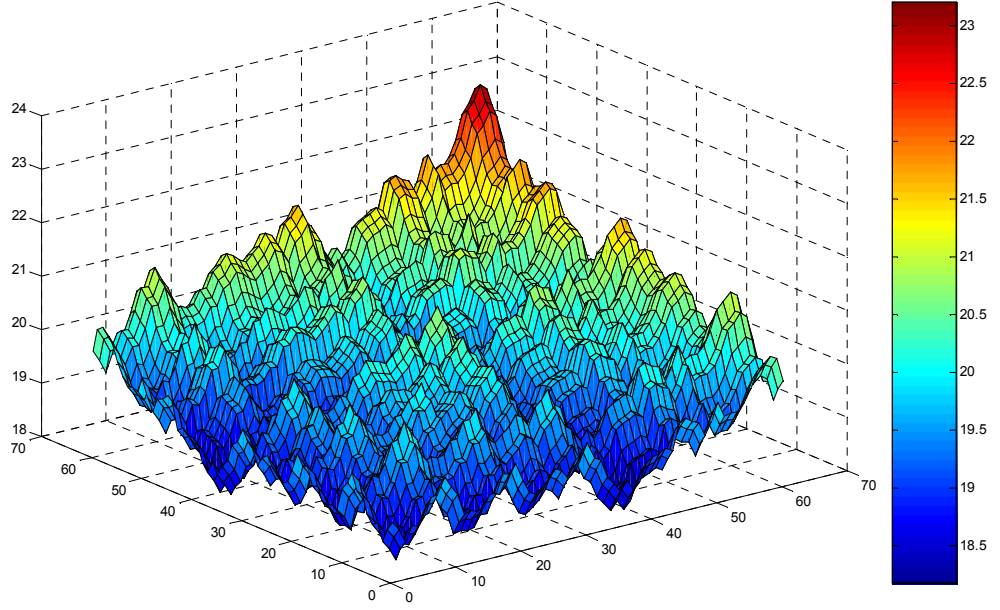
Buna ilave olarak ortalama fonksiyonunun ilgilenilen türev fonksiyonu da Şekil (3.7)de verildiği gibidir.



Şekil(3.7) : Ortalama Türev Fonksiyonu

Şekil (3.7) de verilmiş olan türev fonksiyonu incelendiğinde Şekil (3.4) deki gözlenmesi oldukça güç olan çok küçük iniş çıkışların bile rahatlıkla yakalanabildiği, türevlerin negatif ve pozitif olduğu bölgeler ile türev fonksiyonunun sıfır değerini aldığı maksimum, minimum noktalar rahatlıkla gözlemlenebilmektedir.

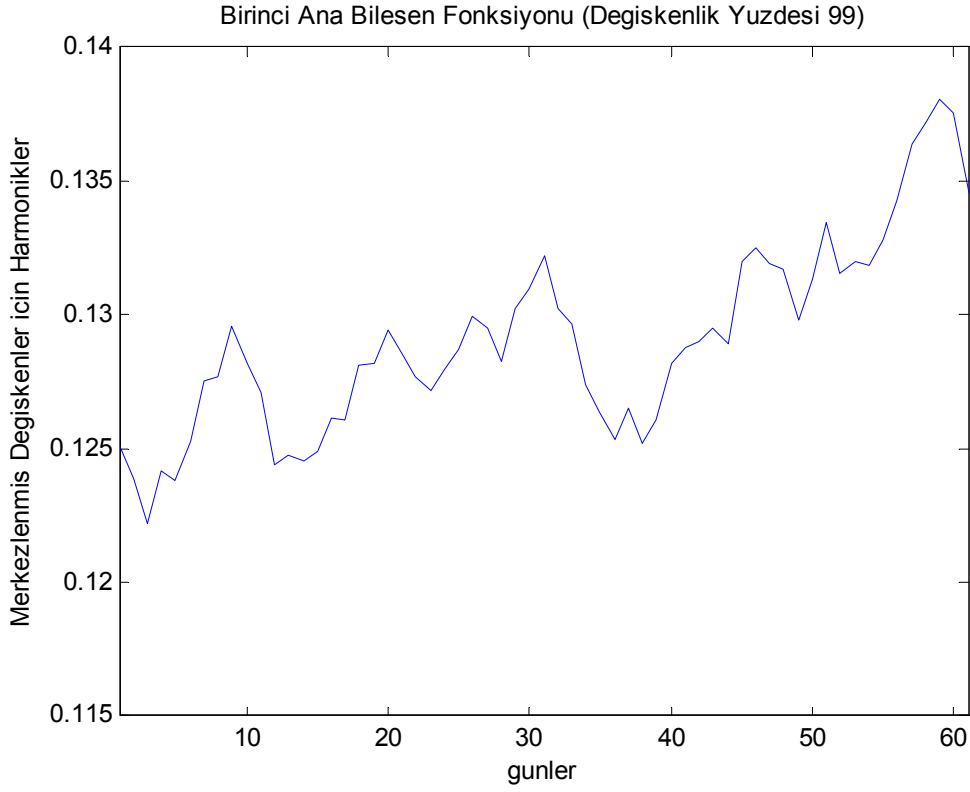
B-Splaynlar, düzgünleştirilmiş fonksiyonlar, birinci türev fonksiyonları, ortalama fonksiyonlar ve İMKB 30 endeksi karşılaştırılması verildikten sonra ilgili kovaryans yüzeyi Şekil (3.8)de olduğu gibi oluşturulmuştur.



Şekil(3.8): Kovaryans Yüzeyi

Şekil (3.8) ile verilen kovaryans yüzeyi 61 ayrıık noktada 30 birey için inceleme yapıldığından dolayı oldukça karmaşıklaşmış ve farklı t zamanlarında gözlemlerin birlikte değişimleri zor incelenebilir hale gelmiştir. Burada güç de olsa 55.inci günlerden itibaren kovaryanslarda ve varyanslarda bir artma olduğu kırmızı bölgeden seçilebilmektedir, ancak bu durum değişken sayısı arttıkça daha da karmaşıklaşmakta ve yüzey üzerinde hiçbir yükseklik tanımlanamaz hale gelmektedir. Kovaryans fonksiyonunda ilgili t değerleri yerine yerleştirildiğinde aldığı değerleri gösteren varyans- kovaryans matrisi EK 3 de verilmektedir.

Bireysel fonksiyonlar, kovaryans yüzeyi, ortalama fonksiyonu ve İMKB 30 endeksi karşılaştırılması verildikten sonra fonksiyonlar birlikte ele alındığında yorumlanması oldukça güç olan fonksiyonlar arasındaki değişimleri belirlemeye yönelik olarak ana bileşen fonksiyonu aşağıdaki biçimde elde edilmiştir.

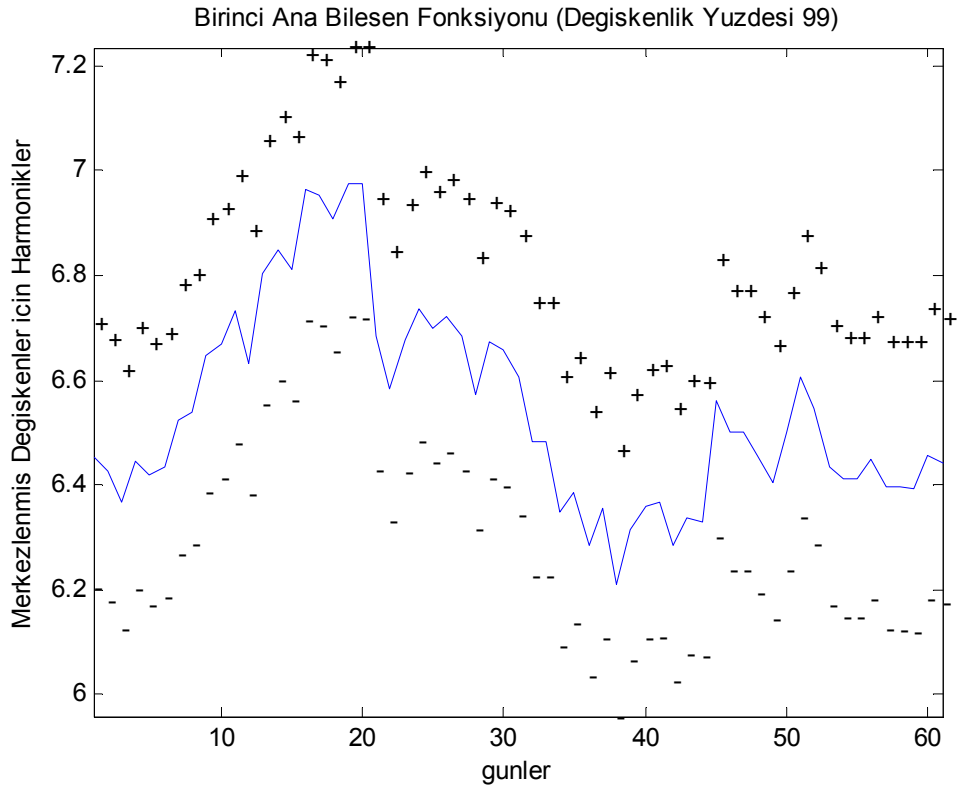


Şekil (3.9): Birinci Ana Bileşen Fonksiyonu

Bu fonksiyon incelediğinde, 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 sıfıra 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. Bunu EK 3 de verilen kovaryans matrisinden de teyit etmek mümkündür. Fonksiyonda sıfırdan mutlak uzaklıklar varyansları verdiği göre bu şekilden özellikle 55.inci günlerden sonra fiyatların değişkenliğinde bir artış olduğu ve zaman noktalarının ardışık olarak birbirleriyle pozitif bir korelasyonda bulunduğu gözlemlenebilmektedir. Şekil (3.8) de verilen ilgili kovaryans yüzeyinin de değişken sayısı ve incelenen birey sayısı arttıkça birlikte değişimi ve değişkenliği göstermesi giderek zorlaşmaktadır ki bu çalışmada

bile iki şekil kıyaslandığında ana bileşen fonksiyonu çok daha rahatlıkla yorumlanabilmektedir. Tüm bunlara ilaveten ana bileşen fonksiyonu %99 gibi çok yüksek bir değişkenlik açıklama yüzdesine sahiptir. Ana bileşen fonksiyonu ile ilgili yorumlamaların daha rahatlıkla anlaşılabilmesi için basit iki örnek Ek 4 de verilmektedir.

Ana bileşen fonksiyonu incelendikten sonra özellikle ana bileşen fonksiyonunun yorumlanmasının güç olduğu durumlarda alternatif yardımcı yöntemler olarak kullanılan ortalama fonksiyona ana bileşen fonksiyonlarının bir çarpanının eklenmesi ve çıkarılmasının etkileri Şekil (3.10) da ve buna ilaveten ana bileşen skorlarının dağılımı Şekil (3.11) de verilmektedir.



Şekil (3.10): Ortalama Fonksiyonu ve Ana Bileşen Fonksiyonu Karşılaştırması

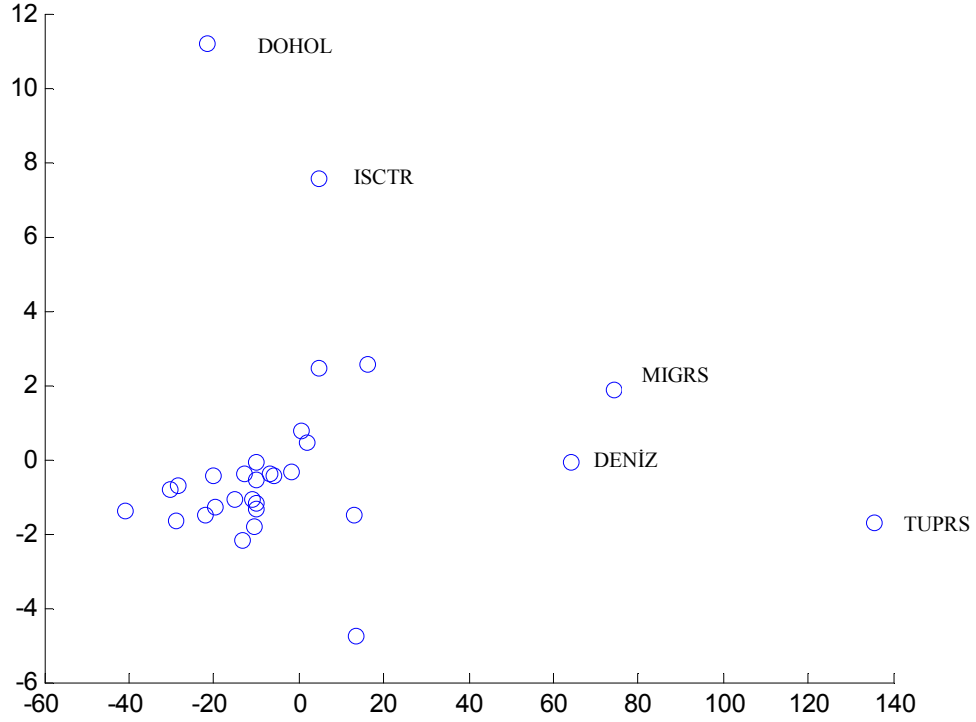
+ ve - noktalar ortalama fonksiyona ana bileşen fonksiyonunun belirli bir sabitle çarpanının eklenmesinin ve çıkarılmasının etkilerini göstermektedir. + ve - noktalar ortalama fonksiyonundan ne kadar uzaksa ortalamadan sapmaların o kadar

yüksek olduğu bu grafik yardımıyla da gözlemlenebilir. Ana bileşen fonksiyonuna benzer şekilde Şekil (3.10)dan da 55.inci günlerden itibaren sapma daha net bir biçimde görülebilmektedir.

Ana Bileşen skorları incelenmek istendiğinde skor değerleri Tablo (3.1)de ve bunların grafiksel dağılımı Şekil (3.11)de verilmektedir.

Şirketler	Birinci Ana Bileşen	İkinci Ana Bileşen	Şirketler	Birinci Ana Bileşen	İkinci Ana Bileşen
ARCLK	16.3222	2.5265	PETKM	-9.8787	-0.5661
DYHOL	-10.0785	-1.3209	SKBNK	-10.6040	-1.7919
EREGL	13.5766	-4.7843	TOASO	-13.1902	-2.1724
GARAN	-12.9261	-0.4061	TCELL	4.8486	2.4699
ISCTR	4.6272	7.5545	VESTL	-21.8127	-1.5059
MIGRS	74.2173	1.8471	AKBNK	13.2221	-1.5172
SAHOL	-6.9899	-0.4167	DOHOL	-21.7199	11.1780
THYAO	-1.8544	-0.3634	HURGZ	-19.5958	-1.3132
TUPRS	135.4085	-1.6921	KCHOL	-10.0598	-1.1884
VAKBN	2.0644	0.4684	PTOFS	-10.7793	-1.0754
DENIZ	64.0042	-0.0701	SISE	-9.7724	-0.0950
DOAS	0.5420	0.7663	TSKB	-28.5403	-0.6985
FINBN	-5.6643	-0.4631	ULKER	-20.1330	-0.4609
GSDHO	-40.7791	-1.3850	YKBNK	-30.4617	-0.7942
ISGYO	-28.7721	-1.6422	FORTIS	-15.2205	-1.0879

Tablo(3.1): Ana Bileşen Skorları



Şekil(3.11): Ana Bileşen Skorlarının Dağılımı

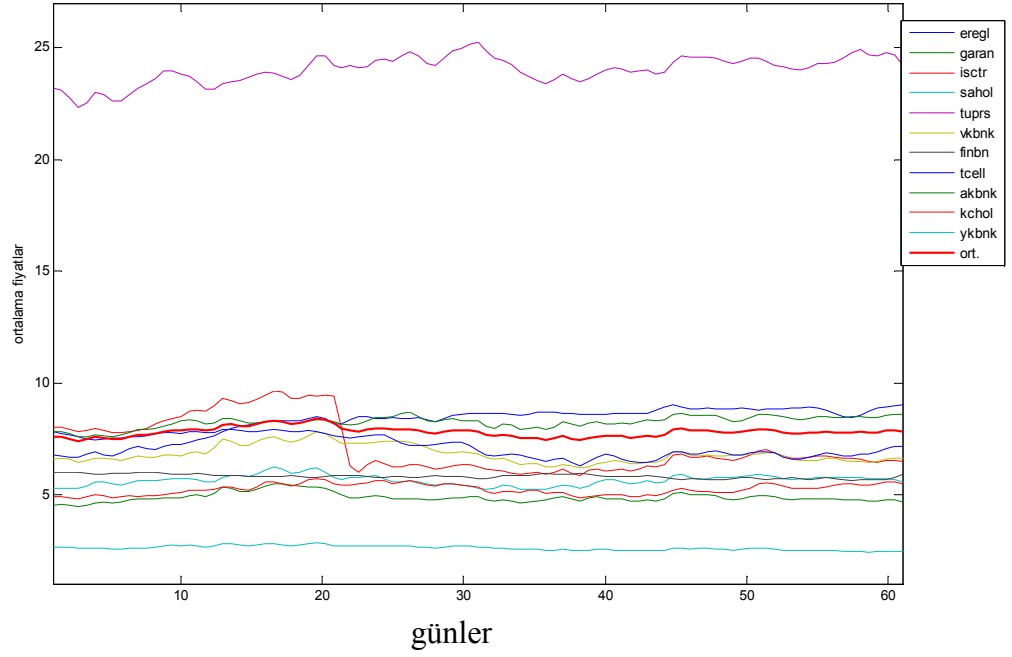
Ana bileşen skorları özellikle %99 varyans açıklayabilme gücüne sahip birinci bileşen açısından incelendiğinde Şekil (3.11) den öncelikle Tüpraş'ın daha sonra Migros ve Denizbank'ın birinci ana bileşen üzerinde oldukça etkili olduğu, yüksek bir temel bileşen skoru değeri taşıdığı görülmektedir. Tüpraş, Migros ve Denizbank'ın bir sapan gözlem gibi diğer şirketlere oranla daha uzak bir konumda yer almasının nedeni Şekil (3.2)de de görülebildiği gibi diğerlerine oranla daha yüksek fiyat ve yüksek değişkenlik içermesi olabilir.

3.2.2. İMKB 30 Endeksinden Seçilen 11 Şirketin İncelenmesi

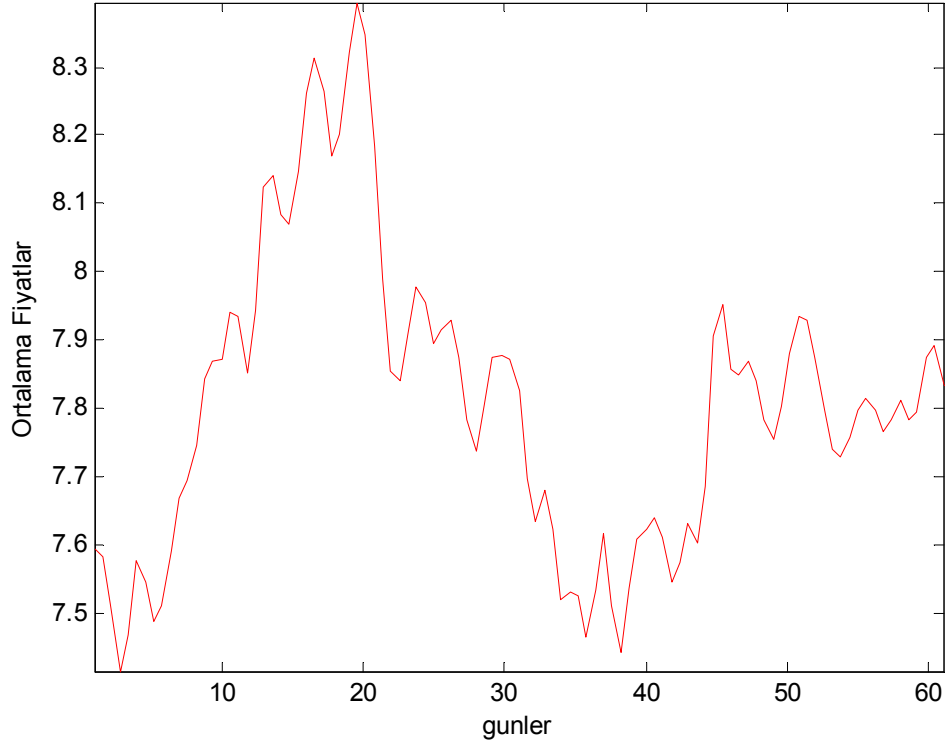
Bu kesimde İMKB 30 endeksindeki ağırlıklarına göre ilk 11 şirket belirlenmiştir. Belirlenen hisse senetleri ve hisse senetleri bazında ağırlıklar sırasıyla Ereğli Demir Çelik (%4,23), Garanti Bankası (%10,47), İş Bankası C (%12,80), Sabancı Holding (%5,34), Tüpraş (%6,04), Vakıflar Bankası (%4,40), Finansbank

(%6,73), Turkcell (%5,50), Akbank (%11,41), Koç Holding (%2,79) ve son olarak da Yapı ve Kredi Bankası (%4,14) şeklindedir.

Bu 11 hisse senedi için öncelikle yine Pürüzlü Ceza Yaklaşımına göre her bir hisse senedi için fonksiyonlar Şekil (3.12) deki biçimde oluşturulmuş ve daha sonra Şekil (3.13) de ortalama fonksiyonu ayrı olarak çizdirilmiştir.



