Ege Akademik Bakış / Ege Academic Review
8 (2) 2008: 915-928

KARŞILAŞTIRMALI OLARAK FONKSİYONEL ANA BİLEŞENLER ANALİZİ VE GSYİH VERİLERİNİN İNCELENMESİ

FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS: INVESTIGATION OF GDP DATA

Araş.Gör. Dr. İstem Köymen KESER, Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, istem.koymen@deu.edu.tr

ÖZET

Bu çalışmada öncelikle Ana Bileşenler ve Fonksiyonel Ana Bileşenler Analizi karşılaştırmalı olarak ele alınmış ve ekonomik gelişmenin bir göstergesi olan Gayri Safi Yurt İçi Hasıla (GSYİH) verileri Fonksiyonel Ana Bileşenler Analizi ile incelenerek verileri fonksiyonel açıdan ele almanın avantajları sunulmuştur. Fonksiyonel Ana Bileşenler Analizi ile 1987-2001 yılları arasında ülkemizdeki 7 bölgenin GSYİH verilerinin değişkenlik yapısı %99 gibi çok yüksek bir varyans açıklama yüzdesine sahip birinci ana bileşen fonksiyonuyla ortaya konulmuştur. Bu ana bileşen fonksiyonun yardımıyla GSYİH açısından bölgeler arasındaki değişkenliğin 1996 yılından sonra bir artışa geçtiği, ancak 2000 yılının ortalarından sonra da tekrar azalmaya başladığı tespit edilmiştir.

Anahtar Kelimeler: Fonksiyonel Veri Analizi, Fonksiyonel Ana Bileşenler Analizi, Ana Bileşenler Analizi

ABSTRACT

In this study, Principal Components and Functional Principal Components Analyses discussed comparatively. Gross Domestic Product (GDP) data is analyzed by Functional Principal Component Analysis and the advantages of dealing data in a functional way is presented. GDP data (between 1987 and 2001) for seven regions of Turkey is studied by Functional Principal Component Analysis, it is found that %99 of the variation structure can be explained by the first principal component function. It is also revealed that the variation between regions began to increase after the year 1996. However, it began to decrease rapidly after year 2000.

Keywords: Functional Data Analysis, Functional Principal Component Analysis, Principal Component Analysis

1. GİRİŞ

Her geçen gün ilerleme hızına ulaşmakta zorluk çekilen teknolojiyle birlikte artan veri toplama kapasiteleri sayesinde verileri klasik tek değişkenli ve çok değişkenli yöntemleri kullanarak analiz etmeye ve yorumlamaya çalışmak giderek yetersiz kalmaya başlamıştır. İncelenen birey veya obje ve inceleme yapılan değişken sayısı arttıkça gözlemleri birbiri ardına gelen sayı noktaları halinde incelemektense incelenen her bir birey veya obje için bireysel girişler olarak ele almak çok çeşitli avantajlar sağlamaktadır. Bireysel girişler olarak ele almayı başarmanın en kolay ve en avantajlı yolu her bir bireyin veya objenin bir fonksiyonel gözlem olarak incelenmesidir. Böylece, örneğin 6 birey için 100 farklı noktada inceleme yapıldığında, elimizde 600 farklı noktanın veya her bir birey için 100 farklı noktanın incelenmesi problemi yerini 6 farklı fonksiyonun incelenmesine bırakır. Ayrık noktalarda gözlenen verileri fonksiyonel verilere döndürmenin bilinen en pratik yolu interpolasyon (interpolation) veya düzgünleştirme (smoothing) dir. Eğer gözlemlenen değerlerin hatasız oldukları varsayılırsa, fonksiyonları oluşturma süreci interpolasyon yöntemi ile yapılır. Fakat verilerde ortadan kaldırılması gereken bazı gözlemsel hatalar varsa, örneğin, verilerin oluşturulması deneysel bir sürece dayanıyorsa, kesikli verilerden fonksiyonel verilere yapılan bu dönüşüm süreci, **düzgünleştirme** olarak adlandırılır(Ramsay ve Silverman,2005; 11).

