

DÜZGÜNLEŞTİRİLMİŞ FONKSİYONEL ANA BİLEŞENLER ANALİZİ VE BİR UYGULAMA

İstem KÖYMEN KESER*

Kadir ERTAŞ**

ÖZET

Fonksiyonel Veri Analizi Yöntemleri, reel bir aralığın sonlu bir alt setinde değerlendirilen eğrilerden veya gözlenen fonksiyonlardan oluşan fonksiyonel verileri analiz etmek üzere geliştirilmiştir. Fonksiyonel Veri Analizi teknikleri, $x_i(t)$ ($i = 1, 2, \dots, N$) şeklinde belirtilen N tane reel fonksiyondan oluşan verilerdeki değişkenliğin veya bu fonksiyonların ilgilenilen dereceden türev fonksiyonlarının incelenerek yorumlanması için 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 gözlem noktaları arttıkça bu fonksiyonları birlikte ele almak ve yorumlayabilmek gittikçe güçleşmektedir. Bunun sonucu olarak ortaya çıkan karmaşıklığı gidermek üzere kullanılan ve fonksiyonlar arasındaki değişimi sadece birkaç fonksiyon bileşenle açıklama gücüne sahip olan Fonksiyonel ve Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi bu çalışmanın kapsamını oluşturmaktadır. Bu çalışmada Fonksiyonel Veri Analizinin İMKB de bir uygulamasına değinilmektedir. Burada İMKB 30 endeksi farklı bir çok değişkenli istatistiksel bakış açısıyla incelenmiştir.

Anahtar Sözcükler: Fonksiyonel Ana Bileşenler Analizi, Düzgünleştirme, Kübik Splayn, Pürüzlü Ceza Yaklaşımı.

1. Giriş

Çoğu istatistiksel analiz, örnek birimleri üzerinde bir veya birden çok özellik bakımından yapılan ölçümlerin sağladığı verilerin değerlendirilmesi ve yorumlanması ile uğraşır. X_1, X_2, \dots, X_N gibi bir rastgele örnek dikkate alınsın. Eğer bu örnekteki gözlemler, $x_i(t)$, $i=1, 2, \dots, N$ şeklinde N tane reel fonksiyondan oluşuyorsa bu verilere Fonksiyonel Veriler adı verilir. Bunun sonucu olarak da, $x_i(t)$, $i=1, 2, \dots, N$ reel fonksiyonları \mathcal{M} gibi ölçülebilir (measurable) örnek uzayında N tane noktayı tanımlar.

* Araş.Gör.Dr., Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler fakültesi, Ekonometri Bölümü, (istem.koymen@deu.edu.tr)

** Doç.Dr., Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler fakültesi, Ekonometri Bölümü, (kadir.ertas@deu.edu.tr)

Herhangi bir örnek birimi $x(t)$ gibi bir reel fonksiyon ile temsil edilir. Bu fonksiyonların sonsuz çeşitlilikte olabilmesi mümkün olduğundan, örnek birimlerinin uzayı (subject space) sonsuz boyutludur. $x_j(t)$ ve $x_k(t)$ reel fonksiyonları için, iç çarpım (inner product),

$$\langle x_j(t), x_k(t) \rangle = \int_a^b x_j(t)x_k(t)dt \quad (1.1)$$

şeklinde yazılabilir. Eğer $j=k$ ise bu iç çarpım $x_k(t)$ fonksiyonunun normunun karesini tanımlar. Eğer üzerinde durulan reel fonksiyonlar sadece sonlu norma sahip olanlarla sınırlandırılırsa, ortaya çıkan uzay Hilbert uzayı olarak bilinir. Hilbert uzaylarının en temel özelliklerinden birisi, bu uzayın herhangi bir elemanı, sayılabilir sayıdaki ortonormal fonksiyonların ağırlıklı bir toplamı olarak ifade edilebilir. Böylece burada Hilbert uzayı sonsuz boyutlu olmasına rağmen, en azından sayılabilir bir uzaydır (Ramsay, 1982).

Fonksiyonel veri analizinde geleneksel olarak, veri matrisi ya da p değişkenli rassal örnekte yer alan N tane gözlem aslında $x_i(t)$ ($i=1,2, \dots, N$) biçiminde reel sürekli fonksiyonlar olmasına rağmen, veriler genellikle birbirlerinden ayrıktaki noktalarda gözlemlenir. N örnek büyüklüğü (hacmi), n_i ise i .inci örnek birimi için yapılan ölçüm sayısı, t_{ij} i .inci örnek birimi için j .inci ölçümlemenin alındığı nokta, y_{ij} , y karakteristiğinin ya da değişkeninin t_{ij} noktasındaki değerini göstermek üzere, çok değişkenli rassal örnekteki i .inci birim için yapılan gözlemler, $y_{i1}, y_{i2}, \dots, y_{in-1}, y_{in}$ şeklinde verilebilir. N hacimlik rassal örneği 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 ve aynı gözlem noktalarında alınmak zorunluluğunda olmadığından bu değişebilirliği göstermek üzere n_i notasyonu kullanılmıştır.

2. Kesikli Verileri Fonksiyonel Verilere Dönüştürme

Fonksiyonel veri analizindeki ilk adım her bir fonksiyonun tanımlı olduğu $t \in [a,b]$ reel aralığında t 'nin tüm değerleri için elde edilebilmesi amacıyla, verilerin fonksiyonel forma dönüştürülmesidir. Bu şekilde elde edilen fonksiyonlardan genellikle fonksiyonel veri objeleri olarak bahsedilir. Bu

dönüşüm verileri farklı yollarla interpolate ederek (interpolation) veya düzgünleştirerek (smoothing) elde edilir.

Periyodik olmayan durumlar için, 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 durumda fonksiyonlar,

$$x(t) = \sum_{i=1}^K c_i B_{i,d} = \underline{\mathbf{c}}^T \underline{\mathbf{B}}(t) \quad (2.1)$$

şeklinde B-Splayn baz fonksiyonların doğrusal bir kombinasyonu olarak ifade edilebilirler. B-Splaynın derecesine bağlı olarak elde edilen fonksiyon, örneğin üçüncü dereceden B-Splayn için bir Kübik Splayn olur. Bu durumda Kübik Splaynın kendisi sürekli olduğu gibi, birinci ve ikinci türev fonksiyonları da sürekli dir. Bu bir Splayn Düzgünleştirme Yaklaşımı olarak bilinir. Burada $\underline{\mathbf{B}}(t)$ (Kx1) boyutlu B-Splayn baz fonksiyonlardan oluşan bir sütun vektörü ve $\underline{\mathbf{c}}$ yine (Kx1) boyutlu katsayılar vektörüdür. Bu çalışmada c_i , (i=1,2,...,K) katsayılarının tahmininde Pürüzlü Ceza Yaklaşımı kullanılmıştır. Ayrıntılı bilgi için Ramsay ve Silverman (1997: 58-62), Ramsay ve Silverman (2005: 81-109) ve Keser (2007) çalışmalarına başvurulabilir. Periyodik durumlar için de baz olarak Fourier serileri tercih edilir.