Şekil(3.12): Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar

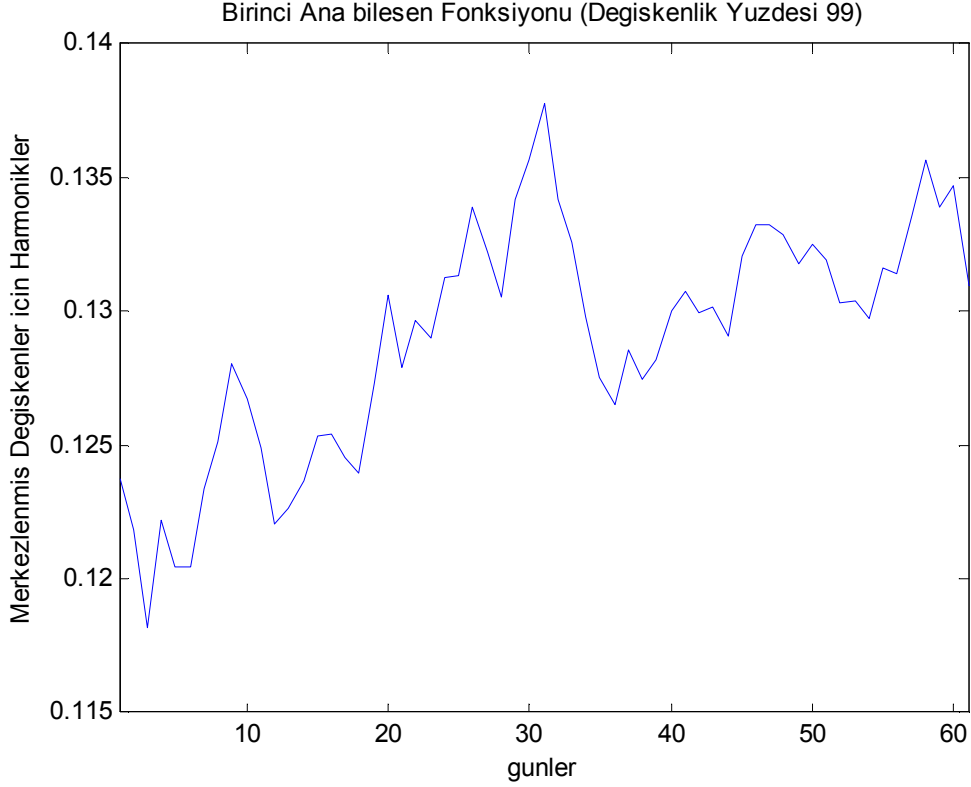


Şekil(3.13): Seçilen 11 Şirketin Ortalama Fonksiyonu

Şekil (3.13) ve Şekil (3.5) deki İMKB 30 endeksi karşılaştırıldığında ortalama fonksiyonunun 30 endeksine ne kadar yakınsadığı rahatlıkla görülebilmektedir. İncelenen bu zaman dilimi için 30 endeksin içerdiği tüm fonksiyonların ortalama fonksiyonunu incelemek yerine en azından bu zaman dilimi için belirlenen 11 şirketin ortalama fonksiyonu bile İMKB 30 endeksine alternatif bir yaklaşım olarak görülebilir.

Ancak burada üzerinde önemle durulması gereken bir konu Şekil (3.13)de verilen 11 şirketin ortalama fonksiyonu İMKB 30 endeksine bir alternatif olarak görüle bile bu belirlenen 11 şirket İMKB 30 da yer alan tüm şirketlerin temsilcisi olarak gösterilemez. Şekil (3.4) ve Şekil (3.13)de verilmekte olan ortalama fonksiyonlar birlikte incelendiğinde de görünüm olarak birbirlerine çok benzer olmalarına rağmen fiyatların değişim aralığının farklı olduğuna dikkat çekilmelidir ki burada aynı değişken açısından inceleme yapılmaktadır. Ayrıca sadece 11 şirketin dikkate alındığı bir durum 30 şirketin varyasyon yapısıyla ilgili doğru bir bilgi

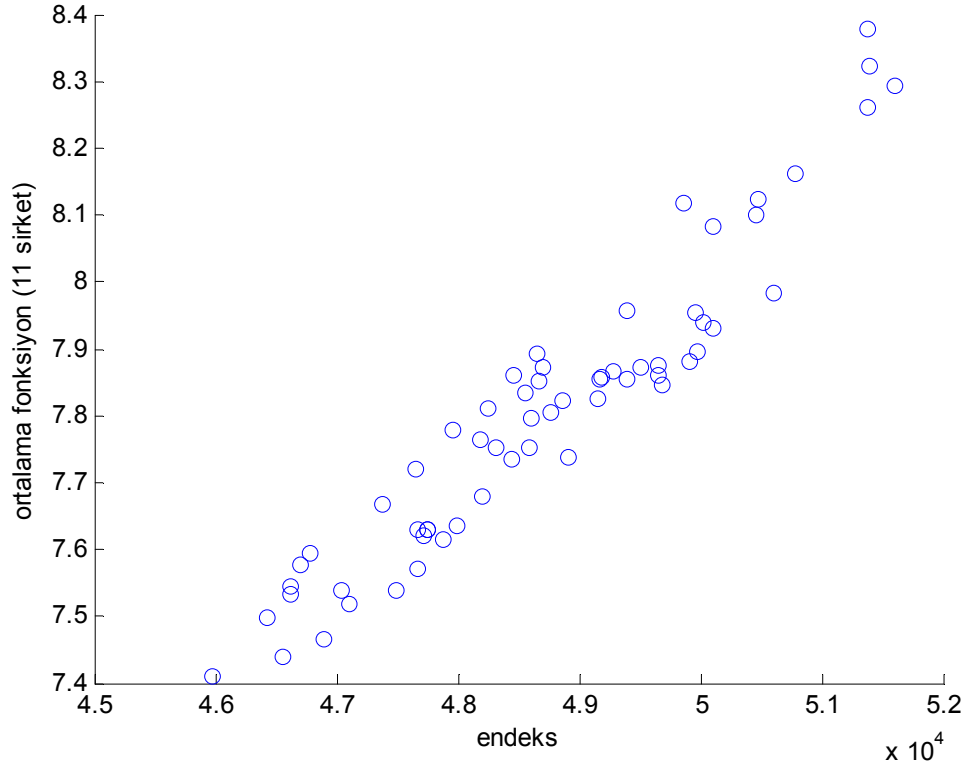
yansıtmayabilir. Bu durum Şekil (3.14)de verilen ana bileşen fonksiyonunun incelenmesi ile ortaya çıkarılacaktır.



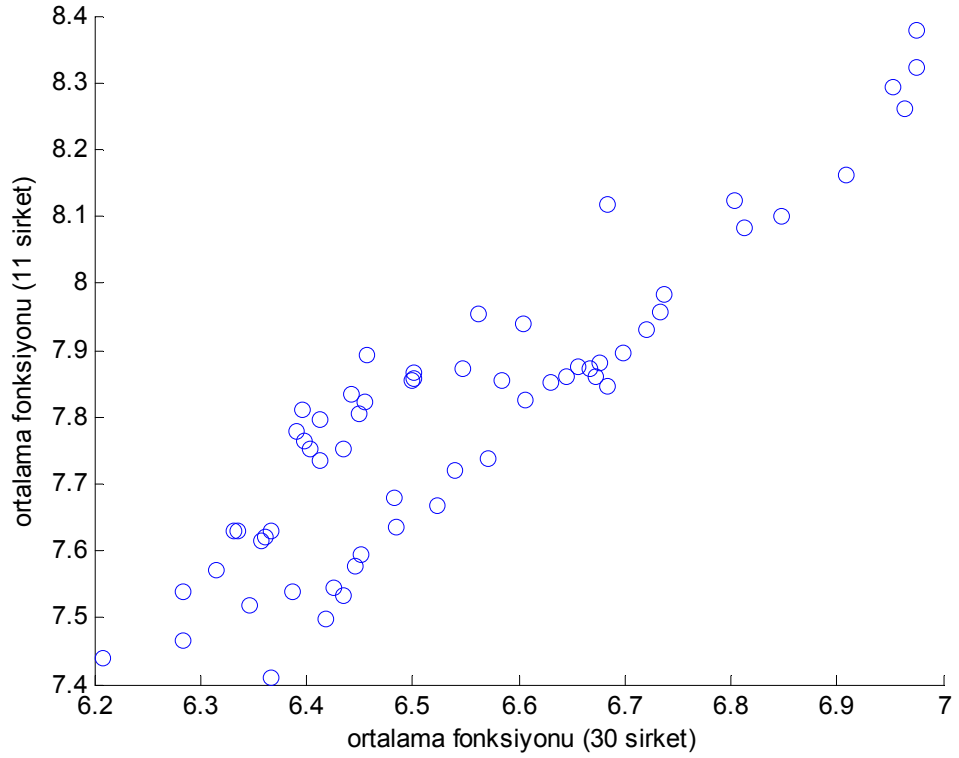
Şekil(3.14): Seçilen 11 Şirketin Ana Bileşen Fonksiyonu

Şekil (3.14)de görüldüğü gibi varyasyon yapıları 30 şirketin birlikte incelenmesinden farklılık göstermektedir ve bu nedenle de ağırlıklarına göre belirlenen 11 şirketin 30 şirketin temsilcisi olarak kullanılması tavsiye edilemez. Şekil (3.14)de verilmiş olan 11 şirketin ana bileşen fonksiyonu Şekil (3.9)da verilmiş olan 30 şirketin ana bileşen fonksiyonu ile karşılaştırıldığında yine Aralık ayında özellikle sonlarına doğru varyasyonda bir artış olduğu ancak Şekil (3.14)de Kasım ayında da varyasyonda Şekil (3.9)a oranla daha belirgin bir artış olduğu görülmektedir.

30 şirketin ortalama fonksiyonu ile 11 şirketin ortalama fonksiyonu ve İMKB 30 endeksi ile 11 şirketin ortalama fonksiyonu arasındaki korelasyonlar sırasıyla Şekil (3.15) ve Şekil (3.16)da verilmektedir.

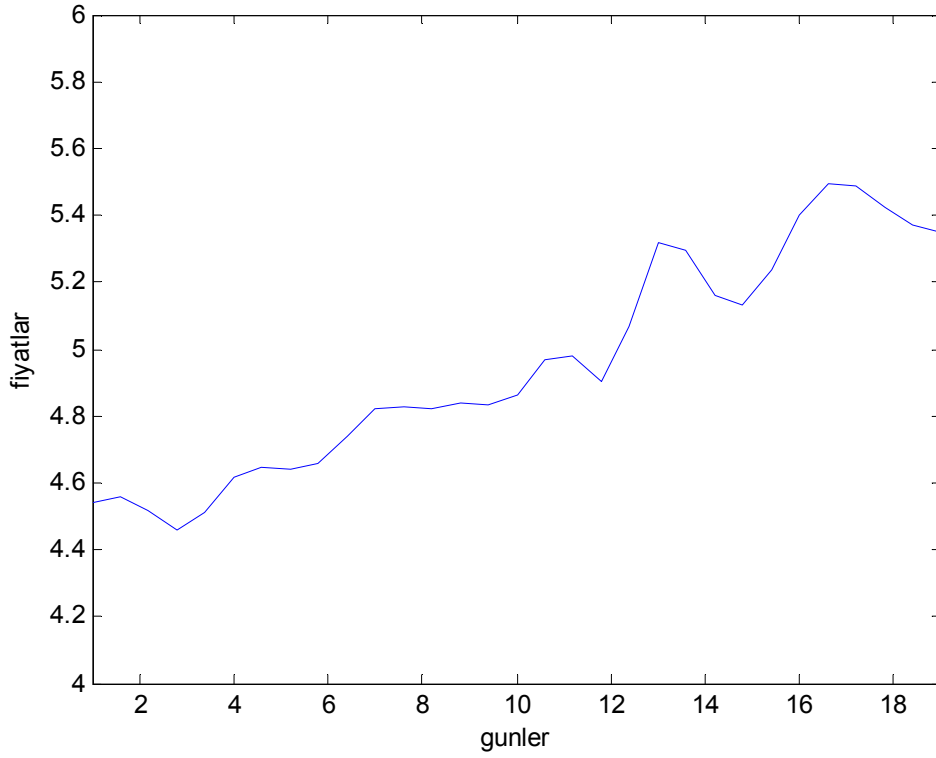


Şekil(3.15): Seçilen 11 Şirket ve İMKB 30 Arasındaki Korelasyon

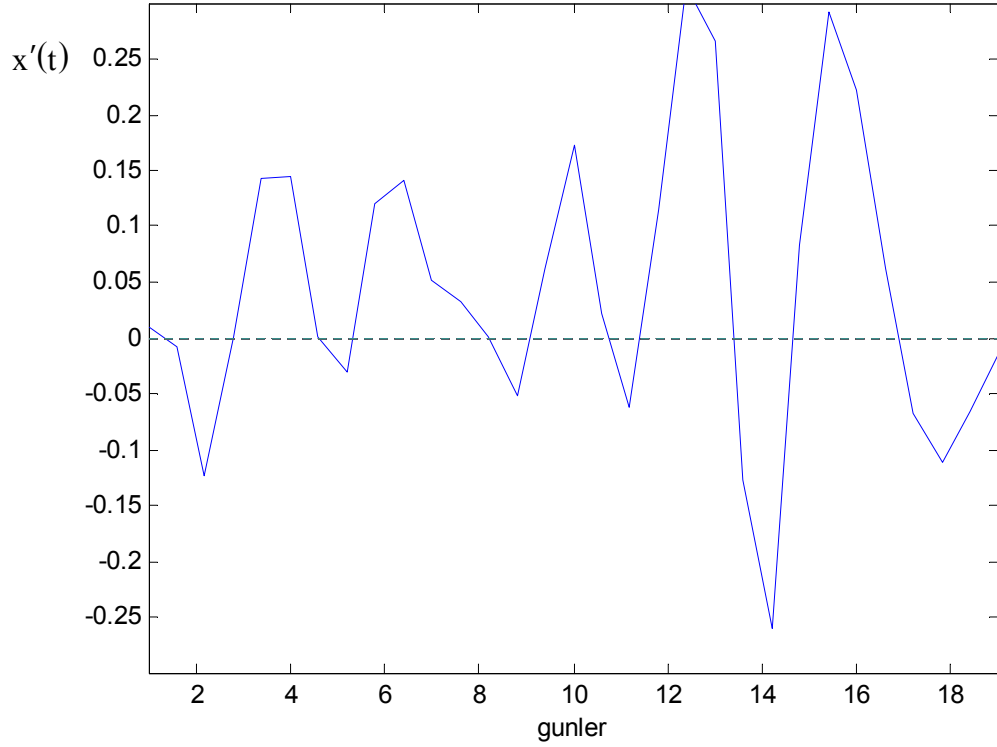


Şekil(3.16): Seçilen 11 Şirket ve 30 Şirket Arasındaki Korelasyon

11 tane şirketin bireysel fonksiyonundan farklı sektörlerden olmak üzere 2 tanesi seçilerek bu fonksiyonlar ve bunların türev fonksiyonları aylar bazında incelenerek, türev fonksiyonlarının yoruma olan katkıları da gösterilmeye çalışılmıştır. İlk olarak Garanti Bankası ele alınmıştır.

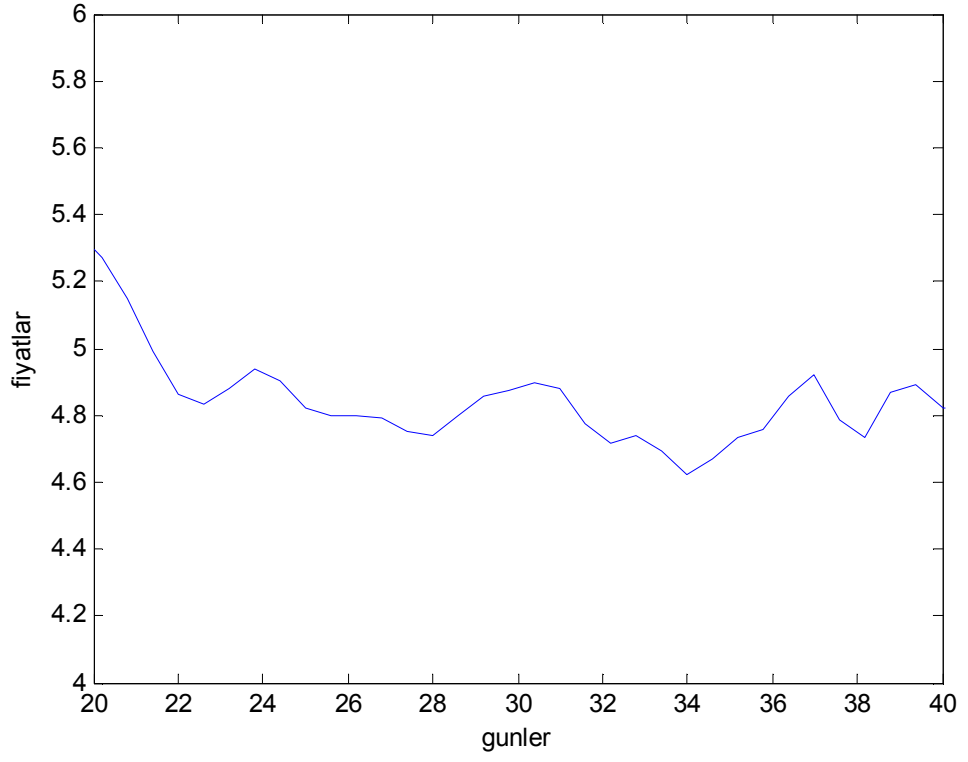


Şekil(3.17): Garanti Bankası Ekim Ayı Grafiği

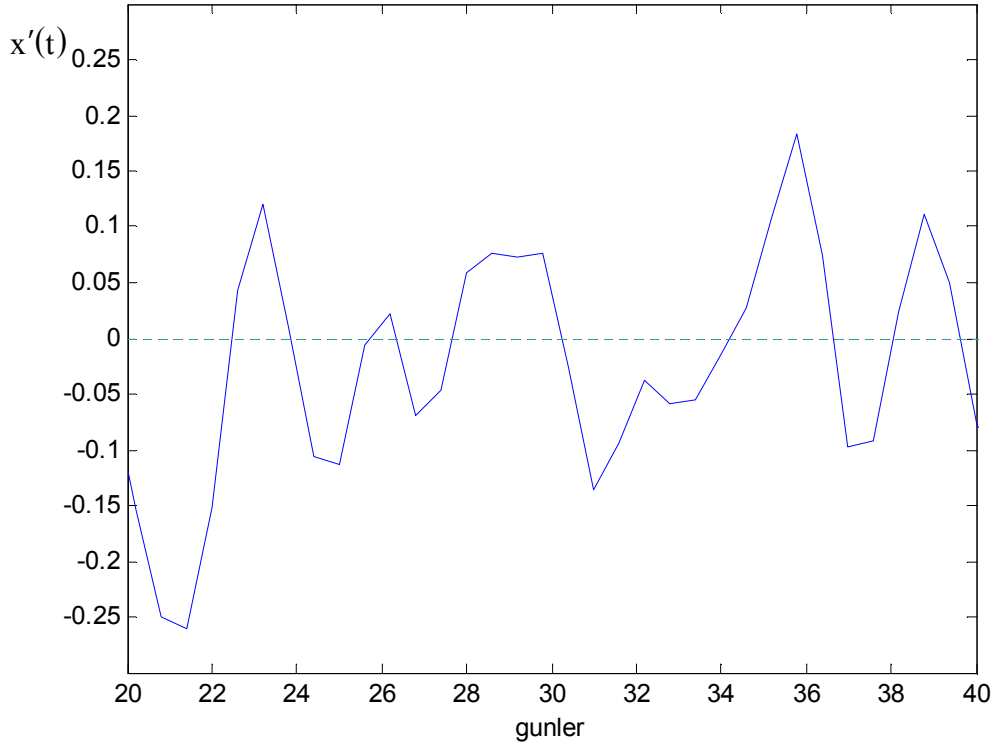


Şekil(3.18): Ekim Ayı Birinci Türev Fonksiyonu

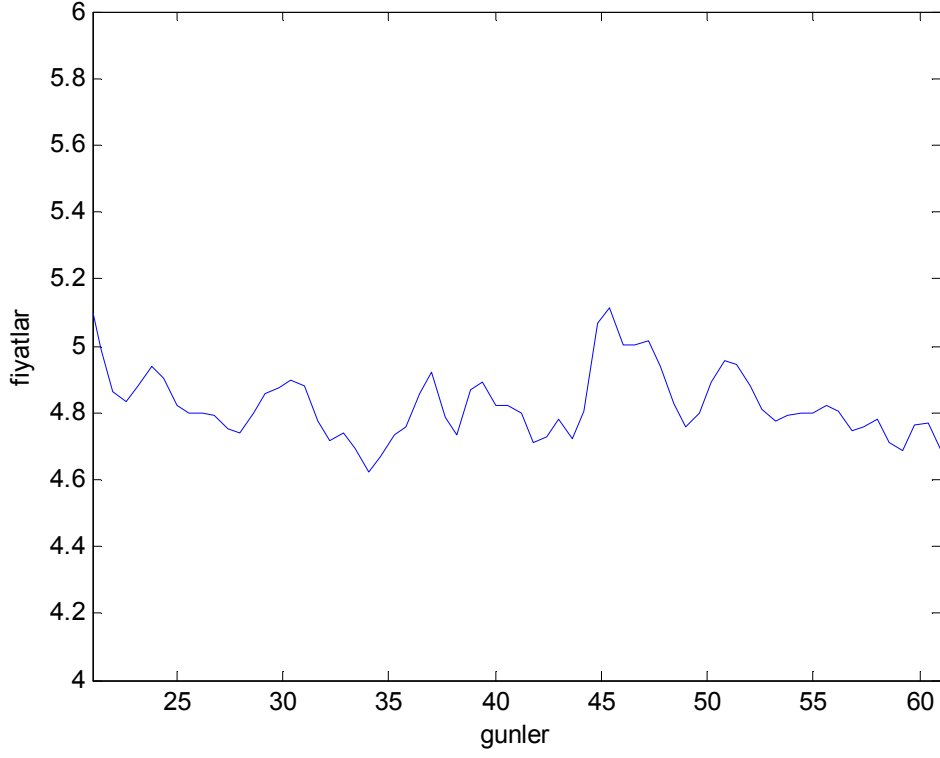
Şirketin fiyatının günlere göre ani değişim oranı (türev) fonksiyonu incelendiğinde fiyatlardaki pozitif ya da negatif değişimin hızı daha rahat bir şekilde görülebilmektedir. Şekil (3.17) ve Şekil (3.18) karşılaştırıldığında Şekil (3.17)de 12.ci güne kadar güçlükle seçilebilen, çok net görülemeyen artış ve azalışlar türev fonksiyonun incelenmesiyle daha rahatlıkla yakalanabilmektedir. Kasım ve Aralık aylarının ve türev fonksiyonlarının grafikleri benzer şekilde Şekil (3.19) ve Şekil (3.22) arasında verilmektedir. Benzer yorumlamalar bu şekiller için de geçerlidir.