Fonksiyonel verileri incelemek için metodlar "Fonksiyonel Veri Analizi" terimi ile adlandırılır(Ramsay ve Dalzell ,1991). Fonksiyonel Veri Analizinde, veri matrisi ya da p değişkenli şans örneğinde yer alan i.inci gözlem aslında $x_i(t)$ (i=1,2, ..., N) biçiminde reel bir sürekli fonksiyon olmasına rağmen, veriler pratikte genellikle birbirlerinden ayrık noktalarda gözlemlenir. Fonksiyonel Veri Analizinde kullanılan genel notasyonda, N, örnek büyüklüğünü (hacmi), n_i, i.inci örnek bireyi için yapılan ölçüm sayısını, t_{ij}, i.inci örnek bireyi için j.inci ölçümlemenin alındığı noktayı, y_{ij}, y karakteristiğinin ya da değişkeninin t_{ij} noktasındaki değerini göstermektedir. i.inci birey için yapılan gözlemler, y_{i1}, y_{i2}, ... , y_{in-1}, y_{in} şeklinde verilebilir. N hacimlik şans örneğini oluşturan veri seti ise,

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

şeklinde özetlenebilir. Burada her bir gözlem için inceleme yapılan nokta sayısı eşit ve aynı noktalarda alınmak zorunluluğunda olmadığından bu değişebilirliği göstermek üzere n_i notasyonu kullanılmıştır. Klasik çok değişkenli kavramda ise üzerinde çalışılan ölçümler klasik olarak

$$y_{ij}, \qquad i=1, 2, \ldots, N, \quad j = 1, 2, \ldots, p$$

biçiminde belirtilen bir veri matrisidir.
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, … , n_i) bir fonksiyonel gözlemden diğerine değişebildiğinden dolayı bu durumda doğrudan çok değişkenli yöntemleri uygulamak mantıklı veya mümkün olmayabilir.

Bir ilave nokta olarak, 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 Analizi tekniklerinde ise, vurgu yörüngelerde ve şekillerdedir.

Gözlemlerin fonksiyonlar olarak ele alınmasının çok önemli bir avantajı da oluşturulan fonksiyonların arzu edilen dereceden türevlerinin incelenebilmesi ve ayrıca interpolasyon yoluyla kayıp verilerle baş edilebilmesidir. Fonksiyonların arzu edilen dereceden türevlerinin alınabilmesi özelliği fonksiyonların Splaynlar aracılığı ile oluşturulması sayesinde sağlanır.

2. ANA BİLEŞENLER VE FONKSİYONEL ANA BİLEŞENLER ANALİZİ

Hem Ana Bileşenler (ABA) ve hem de Fonksiyonel Ana Bileşenler Analizinde (FABA) amaç, bireyler veya objelerin ve ölçüm alınan noktaların sayısı arttıkça oluşan kompleks populasyonların yapısının anlaşılmasıyla ilgili oldukça genel bir probleme etkili bir çözüm sağlamaktır. Veri karmaşıklığını gidermek üzere bireyler veya objeler arasındaki değişimin önemli modlarını tanımlamak için bir boyut indirgeme yöntemi olan Ana Bileşenler ve Fonksiyonel Ana Bileşenler Analizinin kullanımı önerilebilir. Genel anlamda 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 de ele alınan anahtar tekniktir.

Bir diğer bakış açısıyla, 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 ve Silverman, 1997; 85) ve temel fikri, varyans-kovaryans yapısını orijinal değişkenlerin maksimum varyansa sahip doğrusal kombinasyonu aracılığı ile açıklamaya çalışarak veri setinin boyutunu indirgemektir (Mardia, 1989; 213).

Klasik Ana Bileşenler Analizinde $\sum$, varyans-kovaryans matrisi

ve $\boldsymbol{\Gamma}^T \sum \boldsymbol{\Gamma} = \boldsymbol{\Psi}$ olmak üzere,
(2.1)

$\underline{\mathbf{x}}$ şans vektörünün j.inci ana bileşeni,

$$Y_j = \underline{\gamma}_j^T (\underline{\mathbf{x}} - \underline{\mu}) \qquad (j=1,2, \dots ,p)$$
(2.2)

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

Ana Bileşenler Analizinde j.inci doğrusal bileşenin varyansı olan $V(Y_j)$,

$$V(Y_j) = \underline{\gamma}_j^T \sum \underline{\gamma}_j = \psi_j$$
(2.3)

şeklinde,

1) $\left\langle \underline{\gamma}_j , \underline{\gamma}_m \right\rangle = \underline{\gamma}_j^T \underline{\gamma}_m = 1 \qquad$ (j=m ise ; j,m= 1,2, … , p)
(2.4)

2) $\left\langle \underline{\gamma}_j , \underline{\gamma}_m \right\rangle = \underline{\gamma}_j^T \underline{\gamma}_m = 0 \qquad$ ($j \neq m$ ise ; j,m = 1,2, … , p)
(2.5)

kısıtları altında maksimum varyansa sahiptir ve $Cov (Y_j,Y_m) = \underline{\gamma}_j^T \sum \underline{\gamma}_m = 0$ olur. İlk adımda hesaplanan maksimize edilen $V(Y_j)$ değişkenlerdeki değişimin en güçlü ve en önemli modudur.