3. Fonksiyonel Yapıda Temel İstatistiklerin Tanımlanması

$x_1(t), x_2(t), \dots, x_N(t)$ şeklinde belirtilen N tane fonksiyon olduğu varsayıldığında klasik istatistikler benzer biçimde fonksiyonel verilere de uygulanır. Özetlemek gerekirse;

$$\text{Ortalama Fonksiyonu: } \bar{x}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t_j), \quad (3.1)$$

$$\text{Varyans Fonksiyonu: } \text{Var}_x(t) = \frac{1}{N-1} \sum_{i=1}^N (x_i(t) - \bar{x}(t))^2, \quad (3.2)$$

Kovaryans ve Korelasyon Fonksiyonları ise sırasıyla şu şekilde verilebilirler:

$$\text{Cov}_x(s, t) = \frac{1}{N-1} \sum_{i=1}^N \{x_i(s) - \bar{x}(s)\} \{x_i(t) - \bar{x}(t)\}, \quad s, t \in [a, b] \quad (3.3)$$

$$\text{Corr}_x(s,t) = \frac{\text{Cov}_x(s,t)}{\sqrt{\text{Var}_x(s)\text{Var}_x(t)}}, \quad s,t \in [a,b] \quad (3.4)$$

4. Fonksiyonel Ana Bileşenler Analizi

Fonksiyonel veri analizinde örnek birimlerinin 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 olacağından bir tek eğrinin bireysel davranışı diğer birçok eğrinin davranışları arasında kaybolur (Jones ve Rice, 1992). Ana Bileşenler Analizi bu tip karmaşık anakütlelerin yapısının anlaşılmasıyla ilgili oldukça genel bir probleme etkili bir çözüm sağlar. Veri karmaşıklığını gidermek üzere eğriler arasındaki değişimin önemli modlarını tanımlamak için Ana Bileşenler Analizinin (ABA) kullanımı önerilebilir. 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 (Ramsay ve Silverman, 1997: 85). Ana Bileşenler Analizi sistemdeki değişkenlik yapısına daha aydınlatıcı bir biçimde bakabilmeyi ve bu yapıyı yorumlayabilmeyi sağlar.

Fonksiyonel veriler için uygulanan Fonksiyonel Ana Bileşenler Analizinde (FABA) asıl amaç, çok değişkenli veriler için uygulanan Ana Bileşenler Analizi ile benzer olup, bu defa verilerdeki değişkenliği etkili bir biçimde tanımlayan birkaç ortogonal fonksiyonun elde edilmesidir. Aralarındaki temel fark γ_j biçiminde belirtilen ortogonal fonksiyonlar olan ana bileşen ağırlıkları (bunlar genelde harmonik olarak da adlandırılır) şimdi artık zamanın veya ilgili başka bir değişkenin fonksiyonlarıdır. Fonksiyonel Ana Bileşenler Analizinde; anabileşen fonksiyonu, ağırlık fonksiyonu, harmonik fonksiyonu terimleri eşanlamlı olarak kullanılmaktadır. Bu çalışmada verilen tüm ifadelerde yer alan bütün integraller $[a,b]$ sonlu reel aralığında belirli integrali temsil etmektedirler.

$x_i(t)$, ($i=1,2, \dots, N$) şeklinde fonksiyonel veri setine sahip olduğunu varsayalım. Fonksiyonel Ana Bileşenlerin ilk adımında ağırlık fonksiyonu γ_1 ,

$$\|\gamma_1\|^2 = \int \gamma_1(t)^2 dt = 1 \quad (4.1)$$

kısıtı altında, doğrusal bileşenin varyansı olan,

$$\text{Var}(Y_j) = \text{Var}\langle \gamma_j, x - E(x) \rangle = \iint \gamma_j(s) \text{Cov}(s,t) \gamma_j(t) ds dt \quad (4.2)$$

ifadesini maksimum yapacak biçimde belirlenir. Geri kalan ağırlık fonksiyonlarının hesaplanması için Klasik Ana Bileşenler Analizinde olduğu gibi Fonksiyonel Ana Bileşenler Analizinde de ağırlık fonksiyonlarının genel olarak,

$$\langle \gamma_j, \gamma_m \rangle = \int \gamma_j(t) \gamma_m(t) dt = 0 \quad (j \neq m) \quad (4.3)$$

şeklinde ifade edilebilen ortogonalite koşulunu da sağlaması gerekir. Her bir ağırlık fonksiyonunun eğrilerdeki değişimin en önemli modunu tanımlama görevi vardır ve burada her bir modun önceki adımlarda tanımlanan modlara ortogonal olması gerekir (Ramsay ve Silverman, 1997: 88). Ağırlık fonksiyonları her aşamada maksimum değişimi açıklayabilecek biçimde oluşturulan ortogonal baz fonksiyonlar setidir. Doğrusal bileşenin varyansının maksimum yapılması problemi Klasik Ana Bileşenler Analizinde olduğu gibi fonksiyonel veriler için de,

$$\int \text{Cov}(s, t) \gamma(t) dt = \psi \gamma(s) \quad (4.4)$$

şeklindeki özdeğer- özfonksiyon eşitliğinin çözümünü gerektirir. Burada ψ ilgili özdeğeri belirtmektedir.

5. Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi

Fonksiyonel Ana Bileşenler Analizi ile elde edilen özfonksiyonlar bir diğer deyişle ana bileşen ağırlıkları pürüzlülük (roughness) yapısına sahip olabilirler. 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 artmaktadır. Fonksiyonel Ana Bileşenler Analizini düzgünleştirme sadece klasik Ana Bileşenler Analizi ile elde edilen bileşenleri düzgünleştirmek anlamında 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 değil aynı zamanda pürüzlülüğünün (iniş çıkışlarının) de kontrol edilmesi gerekir. İki kez türevi alınabilen bir x eğrisinin pürüzlülüğünü ölçmenin en yaygın olarak kabul edilen bir yolu,

$$PEN_2(x) = \int (D^2 x(t))^2 dt = \|D^2 x\|^2 \quad (5.1)$$

şeklinde ikinci türevinin karesinin integralini almaktır (Eubank, 1999: 227). Bu ifade $x(t)$ deki toplam eğriselliği değerlendirir, veya bir diğer deyişle, $x(t)$ eğrisinin bir doğrudan sapma derecesidir. Bu yüzden, yüksek derecede pürüzlülüğe sahip fonksiyonların yüksek $PEN_2(x)$ değerleri sağlayacağı beklenir, çünkü ikinci türevleri ilgilenilen aralığın çoğunda büyük değerlere sahiptir (Ramsay ve Silverman, 1997: 59). Ayrıca, $\|D^2 x\|^2$, x fonksiyonunun pürüzlülüğünün hızının bir ölçüsü olarak da değerlendirilebilir (Leurgans vd., 1993). 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 'nin pürüzlülüğünü de dikkate alan,

$$\int \gamma_j(t)^2 dt + \lambda \int (D^2 \gamma_j(t))^2 dt = 1 \quad (5.2)$$

kısıtı ile yer değiştirir (Ramsay ve Silverman, 2002: 26). Bu aslında standart Sobolev iç çarpımlarının genelleştirilmiş halidir (Musayev ve Alp, 2000: 114). Bu durumda $\text{Var}(Y_j)$, (5.2) ifadesiyle verilen kısıta bölünerek Cezalı Ana Bileşen Varyansı (CABV),

$$\text{CABV} = \text{Var}(Y_j) = \frac{\iint \gamma_j(s) \text{Cov}(s, t) \gamma_j(t) ds dt}{\int \gamma_j(t)^2 dt + \lambda \int (D^2 \gamma_j(t))^2 dt} \quad (5.3)$$

şeklinde elde edilir (Silverman, 1996). λ parametresi düzgünleştirme parametresidir. Varyansın maksimize edilmesi ve γ 'nin pürüzlülüğü arasında bir denge sağlar. Düzgünleştirme parametresi λ sabit olarak alındığında, ana bileşen ağırlık fonksiyonu γ 'nin pürüzlülüğü arttıkça $PEN_2(\gamma)$ 'nin değeri artacağından, Cezalı Ana Bileşen Varyansı azalır. Ayrıca λ sıfıra yaklaşırken Cezalı Ana Bileşen Varyansı düzgünleştirilmemiş durumdaki varyansa döner. Bir diğer bakış açısıyla ise, λ 'nin değeri arttıkça, Cezalı Ana Bileşen Varyansı, γ 'in pürüzlülüğünden daha çok etkilenecektir. Düzgünleştirme parametresi ile ilgili ayrıntılı bilgi için Green ve Silverman (1994: 29), Silverman (1985), Ramsay ve Silverman (2002: 39), Simonoff (1996: 172), Ramsay ve Li (1998) ve Ulbricht (2004) çalışmalarına başvurulabilir.