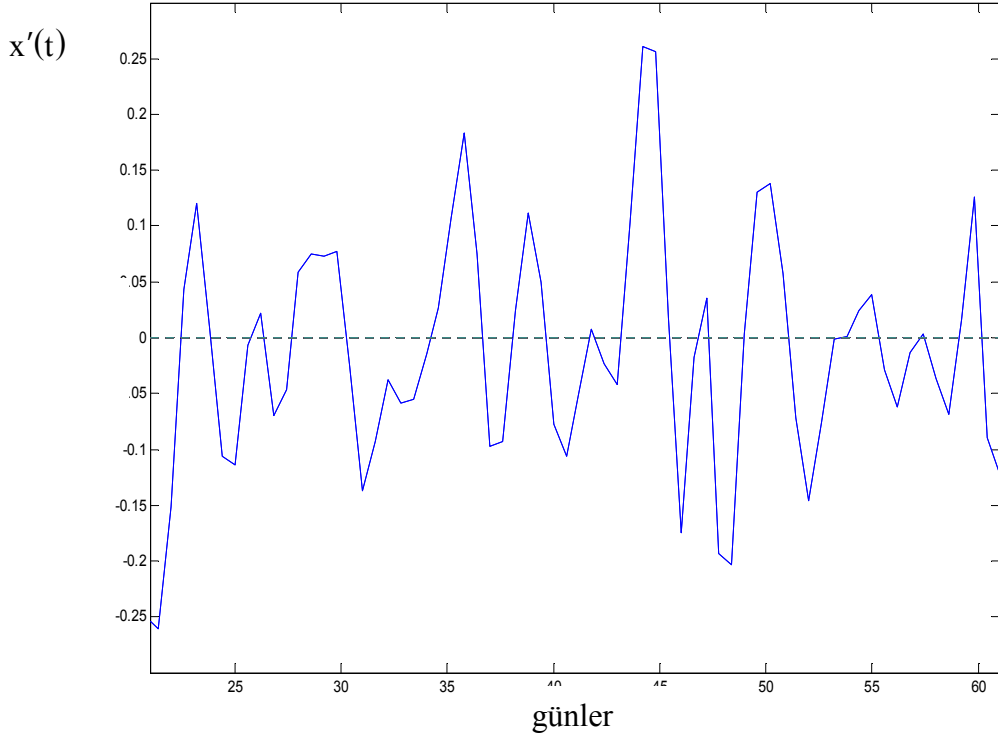
Şekil (3.19): Garanti Bankası Kasım Ayı Grafiği



Şekil(3.20) : Kasım Ayı Birinci Türev Fonksiyonu

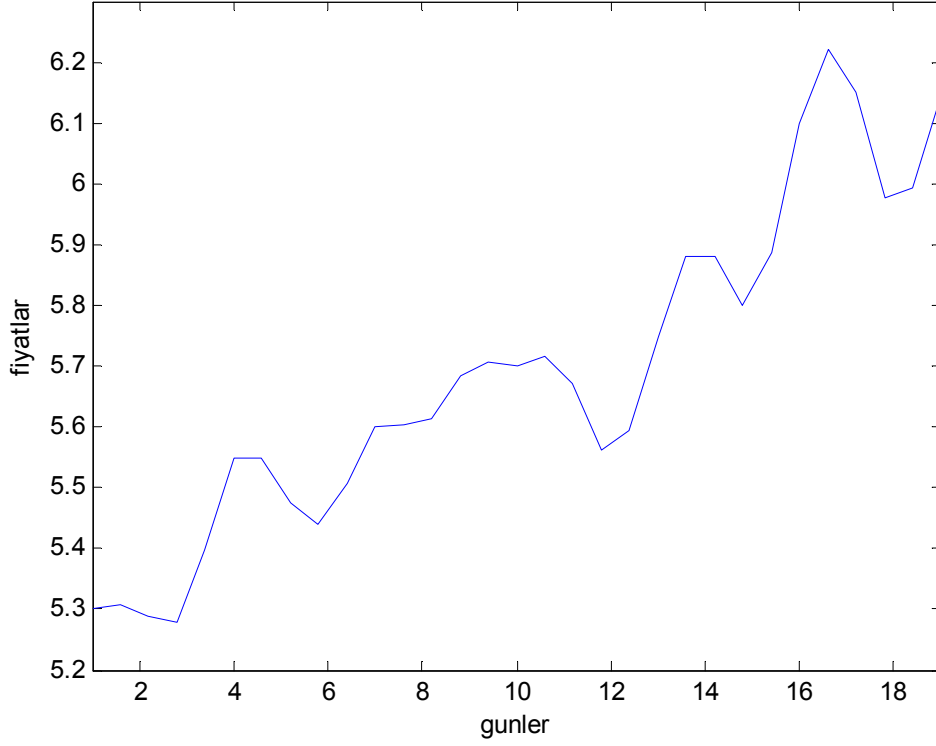


Şekil (3.21): Garanti Bankası Aralık Ayı Grafiği



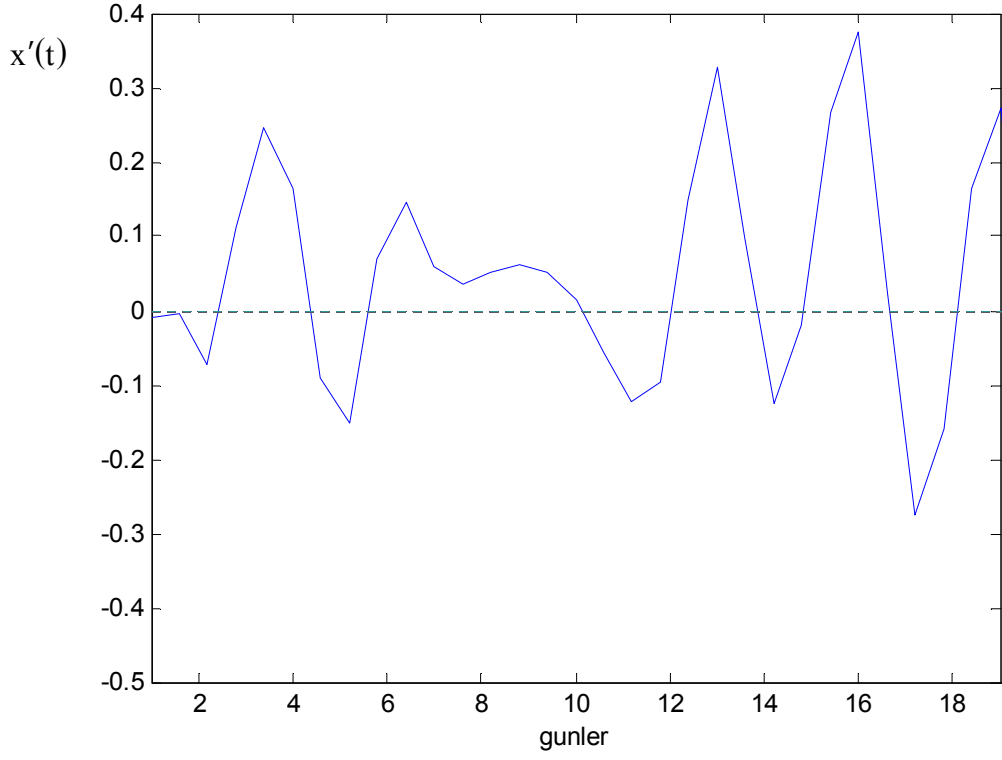
Şekil (3.22): Aralık Ayı Birinci Türev Fonksiyonu

İkinci olarak holdinglerden Sabancı Holding ele alınmıştır. Sabancı Holding'in fonksiyonundaki iniş ve çıkışlar Garanti Bankasına oranla daha belirgindir.

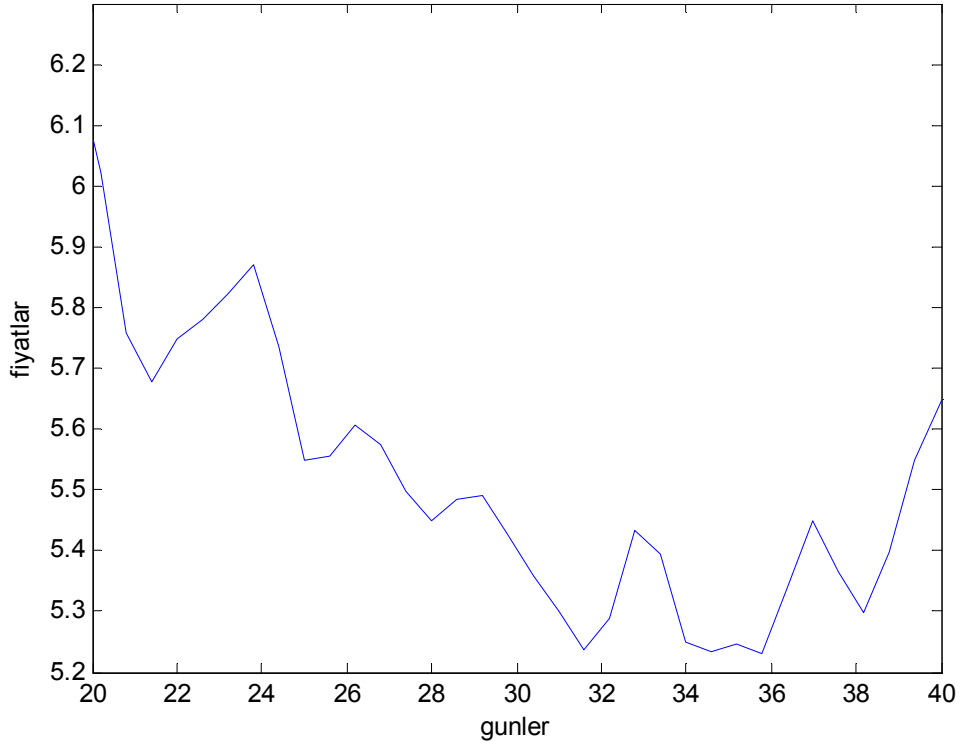


Şekil (3.23) : Sabancı Holding Ekim Ayı Grafiği

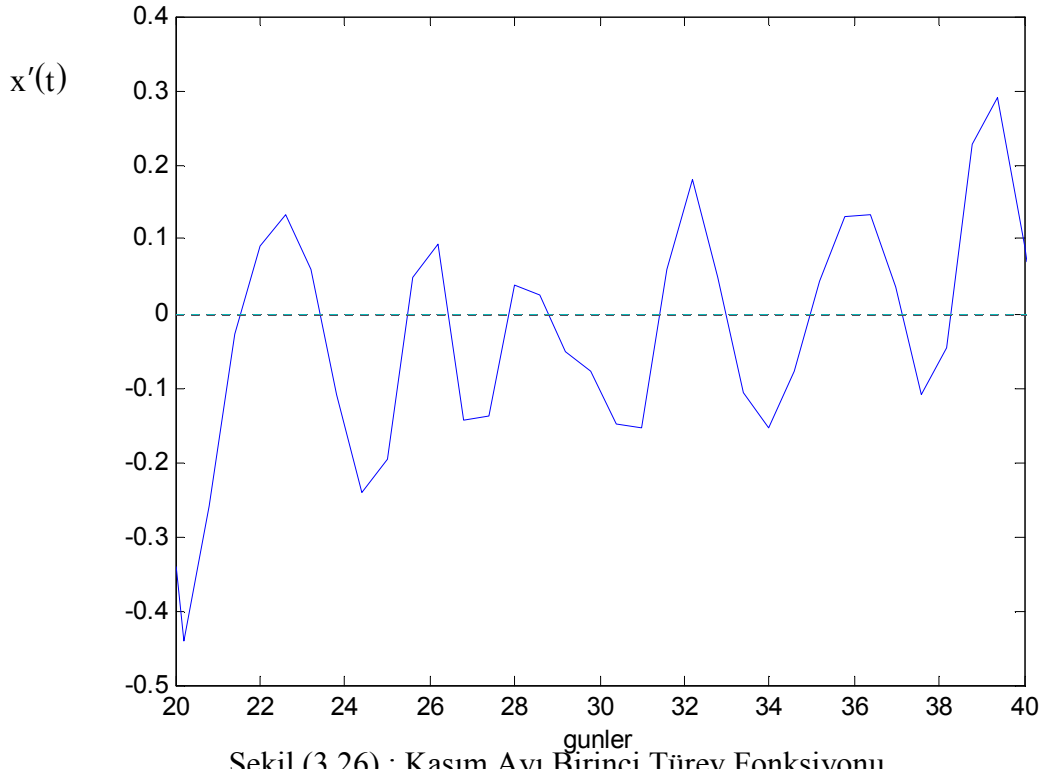
Şekil (3.23) incelendiğinde iniş ve çıkışlar Garanti Bankasına oranla daha net görülebilmeye karşın ani iniş ve çıkışların gücü türev fonksiyonunda daha rahatlıkla yakalanabilmektedir. Örneğin 17. gün civarındaki iniş, fonksiyondan da görülebilmeye karşın 15. gün civarındaki ani iniş türev fonksiyonunda daha net görülebilmekte ve bu iki inişin boyutu daha rahatlıkla kıyaslanabilmektedir.



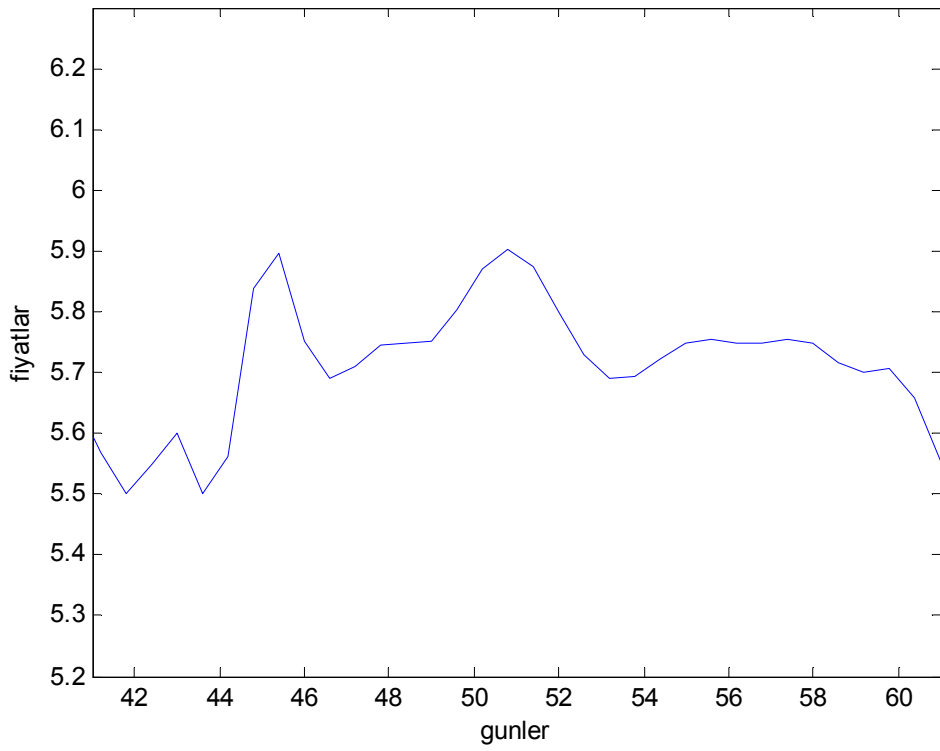
Şekil (3.24) : Ekim Ayı Birinci Türev Fonksiyonu



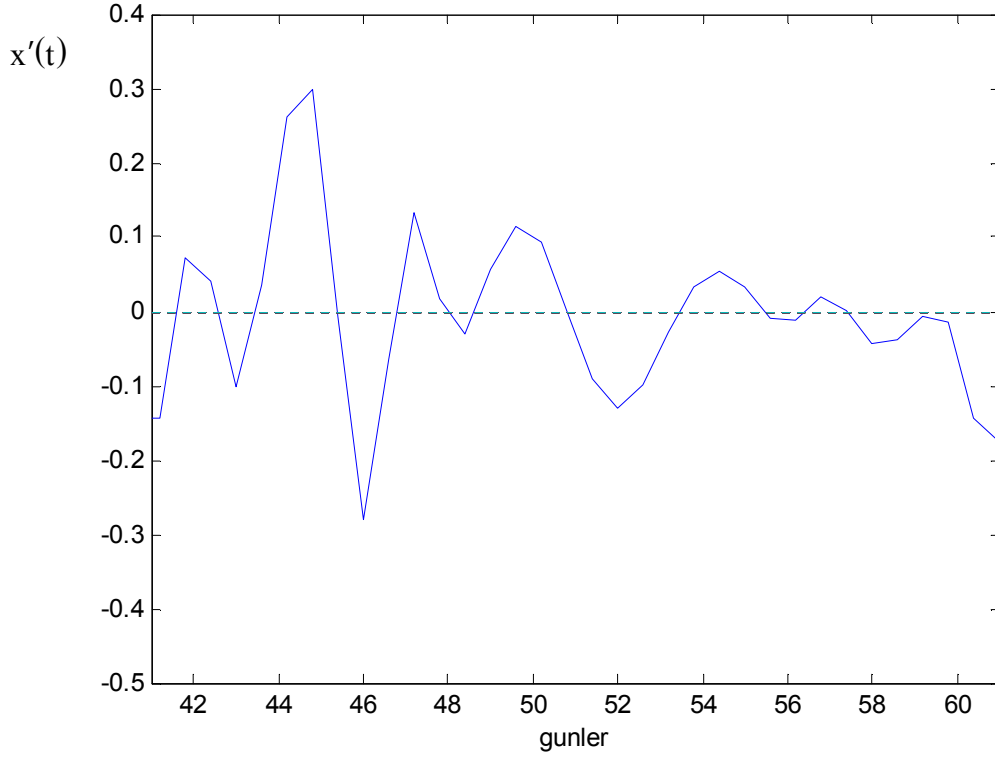
Şekil (3.25) : Sabancı Holding Kasım Ayı Grafiği



Şekil (3.26) : Kasım Ayı Birinci Türev Fonksiyonu



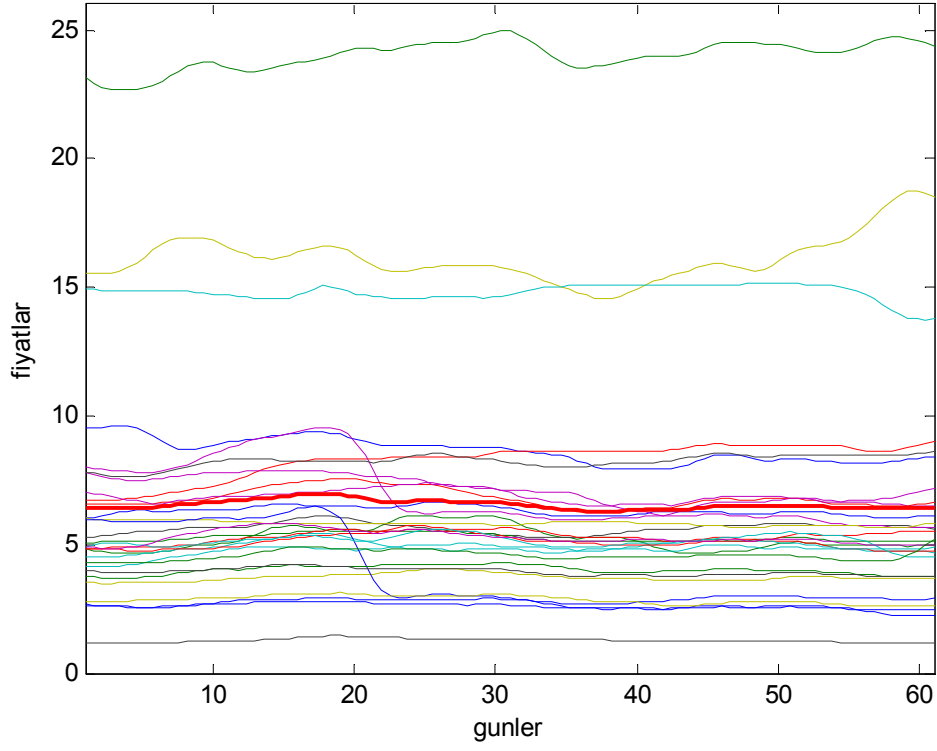
Şekil (3.27) : Sabancı Holding Aralık Ayı Grafiği