Fonksiyonel Ana Bileşenler Analizinde de 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 (γ (t)) ile belirtilir ve doğrusal kombinasyon,

$$Y_j = \langle \gamma_j , x - E(x) \rangle = \int \gamma_j(t) \{ x(t) - E\, x(t) \} dt \tag{2.6}$$

şeklinde tanımlanabilir.

Bundan sonra artık γ_j ile belirtilen ağırlıklar γ (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 ve diğerleri, 1986).

Fonksiyonel Ana Bileşenler Analizinde ise Y_j şeklindeki doğrusal bileşenin varyansı $V(Y_j)$,

$$Var(Y_j) = Var\langle \gamma_j , x - E(x) \rangle = \int\int \gamma_j(s) Cov(s,t) \gamma_j(t)\, ds\, dt = \psi_j \tag{2.7}$$

şeklinde,

1) $\langle \gamma_j , \gamma_m \rangle = \int \gamma_j(t)\gamma_m(t)\, dt = 1$ (j=m ise ; j,m= 1,2, ... , n) (2.8)

2) $\langle \gamma_j , \gamma_m \rangle = \int \gamma_j(t)\gamma_m(t)\, dt = 0$ ($j \neq m$ ise ; j,m= 1,2, ... , n) (2.9)

kısıtları altında, maksimum varyansa sahiptir. (2.7) numaralı denklemin sağ tarafı klasik çok değişkenli analizde şans vektörlerinin doğrusal kombinasyonlarının varyansını gösteren kuadratik forma karşılık gelir(Lober ve Villa, http://www.u-cergy.fr/AFFI_2004/IMG/pdf/MATZNER.pdf, 05.05.2005). Görüldüğü üzere ψ_j biçimindeki özdeğerler, her bir bileşene yüklenebilir varyans miktarını göstermektedir.

Her bir ağırlık fonksiyonunun eğrilerdeki değişimin en önemli modunu tanımlama görevi vardır ve burada (2.9) numaralı kısıtta da gösterildiği gibi 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.

Var(Y_j) ifadesinde yer alan Cov(s,t),

$$\mathrm{Cov}(s,t) = N^{-1}\{ \sum_{i=1}^{N} x_i(s)\, x_i(t) \} \qquad (2.10)$$

şeklindedir. 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 değişimin dominant bileşenlerini ayırmaktır.

Klasik Ana Bileşenler Analizi araştırmanın amacına bağlı olarak genellikle her bir gözlemden ortalamayı çıkardıktan sonra verilerdeki değişimin dominant modlarını bulmak üzere kullanılır. Fonksiyonel Ana Bileşenler Analizi de (2.10) ile ilgili açıklamalardan da anlaşıldığı üzere, 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.

Doğrusal bileşenin varyansının maksimum yapılması problemi Klasik Ana Bileşenler Analizinde,

$$\Sigma\, \underline{\gamma}_j = \psi_j\, \underline{\gamma}_j \qquad (2.11)$$

şeklindeki **özdeğer-özvektör** ayrışımını gerektirir.

Fonksiyonel Ana Bileşenler Analizinde ise doğrusal bileşenin varyansının maksimum yapılması problemi klasik tanıma paralel olarak,

$$\int \mathrm{Cov}(s,t)\, \gamma(t)\, dt = \psi\, \gamma(s) \qquad (2.12)$$

şeklindeki **özdeğer-özfonksiyon** eşitliğinin çözümünü gerektirir. Özetle fonksiyonel açıdan ele alındığında özdenklemi sağlayan artık farklı **özdeğer-özvektör** değil **özdeğer- özfonksiyon** çiftleri vardır.

Klasik Ana Bileşenler Analizinde hesaplanacak anabileşen sayısı için üst limit değişken sayısı olan p'dir. Fonksiyon durumunda ise, "değişkenler" t' nin değerlerine karşılık gelir ve bunun için bir limit yoktur. Bunun yerine ana bileşen sayısı için üst limit gözlem sayısı N ve eğer fonksiyonel gözlemler merkezlenmiş ise N-1 dir(Ramsay, ftp://ego.psych.mcgill.ca, 11.04.2005).

Klasik Ana Bileşenler Analizinde 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 ve Wichern, 1998; 468). Fonksiyonel Ana Bileşenler Analizinde ise kovaryans fonksiyonu yerine korelasyon fonksiyonunun kullanılması için daha az neden vardır. Bunun nedeni fonksiyon değerlerinin hepsinin aynı birim veya ölçekte olmasıdır.