Pürüzlülüğün analize dahil edilmesi ikinci, üçüncü ve daha yüksek dereceli düzgünleştirilmiş ana bileşenlere ilave kısıtlar eklemektedir. j.inci bileşen fonksiyonu (5.2) ile verilen kısıta bağlı olarak Cezalı Ana Bileşen Varyansını,

$$\int \gamma_j(t) \gamma_m(t) dt + \lambda \int (D^2 \gamma_j(t))(D^2 \gamma_m(t)) dt = 0 \quad j \neq m \quad (5.4)$$

ilave kısıtı altında maksimize eder.

Cezalı Ana Bileşen Varyansının maksimum yapılması problemi Klasik Ana Bileşenler Analizinde olduğu gibi fonksiyonel veriler için de,

$$\int \text{Cov}(s, t) \gamma(t) dt = \psi (I + \lambda D^4) \gamma(s) \quad (5.5)$$

şeklindeki özdeğer- özfonksiyon eşitliğinin çözümünü gerektirir.

6. 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 Gayrimenkul 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 Holding (SİSE), Türkiye Sinai 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 ayrı nokta da bir diğer ifadeyle 61 günde gözlenen kapanış fiyatları Splayn Düzgünleştirme ile sürekli bir fonksiyon haline dönüştürülmüş ve öncelikle oluşturulan bu 30 farklı fonksiyon, bir diğer deyişle Kübik Splayn 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 uygulanarak tüm fonksiyonlar birlikte

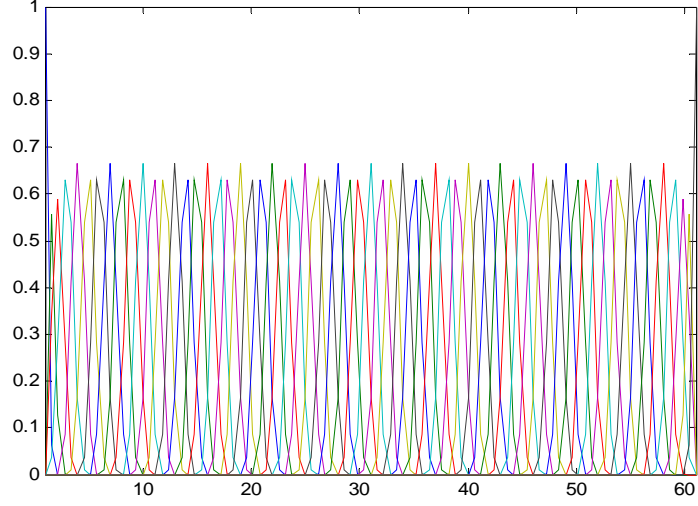
ele alındığında incelenmesi ve yorumlanması güç olan orijinal fonksiyonlar arasındaki değişim ortaya konulmaya çalışılmıştır. İMKB 30 da yer alan şirketlerin içerisinde farklı sektörlerden belirlenen iki tanesinin Ekim ayı bazında fonksiyonları ve türev fonksiyonları verilmiştir. Son olarak da farklı düzgünleştirme parametresi değerleri için fonksiyonlardaki ve ana bileşen fonksiyonlarındaki değişimler incelenmiştir.

Bu çalışmada baz fonksiyon sayısı B-Splayn teorisine göre (Keser, 2007) gözlem noktalarından daha fazladır. Bu durumda, Kübik Splaynın elde edilmesinde, eğrinin iniş çıkışını minimize eden ve pürüzlülük üzerinde sürekli bir kontrol sağlayan Pürüzlü Ceza Yöntemi kullanılmıştır. Hangi günlerin hangi tarihlere karşılık geldiği Tablo (6.1) de verilmektedir.

Tablo (6.1): İMKB-30 Endeks Şirketlerinin Kapanış Fiyatları İçin Dikkate Alınan Günler

Tarih	Gün	Tarih	Gün	Tarih	Gün	Tarih	Gün
02.10.2006	1	26.10.2006	16	16.11.2006	31	08.12.2006	46
03.10.2006	2	27.10.2006	17	20.11.2006	32	11.12.2006	47
04.10.2006	3	30.10.2006	18	21.11.2006	33	12.12.2006	48
05.10.2006	4	31.10.2006	19	22.11.2006	34	13.12.2006	49
06.10.2006	5	01.11.2006	20	23.11.2006	35	14.12.2006	50
09.10.2006	6	02.11.2006	21	24.11.2006	36	15.12.2006	51
10.10.2006	7	03.11.2006	22	27.11.2006	37	18.12.2006	52
11.10.2006	8	06.11.2006	23	28.11.2006	38	19.12.2006	53
12.10.2006	9	07.11.2006	24	29.11.2006	39	20.12.2006	54
13.10.2006	10	08.11.2006	25	30.11.2006	40	21.12.2006	55
16.10.2006	11	09.11.2006	26	01.12.2006	41	22.12.2006	56
17.10.2006	12	10.11.2006	27	04.12.2006	42	25.12.2006	57
18.10.2006	13	13.11.2006	28	05.12.2006	43	26.12.2006	58
19.10.2006	14	14.11.2006	29	06.12.2006	44	27.12.2006	59
20.10.2006	15	15.11.2006	30	07.12.2006	45	28.12.2006	60
						29.12.2006	61

Çalışmada birinci adım B-Splayn baz fonksiyonları elde etmektir.



Şekil (6.1): B-Splayn Baz Fonksiyonlar

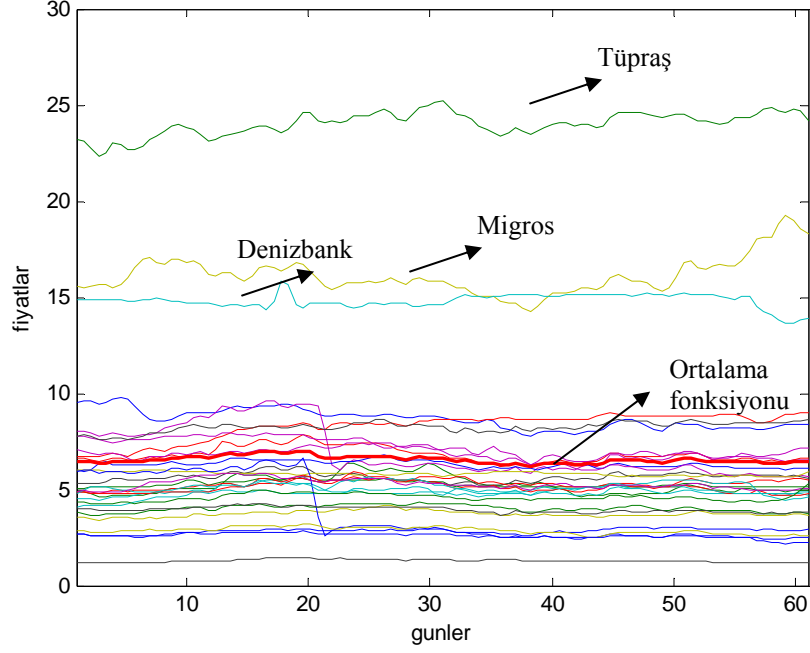
Ramsay ve Li'nin (1998) "curve registration" isimli çalışmasında uygulamalarla ilgili kapsamlı bir araştırma sonucu düzgünleştirme parametresi λ için 10^{-4} , 10^{-3} ve 10^{-2} değerlerinin iyi çalıştığı gözlenmiştir.