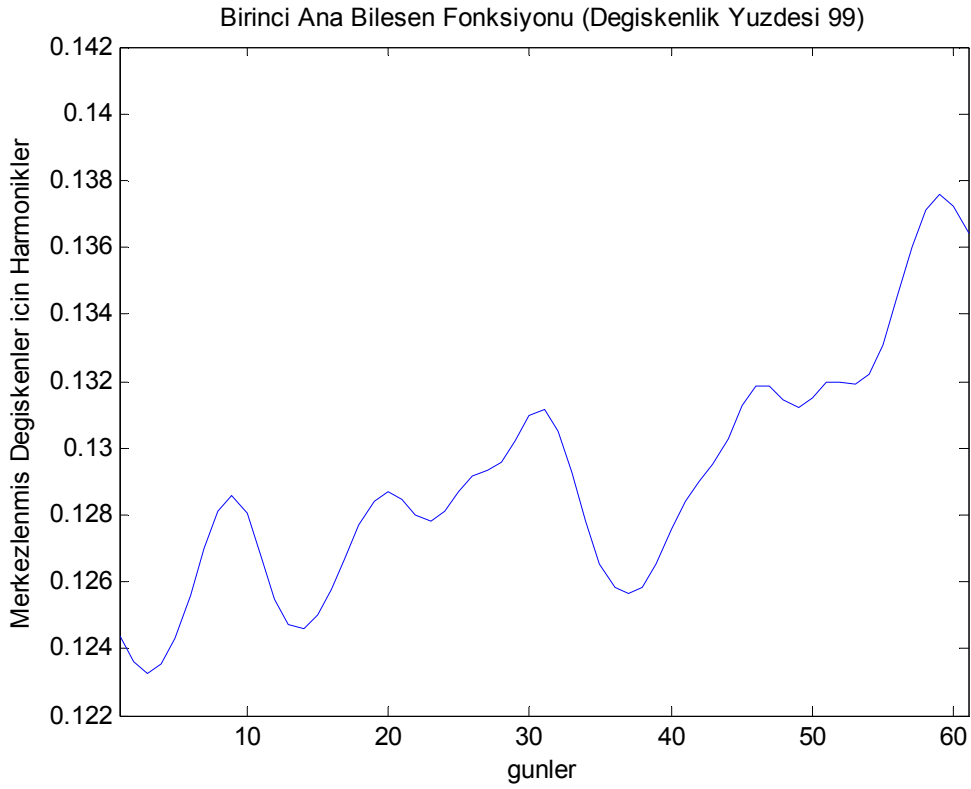
Şekil (3.28) : Aralık Ayı Birinci Türev Fonksiyonu

3.2.3. Düzgünleştirme Parametresinin Etkileri

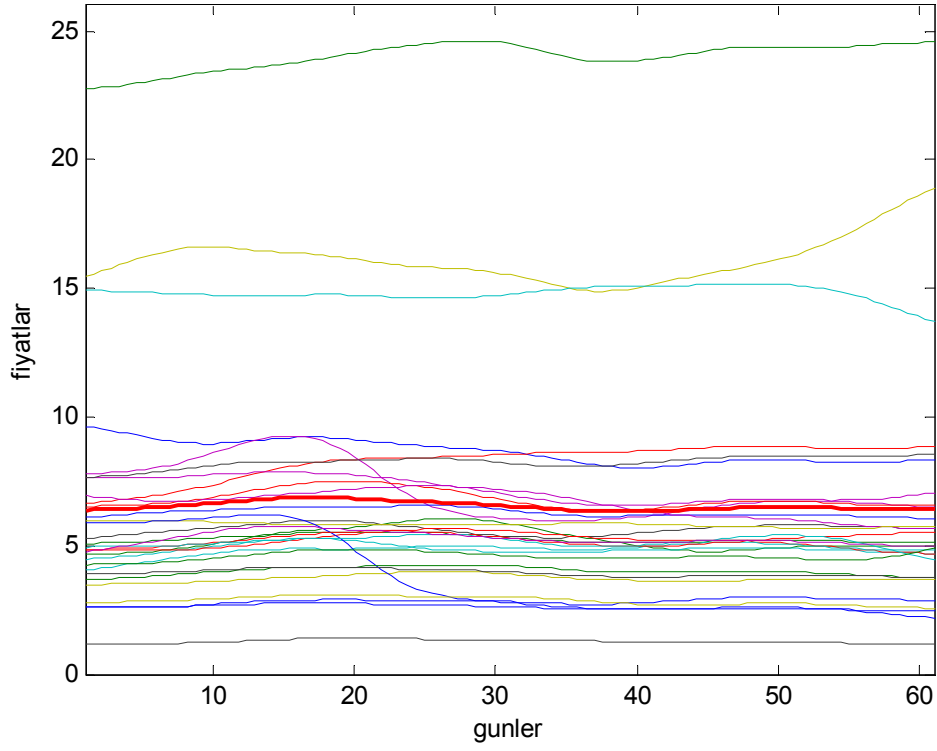
Bu kesimde ele alınacak konu Fonksiyonel Veri Analizinde önemli bir yere sahip olan düzgünleştirme parametresinin etkisinin incelenmesidir. Şimdiye kadar tüm analizler Ulbrict'in (2004) yaklaşımına göre yapılmıştır. Düzgünleştirme parametresinin etkisinin görülebilmesi amacı ile her bir birey için fonksiyonların, ortalama fonksiyonların ve birinci ana bileşen ağırlık fonksiyonlarının çizimleri iki farklı düzgünleştirme parametresi değeri için tekrarlanacaktır. Bu çalışmada sırasıyla $\lambda=1$ ve $\lambda= 50$ değerleri ele alınmıştır.



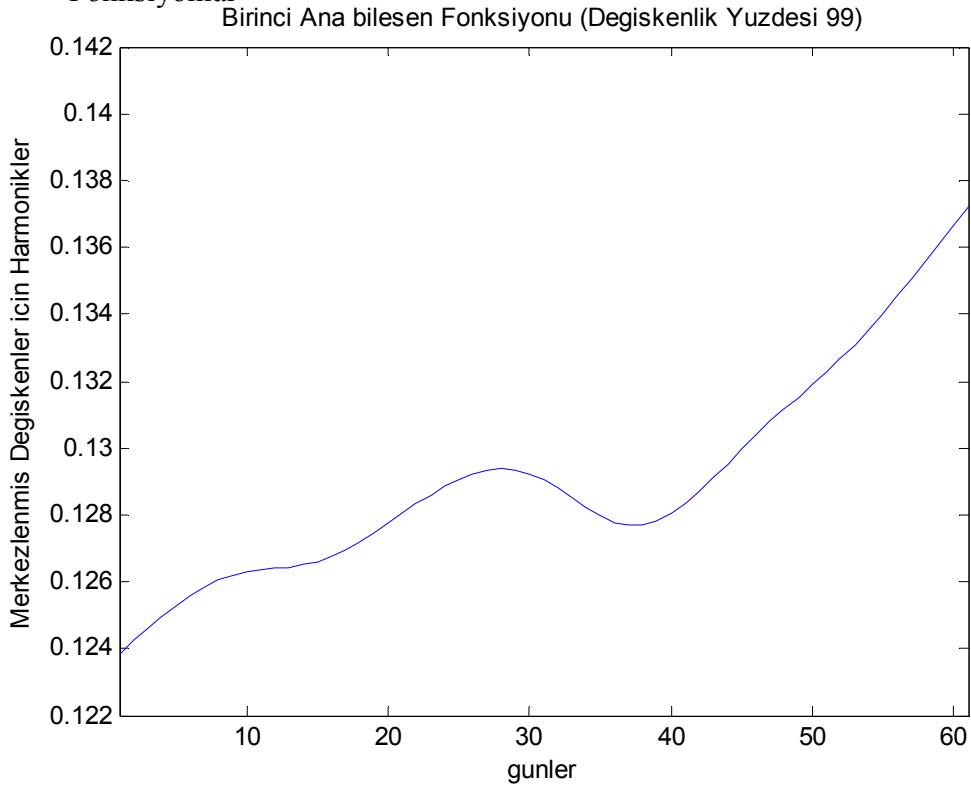
Şekil (3.29) : $\lambda=1$ için Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar



Şekil (3.30) : $\lambda=1$ için Birinci Ana Bileşen Fonksiyonu



Şekil (3.31): $\lambda=50$ için Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar



Şekil(3.32): $\lambda=50$ için Birinci Ana Bileşen Fonksiyonu

Yukarıda subjektif belirlenen çeşitli düzgünleştirme parametreleri için ilgili fonksiyonlar incelendiğinde ana bileşenlerin açıkladığı varyans oranlarında veya fonksiyonların genel seyrinde bir değişiklik olmadığı ancak düzgünleştirme parametresinin yüksek belirlenmesinin, bir diğer deyişle verilerin çok fazla düzgünleştirilmesinin, verilerde bilgi kaybına yol açabileceği açıkça görülmektedir. $\lambda=50$ ve hatta $\lambda=1$ değerleri için bile artık fonksiyonlarda önceden görülen bazı küçük, hatta $\lambda=50$ için oldukça önemli sayılabilecek, iniş çıkışlar görülememekte, kimi durumlarda ciddi bazı bilgi kayıpları yaşanabilmektedir. Buradan da önceden belirtildiği gibi, düzgünleştirme parametresinin belirlenmesinde, Ulbricht'in (2004) yaklaşımı veya Ramsay ve Li'nin (1998) çalışmasında önerdiği düzgünleştirme parametresi değerlerinin görsel olarak uygun olduğu bir kez daha görülmektedir.

SONUÇ

Günümüzde hızla ilerleyen teknolojik gelişmeler sayesinde artık çoğu durumda klasik çok değişkenli istatistiksel yöntemler elde edilen verileri analiz etmek için yetersiz kalmaya başlamıştır. 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) dan 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 orada olması beklenen veya beklenmeyen 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 düzensiz örneklenen fonksiyonlarla ve kayıp verilerle uğraşılmasına imkan vermesi, interpolasyon 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 aralarındaki temel fark fonksiyonlar, ana bileşen fonksiyonları, ortalama fonksiyonu, kovaryans yüzeyleri ve fonksiyonların türevlerinin görsel olarak da incelenebilmesi ve bir çok görünmeyi ortaya çıkarmadaki yeteneğidir.

Bu çalışmada uygulama olarak İMKB 30 kapanış fiyatları ele alınmıştır. Ortalama fonksiyonunun İMKB 30 endeksine bir alternatif olabileceği gösterilmiş ve buna ilave olarak türev fonksiyonu da incelenmiştir. Türev fonksiyonunun incelenmesi ile ortalama fonksiyonunda gözlenmesi oldukça güç olan çok küçük iniş çıkışların bile rahatlıkla yakalanabildiği, türevlerin negatif ve pozitif olduğu bölgeler ile türev fonksiyonunun sıfır değerini aldığı maksimum, minimum noktalar 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. İMKB 30 daki tüm şirketler dikkate alındığında birinci ana bileşen fonksiyonundan özellikle 55.inci günlerden sonra fiyatların değişkenliğinde bir artış olduğu ve zaman noktalarının ardışık olarak birbirleriyle pozitif bir korelasyonda bulunduğu gözlemlenebilmektedir. Ancak İMKB 30 da bulunan ağırlıklarına göre ilk 11 şirket seçildiğinde bu 11 şirketin ortalama fonksiyonu yine İMKB 30 endeksine çok benzer olsa da 30 şirketin temsilcisi olarak kullanılamaz, çünkü ortalama fonksiyonunun değişim aralığı farklıdır, ki burada aynı değişken açısından inceleme yapılmaktadır ve 11 şirketin ana bileşen fonksiyonunda 30 şirketin ana bileşen fonksiyonuna göre Kasım ayında değişkenlikte bir artış olduğu gözlemlenmektedir.

Bu çalışmada tanıtılan Fonksiyonel Veri Analizi gelişime çok açık bir alandır. Özellikle tıp, çevre verileri, eğitim sektörü ve finans sektöründe daha bir çok uygulama yapılabilir. Örneğin tıp verilerinde biyolojik yaşın bireyden bireye değişmesinden dolayı eğri kaydırılması (curve registration) tıp alanındaki uygulamalara çok büyük bir katkı sağlayabilir. Şirketlerin satışları incelenebilir ve

bu satıřlara gre alıcılar gruplandırılabilir. Yine Fonksiyonel Ana Bileřenler Analizi ile đrenci bařarıları inceleyebilir ve đrenciler đrenme gruplarına ayrılabilir. Bu alıřmada B-Splayn baz fonksiyon yaklařımı kullanılmıřtır ancak diđer baz fonksiyon yaklařımları da ayrı bir inceleme konusu olarak ele alınabilir. Fonksiyonel Ana Bileřenler Analizi dıřında yukarıda belirtilen diđer Fonksiyonel Veri Analizi Yntemleri de geliřmekte olup bunlar da ayrı birer alıřma alanı olarak lisansst dzeyinde incelenebilir. Fonksiyonel Ana Bileřenler Analizi, Dzgnleřtirilmiř Fonksiyonel Ana Bileřenler Analizi ve genel anlamda Fonksiyonel Veri Analizinin uluslar arası literatrle eř zamanlı alıřmalar yapılması iin ok msait bir alan olduđu grlmřtir.

KAYNAKLAR

Anderson T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience: USA

Ando T., Imoto S. & Miyano S.. (2004). Functional Data Analysis of The Dynamics of Gene Regulatory Networks, International Symposium, *Springer- Verlag*

Anselone P.M. & Laurent P.J. (1968). A General Method for the Construction of Interpolating or Smoothing Spline-Functions, *Numerische Mathematik*, 12.

Aydın, Dursun.(2005). *Semiparametrik Regresyon Modellemede Splayn Düzeltme Yaklaşımı İle Tahmin ve Çıkarımlar*. Yayınlanmamış Doktora Tezi, Eskişehir Anadolu Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir.

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,DiscussionPaper*,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, *Psychometrika*, 51(2)

Besse P. C., Cardot H., Faivre R., & Goulard M. (2005). Statistical Modelling of Functional Data, *Applied Stochastic Models In Business And Industry*, 21.

Boor C . (1978). *A Practical Guide to Splines*. Springer-Verlag: New-York.

Buchanan J.L. & Turner P.R. (1992). *Numerical Methods and Analysis*. Mc-Graw: NewYork

Burden R.L ve Faires J.D . (1997). *Numerical Analysis*. Brooks/Cole:USA

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.

Craven P. & Wahba G. (1979). Smoothing Noisy Data with Spline Functions. *Numerische Mathematik*. 31.

Dabos M., & Bugallo F. (2000). *Term Structure of Interest Rates Changes During International Financial Crisis: The Case of Argentina*. Eriřim:22.05.2005, http://www.aaep.org.ar/espa/anales/resumen_00/dabos_bugallo.htm

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.

Eiler P.H.C., Marx B.D.(1996). Flexible Smoothing with B-splines and Penalties, *Statistical Science*, 11(2).

Eubank R.L. (1985). Diagnostic for Smoothing Splines. *Journal of the Royal Statistical Society: Series B.*, 47(2).

Eubank R.L. (1999). *Nonparametric Regression and Spline Smoothing*. Marcel Dekker: USA.

Faraway J.J. (2000). *Modeling reach motions using functional regression analysis*, Digital Human Modelling for Design and Engineering Conference. Erişim: 20.02.2006, <http://www.maths.bath.ac.uk/~jjf23/papers/mrmfra.pdf>

Gasser T., Müller H.G, Köhler W., Molinari L., Prader A. (1984) Nonparametric Regression Analysis of Growth Curves. *The Annals of Statistics*, 12(1).

Gasser T., & Kneip A. (1995). Searching for Structure in Curve Sample, *Journal of American Statistical Association*, 90(432)

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).

Hardle W. (1997). *Applied Nonparametric Regression*, Cambridge University Press:USA

He G. (1999). *Functional Canonical Analysis and Linear Modelling*, University of California Doctor Of Philosophy in Statistics with Emphasis in Biostatistics .

Hurvich C.M, Simonoff J.S & Tsai C-L. (1997). Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion. *Journal of the Royal Statistical Society: Series B*. 60(2).

Hutchinson M.F, & de Hoog F.R. (1985). Smoothing Noisy Data with Spline Functions. *Numerische Mathematik*, 47.

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).

Johnson R.A., & Wichern D.W. (1998). *Applied Multivariate Statistical Analysis*. Prentice – Hall: New Jersey.

Joliffe I.T. (2002). *Principal Component Analysis*. Springer – Verlag: New York.

Jones M. C., Rice J. A. (1992). Displaying The Important Features Of Large Collections Of Similar Curves. *The American Statistician*, 46(2).

Joseph Z.B, Gerber G., & Gifford D.K. (2003). Continuous Representation of Time Series Gene Expression Data. *Journal of Computational Biology*, 10(3-4).

Kneip A., & Gasser T. (1992). Statistical Tools to Analyze Data Representing a Sample of Curves. *The Annals of Statistics*, 20(3).

Kreyszig E. (1989). *Introductory Functional Analysis With Applications*, John Wiley & Sons: New York.

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.

Leng X., & Müller G.H. (2006). Classification using functional data analysis for temporal gene expression data. *Bioinformatica*, 22(1).

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

Luan Y., & Li H. (2003). Clustering of Time-Course Gene Expression Data Using A Mixed – Effects Model With B-Splines. *Bioinformatics*, 19(4).

Lyche, T. & Morken, K., (2002). *Spline Methods Draft*, Erişim: 10.07.2005, <http://www.ifi.uio.no/in329/nchap1.pdf>

Mardia K.V., Kent J.T., & Bibby J.M. (1989). *Multivariate Analysis*. Academic Press: London

Musayev B., & Alp M. (2000). *Fonksiyonel Analiz*. Balcı Yayınları: Kütahya

Müller, H. G. (2005). Functional Modelling and Classification of Longitudinal Data, *Scandinavian Journal of Statistics*, 32(2).

Nürnberg G. (1989). *Approximation by Spline Functions*. Springer – Verlag: Berlin.

Ramsay J.O. (1982). When The Data Are Functions. *Psychometrika*, 47(4).

Ramsay, J. O., & Dalzell C. (1991). Some Tools For Functional 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. (1998). Estimating smooth monotone functions. *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. (2001). *Functional data analysis*. In N. J. Smelser and P.B. Baltes (Eds) ,International Encyclopedia of the Social and Behavioral Sciences

Ramsay J. O, & Ramsey J. B. (2002). Functional Data Analysis of the Dynamics of the Monthly Index of Nondurable Goods Production. *Journal of Econometrics*, 107(1-2).

Ramsay J.O, & Silverman B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer – Verlag: New York.

Ramsay J.O., (2003). *Jim Ramsay*, Erişim: 11.04.2005, <http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html>

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).

Raz J., Turesty B., & Fein G. (1989). Selecting the Smoothing Parameter for Estimation of Slowly Changing Evoked Potential Signals. *Biometrics*, 45.

Reinsch C.H. (1967). Smoothing By Spline Functions. *Numerische Mathematik*, 10.

Reinsch C.H. (1971). Smoothing By Spline Functions II. *Numerische Mathematik*, 16.

Rice J.A., & Silverman B.W. (1991). Estimating The Mean and Covariance Structure Nonparametrically When The Data Are Curves. *Journal of the Royal Statistical Society: Series B.*, 53(1).

Rossi F., Delanney N., Conan-Guez B., & Verleysen M. (2005). Representation of Functional Data In Neural Networks. *Neurocomputing*, 64.

Schumaker L.L. (1993). *Spline Functions: Basic Theory*. Krieger Publishing Company: Florida

Schwarz H.R., & Waldvogel J. (1989). *Numerical Analysis*. John Wiley & Sons: New York

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.

Timini A.A. (2003). *Comparison of Smoothers*, CSI 739.

Ulbricht J.(2004). *Representing Functional Data as Smooth Functions*. A Master Thesis, Humboldt University Institute of Statistics and Econometrics, Berlin.

Walker E. (2001), *Determining The Smoothing Parameter in Penalized Regression Splines Using Model Selection Criteria*, Erişim:11.04.2005, <http://bus.utk.edu/Stat/walker/Osaka/OverheadOsaka.pdf>

Wei H.W. (2005). The Smoothing Parameter, Confidence Interval and Robustness for Smoothing Splines. *Nonparametric Statistics*, 00(0).

Wittaker E.T. (1923). On a New Method of Graduation, *Proc. Edinburg Mathematics Society*, 41.

Yamanishi Y., (2004). *Statistical Case Studies: Biostatistics and Geostatistics*, Erişim: 01.07.2005, <http://www.quantlet.com/mdstat/scripts/xcs/pdf/xcspdf.pdf>

Yamanishi Y., & Tanaka Y. (2005). Sensitivity Analysis in Functional Principal Component Analysis, *Computational Statistics*, 20(2).

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.