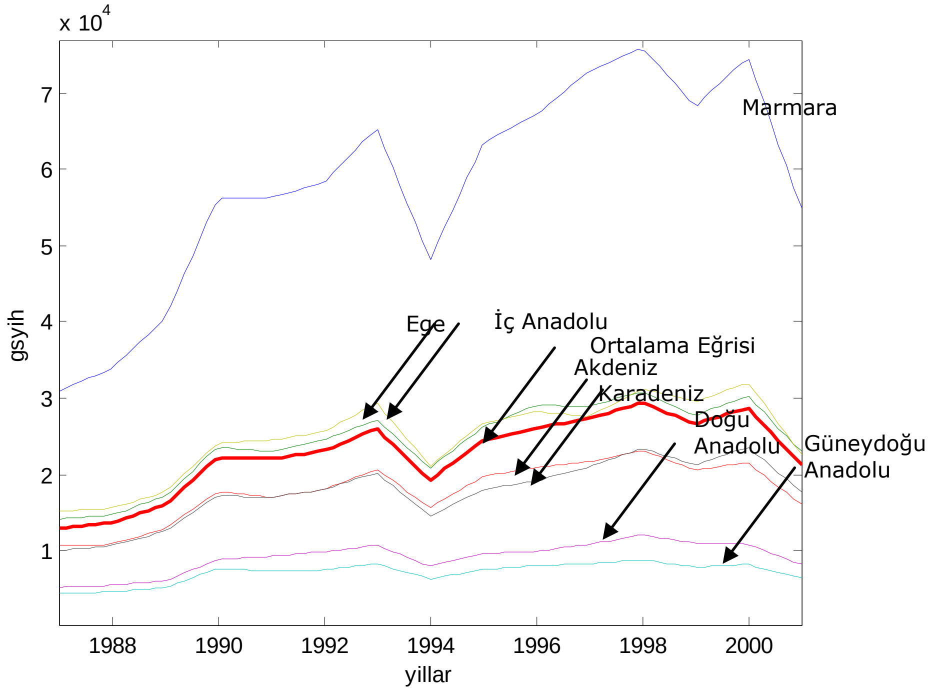
Klasik Ana Bileşenler Analizini içeren bir çok önemli eser bulunmaktadır. Konu ile ilgili ayrıntılı bilgi için Anderson (2003; 459-486), Mardia v.d.(1989; 213-146), Johnson ve Wichern (1998; 458-512) ve Jolliffe (2002) incelenebilir.

Fonksiyonel Ana Bileşenler Analizi ile ilgili ile ilgili ayrıntılı bilgi için Ramsay ve Silverman (1997; 101), Ramsay ve Silverman (2005; 161-163), James ve diğerleri (2000), Lober ve Villa (2004,http://www.u-cergy.fr/AFFI_2004/IMG/pdf/MATZNER.pdf, 05.05.2005), Lee (2004), Barra (2004), Benko (2004) ,Yamanishi ve Tanaka (2005) ve Hall ve Nasab (2006) çalışmalarına başvurulabilir.

3. UYGULAMA

GSYİH bir ülke sınırları içerisinde, belli bir zaman içinde (genellikle bir yıl olmak üzere) üretilen tüm mal ve hizmetlerin para birimi cinsinden değeridir. Bu çalışmada ekonomik büyüklüğün ölçütlerinden biri olarak kabul edilen GSYİH verileri bölgeler bazında 1985-2001 yılları için incelenmiştir. İlgili veriler Türkiye İstatistik Kurumunun (TÜİK) web sitesinden alınmıştır (TÜİK, http://www.tuik.gov.tr, 14.08. 2007). Öncelikle yıllar bazında kesikli olarak gözlenen GSYİH verileri interpole edilerek her bir bölge için bir sürekli fonksiyon bir diğer deyişle birinci dereceden bir Splayn elde edilmiştir. Artık GSYİH verileri her bir bölge için zamanın sürekli bir fonksiyonu olarak dikkate alınmaktadır. Elde edilen 7 farklı bölge için oluşturulan 7 bireysel fonksiyon Şekil (3.1)de verilmektedir.