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

$$\lambda=10^{-4} [\text{tr} (\mathbf{B}^T \mathbf{B}) / \text{tr} \mathbf{R}] \quad (6.1)$$

şeklinde kullanılması önerilmiştir.

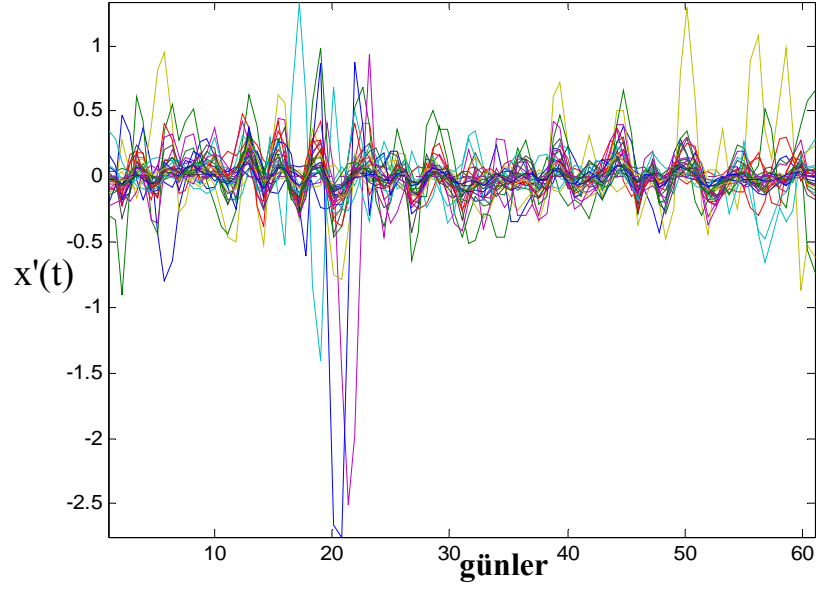
Bu çalışmada, düzgünleştirme parametresinin değeri Ulbricht'in (2004) yaklaşımına göre $\lambda=10^{-4} \times 0.1346$ olarak alınmıştır. Burada \mathbf{B} $B_i(t_j)$ baz fonksiyon değerlerinden oluşan $(n \times K)$ boyutlu bir reel matrisi, ve \mathbf{R} $(K \times K)$ boyutlu reel simetrik bir pürüzlü ceza matrisini göstermektedir(Keser, 2007).



Şekil (6.2): Pürüzlü Ceza Yöntemine göre Oluşturulan Fonksiyonlar

Şekil (6.2) den 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 seyretmektedirler. Bu şirketlerin işlem gördüğü fiyatlar da diğer şirketlerin fiyatlarına yakın olsaydı, bunların fonksiyonları da diğer fonksiyonların arasında yer alacaktı ve her zaman karşılaşılan genel bir zorluk olan yorumlama güçlükleri bu şirketler için de geçerli olacaktı. Burada ortalama eğrisi belirgin olması açısından diğerlerine göre daha kalın olarak çizdirilmiştir. 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 yorum yapılabilmesinin zorlaşacağı açıkça görülmektedir.

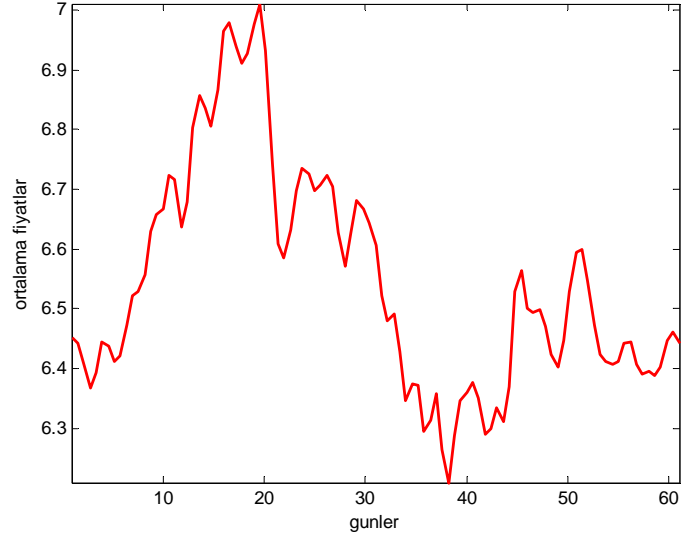
Ayrıca fonksiyonların zamana göre birinci türevleri incelenmiştir ve bunlar Şekil (6.3) de verilmektedir. Zamana göre birinci türev fonksiyonları, orijinal fonksiyonları oluşturan kapanış fiyatlarının artış veya azalış hızlarını yansıtmaktadır.



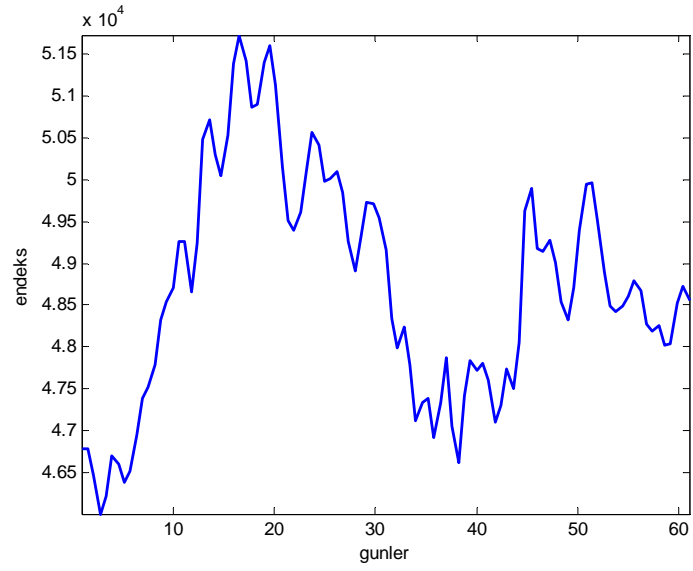
Şekil (6.3): İMKB 30 Endeksi Şirketleri İçin Birinci Türev Fonksiyonları

Şekil (6.2) 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 ayrıca Ekim ayı bazında yapılmıştır.

B-Splaynlar, düzgünleştirilmiş fonksiyonlar ve birinci türev fonksiyonları oluşturulduktan sonra fonksiyonel verilerin ortalama fonksiyonu ayrı olarak çizdirilmiştir ve Şekil (6.4) de verilmiştir.