EK1

```
function basisobj = create_bspline_basis(rangeval, nbasis, norder, ...
                                         breaks, dropind, quadvals, values)

if length(rangeval) == 1
    if rangeval <= 0
        error('RANGEVAL a single value that is not positive.');
```

```
    end
    rangeval = [0,rangeval];
end

if rangechk(rangeval) ~= 1
    error('RANGEVAL is not a legitimate range.');
```

```
end

% set some default values

% number of basis functions
if nargin < 2
    nbasis = 1;
end
% order of splines
if nargin < 3
    norder = min(4,nbasis);
end
% knots
if nargin < 4
    breaks = [];
else
    if size(breaks,1) > 1, breaks = breaks'; end
    if size(breaks,1) > 1
        error('BREAKS must be a vector.');
```

```
    end
end
% indices of basis functions to be dropped
if nargin < 5
    dropind = [];
end
% quadrature points and weights
if nargin < 6
    quadvals = [];
end
% basis derivative values at quadrature points
if nargin < 7
    values{1} = [];
end
end
```

```

type = 'bspline';

% Determine what to do if some arguments are empty

% If both NBASIS and BREAKS are missing, but NORDER IS NOT,
% use 21 equally spaced knots, and determine NBASIS.
% by NBASIS = NORDER + 19.
if isempty(nbasis) & isempty(breaks)
    nbreaks = 21;
    nbasis = 19 + norder;
    breaks = linspace(rangeval(1), rangeval(2), nbreaks);
end
% If NBASIS is empty but BREAKS are supplied, determine
% NBASIS = NORDER + number of interior knots.
if isempty(nbasis) & ~isempty(breaks)
    nbreaks = length(breaks);
    nbasis = nbreaks + norder - 2;
end
% If NBASIS and NORDER are present but no BREAKS supplied,
% set up NBASIS - NORDER + 2 equally spaced breaks.
if ~isempty(nbasis) & isempty(breaks)
    nbreaks = nbasis - norder + 2;
    breaks = linspace(rangeval(1), rangeval(2), nbreaks);
end
nbreaks = length(breaks);

% check the compatibility of NBASIS, NBREAKS and RANGEVAL

if (nbreaks < 2)
    error('Number of values in BREAKS less than 2.');
```

```

end
if (nbasis < nbreaks-1)
    error('NBASIS is less than number of values=BREAKS.');
```

```

end
if (breaks(1) ~= rangeval(1))
    error('Smallest value in BREAKS not equal to RANGEVAL(1).');
```

```

end
if (breaks(nbreaks) ~= rangeval(2))
    error('Largest value in BREAKS not equal to RANGEVAL(2).');
```

```

end

% The PARAMS field contains only the interior knots

if nbreaks > 2
    params = breaks(2:(nbreaks-1));
else
    params = [];

```

```

end

% construct basis object

basisobj = basis(type, rangeval, nbasis, params, ...
                dropind, quadvals, values);

function [fdobj, coef, SSE, penmat, y2cMap, basismat, yhat, basisobj, fdnames] = ...
    smooth_basis_ist(argvals, y, fdParobj, wtvec, dffactor, fdnames)

if nargin < 3
    error('There is not at least three arguments. ');
end

% check ARGVALS

if ~strcmp(class(argvals), 'double')
    error('ARGVALS is not of class double. ');
end

if size(argvals,1) == 1
    argvals = argvals';
end

[n, ncl] = size(argvals); % number of observations
if ncl > 1
    error('ARGVALS is not a vector. ')
end
if n < 2
    error('ARGVALS does not contain at least two values. ');
end

% check Y

```

```

if ~strcmp(class(y), 'double')
    error('Y is not of class double.');
```

end

```

ydim = size(y);
if length(ydim) == 2 & ydim(1) == 1
    y = y';
end
```

```

ydim = size(y); % number of observations
if ydim(1) ~= n
    error('Y is not the same length as ARGVALS.');
```

end

```

% set default argument values
```

```

if nargin < 6
    fdnames{1} = 'arguments';
    fdnames{2} = 'replications';
    fdnames{3} = 'variables';
end
```

```

if nargin < 5, dffactor = 1;    end
```

```

if nargin < 4, wtvec = ones(n,1); end
```

```

% check fdParobj
```

```

if ~isa_fdPar(fdParobj)
    if isa_fd(fdParobj) | isa_basis(fdParobj)
        fdParobj = fdPar(fdParobj);
    end
end
```

```

else
    error(['FDPAROBJ is not a functional parameter object, ', ...
        'not a functional data object, and ', ...
        'not a basis object.']);
end
end
end

% check LFD

Lfdobj = getLfd(fdParobj);
Lfdobj = int2Lfd(Lfdobj);
nderiv = getnderiv(Lfdobj);

% check BASIS

fdobj = getfd(fdParobj);
basisobj = getbasis(fdobj);
if ~isa_basis(basisobj)
    error('BASIS is not a basis object.');
```

```

end

nbasis = getnbasis(basisobj);
onebasis = ones(1,nbasis);

% check WTVEC

sizew = size(wtvec);
if (length(sizew) > 1 & sizew(1) > 1 & sizew(2) > 1) | ...
    length(sizew) > 2
    error ('WTVEC must be a vector.');
```

```

end
if length(sizew) == 2 & sizew(1) == 1
```

```

    wtvec = wtvec';
end
if length(wtvec) ~= n
    error('WTVEC of wrong length');
end
if min(wtvec) <= 0
    error('All values of WTVEC must be positive.');
```

end

% check LAMBDA

```

lambda = getlambda(fdParobj);
if lambda < 0
    warning('Value of LAMBDA was negative, and 0 used instead.');
```

lambda = 0;

end

% set number of curves and number of variables

```

sizey = size(y);
ndim = length(sizey);
switch ndim
    case 1
        ncurves = 1;
        nvar = 1;
    case 2
        ncurves = sizey(2);
        nvar = 1;
    case 3
        ncurves = sizey(2);
        nvar = sizey(3);
    otherwise
```

```

        error('Second argument must not have more than 3 dimensions');
    end

    basismat = eval_basis(argvals, basisobj);

    if n >= nbasis | lambda > 0

        % The following code is for the coefficients completely determined

        basisw = basismat .* (wtvec * ones(1,nbasis));
        Bmat = basisw' * basismat;
        Bmat0 = Bmat;

        if ndim < 3
            Dmat = basisw' * y;
        else
            Dmat = zeros(nbasis,ncurves,nvar);
            for ivar = 1:nvar
                Dmat(:, :, ivar) = basisw' * y(:, :, ivar);
            end
        end
    end

    if lambda > 0
        % it was thought better of using forcing functions in
        % the definition of a linear differential operator object.
        % smoothing required, set up coefficient matrix for normal equations
        % afdcell = getafdcell(Lfdobj); % multiplier(s) of forcing function(s)
        % ufdcell = getufdcell(Lfdobj); % forcing function(s)
        % if ~isempty(afdcell) & ~isempty(ufdcell)
        % % the linear differential operator is not homogeneous.
        % % first set up the homogeneous counterpart
        % wtfldcell = getwfdcell(Lfdobj);
    end

```

```

%      Lfdhom = Lfd(nderiv, wfdcell);
%      % evaluate the penalty matrix for the homogeneous operator
%      penmat = eval_penalty(basisobj, Lfdhom);
%      % set up the part of the roughness penalty affected by the
%      % presence of forcing function(s)
%      penvec = zeros(nbasis,1);
%      nforce = length(ufdcell);
%      for k=1:nforce
%          afdk = getfd(afdcell{k});
%          ufdk = ufdcell{k};
%          ubasis = getbasis(ufdk);
%          if strcmp(getbasistype(ubasis), 'bspline') & ...
%              getnbasis(ubasis) - length(getbasispar(ubasis)) == 1
%              range = getbasisrange(ubasis);
%              breaks = [range(1), getbasispar(ubasis), range(2)];
%              nbreaks = length(breaks);
%              for ibreak=2:nbreaks
%                  uval = ...
%                      eval_fd((breaks(ibreak)+breaks(ibreak-1))/2, ufdk);
%                  if uval ~= 0
%                      rangei = [breaks(ibreak-1),breaks(ibreak)];
%                      auval = inprod(basisobj, uval.*afdk, Lfdhom, int2Lfd(0),
rangei);
%                      penvec = penvec - auval;
%                  end
%              end
%          else
%              ffdk = times(afdk, ufdk, basisobj);
%              penvec = penvec - inprod(basisobj, ffdk, Lfdhom, int2Lfd(0));
%          end
%      end
%  else

```

```

        % here the linear differential operator is homogeneous
        % only the penalty matrix is needed.
        penmat = eval_penalty(basisobj, Lfdobj);
        penvec = zeros(nbasis,1);
%     end
        Bnorm = sqrt(sum(sum(Bmat.^2)));
        pennorm = sqrt(sum(sum(penmat.^2)));
        condno = pennorm/Bnorm;
        if lambda*condno > 1e12
            lambda = 1e12/condno;
            warning(['lambda reduced to ',num2str(lambda), ...
                ' to prevent overflow']);
        end
        Bmat = Bmat + lambda .* penmat;
        if ~all(penvec == 0)
            % if the linear differential operator is nonhomogeneous
            % use PENVEC to alter the right side of the equation.
            Dmat = Dmat + lambda.*(penvec*ones(1,ncurves));
        end

else

        penmat = zeros(nbasis);

end

% compute inverse of Bmat

if is_diag(Bmat)
    Bmatinv = diag(1./diag(Bmat));
else
    Bmatinv = inv(Bmat);

```

```

end

% compute map from y to c

y2cMap = Bmatinv * basisw';

% compute degrees of freedom of smooth

df = full(sum(diag(Bmatinv * Bmat0)));

% solve normal equations for each observation

if ndim < 3
    coef = Bmatinv * Dmat;
else
    coef = zeros(nbasis, ncurves, nvar);
    for ivar = 1:nvar
        coef(:, :, ivar) = Bmatinv * Dmat(:, :, ivar);
    end
end

end

else
    error(['The number of basis functions exceeds the number of ', ...
        'points to be smoothed.']);

% % The following code is for the underdetermined coefficients:
% % the number of basis functions exceeds the number of argument values.
% % No smoothing is used.
%
% [Qmat,Rmat] = qr(basismat');
% Q1mat = Qmat(:,1:n);
% Q2mat = Qmat(:,((n+1):nbasis));

```

```

% Hmat = eval_penalty(basisobj);
% Q2tHmat = Q2mat' * Hmat;
% Q2tHQ2mat = Q2tHmat * Q2mat;
% Q2tHQ1mat = Q2tHmat * Q1mat;
% if ndim < 3
%     z1mat = symsolve(Rmat, y);
%     z2mat = symsolve(Q2tHQ2mat, Q2tHQ1mat*z1mat);
%     coef = Q1mat * z1mat + Q2mat * z2mat;
% else
%     for ivar = 1:nvar
%         z1mat = symsolve(Rmat, y(:, :, ivar));
%         z2mat = symsolve(Q2tHQ2mat, Q2tHQ1mat*z1mat);
%         coef(:, :, ivar) = Q1mat * z1mat + Q2mat * z2mat;
%     end
% end
% y2cMap = eye(n);
% df = n;
end

% compute error sum of squares

if ndim < 3
    yhat = basismat * coef;
    SSE = sum((y - yhat).^2);
else
    SSE = zeros(nvar, ncurves);
    for ivar = 1:nvar
        coefi = squeeze(coef(:, :, ivar));
        yhati = basismat * coefi;
        yi = squeeze(y(:, :, ivar));
        SSE(ivar, :) = sum((yi - yhati).^2);
    end
end

% compute GCV index

if df < n
    gcv = (SSE./n)/((n - dffactor*df)/n)^2;
else
    gcv = NaN;
end

fdobj = fd(coef, basisobj, fdnames);

```

```

function [harmcoef,harmscr, eigvecs, Lmat, Jmat, Cmat,pcastr, covC] =
istem_pca(fdobj, nharm, lambda, Lfdobj, centerfns)

if ~isa_fd(fdobj)
    error('First argument is not a functional data object.');
```

end

% set up default values

```

if nargin < 5
    centerfns = 1;
end
if nargin < 4
    Lfdobj = int2Lfd(2);
end
if nargin < 3
    lambda = 0;
end
if nargin < 2
    nharm = 2;
end

% check LFDOBJ

Lfdobj = int2Lfd(Lfdobj);

% compute mean function and center if required

meanfd = mean(fdobj);
if centerfns ~= 0
    fdobj = center(fdobj);
end

coef = getcoef(fdobj);
coefd = size(coef);
nbasis = coefd(1);
nrep = coefd(2);
ndim = length(coefd);

if nrep < 2
    error('PCA not possible without replications');
end

fdbasis = getbasis(fdobj);
type = getbasistype(fdbasis);
rangeval = getbasisrange(fdbasis);

if ndim == 3
```

```

nvar = coefd(3);
ctemp = zeros(nvar*nbasis,nrep);
for j = 1:nvar
    index = (1:nbasis) + (j-1)*nbasis;
    ctemp(index,:) = coef(:,j);
end
else
    nvar = 1;
    ctemp = coef;
end

% set up cross product and penalty matrices

Cmat = ctemp*ctemp'./nrep
aaaa='ilk cmat'
covC=Cmat;
Jmat = eval_penalty(fdbasis, int2Lfd(0));
if lambda > 0
    Kmat = eval_penalty(fdbasis, Lfdobj);
    Wmat = Jmat + lambda .* Kmat;
else
    Wmat = Jmat;
end

% compute the Choleski factor of Wmat

Lmat = chol(Wmat);
Lmatinv = inv(Lmat);

% set up matrix for eigenanalysis

if nvar == 1
    if lambda > 0
        Cmat = Lmatinv' * Jmat * Cmat * Jmat * Lmatinv;
    else
        Cmat = Lmat * Cmat * Lmat'
        aaaa='yeni cmat'
    end
else
    for i = 1:nvar
        indexi = (1:nbasis) + (i-1)*nbasis;
        for j = 1:nvar
            indexj = (1:nbasis) + (j-1)*nbasis;
            if lambda > 0
                Cmat(indexi,indexj) = ...
                    Lmatinv' * Jmat * Cmat(indexi,indexj) * Jmat * Lmatinv;
            else
                Cmat(indexi,indexj) = Lmat * Cmat(indexi,indexj) * Lmat';
            end
        end
    end
end

```

```

        end
    end
end
end

% Eigenanalysis

Cmat = (Cmat + Cmat')./2
aa='son cmat'
[eigvecs, eigvals] = eig(Cmat);
[eigvals, indsrt ] = sort(diag(eigvals));
eigvecs = eigvecs(:,indsrt);
neig    = nvar*nbasis;
indx    = neig + 1 - (1:nharm);
eigvals = eigvals(neig + 1 - (1:neig));
eigvecs = eigvecs(:,indx);
sumvecs = sum(eigvecs);
eigvecs(:,sumvecs < 0) = -eigvecs(:,sumvecs < 0);

varprop = eigvals(1:nharm)./sum(eigvals)

if nvar == 1
    harmcoef = Lmatinv * eigvecs;
    harmscr  = ctemp * Lmat' * eigvecs;

else
    harmcoef = zeros(nbasis,nharm,nvar)
    harmscr  = zeros(nrep,nharm);
    for j = 1:nvar
        index = (1:nbasis) + (j-1)*nbasis;
        temp  = eigvecs(index,:);
        harmcoef(:,j) = Lmatinv * temp;
        harmscr = harmscr + ctemp(index,:) * Lmat' * temp;
    end
end

harmnames = getnames(fdobj);
harmnames{2} = 'Harmonics';
harmnames{3} = ['Harmonics for',harmnames{3}];

harmfd = fd(harmcoef, fdbasis, harmnames);

pcastr.harmfd = harmfd;
pcastr.eigvals = eigvals;
pcastr.harmscr = harmscr;
pcastr.varprop = varprop;
pcastr.meanfd = meanfd;

```

```

function istem_plotpca_mean(pcastr, nx, pointplot, harm, expand, cycle)

if nargin < 6
    cycle = 0;
end
if nargin < 5
    expand = 0;
end
if nargin < 4
    harm = 0;
end
if nargin < 3
    pointplot = 1;
end
if nargin < 2
    nx = 101;
end

harmfd = pcastr.harmfd;
basis = getbasis(harmfd);
rangex = getbasisrange(basis);
fdnames = getnames(harmfd);
x = linspace(rangex(1), rangex(2), nx)
fdmat = eval_fd(harmfd, x)
meanmat = squeeze(eval_fd(pcastr.meanfd, x))
dimfd = size(fdmat);
nharm = dimfd(2);
if harm == 0
    harm = (1:nharm);
end
if length(dimfd) == 2
    % plotting for univariate functions
    for iharm = harm
        if expand == 0
            fac = sqrt(pcastr.eigvals(iharm));
        else
            fac = expand;
        end
        vecharm = fdmat(:,iharm);
        percentvar = round(100 * pcastr.varprop(iharm));

        %ortalama fonksiyonun çizimi

        plot(x, meanmat, '-');
        xlabel(fdnames{1});
        ylabel(fdnames{3});
        title(['ortalama fonksiyon '])
    end
end

```

```

pause;

% çarpan eklenmiş- çıkarılmış halinin çizimi

meanplus = meanmat+fac.*vecharm;
meanminus = meanmat-fac.*vecharm;
plottop = max([meanplus;meanminus]);
plotbot = min([meanplus;meanminus]);
plot(x, meanmat, '-')
text(x, meanplus, '+')
text(x, meanminus, '-')
xlabel(fdnames{1});
ylabel(fdnames{3});
axis([rangex(1), rangex(2), plotbot, plottop])
pluss(iharm,:)=meanplus;
minuss(iharm,:)=meanminus;
vecharmm(iharm,:)=vecharm;
face(iharm,:)=fac;

title(['PCA function ', num2str(iharm), ...
      ' (Percentage of variability ', num2str(percentvar), ')'])
disp('Press any key to continue')
pause;
end
pluss
minuss
vecharmm
face
end

```

Tarih		ARCLK	DYHOL	EREGL	GARAN	ISCTR	MIGRS	SAHOL	THYAO	TUPRS	VAKBN
02.10.2006	1	9,5	4,84	6,75	4,54	8,0000	15,6000	5,3000	6,0500	23,2000	6,6000
03.10.2006	2	9,6	4,8	6,7	4,54	7,9500	15,5000	5,3000	5,9500	22,9000	6,6000
04.10.2006	3	9,45	4,62	6,7	4,46	7,8000	15,7000	5,3000	6,4000	22,3000	6,4500
05.10.2006	4	9,65	4,68	6,9	4,62	7,9500	15,5000	5,5500	6,3000	23,0000	6,6500
06.10.2006	5	9,8	4,7	6,7	4,64	7,8000	15,9000	5,5000	6,3000	22,7000	6,6000
09.10.2006	6	9,15	4,68	6,95	4,68	7,8000	16,7000	5,4500	6,3000	22,7000	6,5500
10.10.2006	7	8,6	4,74	7,15	4,82	7,8500	17,1000	5,6000	6,2500	23,2000	6,7000
11.10.2006	8	8,55	4,88	7	4,82	8,1000	16,7000	5,6000	6,3500	23,5000	6,6500
12.10.2006	9	8,7	5,06	7,2	4,84	8,3500	17,0000	5,7000	6,4000	24,0000	6,8000
13.10.2006	10	9	5	7,25	4,86	8,5000	16,8000	5,7000	6,3000	23,8000	6,7000
16.10.2006	11	9	5,1	7,4	5	8,8000	16,8000	5,7000	6,3000	23,6000	6,9500
17.10.2006	12	9	5	7,55	4,92	8,7500	16,1000	5,5500	6,3500	23,1000	6,8500
18.10.2006	13	9,05	5,35	7,8	5,32	9,3000	16,3000	5,7500	6,4500	23,4000	7,5000
19.10.2006	14	9,3	5,8	8,05	5,2	9,1000	16,1000	5,9000	6,6500	23,5000	7,2500
20.10.2006	15	9,25	5,7	8,15	5,15	9,1500	15,9000	5,8000	6,6000	23,7000	7,2500
26.10.2006	16	9,3	5,75	8,25	5,4	9,4500	16,5000	6,1000	6,5000	23,9000	7,5500
27.10.2006	17	9,35	5,65	8,3	5,5	9,6500	16,6000	6,2000	6,5000	23,8000	7,5000
30.10.2006	18	9,45	5,5	8,3	5,4	9,2500	16,4000	5,9500	6,5000	23,6000	7,3500
31.10.2006	19	9,35	5,6	8,4	5,35	9,4500	16,8000	6,1500	6,5500	24,2000	7,6500
01.11.2006	20	9,1	5,5	8,45	5,3	9,4000	16,5000	6,1000	6,5500	24,7000	7,8000
02.11.2006	21	8,9	5,4	8,1	5,1	9,1000	15,8000	5,7000	6,4000	24,1000	7,4000
03.11.2006	22	8,8	5,55	8,35	4,86	6,3000	15,4000	5,7500	6,4000	24,2000	7,3000
06.11.2006	23	8,9	6,25	8,5	4,86	6,2500	15,7000	5,8000	6,4500	24,1000	7,3000
07.11.2006	24	8,85	6,3	8,4	4,94	6,5000	15,7000	5,8500	6,5000	24,5000	7,3500
08.11.2006	25	8,8	6	8,45	4,82	6,2500	15,8000	5,5500	6,6000	24,4000	7,4000
09.11.2006	26	8,85	6	8,4	4,8	6,3000	15,8000	5,6000	6,8000	24,8000	7,3500
10.11.2006	27	8,85	6,1	8,45	4,78	6,3000	16,0000	5,5500	6,6500	24,5000	7,2000
13.11.2006	28	8,7	5,95	8,25	4,74	6,1500	15,7000	5,4500	6,4500	24,2000	6,9500
14.11.2006	29	8,8	6,05	8,5	4,84	6,2500	16,0000	5,5000	6,5500	24,8000	6,8500
15.11.2006	30	8,8	6,35	8,6	4,88	6,3500	15,8000	5,4000	6,5000	25,0000	6,9000
16.11.2006	31	8,75	6	8,65	4,88	6,2500	15,8000	5,3000	6,4500	25,2500	6,8000
20.11.2006	32	8,6	5,7	8,65	4,72	6,1000	15,5000	5,2500	6,3000	24,6000	6,6000
21.11.2006	33	8,45	5,5	8,65	4,74	6,0500	15,4000	5,4500	6,2000	24,4000	6,6000
22.11.2006	34	8,5	5,3	8,55	4,62	5,9000	15,0000	5,2500	6,0500	23,9000	6,3500
23.11.2006	35	8,6	5,2	8,65	4,72	6,0000	15,1000	5,2500	6,1500	23,6000	6,4000
24.11.2006	36	8,15	5,25	8,7	4,78	5,9000	14,7000	5,2500	6,1000	23,4000	6,2000
27.11.2006	37	8,05	5,35	8,65	4,92	6,1500	14,7000	5,4500	6,1500	23,8000	6,3500
28.11.2006	38	7,8	5,15	8,6	4,72	5,8500	14,3000	5,3000	6,1000	23,5000	6,2000
29.11.2006	39	7,95	5,4	8,6	4,9	6,1500	14,6000	5,4500	6,0500	23,7000	6,3000
30.11.2006	40	8,05	5,1	8,6	4,82	6,0500	15,2000	5,6500	6,1500	24,0000	6,4500
01.12.2006	41	7,95	5,05	8,65	4,82	6,1500	15,2000	5,6000	6,3500	24,1000	6,5000
04.12.2006	42	7,85	4,66	8,65	4,7	6,0500	15,4000	5,5000	6,3000	23,9000	6,3500
05.12.2006	43	8	4,66	8,65	4,78	6,3000	15,5000	5,6000	6,2500	24,0000	6,4500
06.12.2006	44	8,1	4,66	8,9	4,74	6,2500	15,5000	5,5000	6,1000	23,8000	6,5000
07.12.2006	45	8,45	4,76	9	5,12	6,8500	16,1000	5,9000	6,3500	24,5000	6,9500
08.12.2006	46	8,65	4,74	8,85	5	6,6500	16,0000	5,7500	6,2500	24,6000	6,7500
11.12.2006	47	8,55	4,74	8,85	5,02	6,7000	15,8000	5,7000	6,2000	24,6000	6,7500
12.12.2006	48	8,25	4,8	8,85	4,9	6,6000	15,7000	5,7500	6,1500	24,5000	6,7000
13.12.2006	49	8,05	4,76	8,85	4,76	6,5500	15,3000	5,7500	6,1500	24,3000	6,7500
14.12.2006	50	8,4	5	8,85	4,86	6,7000	15,8000	5,8500	6,2000	24,5000	6,8000
15.12.2006	51	8,4	5,25	8,8	4,96	7,0000	16,9000	5,9000	6,3000	24,5000	6,8500
18.12.2006	52	8,4	5,1	8,85	4,88	6,8500	16,5000	5,8000	6,4000	24,2000	6,8500
19.12.2006	53	8,15	5	8,85	4,78	6,6500	16,6000	5,7000	6,2000	24,1000	6,6000
20.12.2006	54	8,2	4,9	8,9	4,8	6,6500	16,7000	5,7000	6,2000	24,0000	6,5500
21.12.2006	55	8,2	4,9	8,9	4,8	6,7000	16,6000	5,7500	6,1000	24,3000	6,5500
22.12.2006	56	8,2	5,06	8,7	4,82	6,7000	17,2000	5,7500	6,1000	24,3000	6,6500
25.12.2006	57	8,15	4,94	8,45	4,74	6,6000	18,1000	5,7500	6,0500	24,6000	6,5000
26.12.2006	58	8,2	4,98	8,55	4,78	6,6000	18,1000	5,7500	6,1000	24,9000	6,5000
27.12.2006	59	8,35	4,9	8,85	4,68	6,4500	19,2000	5,7000	6,0500	24,6000	6,4500
28.12.2006	60	8,4	4,94	8,95	4,78	6,5500	18,8000	5,7000	6,1000	24,8000	6,6000
29.12.2006	61	8,35	4,98	9	4,68	6,5000	18,3000	5,5500	6,1000	24,2000	6,6500