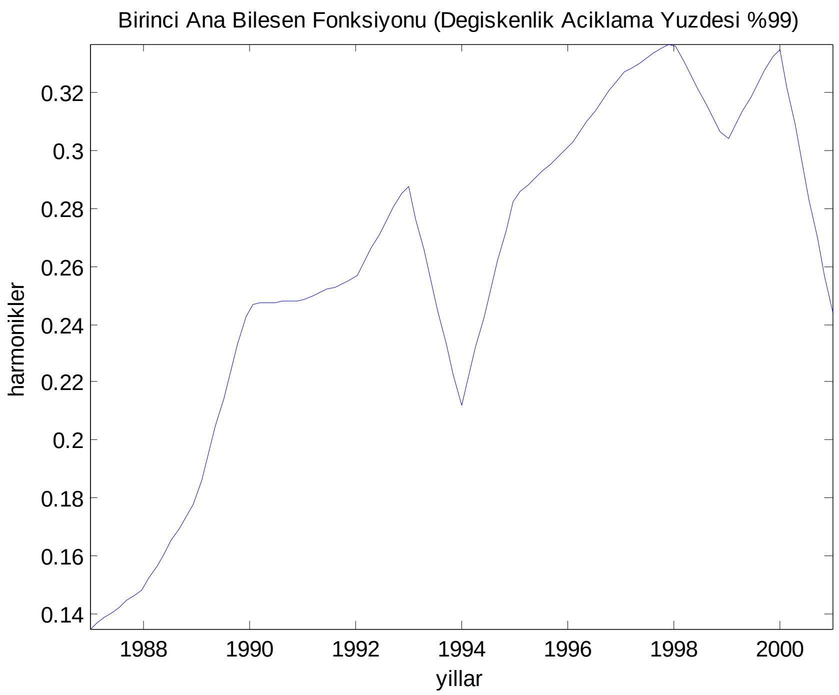
Şekil (3.1): 7 Bölgenin Bireysel Fonksiyonu



Şekil (3.1)de oluşturulan 7 farklı bölgeyi gösteren bireysel fonksiyonlar incelendiğinde Marmara Bölgesinin ekonomik gelişmişlik bakımından en önde geldiği açıkça görülmektedir. Diğerlerinden uç bir noktada bulunduğu için yıllara göre GSYİH' nın 1987-2001 yılları arasındaki seyri, hangi yıllarda iniş ve çıkışlara rastlandığı, bu bölge için diğerlerine oranla çok daha rahatlıkla seçilebilmektedir. Ege ve Akdeniz bölgeleri birbirlerini çok yakın izlemekte ve bu bölgeler de Türkiye ortalamasının incelenen tüm yıllar boyunca üzerinde bulunmaktadır. Türkiye ortalamasının hemen altında yine birbirlerini çok yakın izleyen Karadeniz ve Akdeniz Bölgeleri ve ardından da Doğu Anadolu ve son olarak da ekonomik gelişmişliğin tüm yıllar boyunca en düşük olduğu Güneydoğu Anadolu Bölgesi gelmektedir. Bu fonksiyonların her biri ayrık noktalarda gözlenen verileri interpole ederek birinci dereceden Splaynlar aracılığı ile oluşturulmuştur. Birinci dereceden Splaynlar şekilden de görüldüğü üzere sürekli fonksiyonlardır. Burada birinci değil de örneğin üçüncü dereceden Splaynlar seçilmiş olsaydı, oluşturulan bu üçüncü dereceden Splaynların kendisi, birinci ve ikinci türev fonksiyonları da sürekli olabilecek (Lyche ve Morken, http://www.ifi.uio.no/in329/nchap1.pdf, 10.07.2005) ve her bir bölgenin bireysel fonksiyonunun zamana göre birinci türevi alınıp elde edilen sürekli birinci türev fonksiyonundan GSYİH'nın 1987-2001 yılları arası büyüme hızları ve sürekli ikinci türev fonksiyonundan da ivmeleri incelenebilecekti. GSYİH verilerinin fonksiyonel açıdan ele alınmasıyla 7 farklı bölge için 15 nokta veya toplamda 105 nokta yerine 7 bireysel bölge fonksiyonu daha rahat bir biçimde incelenebilmektedir. Ayrıca 7 bölgeden herhangi bir yılla ilgili kayıp veri olsa bile interpolasyon yoluyla bu sorun çözülmüş olacaktır.

Şekil (3.1) incelendiğinde Marmara Bölgesi dışında, fonksiyonların bireysel seyirleri zaman zaman diğer fonksiyonların arasında kaybolabilmekte ve böylece fonksiyonların bireysel ve genel seyirleri hakkında bir yorumlama yapılabilmesi güç olmaktadır. Bu durum incelenen bölge veya yıl sayısı arttığında çok daha karmaşık bir hal alabilecektir. Bu amaçla, veri karmaşıklığını gidermek üzere bölgeler arasındaki değişimin önemli modlarını tanımlamak için bir boyut indirgeme yöntemi olan Ana Bileşenler ve Fonksiyonel Ana Bileşenler Analizinin kullanımı önerilebilir. Ancak burada incelenen birey sayısı gözlem sayısından az olduğu için Klasik Ana Bileşenler Analizinin kullanılması uygun değildir. Bu durumda Fonksiyonel Ana Bileşenler Analiziyle, bölgeler arasındaki değişimi açıklamaya yönelik olarak oluşturulan bir özfonksiyon olan ana bileşen ağırlık fonksiyonu Şekil (3.2)deki gibi elde edilmiştir.