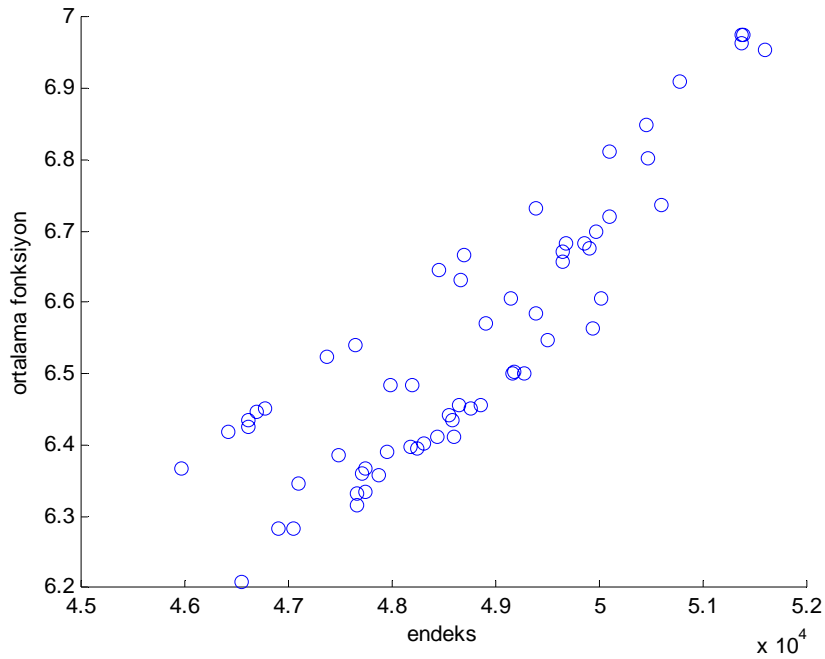
Şekil (6.4): Ortalama Fonksiyonu



Şekil (6.5) : İMKB 30 Endeksi

Şekil (6.5) de verilen İMKB-30 endeks grafiğinin ortalama fonksiyonu ile bir kıyaslaması yapıldığında, ortalama fonksiyonunun piyasanın genel seyrini göstermedeki yeterliliği açık bir şekilde ortaya konulmuş olmaktadır.

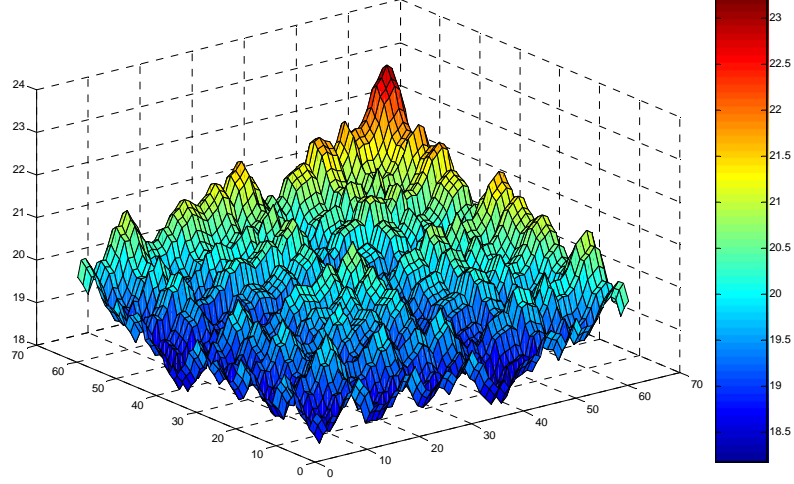
Şekil (6.4) ve Şekil (6.5) birlikte ele alındığında iki şeklin birbirine çok büyük benzerlikler gösterdiği, aynı dönemlerde artışlara ve azalışlara rastlandığı yine İMKB 30 endeks fonksiyonunun da ortalama fonksiyonu gibi 3 temel bölgeye ayrıldığı görülebilmektedir. Ortalama fonksiyonu ile İMKB 30 endeksi arasındaki korelasyon oldukça yüksektir ve Şekil (6.6) ile verilmiştir.



Şekil (6.6) : 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.

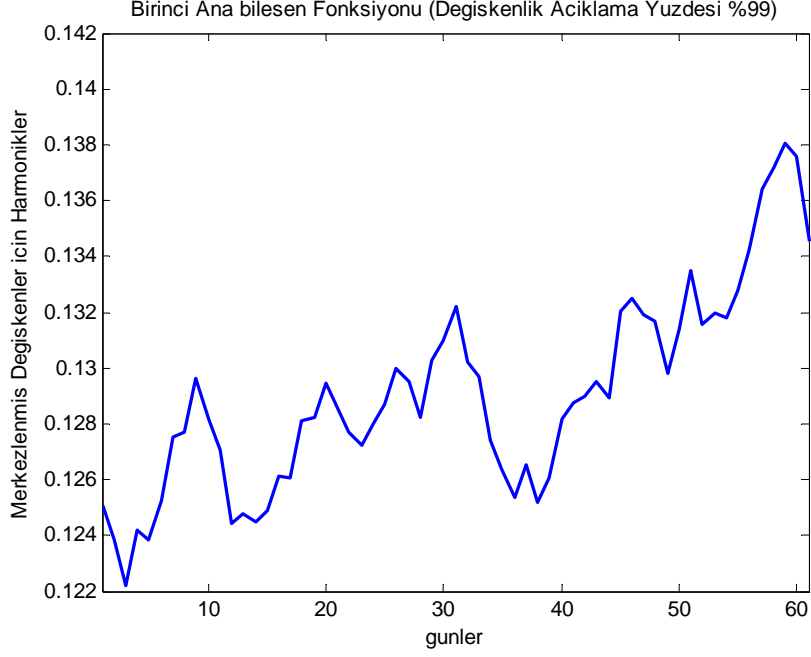
B-Splaynlar, düzgünleştirilmiş fonksiyonlar, birinci türev fonksiyonları, ortalama fonksiyonu ve İMKB 30 endeksi karşılaştırılması verildikten sonra ilgili kovaryans yüzeyi Şekil (6.7) de olduğu gibi elde edilmiştir.



Şekil (6.7): Kovaryans Yüzeyi

Şekil (6.7) ile verilen kovaryans yüzeyi 61 ayrı noktada 30 birim için inceleme yapıldığından dolayı oldukça karmaşıktır ve farklı t zamanlarında gözlemlerin birlikte değişimlerinin incelenmesi güçtür. Burada güç de olsa 55.inci günlerden itibaren kovaryanslarda ve varyanslarda bir artma olduğu en koyu uç bölgeden seçilebilmektedir, ancak bu durum değişken sayısı arttıkça daha da karmaşıklaşmakta ve yüzey üzerindeki yükseklikler yorumlanamaz hale gelmektedir.

Bireysel fonksiyonlar, kovaryans yüzeyi, ortalama fonksiyonu ve İMKB 30 endeksi karşılaştırması verildikten sonra, orijinal fonksiyonlar arasındaki değişkenlik yapısını araştırmaya ve yorumlamaya yönelik olarak ana bileşen fonksiyonu aşağıdaki biçimde elde edilmiştir.