Tarih		DENİZ	DOAS	FINBN	GSDHO	ISGYO	PETKM	SKBNK	TOASO	TCELL	VESTL
02.10.2006	1	14,9000	7,1000	6,0000	1,2000	2,6800	5,0600	4,8800	4,1200	7,8000	3,5400
03.10.2006	2	14,9000	6,9000	6,0000	1,1700	2,6400	5,1000	4,8000	4,1800	7,7000	3,5000
04.10.2006	3	14,9000	6,8000	6,0000	1,1600	2,5800	5,1500	4,8000	4,1400	7,5500	3,4200
05.10.2006	4	14,9000	6,9000	5,9000	1,1800	2,6000	5,1500	4,8200	4,1400	7,4500	3,5600
06.10.2006	5	14,8000	6,6500	5,9500	1,1700	2,5800	5,1500	4,7600	4,2600	7,5500	3,5600
09.10.2006	6	14,8000	6,4500	5,9500	1,1700	2,5200	5,1500	4,7600	4,4000	7,5500	3,5000
10.10.2006	7	14,9000	6,7000	6,0000	1,1700	2,5800	5,1500	4,8200	4,5200	7,6000	3,5200
11.10.2006	8	14,9000	6,7000	6,0000	1,2100	2,6000	5,2000	4,8400	4,6000	7,7000	3,5000
12.10.2006	9	14,8000	6,9000	5,9500	1,2900	2,6400	5,2500	4,8600	4,5600	7,8000	3,5600
13.10.2006	10	14,7000	6,8000	5,9000	1,2900	2,7400	5,4500	4,8400	4,7000	7,7500	3,5800
16.10.2006	11	14,7000	6,9500	5,9500	1,2900	2,7800	5,4000	4,8800	4,9000	7,8500	3,7600
17.10.2006	12	14,7000	6,8000	5,9000	1,2400	2,8400	5,2500	4,9000	4,7800	7,7500	3,6000
18.10.2006	13	14,5000	6,9000	5,8500	1,2700	2,8800	5,4000	5,0800	4,9600	7,9000	3,7400
19.10.2006	14	14,6000	7,0000	5,8500	1,3200	2,8600	5,4000	5,1500	4,8600	7,9000	3,7200
20.10.2006	15	14,5000	6,9000	5,8000	1,3400	2,8600	5,4000	5,1500	4,8600	7,8000	3,7800
26.10.2006	16	14,6000	7,1000	5,8000	1,3600	2,8800	5,5500	5,3500	4,9400	7,8500	3,8000
27.10.2006	17	14,5000	7,0500	5,8000	1,4500	2,9200	5,5000	5,3000	4,8800	7,9000	3,8000
30.10.2006	18	16,0000	6,9500	5,8500	1,4400	2,9000	5,5000	5,2500	4,7800	7,8000	3,7200
31.10.2006	19	14,6000	7,1500	5,7500	1,4600	2,9000	5,6000	5,5500	4,8200	7,8500	3,8600
01.11.2006	20	14,6000	7,3000	5,8000	1,4500	2,9200	5,6500	5,5000	4,7800	7,7500	3,9000
02.11.2006	21	14,7000	7,2000	5,8500	1,3800	2,8200	5,4500	5,3500	4,8000	7,6000	3,8000
03.11.2006	22	14,7000	7,1500	5,8500	1,3900	2,8400	5,5000	5,3000	4,8200	7,5500	3,8600
06.11.2006	23	14,6000	7,3500	5,8500	1,4100	2,8400	5,5000	5,6500	4,8400	7,6000	3,9800
07.11.2006	24	14,4000	7,4000	5,7500	1,3500	2,9000	5,5000	5,9000	5,1000	7,7000	3,9400
08.11.2006	25	14,7000	7,3500	5,8000	1,3200	2,8800	5,4500	5,7000	4,9600	7,5000	4,0600
09.11.2006	26	14,6000	7,4000	5,8000	1,3500	2,8800	5,5500	5,7000	4,9600	7,2000	4,1400
10.11.2006	27	14,7000	7,3500	5,7500	1,3500	2,9000	5,5000	5,5500	5,0200	7,2000	4,1400
13.11.2006	28	14,7000	7,1500	5,8500	1,2900	2,8400	5,4000	5,4500	5,0000	7,2500	3,9400
14.11.2006	29	14,5000	7,2500	5,8000	1,3400	2,9000	5,4500	5,6000	5,1500	7,3500	4,0000
15.11.2006	30	14,6000	7,1500	5,8000	1,3300	2,9000	5,4000	5,7000	4,9500	7,3000	3,9400
16.11.2006	31	14,6000	7,0500	5,7000	1,3300	2,9000	5,4000	5,6000	4,7200	6,9500	3,9000
20.11.2006	32	14,9000	7,1500	5,7500	1,3100	2,8000	5,1500	5,6000	5,0000	6,7000	3,8200
21.11.2006	33	15,1000	7,0500	5,8500	1,3100	2,7800	5,2000	5,6000	4,9600	6,8000	3,8200
22.11.2006	34	15,0000	6,8000	5,8500	1,2900	2,7200	5,1500	5,3500	4,8000	6,7500	3,7400
23.11.2006	35	15,0000	7,2500	5,9000	1,3400	2,7600	5,1500	5,4000	4,8200	6,6000	3,7600
24.11.2006	36	15,1000	6,9000	5,9500	1,3500	2,6800	5,1000	5,2500	4,7600	6,5000	3,6600
27.11.2006	37	15,1000	6,8000	5,9000	1,3500	2,6600	5,1500	5,3000	4,8200	6,6500	3,7000
28.11.2006	38	15,1000	6,3500	5,9500	1,2600	2,6000	5,1000	5,2500	4,7400	6,3000	3,6000
29.11.2006	39	15,1000	6,3500	5,9000	1,2500	2,7400	5,1500	5,3000	4,7800	6,5500	3,7000
30.11.2006	40	15,0000	6,3500	5,8000	1,2500	2,8000	5,1500	5,3000	5,0000	6,8000	3,6600
01.12.2006	41	15,1000	6,3000	5,8000	1,2500	2,8200	5,1500	5,2000	5,0500	6,6000	3,6800
04.12.2006	42	15,1000	6,0500	5,8000	1,2300	2,7400	5,1500	5,1000	4,9500	6,4500	3,6000
05.12.2006	43	15,1000	6,0500	5,8500	1,2400	2,8400	5,1000	5,1000	5,1200	6,4500	3,6400
06.12.2006	44	15,1000	6,0500	5,8000	1,2400	2,9200	5,1500	5,1500	5,1500	6,6500	3,5400
07.12.2006	45	15,1000	6,3000	5,7000	1,2800	2,9800	5,2500	5,3000	5,2500	6,9500	3,6600
08.12.2006	46	15,1000	6,2000	5,7000	1,2600	3,0200	5,1500	5,3000	5,1000	6,8000	3,6200
11.12.2006	47	15,1000	6,1500	5,7000	1,2600	3,0200	5,1500	5,2500	5,1000	6,9000	3,6600
12.12.2006	48	15,2000	6,0000	5,6500	1,2400	2,9800	5,1500	5,2000	5,1000	6,9500	3,6200
13.12.2006	49	15,1000	6,0000	5,7000	1,2400	3,0000	5,2000	5,1000	5,1000	6,7500	3,6200
14.12.2006	50	15,1000	6,1000	5,7500	1,2600	3,0200	5,2500	5,1500	5,1000	6,8500	3,6600
15.12.2006	51	15,2000	6,2500	5,7000	1,2700	3,0000	5,3000	5,2000	5,1000	6,9000	3,7600
18.12.2006	52	15,1000	6,1000	5,7000	1,2600	2,9600	5,3500	5,1000	4,9800	6,8500	3,8600
19.12.2006	53	15,1000	5,8000	5,7500	1,2300	2,9200	5,1500	5,0000	4,9800	6,6000	3,7400
20.12.2006	54	15,0000	5,8000	5,7000	1,2100	2,9200	5,1500	4,8500	4,8600	6,6000	3,7200
21.12.2006	55	14,9000	5,7000	5,7000	1,2100	2,9400	5,1000	4,7800	4,8200	6,7500	3,7000
22.12.2006	56	15,1000	5,8000	5,7500	1,2000	2,9200	5,1000	4,7600	4,8200	6,8500	3,7000
25.12.2006	57	14,4000	5,7000	5,6500	1,1900	2,8800	5,1000	4,9400	4,8000	6,7000	3,6800
26.12.2006	58	14,1000	5,7000	5,6500	1,1900	2,9200	5,1500	4,6600	4,8000	6,8000	3,6800
27.12.2006	59	13,7000	5,6500	5,6500	1,1700	2,8400	5,1000	4,7000	4,7800	6,8500	3,6600
28.12.2006	60	13,7000	5,7000	5,7000	1,1800	2,8800	5,1500	4,7200	4,8200	7,1000	3,7200
29.12.2006	61	13,9000	5,7000	5,9000	1,1800	2,9000	5,1000	4,7600	4,9000	7,1500	3,6800

Tarih		AKBNK	DOHOL	HURGZ	KCHOL	PTOFS	SISE	TSKB	ULKER	YKBNK	FORTIS
02.10.2006	1	7,8	5,95	3,8000	4,8800	4,7800	4,9000	2,8200	3,9800	2,6600	4,32
03.10.2006	2	7,8	6	3,6800	4,8600	5,0800	4,9000	2,8000	3,9600	2,6400	4,3
04.10.2006	3	7,55	5,85	3,6600	4,8000	5,1200	4,8200	2,7800	3,9200	2,6000	4,24
05.10.2006	4	7,7	5,95	3,7400	5,0000	5,0500	4,9000	2,8000	3,9400	2,6200	4,28
06.10.2006	5	7,6	5,85	3,7200	4,8800	4,9400	4,9000	2,8200	3,9000	2,5600	4,3
09.10.2006	6	7,7	5,85	3,7200	4,9400	4,9000	5,1000	2,8600	3,9000	2,5800	4,3
10.10.2006	7	7,9	5,9	3,8800	4,9200	4,9000	5,4000	2,9000	3,9800	2,6000	4,34
11.10.2006	8	7,95	5,85	3,9000	4,9600	4,9000	5,4000	2,8600	3,9800	2,6400	4,34
12.10.2006	9	8,1	6,1	3,9800	5,0200	5,0200	5,4000	2,8800	4,0200	2,7200	4,48
13.10.2006	10	8,3	6,05	4,0800	5,1000	5,1000	5,5500	2,9400	4,0200	2,7200	4,5
16.10.2006	11	8,35	6,1	4,1000	5,2000	5,2000	5,7000	2,9400	4,0800	2,7200	4,5
17.10.2006	12	8,15	5,9	3,9000	5,2000	5,1000	5,7500	2,8200	4,0400	2,6600	4,48
18.10.2006	13	8,4	6	4,0400	5,3500	5,2500	5,8000	2,9600	4,1400	2,7800	4,68
19.10.2006	14	8,35	6,2	4,2200	5,2500	5,2000	5,8500	3,1400	4,2200	2,7600	4,76
20.10.2006	15	8,2	6,05	4,1400	5,2000	5,2000	5,7000	3,0800	4,2000	2,7200	4,84
26.10.2006	16	8,25	6,45	4,2800	5,5500	5,4500	5,8500	3,1400	4,2600	2,7800	5
27.10.2006	17	8,3	6,45	4,1000	5,5000	5,3500	5,8000	3,1000	4,2000	2,7800	4,88
30.10.2006	18	8,15	6,25	4,0800	5,4000	5,3000	5,5500	3,0600	4,0800	2,7200	4,8
31.10.2006	19	8,3	6,25	4,1800	5,6500	5,2500	5,6500	3,1400	4,1600	2,8000	4,86
01.11.2006	20	8,35	6,25	4,1200	5,7000	5,2500	5,5500	3,1800	4,1400	2,8200	4,86
02.11.2006	21	8,2	2,96	4,1000	5,4500	4,9000	5,4000	3,0400	4,0600	2,7000	4,76
03.11.2006	22	8,1	2,86	4,2000	5,4500	5,1600	5,4500	2,9400	4,0400	2,7000	4,76
06.11.2006	23	8,25	3,02	4,2800	5,5000	5,4500	5,5000	2,9800	4,1000	2,6800	4,8
07.11.2006	24	8,45	3,12	4,2800	5,6500	5,4500	5,6000	3,0600	4,1000	2,7200	4,86
08.11.2006	25	8,5	3,12	4,2200	5,5000	5,6500	5,6000	3,0200	4,0800	2,6800	4,8
09.11.2006	26	8,7	3,1	4,2200	5,6000	5,5500	5,5500	3,0000	4,1000	2,7000	4,8
10.11.2006	27	8,4	3,06	4,2200	5,5000	5,5500	5,4500	3,0000	4,0800	2,6600	4,76
13.11.2006	28	8,25	2,94	4,1400	5,4000	5,4500	5,2500	2,9600	4,0400	2,6200	4,68
14.11.2006	29	8,4	2,98	4,3400	5,5000	5,5500	5,4000	3,0600	4,0600	2,6800	4,72
15.11.2006	30	8,3	2,98	4,2200	5,4000	5,3000	5,3000	3,0600	4,1000	2,7000	4,7
16.11.2006	31	8,3	2,98	4,2800	5,3500	5,2000	5,3000	3,1200	4,0600	2,6600	4,64
20.11.2006	32	7,95	2,82	4,2200	5,0500	5,0600	5,2000	2,9600	3,9200	2,6200	4,52
21.11.2006	33	8,2	2,76	4,2400	5,1500	5,0200	5,2500	3,0000	3,9200	2,5800	4,48
22.11.2006	34	7,9	2,64	4,1400	5,1000	4,9000	5,1000	2,9600	3,8000	2,5400	4,46
23.11.2006	35	8,05	2,64	4,1200	5,2000	4,9600	5,1500	2,9600	3,8200	2,5600	4,5
24.11.2006	36	7,9	2,48	4,0200	5,0500	4,8600	5,0500	2,8200	3,6800	2,5000	4,48
27.11.2006	37	8,25	2,54	3,9600	5,1000	4,8400	5,2000	2,8600	3,9000	2,5600	4,52
28.11.2006	38	8,05	2,47	3,8000	4,8600	4,9400	4,9500	2,6800	3,7800	2,5000	4,46
29.11.2006	39	8,25	2,54	3,8600	4,9400	5,1000	5,2000	2,7600	3,8200	2,5400	4,54
30.11.2006	40	8,15	2,54	4,0200	4,9800	5,0500	5,2500	2,7800	3,8000	2,5400	4,54
01.12.2006	41	8,2	2,52	3,9200	5,0000	5,0500	5,3500	2,7600	3,8400	2,5000	4,56
04.12.2006	42	8,15	2,44	3,9600	4,8800	5,0500	5,1000	2,6000	3,7600	2,5000	4,52
05.12.2006	43	8,35	2,49	4,0200	4,9800	5,1000	5,1500	2,5200	3,7400	2,5200	4,52
06.12.2006	44	8,3	2,56	3,9200	4,9800	5,0500	5,0000	2,5600	3,7600	2,5000	4,5
07.12.2006	45	8,65	2,72	4,0200	5,2500	5,3500	5,2500	2,7400	3,8600	2,6200	4,68
08.12.2006	46	8,55	2,6	4,0600	5,2000	5,2000	5,2000	2,7400	3,8600	2,5800	4,58
11.12.2006	47	8,55	2,6	4,1600	5,1500	5,2500	5,2000	2,8000	3,8400	2,6000	4,68
12.12.2006	48	8,5	2,56	4,0400	5,1000	5,3500	5,1000	2,7600	3,8600	2,5600	4,56
13.12.2006	49	8,25	2,54	3,9000	5,1000	5,4500	5,1500	2,7800	3,8600	2,5200	4,52
14.12.2006	50	8,4	2,64	4,0000	5,2500	5,4500	5,2500	2,7800	3,8600	2,6000	4,58
15.12.2006	51	8,6	2,66	4,0800	5,5000	5,5000	5,2500	2,8400	3,9400	2,6200	4,62
18.12.2006	52	8,55	2,62	4,0000	5,5000	5,5000	5,3000	2,8200	3,9200	2,5800	4,54
19.12.2006	53	8,45	2,56	3,9000	5,3000	5,3000	5,1000	2,7200	3,8600	2,5000	4,46
20.12.2006	54	8,35	2,58	3,9000	5,3000	5,3000	5,1000	2,6800	3,8200	2,5200	4,4
21.12.2006	55	8,5	2,52	3,8800	5,3000	5,4000	5,0500	2,6400	3,8400	2,5100	4,34
22.12.2006	56	8,45	2,5	3,9200	5,4000	5,3500	5,0000	2,6400	3,8200	2,5000	4,46
25.12.2006	57	8,45	2,32	3,7800	5,5000	4,8500	4,9800	2,5800	3,7600	2,4600	4,32
26.12.2006	58	8,5	2,27	3,7400	5,4500	4,6000	5,0000	2,6000	3,7800	2,4400	4,4
27.12.2006	59	8,45	2,18	3,7200	5,4500	4,4400	4,9400	2,5800	3,7400	2,4400	4,48
28.12.2006	60	8,55	2,23	3,8200	5,6000	4,5200	5,0000	2,6400	3,8200	2,4800	4,76
29.12.2006	61	8,6	2,23	3,7400	5,5000	4,5800	5,0000	2,5800	3,7800	2,4600	5,34