Şekil (3.2): Birinci Ana Bileşen Fonksiyonu

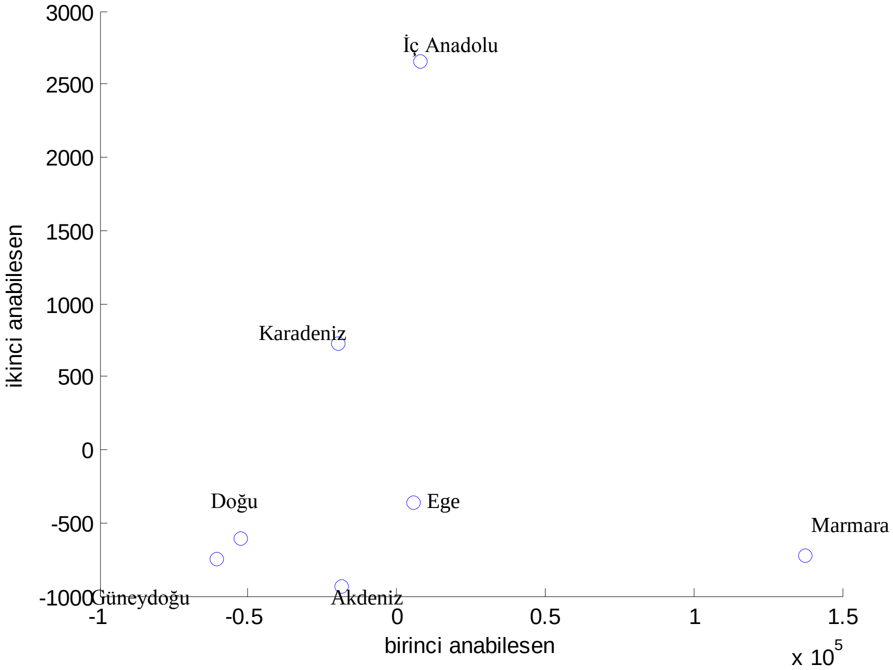


Bölgeler arasındaki değişkenliğin yüksek olduğu ana bileşen fonksiyonundan rahatlıkla gözlemlenebilmektedir. Bölgeler arasındaki değişim dalgalı bir seyir izlemektedir. Birinci ana bileşen fonksiyonu γ_1 tüm yıllar boyunca pozitif olmasına rağmen, 96 yılı sonrası yaklaşık 2000 yılı ortalarına kadar ağırlıklar 96 yılı öncesi ağırlıkların yaklaşık 2 katıdır. Bu da GSYİH açısından bölgeler arasında değişkenliğin 96 yılından sonra bir artışa geçtiğini göstermektedir. Ancak 2000 yılının ortalarından sonra da hızla azalmaya başladığı dikkati çekmektedir. Burada dikkati çeken diğer bir nokta 1993-1994 ekonomik krizi ve 1999 depremi sırasında GSYİH verilerinin bölgeler arasındaki değişkenliğinde bir azalma görülmektedir. Bunun nedeni büyük olasılıkla bu kriz dönemlerinde tüm bölgelerde GSYİH nın birlikte azalma eğilimine girmiş olmasıdır. Birinci ana bileşen fonksiyonu yaklaşık %99 gibi çok yüksek bir değişkenlik açıklama yüzdesine sahiptir. Aynı verilere Klasik Ana Bileşenler uygulansaydı ki uygulamamızda bu durum mümkün değildir, yalnızca özvektörlere veya korelasyonlara bakarak bu tip bir yorumlama yapmak çok kolay olmayacaktır. Özellikle de incelenen değişken sayısı arttıkça yorumlama iyice karmaşık hale gelecektir. Fonksiyonel Ana Bileşenler Analizinde, ana bileşen fonksiyonlarının çizimiyle görselliğin de işin içine girmesi yorumlama kolaylıkları yaratmaktadır.

Fonksiyonel Ana Bileşenler Analizinde de Klasik Ana Bileşenler Analizine benzer olarak ana bileşen skorlarının işaretlenmesi önemli bir özelliktir. γ_{ij} şeklinde belirtilen 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 ve diğerleri, http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2006-010.pdf, 10.11.2006). Ana bileşen skorlarının dağılımı sapan gözlemlerin (outlier) saptanmasında da kullanılabilir.

Burada ilk iki ana bileşen skorları işaretlenecektir, ancak ikinci ana bileşenin değişkenlik açıklama yüzdesi %1 gibi çok düşük bir değer olduğundan yorumlamalar birinci ana bileşen açısından yapılacaktır.