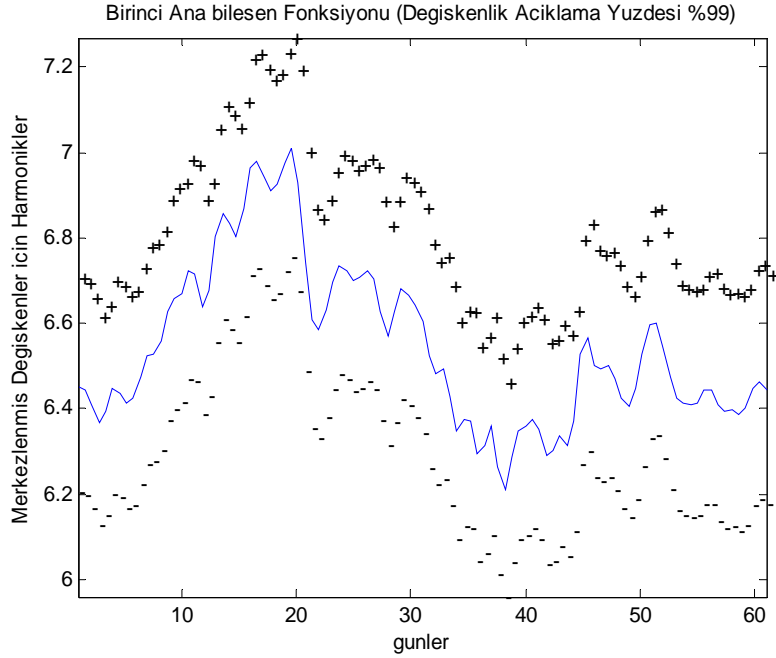


Şekil (6.8): Birinci Ana Bileşen Fonksiyonu

Bu fonksiyon incelediğinde, ardışık inişler ve çıkışlar birbirini izleyen iki rassal değişken (burada günlük kapanış fiyatları) arasındaki kovaryansın yönü tarafından belirlenir. Bu artış yada azalışların büyüklüğü de kovaryansın mutlak değeri tarafından belirlenir, ardışık olarak iki değişken arasında ters ilişki varsa, fonksiyon x eksenine göre pozisyon değiştirir. Ayrıca, noktaların (günlük kapanış fiyatlarının) x eksenine olan uzaklıkları ise, o değişkenin varyansının büyüklüğü ile doğrudan ilişkilidir. Bilindiği gibi, Cauchy-Schwarz eşitsizliği gereği, ardışık değişkenlerin varyanslarının yüksekliği bunlar arasındaki kovaryansın yüksek bir değer alabilmesini mümkün kılmaktadır. Burada şekilden de anlaşıldığı gibi, tüm ardışık kovaryans değerleri pozitiftir. Fonksiyonda x ekseninden uzaklıklar (yani belirli bir gün için harmonik değeri) varyanslarla ilgili olduğundan bu şekilden özellikle 55.inci günlerden sonra, Aralık ayının sonlarına doğru, fiyatların değişkenliğinde bir artış olduğu ve zaman noktalarındaki fiyatların ardışık olarak birbirleriyle pozitif korelasyonlu oldukları gözlemlenebilmektedir. Şekil (6.7) de verilen ilgili kovaryans yüzeyinin de, değişken sayısı ve incelenen örnek birimi sayısı arttıkça birlikte değişimi ve değişkenliği göstermesi giderek zorlaşmaktadır ki, bu çalışmada

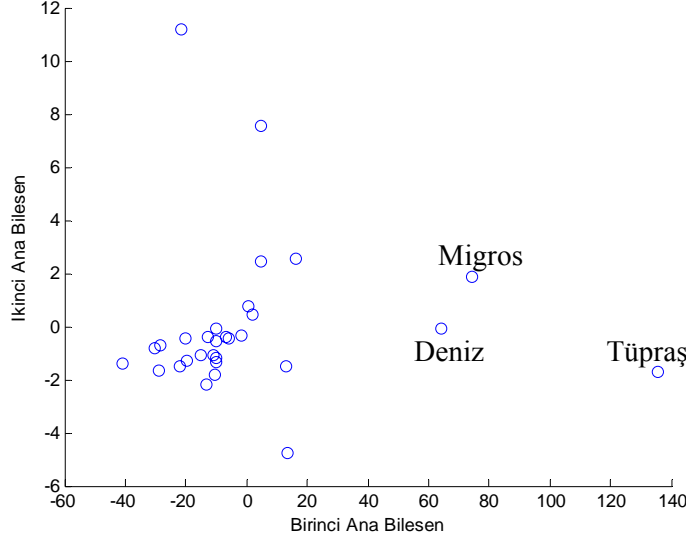
bile iki şekil kıyaslandığında ana bileşen fonksiyonu çok daha rahatlıkla yorumlanabilmektedir. Tüm bunlara ilaveten ana bileşen fonksiyonu sistemdeki toplam değişkenliğin %99 u gibi çok yüksek bir değişkenlik açıklama gücüne sahiptir.

Ana bileşen fonksiyonu incelendikten sonra özellikle ana bileşen fonksiyonunun yorumlanmasının güç olduğu bazı durumlarda alternatif yardımcı yöntemler olarak kullanılan ortalama fonksiyonuna ana bileşen fonksiyonlarının bir çarpanının eklenmesi ve çıkarılmasının etkileri Şekil (6.9) da ve ana bileşen skorlarının dağılımı Şekil (6.10) da verilmektedir.



Şekil (6.9) : Ortalama Fonksiyon ile Ana Bileşen Fonksiyonunun Karşılaştırılması

Şekil (6.9) daki + ve - noktalar ortalama fonksiyonuna 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 (6.9) dan da, 55.inci günlerden itibaren sapma daha net bir biçimde görülebilmektedir.



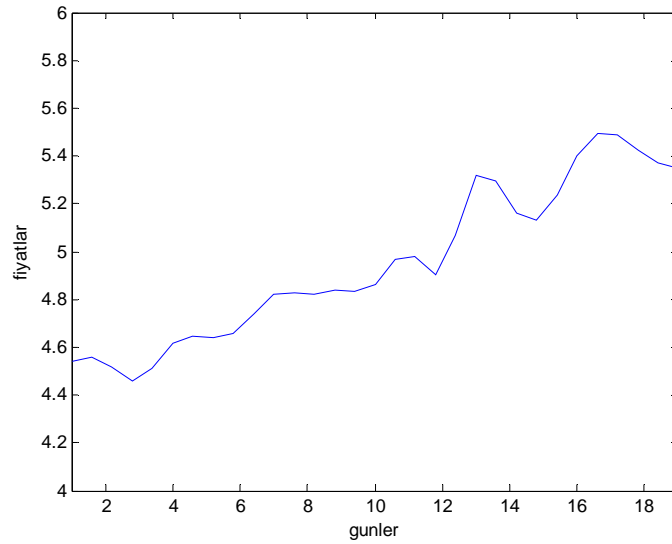
Şekil (6.10): Ana Bileşen Skorlarının Dağılımı

Ana Bileşen skorlarının incelenebilmesi için, skor değerleri Tablo (6.2) de verilmektedir. Ana bileşen skorları özellikle %99 varyans açıklayabilme gücüne sahip birinci bileşen açısından incelendiğinde, Şekil (6.10) dan, 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 diğer şirketlere oranla daha uzak bir konumda yer almalarının nedeni Şekil (6.2) de de görülebildiği gibi diğerlerine oranla bu şirketlerin daha yüksek birim fiyatlara sahip olmalarıdır.

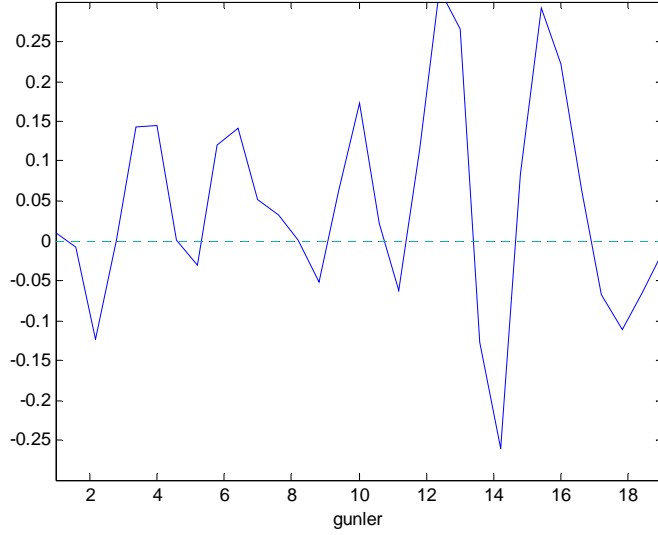
Tablo (6.2): Ana Bileşen Skorları

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
DENİZ	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

Daha sonra İMKB 30 da yer alan şirketlerden Garanti Bankası ve Sabancı Holding şirketlerinin hisse senetleri seçilerek bu fonksiyonlar ve bunların türev fonksiyonları Ekim ayı bazında incelenmişler, türev fonksiyonlarının yoruma olan katkıları da gösterilmeye çalışılmıştır. İlk olarak Garanti Bankası ele alınmıştır.