EK 3

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	19,09	18,91	18,64	18,94	18,88	19,05	19,36	19,38	19,69	19,47	19,30	18,89	18,93	18,90	18,94	19,14	19,14	19,46	19,44	19,61
2	18,91	18,73	18,47	18,76	18,71	18,87	19,17	19,20	19,49	19,28	19,12	18,71	18,75	18,72	18,76	18,96	18,95	19,28	19,25	19,41
3	18,64	18,47	18,23	18,50	18,46	18,63	18,93	18,95	19,23	19,02	18,86	18,46	18,49	18,45	18,49	18,69	18,69	19,02	18,97	19,13
4	18,94	18,76	18,50	18,80	18,75	18,91	19,21	19,23	19,53	19,32	19,15	18,75	18,79	18,76	18,80	19,00	19,00	19,32	19,29	19,46
5	18,88	18,71	18,46	18,75	18,71	18,88	19,17	19,19	19,48	19,27	19,11	18,70	18,73	18,70	18,74	18,94	18,94	19,27	19,23	19,39
6	19,05	18,87	18,63	18,91	18,88	19,09	19,41	19,41	19,71	19,49	19,33	18,90	18,94	18,90	18,93	19,15	19,15	19,47	19,45	19,60
7	19,36	19,17	18,93	19,21	19,17	19,41	19,75	19,76	20,07	19,83	19,67	19,23	19,27	19,22	19,26	19,49	19,49	19,79	19,79	19,95
8	19,38	19,20	18,95	19,23	19,19	19,41	19,76	19,78	20,09	19,86	19,69	19,26	19,30	19,26	19,29	19,52	19,51	19,82	19,82	19,99
9	19,69	19,49	19,23	19,53	19,48	19,71	20,07	20,09	20,41	20,17	20,00	19,56	19,62	19,57	19,61	19,84	19,84	20,13	20,14	20,32
10	19,47	19,28	19,02	19,32	19,27	19,49	19,83	19,86	20,17	19,95	19,78	19,35	19,40	19,35	19,39	19,62	19,62	19,91	19,92	20,09
11	19,30	19,12	18,86	19,15	19,11	19,33	19,67	19,69	20,00	19,78	19,63	19,20	19,26	19,21	19,25	19,47	19,48	19,76	19,77	19,94
12	18,89	18,71	18,46	18,75	18,70	18,90	19,23	19,26	19,56	19,35	19,20	18,79	18,85	18,80	18,84	19,06	19,06	19,35	19,35	19,51
13	18,93	18,75	18,49	18,79	18,73	18,94	19,27	19,30	19,62	19,40	19,26	18,85	18,93	18,88	18,92	19,14	19,15	19,41	19,44	19,61
14	18,90	18,72	18,45	18,76	18,70	18,90	19,22	19,26	19,57	19,35	19,21	18,80	18,88	18,84	18,89	19,10	19,10	19,37	19,39	19,56
15	18,94	18,76	18,49	18,80	18,74	18,93	19,26	19,29	19,61	19,39	19,25	18,84	18,92	18,89	18,94	19,15	19,15	19,42	19,44	19,62
16	19,14	18,96	18,69	19,00	18,94	19,15	19,49	19,52	19,84	19,62	19,47	19,06	19,14	19,10	19,15	19,38	19,38	19,64	19,68	19,85
17	19,14	18,95	18,69	19,00	18,94	19,15	19,49	19,51	19,84	19,62	19,48	19,06	19,15	19,10	19,15	19,38	19,39	19,64	19,68	19,85
18	19,46	19,28	19,02	19,32	19,27	19,47	19,79	19,82	20,13	19,91	19,76	19,35	19,41	19,37	19,42	19,64	19,64	19,99	19,94	20,11
19	19,44	19,25	18,97	19,29	19,23	19,45	19,79	19,82	20,14	19,92	19,77	19,35	19,44	19,39	19,44	19,68	19,68	19,94	19,99	20,17
20	19,61	19,41	19,13	19,46	19,39	19,60	19,95	19,99	20,32	20,09	19,94	19,51	19,61	19,56	19,62	19,85	19,85	20,11	20,17	20,36
21	19,26	19,06	18,79	19,11	19,05	19,24	19,59	19,63	19,94	19,72	19,57	19,17	19,26	19,20	19,27	19,46	19,46	19,75	19,79	19,98
22	19,02	18,82	18,56	18,88	18,81	19,00	19,35	19,38	19,67	19,45	19,28	18,88	18,93	18,91	18,97	19,14	19,11	19,44	19,46	19,66
23	18,93	18,73	18,47	18,79	18,73	18,92	19,27	19,30	19,59	19,36	19,20	18,80	18,85	18,83	18,89	19,07	19,04	19,36	19,39	19,58
24	19,05	18,85	18,58	18,90	18,84	19,04	19,39	19,42	19,72	19,50	19,33	18,92	18,99	18,97	19,03	19,21	19,18	19,48	19,53	19,73
25	19,15	18,96	18,70	19,01	18,95	19,15	19,51	19,53	19,83	19,60	19,43	19,03	19,08	19,06	19,12	19,29	19,26	19,58	19,62	19,82
26	19,33	19,13	18,87	19,19	19,12	19,32	19,68	19,72	20,02	19,79	19,61	19,20	19,26	19,24	19,30	19,48	19,44	19,76	19,81	20,01
27	19,26	19,07	18,81	19,12	19,06	19,27	19,62	19,65	19,95	19,72	19,55	19,14	19,19	19,17	19,23	19,41	19,38	19,70	19,73	19,93
28	19,08	18,89	18,63	18,94	18,88	19,08	19,43	19,46	19,75	19,53	19,35	18,95	19,00	18,97	19,04	19,21	19,18	19,51	19,53	19,73
29	19,36	19,16	18,90	19,22	19,16	19,37	19,73	19,76	20,06	19,83	19,65	19,24	19,29	19,27	19,34	19,51	19,48	19,79	19,84	20,04
30	19,48	19,27	19,00	19,33	19,26	19,47	19,83	19,86	20,17	19,94	19,76	19,34	19,40	19,39	19,46	19,63	19,60	19,92	19,96	20,17
31	19,65	19,44	19,17	19,50	19,43	19,64	20,00	20,04	20,35	20,11	19,93	19,51	19,57	19,56	19,63	19,80	19,77	20,09	20,14	20,35
32	19,36	19,15	18,89	19,21	19,14	19,35	19,71	19,74	20,03	19,80	19,62	19,22	19,26	19,25	19,32	19,49	19,45	19,80	19,81	20,02
33	19,27	19,07	18,80	19,12	19,06	19,26	19,62	19,65	19,94	19,71	19,53	19,13	19,17	19,16	19,22	19,39	19,36	19,72	19,72	19,92
34	18,94	18,74	18,48	18,80	18,73	18,92	19,27	19,30	19,58	19,36	19,18	18,80	18,83	18,81	18,88	19,04	19,01	19,37	19,36	19,56
35	18,78	18,59	18,34	18,65	18,58	18,78	19,12	19,14	19,42	19,20	19,03	18,65	18,68	18,66	18,72	18,89	18,86	19,22	19,21	19,40
36	18,60	18,41	18,17	18,47	18,40	18,60	18,94	18,97	19,24	19,02	18,85	18,48	18,50	18,49	18,56	18,71	18,68	19,06	19,03	19,23
37	18,78	18,58	18,33	18,64	18,57	18,77	19,12	19,16	19,43	19,21	19,04	18,66	18,70	18,68	18,75	18,90	18,88	19,24	19,22	19,43
38	18,56	18,38	18,13	18,44	18,36	18,56	18,91	18,94	19,21	18,99	18,82	18,45	18,48	18,46	18,53	18,68	18,65	19,03	18,99	19,20
39	18,70	18,51	18,26	18,57	18,50	18,70	19,05	19,08	19,36	19,14	18,97	18,59	18,63	18,61	18,68	18,84	18,81	19,18	19,15	19,35
40	19,01	18,82	18,56	18,88	18,81	19,03	19,39	19,42	19,70	19,48	19,30	18,91	18,95	18,92	18,99	19,15	19,13	19,48	19,47	19,68
41	19,09	18,90	18,64	18,96	18,89	19,11	19,47	19,50	19,78	19,56	19,39	19,00	19,03	19,01	19,08	19,24	19,21	19,58	19,56	19,77
42	19,12	18,93	18,69	18,99	18,93	19,16	19,52	19,55	19,82	19,60	19,43	19,03	19,06	19,03	19,10	19,27	19,24	19,61	19,59	19,80
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
43	19,20	19,01	18,76	19,07	19,01	19,24	19,60	19,63	19,91	19,69	19,51	19,12	19,15	19,12	19,19	19,36	19,33	19,70	19,68	19,88
44	19,12	18,93	18,68	18,99	18,93	19,16	19,52	19,54	19,82	19,60	19,43	19,04	19,08	19,04	19,11	19,28	19,26	19,62	19,60	19,80
45	19,59	19,40	19,14	19,46	19,40	19,63	20,00	20,03	20,32	20,10	19,93	19,52	19,58	19,53	19,60	19,79	19,77	20,11	20,12	20,32
46	19,67	19,48	19,21	19,54	19,48	19,70	20,07	20,09	20,39	20,17	19,99	19,58	19,63	19,59	19,66	19,84	19,82	20,17	20,18	20,38
47	19,58	19,39	19,12	19,45	19,39	19,61	19,97	20,00	20,29	20,08	19,90	19,49	19,54	19,50	19,58	19,75	19,73	20,08	20,09	20,29
48	19,53	19,34	19,08	19,40	19,34	19,56	19,93	19,96	20,25	20,03	19,85	19,45	19,50	19,46	19,53	19,71	19,69	20,04	20,04	20,24
49	19,25	19,06	18,80	19,13	19,06	19,27	19,64	19,67	19,95	19,74	19,56	19,17	19,22	19,18	19,25	19,42	19,40	19,76	19,75	19,96
50	19,49	19,30	19,04	19,37	19,30	19,53	19,89	19,92	20,21	19,99	19,82	19,41	19,46	19,42	19,49	19,67	19,65	20,00	20,01	20,21
51	19,81	19,61	19,36	19,68	19,62	19,88	20,25	20,27	20,57	20,34	20,17	19,74	19,80	19,75	19,81	20,01	19,99	20,34	20,35	20,54
52	19,52	19,33	19,08	19,40	19,34	19,58	19,95	19,97	20,26	20,04	19,87	19,45	19,51	19,46	19,53	19,71	19,70	20,04	20,05	20,24
53	19,56	19,37	19,13	19,44	19,39	19,64	20,01	20,03	20,32	20,09	19,92	19,50	19,55	19,50	19,57	19,76	19,74	20,09	20,10	20,29
54	19,55	19,36	19,12	19,42	19,38	19,63	20,00	20,01	20,30	20,08	19,91	19,49	19,54	19,49	19,55	19,75	19,73	20,08	20,08	20,27
55	19,69	19,50	19,25	19,56	19,51	19,76	20,13	20,15	20,45	20,22	20,05	19,63	19,68	19,63	19,70	19,89	19,88	20,22	20,23	20,42
56	19,92	19,72	19,48	19,78	19,74	20,01	20,38	20,39	20,69	20,46	20,29	19,85	19,90	19,85	19,91	20,11	20,10	20,45	20,46	20,64
57	20,22	20,02	19,77	20,08	20,05	20,34	20,73	20,73	21,05	20,8										

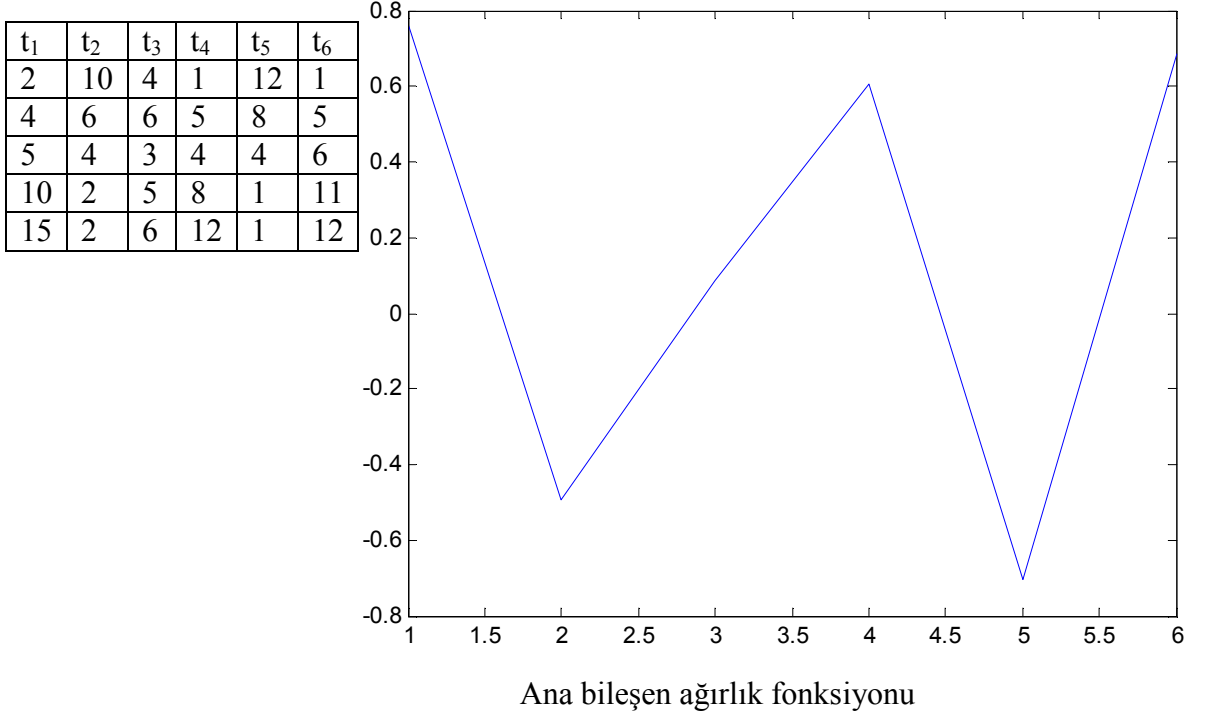
	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	19,65	19,36	19,27	18,94	18,78	18,60	18,78	18,56	18,70	19,01	19,09	19,12	19,20	19,12	19,59	19,67	19,58	19,53	19,25	19,49
2	19,44	19,15	19,07	18,74	18,59	18,41	18,58	18,38	18,51	18,82	18,90	18,93	19,01	18,93	19,40	19,48	19,39	19,34	19,06	19,30
3	19,17	18,89	18,80	18,48	18,34	18,17	18,33	18,13	18,26	18,56	18,64	18,69	18,76	18,68	19,14	19,21	19,12	19,08	18,80	19,04
4	19,50	19,21	19,12	18,80	18,65	18,47	18,64	18,44	18,57	18,88	18,96	18,99	19,07	18,99	19,46	19,54	19,45	19,40	19,13	19,37
5	19,43	19,14	19,06	18,73	18,58	18,40	18,57	18,36	18,50	18,81	18,89	18,93	19,01	18,93	19,40	19,48	19,39	19,34	19,06	19,30
6	19,64	19,35	19,26	18,92	18,78	18,60	18,77	18,56	18,70	19,03	19,11	19,16	19,24	19,16	19,63	19,70	19,61	19,56	19,27	19,53
7	20,00	19,71	19,62	19,27	19,12	18,94	19,12	18,91	19,05	19,39	19,47	19,52	19,60	19,52	20,00	20,07	19,97	19,93	19,64	19,89
8	20,04	19,74	19,65	19,30	19,14	18,97	19,16	18,94	19,08	19,42	19,50	19,55	19,63	19,54	20,03	20,09	20,00	19,96	19,67	19,92
9	20,35	20,03	19,94	19,58	19,42	19,24	19,43	19,21	19,36	19,70	19,78	19,82	19,91	19,82	20,32	20,39	20,29	20,25	19,95	20,21
10	20,11	19,80	19,71	19,36	19,20	19,02	19,21	18,99	19,14	19,48	19,56	19,60	19,69	19,60	20,10	20,17	20,08	20,03	19,74	19,99
11	19,93	19,62	19,53	19,18	19,03	18,85	19,04	18,82	18,97	19,30	19,39	19,43	19,51	19,43	19,93	19,99	19,90	19,85	19,56	19,82
12	19,51	19,22	19,13	18,80	18,65	18,48	18,66	18,45	18,59	18,91	19,00	19,03	19,12	19,04	19,52	19,58	19,49	19,45	19,17	19,41
13	19,57	19,26	19,17	18,83	18,68	18,50	18,70	18,48	18,63	18,95	19,03	19,06	19,15	19,08	19,58	19,63	19,54	19,50	19,22	19,46
14	19,56	19,25	19,16	18,81	18,66	18,49	18,68	18,46	18,61	18,92	19,01	19,03	19,12	19,04	19,53	19,59	19,50	19,46	19,18	19,42
15	19,63	19,32	19,22	18,88	18,72	18,56	18,75	18,53	18,68	18,99	19,08	19,10	19,19	19,11	19,60	19,66	19,58	19,53	19,25	19,49
16	19,80	19,49	19,39	19,04	18,89	18,71	18,90	18,68	18,84	19,15	19,24	19,27	19,36	19,28	19,79	19,84	19,75	19,71	19,42	19,67
17	19,77	19,45	19,36	19,01	18,86	18,68	18,88	18,65	18,81	19,13	19,21	19,24	19,33	19,26	19,77	19,82	19,73	19,69	19,40	19,65
18	20,09	19,80	19,72	19,37	19,22	19,06	19,24	19,03	19,18	19,48	19,58	19,61	19,70	19,62	20,11	20,17	20,08	20,04	19,76	20,00
19	20,14	19,81	19,72	19,36	19,21	19,03	19,22	18,99	19,15	19,47	19,56	19,59	19,68	19,60	20,12	20,18	20,09	20,04	19,75	20,01
20	20,35	20,02	19,92	19,56	19,40	19,23	19,43	19,20	19,35	19,68	19,77	19,80	19,88	19,80	20,32	20,38	20,29	20,24	19,96	20,21
21	20,30	19,99	19,90	19,55	19,40	19,24	19,43	19,21	19,35	19,67	19,76	19,78	19,87	19,78	20,28	20,35	20,27	20,22	19,94	20,18
22	20,27	19,97	19,89	19,54	19,38	19,22	19,40	19,20	19,32	19,64	19,73	19,75	19,82	19,73	20,19	20,27	20,19	20,15	19,87	20,10
23	20,21	19,90	19,81	19,46	19,31	19,15	19,32	19,12	19,24	19,56	19,64	19,66	19,72	19,64	20,10	20,18	20,09	20,05	19,78	20,01
24	20,34	20,02	19,93	19,57	19,41	19,25	19,43	19,22	19,36	19,68	19,76	19,77	19,84	19,75	20,22	20,30	20,21	20,18	19,90	20,13
25	20,44	20,14	20,05	19,69	19,53	19,37	19,55	19,34	19,47	19,79	19,88	19,90	19,96	19,88	20,34	20,42	20,33	20,30	20,02	20,25
26	20,66	20,34	20,25	19,88	19,73	19,56	19,75	19,54	19,67	19,99	20,08	20,10	20,16	20,07	20,54	20,63	20,54	20,50	20,22	20,45
27	20,57	20,26	20,17	19,81	19,65	19,49	19,66	19,46	19,59	19,91	20,00	20,02	20,09	20,00	20,46	20,55	20,46	20,42	20,14	20,37
28	20,37	20,07	19,98	19,62	19,47	19,31	19,48	19,28	19,41	19,72	19,81	19,83	19,90	19,81	20,27	20,35	20,26	20,23	19,95	20,18
29	20,71	20,39	20,30	19,93	19,77	19,61	19,79	19,58	19,71	20,04	20,12	20,15	20,21	20,12	20,59	20,68	20,59	20,55	20,27	20,50
30	20,83	20,52	20,42	20,05	19,88	19,73	19,91	19,70	19,83	20,15	20,24	20,26	20,32	20,23	20,70	20,79	20,70	20,66	20,38	20,62
31	21,03	20,70	20,61	20,24	20,07	19,91	20,09	19,89	20,02	20,34	20,43	20,45	20,51	20,42	20,90	20,99	20,90	20,86	20,57	20,81
32	20,70	20,41	20,32	19,96	19,80	19,65	19,82	19,62	19,74	20,06	20,15	20,17	20,23	20,15	20,60	20,69	20,60	20,56	20,28	20,51
33	20,61	20,32	20,23	19,88	19,72	19,57	19,75	19,55	19,67	19,98	20,07	20,10	20,16	20,08	20,52	20,61	20,52	20,49	20,21	20,43
34	20,24	19,96	19,88	19,54	19,38	19,24	19,40	19,21	19,33	19,63	19,72	19,75	19,81	19,73	20,16	20,25	20,16	20,13	19,86	20,08
35	20,07	19,80	19,72	19,38	19,24	19,09	19,25	19,06	19,17	19,47	19,56	19,59	19,65	19,57	20,00	20,09	20,00	19,96	19,69	19,91
36	19,91	19,65	19,57	19,24	19,09	18,95	19,12	18,93	19,04	19,33	19,42	19,46	19,51	19,43	19,85	19,94	19,85	19,82	19,56	19,77
37	20,09	19,82	19,75	19,40	19,25	19,12	19,29	19,11	19,22	19,51	19,61	19,63	19,70	19,61	20,04	20,12	20,04	20,01	19,74	19,95
38	19,89	19,62	19,55	19,21	19,06	18,93	19,11	18,94	19,04	19,33	19,42	19,45	19,51	19,43	19,85	19,93	19,85	19,82	19,56	19,76
39	20,02	19,74	19,67	19,33	19,17	19,04	19,22	19,04	19,16	19,45	19,54	19,57	19,63	19,55	19,98	20,06	19,98	19,95	19,68	19,89
40	20,34	20,06	19,98	19,63	19,47	19,33	19,51	19,33	19,45	19,76	19,85	19,89	19,95	19,87	20,31	20,39	20,31	20,28	20,00	20,22

41 | 20,43 | 20,15 | 20,07 | 19,72 | 19,56 | 19,42 | 19,61 | 19,42 | 19,54 | 19,85 | 19,95 | 19,99 | 20,05 | 19,96 | 20,41 | 20,49 | 20,40 | 20,37 | 20,10 | 20,32