Şekil(3.3): Ana Bileşen Skorları Dağılımı



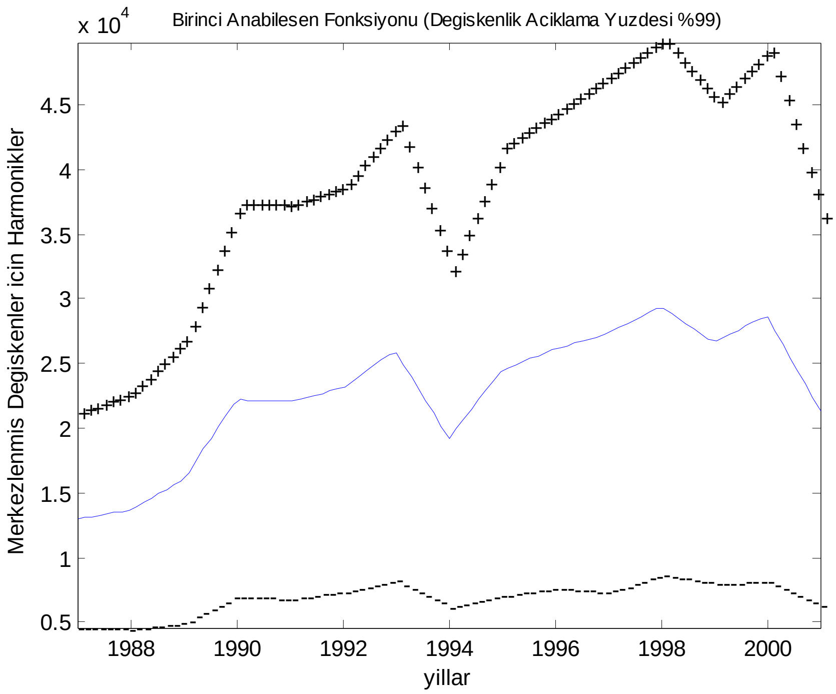
Şekil (3.3)den de görüldüğü üzere birinci ana bileşen açısından en yüksek değere Marmara Bölgesi ve en düşük değere ise Güneydoğu Anadolu Bölgesi sahiptir. Bu değişkenliğin en yüksek olduğu zaman aralıklarında da değişmemektedir. Bunun nedeni Marmara Bölgesinin tüm zaman aralığı boyunca GSYİH bakımından en yüksek ve Güneydoğu Anadolu Bölgesinin ise en düşük değerlere sahip olmasıdır.

Fonksiyonel Ana Bileşenler Analizinde Klasik Ana Bileşenler Analizine ilave olarak ana bileşen ağırlık fonksiyonlarının yorumlanmasına yardımcı olacak bir yöntem, ağırlık fonksiyonunun bir çarpanıyla ortalama fonksiyonun karşılaştırılmasıdır. Burada ortalama fonksiyonuna ilgilenilen, uygun bir çarpanla çarpılmış, ana bileşen fonksiyonu eklenerek ve çıkarılarak elde edilen fonksiyonlarla ortalama fonksiyonu aynı şekil üzerinde çizdirilir ve karşılaştırmalar yapılarak çeşitli yorumlamalarda bulunulabilir.

Burada hangi çarpanın kullanılacağını seçmek önemlidir. Ramsay (Ramsay, http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html, 11.04.2005) çarpan olarak ana bileşen fonksiyonunun oluşturulmasında her bir özfonksiyona karşılık gelen özdeğerlerin kareköklerini almayı önermiştir. Böylece her bir bileşen için

$\hat{\mu} \pm \sqrt{\psi_j}\,\hat{\gamma}$ plot edilir. Ancak bu konu araştırmaya açıktır ve konuyla ilgili farklı önerilerde bulunulabilinir. Tüm bileşenler için aynı sabitin kullanımı da önerilebilir(Ramsay ve Silverman, 1997; 93).

Şekil(3.4): Ortalama Fonksiyonu ve Birinci Ana Bileşen Fonksiyonu Karşılaştırması



Şekil (3.4)de görülen düz çizgi GSYİH bakımından Türkiye'nin ortalama fonksiyonunu belirtmektedir. + ve – noktalar ise ortalama fonksiyona ana bileşen fonksiyonunun belirli bir sabitle çarpanının eklenmesinin ve çıkarılmasının etkilerini göstermektedir. Burada sabit olarak Ramsay (http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html, 11.04.2005) yaklaşımı kullanılmıştır. + ve – noktalar ortalama fonksiyonundan ne kadar uzaksa ortalamadan sapmaların o kadar yüksek olduğu bu grafik yardımıyla da gözlemlenebilir. Şekil (3.2) de verilen ana bileşen fonksiyonuna benzer şekilde, Şekil (3.4)den de, değişkenliğin yüksek olduğu bölgelerde ortalamadan sapmaların arttığı görülmektedir.