Şekil (6.11): Garanti Bankası Ekim Ayı Grafiği

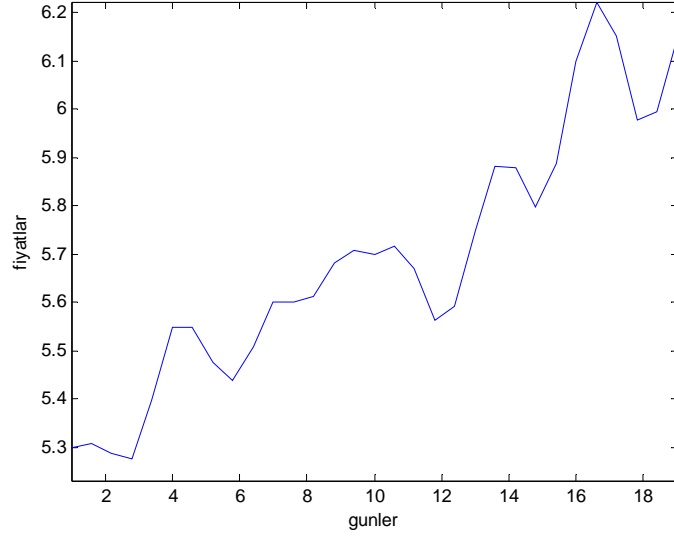


Şekil (6.12): Garanti Bankası Ekim Ayı BirinciTürev Fonksiyonu

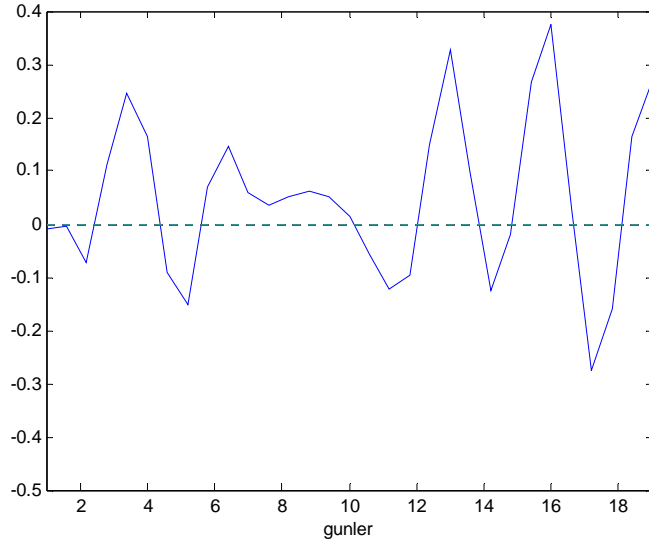
Şirketin fiyatının günlere göre ani deęişim oranı (türev) fonksiyonu incelendiğinde fiyatlardaki pozitif ya da negatif deęişimin hızı daha rahat bir şekilde görülebilmektedir. Şekil (6.11)'de verilen Ekim ayı grafięi ve Şekil (6.12) ile verilen türev fonksiyonu karşılaştırıldığında, orijinal fonksiyonda 12.ci güne kadar güçlükle seçilebilen, çok net görülemeyen artış ve azalışlar türev fonksiyonun incelenmesiyle daha rahatlıkla yakalanabilmektedir.

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

İstem Köymen Keser-Kadir Ertaş



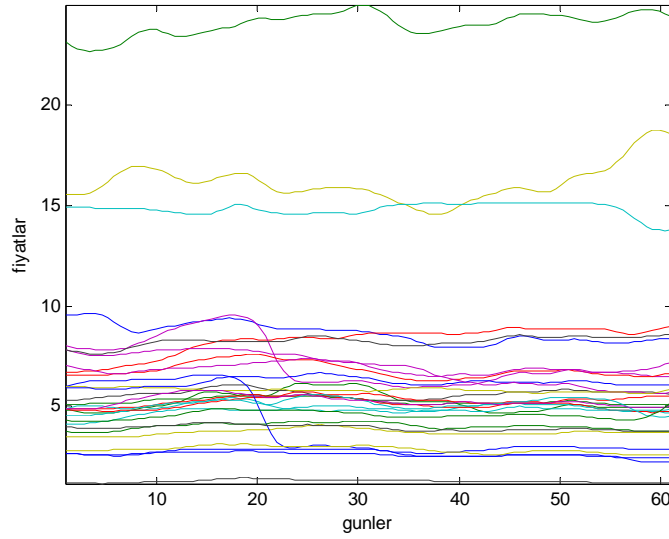
Şekil (6.13) : Sabancı Holding Ekim Ayı Grafiği



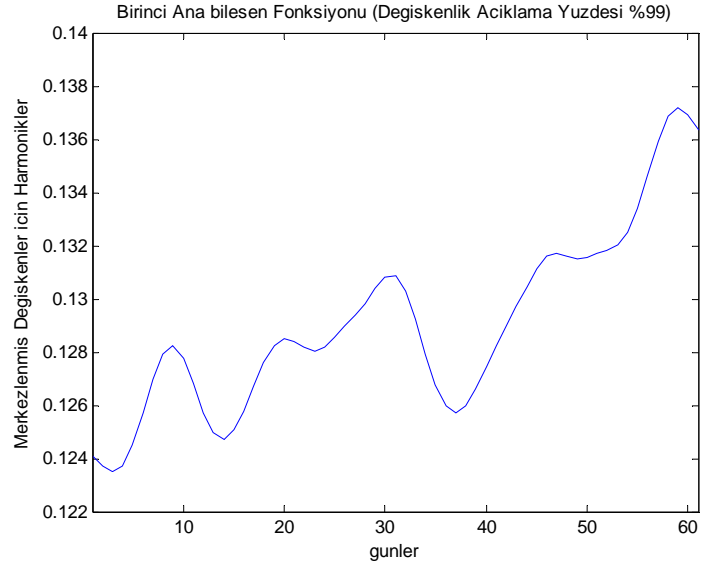
Şekil (6.14) : Ekim Ayı Birinci Türev Fonksiyonu

Şekil (6.13) incelendiğinde fonksiyondaki iniş ve çıkışlar Garanti Bankasına oranla daha kolaylıkla görülebilir durumdadır. Bununla birlikte, türev fonksiyonunun incelenebilmesi bu iniş ve çıkışları çok daha rahat yakalanabilir hale getirmektedir. Örneğin 17. gün civarındaki iniş, fonksiyondan da görülebilmesine karşın 15. gün civarındaki ani iniş türev fonksiyonunda daha net görülebilmekte ve bu iki inişin boyutu x-eksenine göre daha rahatlıkla kıyaslanabilmektedir.