	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
42	20,45	20,17	20,10	19,75	19,59	19,46	19,63	19,45	19,57	19,89	19,99	20,03	20,10	20,01	20,46	20,54	20,45	20,42	20,14	20,
43	20,51	20,23	20,16	19,81	19,65	19,51	19,70	19,51	19,63	19,95	20,05	20,10	20,17	20,08	20,53	20,61	20,52	20,49	20,22	20,
44	20,42	20,15	20,08	19,73	19,57	19,43	19,61	19,43	19,55	19,87	19,96	20,01	20,08	20,00	20,45	20,52	20,44	20,41	20,13	20,
45	20,90	20,60	20,52	20,16	20,00	19,85	20,04	19,85	19,98	20,31	20,41	20,46	20,53	20,45	20,93	21,00	20,91	20,88	20,59	20,
46	20,99	20,69	20,61	20,25	20,09	19,94	20,12	19,93	20,06	20,39	20,49	20,54	20,61	20,52	21,00	21,08	20,99	20,95	20,67	20,
47	20,90	20,60	20,52	20,16	20,00	19,85	20,04	19,85	19,98	20,31	20,40	20,45	20,52	20,44	20,91	20,99	20,90	20,87	20,58	20,
48	20,86	20,56	20,49	20,13	19,96	19,82	20,01	19,82	19,95	20,28	20,37	20,42	20,49	20,41	20,88	20,95	20,87	20,84	20,55	20,
49	20,57	20,28	20,21	19,86	19,69	19,56	19,74	19,56	19,68	20,00	20,10	20,14	20,22	20,13	20,59	20,67	20,58	20,55	20,28	20,
50	20,81	20,51	20,43	20,08	19,91	19,77	19,95	19,76	19,89	20,22	20,32	20,36	20,44	20,35	20,82	20,90	20,81	20,78	20,50	20,
51	21,11	20,81	20,73	20,36	20,20	20,04	20,23	20,03	20,17	20,51	20,61	20,66	20,73	20,64	21,13	21,21	21,11	21,08	20,79	21,
52	20,82	20,52	20,44	20,08	19,92	19,77	19,96	19,76	19,89	20,23	20,33	20,38	20,45	20,36	20,84	20,92	20,82	20,79	20,50	20,
53	20,88	20,58	20,50	20,14	19,98	19,83	20,02	19,82	19,96	20,29	20,39	20,45	20,52	20,43	20,91	20,98	20,89	20,86	20,57	20,
54	20,85	20,55	20,47	20,11	19,95	19,80	19,98	19,79	19,92	20,26	20,36	20,42	20,49	20,41	20,88	20,96	20,86	20,83	20,54	20,
55	21,00	20,69	20,62	20,26	20,09	19,94	20,13	19,93	20,07	20,41	20,51	20,56	20,64	20,55	21,04	21,11	21,02	20,99	20,70	20,
56	21,22	20,91	20,83	20,47	20,30	20,14	20,33	20,13	20,27	20,62	20,72	20,78	20,85	20,77	21,25	21,33	21,23	21,20	20,90	21,
57	21,57	21,23	21,15	20,77	20,60	20,43	20,62	20,40	20,55	20,93	21,03	21,09	21,17	21,08	21,59	21,67	21,56	21,52	21,21	21,
58	21,70	21,36	21,27	20,89	20,72	20,54	20,74	20,52	20,67	21,05	21,15	21,21	21,29	21,20	21,71	21,79	21,68	21,64	21,33	21,
59	21,83	21,48	21,38	21,00	20,83	20,64	20,83	20,60	20,76	21,16	21,25	21,33	21,41	21,32	21,85	21,93	21,81	21,77	21,44	21,
60	21,76	21,41	21,31	20,93	20,76	20,58	20,77	20,54	20,70	21,09	21,19	21,25	21,34	21,25	21,77	21,85	21,74	21,69	21,37	21,
61	21,27	20,94	20,85	20,48	20,32	20,15	20,33	20,11	20,26	20,64	20,74	20,80	20,88	20,80	21,30	21,38	21,27	21,23	20,91	21,

EK 4

ÖRNEK 1:



İlgili Kovaryans Matrisi aşağıda verilmektedir:

22.1459	-11.7423	2.88446	17.1586	-17.1965	17.9805
-11.7423	8.9427	-1.0411	-9.5663	12.6037	-11.5792
2.8446	-1.0411	1.3599	2.8003	-1.1610	2.2030
17.1586	-9.5663	2.8003	13.9439	-13.5377	14.3578
-17.1965	12.6037	-1.1610	-13.5377	18.0906	-16.5540
17.9805	-11.5792	2.2060	14.3578	-16.5540	16.3760

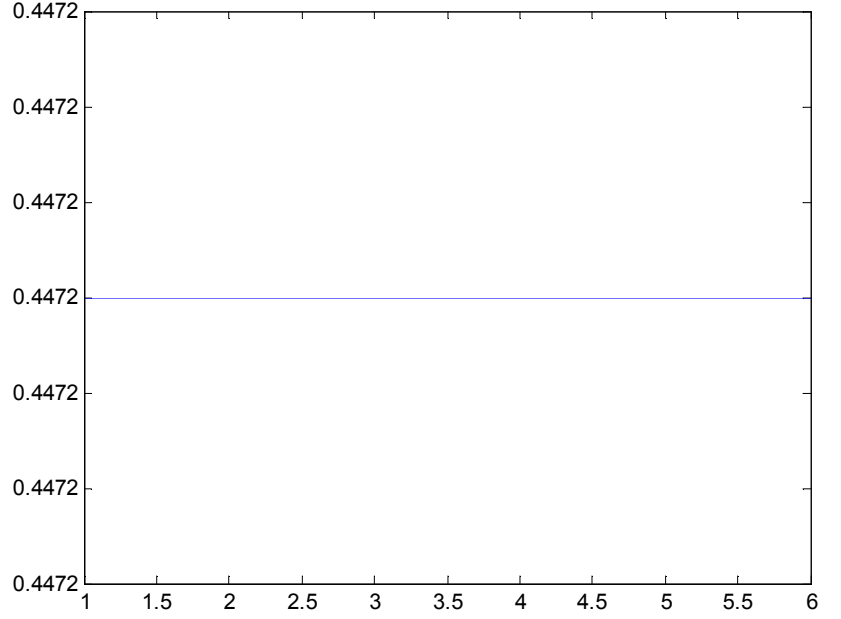
Ana bileşen fonksiyonunda ardışık inişler ve çıkışlar kovaryansın yönü tarafından belirlenir. Bu artış yada azalışların uzunluğu da kovaryansın mutlak değeri tarafından belirlenir, ardışık olarak ters ilişki varsa fonksiyon sifıra göre pozisyon değiştirir.

Ayrıca sıfırdan y eksenini boyutundaki mutlak uzaklıklar değişkenin varyansı ile doğrudan ilişkilidir.

ÖRNEK 2:

$y=[2\ 4\ 6\ 8\ 10;3\ 5\ 7\ 9\ 11;4\ 6\ 8\ 10\ 12;5\ 7\ 9\ 11\ 13;6\ 8\ 10\ 12\ 14;7\ 9\ 11\ 13\ 15]$

t_1	t_2	t_3	t_4	t_5	t_6
2	3	4	5	6	7
4	5	6	7	7	9
6	7	8	9	9	11
8	9	10	11	11	13
10	11	12	13	13	15



Ana bileşen ağırlık fonksiyonu

İlgili Kovaryans Matrisi aşağıda verilmektedir.

8	8	8	8	8	8
8	8	8	8	8	8
8	8	8	8	8	8
8	8	8	8	8	8
8	8	8	8	8	8
8	8	8	8	8	8

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
DOKTORA TEZİ

**ÇOK DEĞİŞKENLİ İSTATİSTİKSEL BOYUT
İNDİRGEME YÖNTEMİ OLARAK
DÜZGÜNLEŞTİRİLMİŞ FONKSİYONEL ANA
BİLEŞENLER ANALİZİ ÜZERİNE BİR ARAŞTIRMA**

İstem KESER

Danışman
Doç.Dr.Kadir ERTAŞ

2007

YEMİN METNİ

Doktora Tezi olarak sunduđum “ Çok Deđişkenli İstatistiksel Boyut İndirgeme Yöntemi Olarak Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi Üzerine Bir Araştırma” adlı çalışmanın, tarafımdan, bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım eserlerin bibliyografyada gösterilenlerden oluştuđunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

Tarih

.../.../.....

İstem KESER

DOKTORA TEZ SINAV TUTANAĞI

Öğrencinin

Adı ve Soyadı : İstem KESER
Anabilim Dalı : Ekonometri
Programı : Doktora
Tez Konusu : Çok Değişkenli İstatistiksel Boyut İndirgeme
Yöntemi Olarak Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi Üzerine Bir Araştırma
Sınav Tarihi ve Saati :

Yukarıda kimlik bilgileri belirtilen öğrenci Sosyal Bilimler Enstitüsü'nün tarih ve Sayılı toplantısında oluşturulan jürimiz tarafından Lisansüstü Yönetmeliğinin 30.maddesi gereğince doktora tez sınavına alınmıştır.

Adayın kişisel çalışmaya dayanan tezini dakikalık süre içinde savunmasından sonra jüri üyelerince gerek tez konusu gerekse tezin dayanağı olan Anabilim dallarından sorulan sorulara verdiği cevaplar değerlendirilerek tezin,

BAŞARILI	<input type="radio"/>	OY BİRLİĞİ ile	<input type="radio"/>
DÜZELTME	<input type="radio"/>	OY ÇOKLUĞU	<input type="radio"/>
RED edilmesine	<input type="radio"/>	ile karar verilmiştir.	

Jüri teşkil edilmediği için sınav yapılamamıştır. ***
Öğrenci sınava gelmemiştir. **

* Bu halde adaya 6 ay süre verilir.
** Bu halde adayın kaydı silinir.
*** Bu halde sınav için yeni bir tarih belirlenir.

	Evet
Tez, burs, ödül veya teşvik programlarına (Tüba, Fullbright vb.) aday olabilir.	<input type="radio"/>
Tez, mevcut hali ile basılabilir.	<input type="radio"/>
Tez, gözden geçirildikten sonra basılabilir.	<input type="radio"/>
Tezin, basımı gerekliliği yoktur.	<input type="radio"/>

JÜRİ ÜYELERİ

İMZA

.....	<input type="checkbox"/> Başarılı	<input type="checkbox"/> Düzeltme	<input type="checkbox"/> Red
.....	<input type="checkbox"/> Başarılı	<input type="checkbox"/> Düzeltme	<input type="checkbox"/> Red
.....	<input type="checkbox"/> Başarılı	<input type="checkbox"/> Düzeltme	<input type="checkbox"/> Red
.....	<input type="checkbox"/> Başarılı	<input type="checkbox"/> Düzeltme	<input type="checkbox"/> Red
.....	<input type="checkbox"/> Başarılı	<input type="checkbox"/> Düzeltme	<input type="checkbox"/> Red

TEŞEKKÜR

Öncelikle çalıştığım tez konusunu seçmemde beni yönlendiren ve çalışmamım her aşamasında büyük emeği olan değerli hocam Doç.Dr.Kadir ERTAŞ'a, önerileriyle beni yönlendiren değerli hocalarım Prof.Dr.Levent ŞENYAY ve Prof.Dr.Serdar KORUKOĞLU'na, sorularımı yanıtlamaya daima vakit bulan sevgili hocam Yrd.Doç.Dr. Ali Kemal ŞEHİRLİOĞLU'na, Matlab konusundaki ve özellikle manevi desteğinden dolayı dostum İpek DEVECİ KOCAKOÇ'a, çalışmamım her aşamasında tüm sıkıntılarımı ve mutluluklarımı benimle paylaşan değerli aileme ve sevgili eşim Hakan'a ve en büyük moral kaynağıma, oğluma, teşekkürlerimi iletmeyi bir borç bilirim.

İstem KESER

ÖZET

Doktora Tezi

Çok Değişkenli İstatistiksel Boyut İndirgeme Yöntemi Olarak Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi Üzerine Bir Araştırma İstem KESER

Dokuz Eylül Üniversitesi

Sosyal Bilimleri Enstitüsü

Ekonometri Anabilim Dalı

Doktora Programı

Teknolojinin gelişimiyle artık büyük veri setleri kolaylıkla fonksiyonel veriler formunda taranabilir. Fonksiyonel veriler bir aralığın sonlu alt setinde değerlendirilen eğrilerden veya gözlenen fonksiyonlardan oluşur. Ramsay ve Silverman'ın öncülük ettiği Fonksiyonel Veri Analizi Yöntemleri bu tip verileri analiz etmek üzere geliştirilmiştir. Fonksiyonel Veri Analizi fonksiyonel verilere geleneksel veri analizlerinin bir uzantısıdır. Fonksiyonel Veri Analizindeki teknikler x_i ($i = 1, 2, \dots, N$) şeklinde belirtilen fonksiyonlardan ve onların türevlerinden oluşan örnekteki değişimi çalışmak üzere kullanılabilir. Pratikte bu fonksiyonlar sıklıkla ayrık noktalarda gözlenen verilere uygulanan düzgünleştirme proseslerinin bir sonucu olarak ortaya çıkarlar. Bu çalışmada da Splayn Düzgünleştirme Yöntemleri ele alınmıştır.

Elde edilen fonksiyonların sayısı ve incelenen saha genişledikçe bu fonksiyonları birlikte ele almak gittikçe güçleşmektedir. Bu durumda bu karmaşıklığı gidermek üzere kullanılan ve fonksiyonlar arasındaki değişimi sadece birkaç bileşenle açıklama gücüne sahip olan Fonksiyonel ve Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi bu tez çalışmasının kapsamını içermektedir.

Fonksiyonel Ana Bileşenler Analizi kullanım esnekliği bakımından kapsamlı bir uygulama alanına sahiptir. Bu çalışmada Fonksiyonel Veri

Analizinin İMKB de bir uygulamasına değinilmektedir. Burada İMKB 30 endeksi farklı bir bakış açısıyla incelenmiş ve piyasaların genel seyrini göstermede İMKB 30 endeksi yerine ortalama fonksiyonu yeni bir yaklaşım olarak ortaya konmuştur.

Anahtar Kelimeler: 1) Fonksiyonel Veri Analizi, 2) Fonksiyonel Ana Bileşenler Analizi, 3) Düzgünleştirme, 4) Pürüzlü Ceza Yöntemi, 5) Kübik Splayn.

ABSTRACT

Dissertation

A Study on Regularized Functional Principal Component Analysis

As a Statistical Dimension Reduction Technique

İstem KESER

Dokuz Eylül University
Institute of Social Science
Department of Econometrics

With modern technology, massive datasets can easily be scanned in the form of functional data. Functional data refer to data which consist of observed functions or curves evaluated at a finite subset of some interval. Functional data analysis techniques pioneered by Ramsay and Silverman are improved to analyses such type of data. Functional data analysis is an extension of traditional data analysis to functional data. Techniques in functional data analysis can be used to study the variation in a sample of functions $x_i, i=1, 2, \dots, N$, and their derivatives. In practice these functions are often a consequence of a preliminary smoothing process applied to discrete data. In this work, Spline Smoothing Methods are used.

As the number of functions and the number of fields are increasing, it would be difficult to handle the functions altogether. Regarding this fact, in order to overcome this complexity, in this work, we utilize Functional and Regularized Functional Principal Component Analyses that have the capability of explaining a high percentage of variation with only a few components.

Because of its flexibility of usage, functional data analysis has a comprehensive application area. In this dissertation an application in İMKB is given. Here İMKB 30 indice is examined from a different aspect and a new approach, the mean function, is proposed to demonstrate the general trend of the market.

Keywords: 1) Functional Data Analysis, 2) Functional Principal Component Analysis, 3) Smoothing, 4) Roughness Penalty Methods, 5) B-Spline.

**ÇOK DEĞİŞKENLİ İSTATİSTİKSEL BOYUT İNDİRGEME YÖNTEMİ
OLARAK DÜZGÜNLEŞTİRİLMİŞ FONKSİYONEL ANA BİLEŞENLER
ANALİZİ ÜZERİNE BİR ARAŞTIRMA**

YEMİN METNİ	II
TUTANAK	III
TEŞEKKÜR	IV
ÖZET	V
ABSTRACT	VI
İÇİNDEKİLER	IX
ŞEKİL VE TABLO LİSTESİ	XI
GİRİŞ	1

BİRİNCİ BÖLÜM

KESİKLİ VERİLERİ FONKSİYONEL VERİLERE DÖNÜŞTÜRME

1.1. FONKSİYONEL VERİLERİ DÜZGÜNLEŞTİRME	10
1.1.1. Baz Fonksiyon Yaklaşımı	10
1.1.1.1. Baz Fonksiyonlar	12
1.1.1.2. B-Splayn	15
1.1.2. Pürüzlü Ceza Yaklaşımı	25
1.1.2.1. Düzgünleştirme Parametresinin Belirlenmesi	36
1.2. FONKSİYONEL VERİ İÇİN TANIMLAYICI İSTATİSTİKLER	39
1.2.1. Ortalama Fonksiyonu	39
1.2.2. Varyans ve Kovaryans Fonksiyonları	40

İKİNCİ BÖLÜM

FONKSİYONEL ANA BİLEŞENLER ANALİZİ

2.1. ÇOK DEĞİŞKENLİ ANA BİLEŞENLER ANALİZİ	45
2.2. FONKSİYONEL ANA BİLEŞENLER ANALİZİ	48
2.2.1. Fonksiyonel Ana Bileşenlerin Elde Edilmesinde Baz Fonksiyon Yaklaşımı	51
2.3. DÜZGÜNLEŞTİRİLMİŞ FONKSİYONEL ANA BİLEŞENLER ANALİZİ	56

2.3.1. Düzgünleştirilmiş Fonksiyonel Ana Bileşenlerin Elde Edilmesinde Baz Fonksiyon Yaklaşımı	63
2.4. FONKSİYONEL ANA BİLEŞENLER ANALİZİNE LİTERATÜRDEN BAZI ÖRNEKLER	65

ÜÇÜNCÜ BÖLÜM

UYGULAMA

3.1. KULLANILAN PROGRAM	68
3.2. PROGRAMIN SONUÇLARI VE YORUMLAR	74
3.2.1. İMKB 30 Endeksi ve İMKB 30 da Yer Alan Şirketlerin Fonksiyonel Analizi	74
3.2.2. İMKB 30 Endeksinden Seçilen 11 Şirketin İncelenmesi	85
3.2.3. Düzgünleştirme Parametresinin Etkileri	97
SONUÇ	101
KAYNAKLAR	104
EKLER	113

ŞEKİL LİSTESİ

Şekil 1.1: Sıfırıncı Dereceden B-Splayn	18
Şekil 1.2: Birinci Dereceden B-Splayn	19
Şekil 1.3: İkinci Dereceden B-Splayn	20
Şekil 1.4: Üçüncü Dereceden B-Splayn	21
Şekil 1.5: Tekrarlı Düğüm Değerli Üçüncü Dereceden B-Splayn	23
Şekil 1.6: Kovaryans Yüzeyi	42
Şekil 3.1: B-Splayn Baz Fonksiyonlar	74
Şekil 3.2: Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar	75
Şekil 3.3: İMKB 30 Endeksi Birinci Türev Fonksiyonları	76
Şekil 3.4: Ortalama Fonksiyonu	77
Şekil 3.5: İMKB 30 Endeksi	78
Şekil 3.6: İMKB 30 Endeksi ile Ortalama Fonksiyonu Arasındaki Korelasyon	79
Şekil 3.7: Ortalama Türev Fonksiyonu	80
Şekil 3.8: Kovaryans Yüzeyi	81
Şekil 3.9 : Birinci Ana Bileşen Fonksiyonu	82
Şekil 3.10 : Ortalama Fonksiyonu ve Ana Bileşen Fonksiyonu Karşılaştırılması	83
Şekil 3.11: Ana Bileşen Skorlarının Dağılımı	85
Şekil 3.12: Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar	86
Şekil 3.13: Seçilen 11 Şirketin Ortalama Fonksiyonu	87
Şekil 3.14: Seçilen 11 Şirketin Ana Bileşen Fonksiyonu	88
Şekil 3.15: Seçilen 11 Şirket ve İMKB 30 Arasındaki Korelasyon	89
Şekil 3.16 : Seçilen 11 Şirket ve 30 Şirket Arasındaki Korelasyon	89
Şekil 3.17: Garanti Bankası Ekim Ayı Grafiği	90
Şekil 3.18: Ekim Ayı Birinci Türev Fonksiyonu	91
Şekil 3.19: Garanti Bankası Kasım Ayı Grafiği	92
Şekil 3.20 : Kasım Ayı Birinci Türev Fonksiyonu	92
Şekil 3.21: Garanti Bankası Aralık Ayı Grafiği	93

Şekil 3.22: Aralık Ayı Birinci Türev Fonksiyonu	93
Şekil 3.23 : Sabancı Holding Ekim Ayı Grafiği	94
Şekil 3.24 : Ekim Ayı Birinci Türev Fonksiyonu	95
Şekil 3.25 : Sabancı Holding Kasım Ayı Grafiği	95
Şekil 3.26 : Kasım Ayı Birinci Türev Fonksiyonu	96
Şekil 3.27 : Sabancı Holding Aralık Ayı Grafiği	96
Şekil 3.28 : Aralık Ayı Birinci Türev Fonksiyonu	97
Şekil 3.29 : $\lambda=1$ için Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar	98
Şekil 3.30 : $\lambda=1$ için Birinci Ana Bileşen Fonksiyonu	98
Şekil 3.31 : $\lambda=50$ için Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar	99
Şekil 3.32 : $\lambda=50$ için Birinci Ana Bileşen Fonksiyonu	99

TABLO LİSTESİ

Tablo 1.1: Varyans Kovaryans Değerleri	42
Tablo3.1: Ana Bileşen Skorları	84