4. SONUÇ

İncelenen birey veya obje ve inceleme veya deney yapılan nokta sayısı arttıkça verileri fonksiyonel açıdan ele almak çok çeşitli avantajlar doğurmaktadır ve ayrıca artan veri kapasiteleriyle birlikte klasik istatistiksel yöntemler de çoğu zaman gerekli analizler için yetersiz kalmaya başlamıştır. Bunun sonucu ortaya çıkan karmaşık populasyonlardaki temel değişim modlarını tanımlamak için Fonksiyonel Ana Bileşenler Analizinden yararlanılabilinir. Fonksiyonel Ana Bileşenler Analizi hem Klasik Ana Bileşenler Analizinin kullanılmasının uygun olmadığı durumlarda ve hem de fonksiyonların türevlerinin alınabilmesi, kayıp verilerle interpolasyon yoluyla baş edebilmesi ve görselliğin de yorumlara dahil edilmesiyle sağlanan yeni bakış açıları ve benzeri avantajlarından dolayı tercih edilmektedir.

Bu çalışmada da Fonksiyonel Veri Analizi, Ana Bileşenler Analizi ile karşılaştırmalı olarak Fonksiyonel Ana Bileşenler Analizi kısaca tanıtılmış ve 1987-2001 yılları arasında bölgeler bazında GSYİH verilerindeki değişkenlik yapısı ortaya konulmuştur.

KAYNAKÇA

ANDERSON T.W. (2003). "An Introduction to Multivariate Statistical Analysis",Wiley-Interscience, USA

BARRA V. (2004). "Analysis of Gene Expression Data Using Functional Principal Components", Computer Methods and Programs in Biomedicine, 75(1):1-9.

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., and KNEİP A.(2006). "Common Functional Principal Components" SFB 649, DiscussionPaper http://ideas.repec.org/p/hum/wpaper/sfb649dp2006-010.html, (10.11.2006)

CASTRO P. E, LAWTON W. H., and SYLVESTRE E. A. (1986). "Principal Modes Of Variation for Processes with Continuous Sample Curves" ,*Technometrics,* 28(4):329-337.

HALL P. and NASAB H. M. (2006). "On Properties Of Functional Principal Components Analysis", Journal of the Royal Statistical Society: Series *B*, 68(1):109-126.

JAMES G. M., HASTİE T.J., and SUGAR C.A. (2000), "Principal Components Models For Sparse Functional Data", Biometrica, 87(3):587-602.

JOLIFFE I.T. (2002). “Principal Component Analysis”, Springer – Verlag, New York.

LEE H.J. (2004). “Functional Data Analysis: Classification and Regression”. Doctor of Philosophy, Texas A&M University.

LOBER E.M., and VİLLA C. (2004). “*Functional Principal Component Analysis of the Yield Curve*”, http://www.u-cergy.fr/AFFI_2004/IMG/pdf/MATZNER.pdf, (05.05.2005).

LOBER E.M., and VİLLA C. (2004). “Functional Principal Component Analysis of the Yield Curve”, http://www.u-cergy.fr/AFFI_2004/IMG/pdf/MATZNER.pdf, (05.05.2005).

LYCHE, T. and MORKEN, K., (2002). “Spline Methods Draft”, http://www.ifi.uio.no/in329/nchap1.pdf ,(10.07.2005).

MARDİA K.V., KENT J.T., and BİBBY J.M. (1989). “Multivariate Analysis”, Academic Press, London.

RAMSAY J.O., & SİLVERMAN B.W. (2005). “Functional Data Analysis” Second Edition, Springer , USA.

RAMSAY J.O., (2000). “Functional Principal Component Analysis”, , ftp://ego.psych.mcgill.ca, (11.04.2005).

RAMSAY J.O., and SİLVERMAN B.W. (1997). “Functional Data Analysis”, Springer – Verlag, New York.

RAMSAY, J. O., and DALZELL C. (1991). “Some Tools For Functional Data Analysis”, Journal of the Royal Statistical Society: Series B.,53 (3):539-572.

RAMSAYJ.O.,(2003).,http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html, (11.04.2005).

TÜİK (2007), http://www.tuik.gov.tr, (14.08. 2007).

YAMANİSHİ Y., & TANAKA Y. (2005). “Sensitivity Analysis in Functional Principal Component Analysis” *Computational Statistics,* 20(2):313-329.