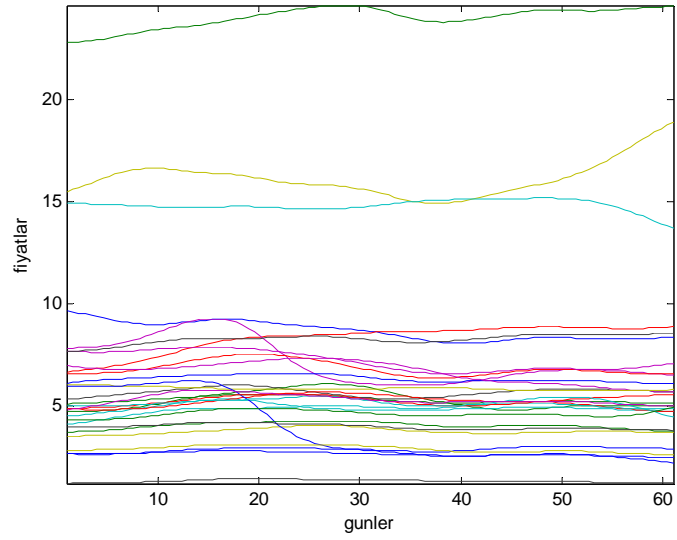
Son olarak Fonksiyonel Veri Analizinde önemli bir yere sahip olan düzgünleştirme parametresinin etkisi incelenmektedir. Şimdiye kadar tüm analizler Ulbricht'in (2004) yaklaşımına göre yapılmıştır. Düzgünleştirme parametresinin etkisinin görülebilmesi amacı ile her bir örnek birimi için fonksiyonların, ortalama fonksiyonunun ve birinci ana bileşen ağırlık fonksiyonlarının çizimleri iki farklı düzgünleştirme parametresi değeri için tekrarlanmıştır. Bu çalışmada sırasıyla $\lambda=1$ ve $\lambda= 50$ değerleri ele alınmıştır.



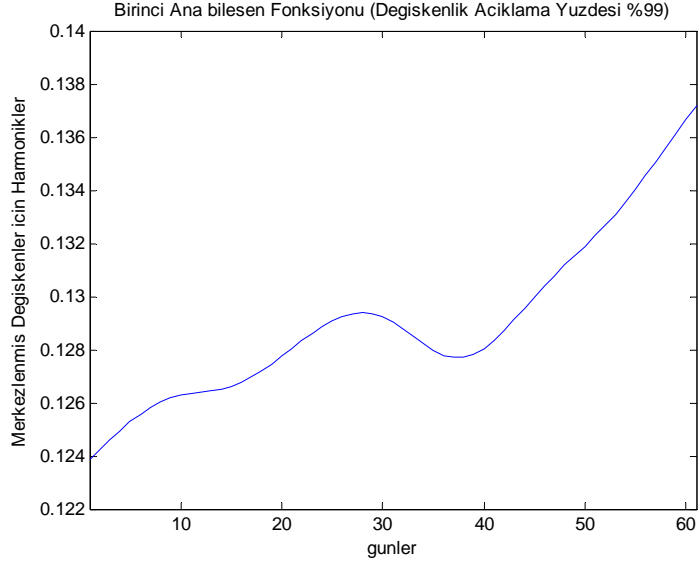
Şekil (6.15) : $\lambda=1$ için Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar



Şekil (6.16) : $\lambda=1$ için Birinci Ana Bileşen Fonksiyonu



Şekil (6.17): $\lambda= 50$ için Pürüzlü Ceza Yöntemine Göre Oluşturulan Fonksiyonlar



Şekil(6.18): $\lambda=50$ 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, verilerin yorumlanmasında 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.

7. Sonuç

Gittikçe ilerleyen bilgisayar teknolojisindeki gelişmeyle birlikte elde edilen verilerin analizi için klasik istatistiksel yöntemlerin yetersiz kalması, fonksiyonel veri analizinin düzensiz örneklenen fonksiyonlarla ve eksik verilerle (missing data) uğraşılmasına imkan vermesi, interpolasyon ve düzgünleştirme ile oluşturulan orijinal fonksiyonların arzu edilen dereceden türevlerinin de incelenebilmesi gibi görsel olarak da çoğu konuda

araştırmacılara yardımcı olması ve bu açıdan veri analizine yeni bir bakış açısı getirmesi Fonksiyonel Veri Analizinin öneminin artmasına sebep olmuştur. 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 incelenebilmesini olanaklı hale getirmesi ve bir çok görünmeyeni ortaya çıkarmadaki yeteneğidir.

Yapılan uygulama ile ise, ortalama fonksiyonunun İMKB 30 endeksine bir alternatif olabileceği gösterilmiştir. Türev fonksiyonlarının incelenmesi ile fonksiyonlarda 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 sayesinde 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 noktalarındaki hisse senedi kapanış fiyatlarının ardışık olarak birbirleriyle pozitif korelasyonlu bulunduğu gözlemlenebilmektedir.

ABSTRACT

REGULARIZED FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS AND AN APPLICATION ON THE SHARE PRICES OF THE COMPANIES BELONGING TO THE ISE-30 INDEX

The objective of the Functional Data Analysis techniques is to study such type of data which consist of observed functions or curves evaluated at a finite subset of some real interval. Techniques in Functional Data Analysis can be used to study the variation in a random sample of real functions, $x_i(t)$, $i=1, 2, \dots, N$ and their derivatives. In practice, these functions are often a consequence of a preliminary smoothing process applied to discrete data and in this work, Spline Smoothing Methods are used. As the number of functions and the number of observation points increases, it would be difficult to handle the functions altogether. In order to overcome this complexity, we utilize

Functional and Regularized Functional Principal Component Analyses where a high percentage of total variation could be accounted for with only a few component functions. Finally, an application on the daily closing data for the share prices of the companies belonging to the ISE-30 index is also given.

Keywords: Functional Data Analysis, Smoothing, Cubic Spline, Roughness Penalty Approach.

KAYNAKÇA

- EUBANK, R.L. (1999), *Nonparametric Regression and Spline Smoothing*, USA:Marcel Dekker.
- GREEN, P.J. ve B.W. SILVERMAN (1994), *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, London: Chapman & Hall.
- JONES, M. C. ve J. A. RICE (1992), “Displaying The Important Features Of Large Collections Of Similar Curves”, *The American Statistician*, 46(2), 140-145.
- KESER, İ. (2007), *Çok Değişkenli İstatistiksel Boyut İndirgeme Yöntemi Olarak Düzgünleştirilmiş Fonksiyonel Ana Bileşenler Analizi Üzerine Bir Araştırma*,Yayınlanmamış Doktora Tezi, Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, İzmir.
- LEURGANS, S.E., MOYEED, R.A ve SILVERMAN B.W. (1993), “Canonical Correlation Analysis when the Data are Curves”, *Journal of the Royal Statistical Society: Series B.*, 55(3), 725-740.
- MUSAYEV B. ve M. ALP (2000), *Fonksiyonel Analiz*, Kütahya : Balcı Yayınları.
- RAMSAY, J.O. (1982), “When The Data Are Functions” *Psychometrica*, 47(4),379-396.
- RAMSAY, J.O. ve B.W. SILVERMAN (1997), *Functional Data Analysis*, New York : Springer – Verlag.
- RAMSAY, J. O. ve X. LI (1998), “Curve Registration”, *Journal of the Royal Statistical Society: Series B*, 60(2), 351-363.
- RAMSAY, J.O. ve B.W. SILVERMAN (2002), *Applied Functional Data Analysis: Methods and Case Studies*, New-York: Springer – Verlag.
- RAMSAY J.O. ve SILVERMAN B.W. (2005), *Functional Data Analysis. Second Edition*, USA : Springer .

- 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), 1-52.
- SILVERMAN, B.W. (1996), “Smoothed Functional Principal Component Analysis By Choice Of Norm”, *The Annals of Statistics*, 24(1), 1-24.
- SIMONOFF, J.S. (1996), *Smoothing Methods in Statistics*, New- York: Springer- Verlag.
- ULBRICHT, J. (2004), *Representing Functional Data as Smooth Functions*, A Master Thesis, Institute of Statistics and Econometrics, Humboldt University, Berlin.