HANEHALKI TÜKETİM HARCAMALARININ PARAMETRİK OLMAYAN REGRESYON YÖNTEMİ İLE TAHMİNİ: TÜRKİYE ÖRNEĞİ

(Yüksek Lisans Tezi)

Taner AKMERCAN

Kütahya - 2016

T.C.
DUMLUPINAR ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
Ekonometri Anabilim Dalı

Yüksek Lisans Tezi

HANEHALKI TÜKETİM HARCAMALARININ PARAMETRİK OLMAYAN REGRESYON YÖNTEMİ İLE TAHMİNİ: TÜRKİYE ÖRNEĞİ

Danışman:
Yrd. Doç. Dr. Noyan AYDIN

Hazırlayan:
Taner AKMERCAN

Kütahya – 2016

Kabul ve Onay

Taner AKMERCAN'ın hazırladığı "Hanehalkı Tüketim Harcamalarının Parametrik Olmayan Regresyon Yöntemi İle Tahmini: Türkiye Örneği" başlıklı Yüksek Lisans tez çalışması, jüri tarafından lisansüstü yönetmeliğinin ilgili maddelerine göre değerlendirilip oybirliği / oyçokluğu ile kabul edilmiştir.

...../....../2016

Tez Jürisi	**İmza**	
	Kabul	**Red**
Prof. Dr. Mahmut ZORTUK		
Doç. Dr. Mustafa Kemal BEŞER		
Yrd. Doç. Dr. Noyan AYDIN (Danışman)		

Doç. Dr. Niyazi KURNAZ

Sosyal Bilimler Enstitüsü Müdürü

Yemin Metni

Yüksek lisans tezi olarak sunduğum "Hanehalkı Tüketim Harcamalarının Parametrik Olmayan Regresyon Yöntemi İle Tahmini: Türkiye Örneği" adlı çalışmamın, tarafımdan bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın yazıldığını ve yararlandığım kaynakların kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

....../....../ 2016

Taner AKMERCAN

Özgeçmiş

Taner AKMERCAN 15 Mayıs 1988 tarihinde Malatya' da doğdu. İlköğretim eğitimini Hoca Ahmet Yesevi İlköğretim Okulunda, Lise eğitimini ise Bağcılar Lisesinde 2005 yılında tamamladı. 2006 yılında okumaya hak kazandığı Dumlupınar Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümünden 2010 yılında mezun oldu.

ÖZET

HANEHALKI TÜKETİM HARCAMALARININ PARAMETRİK OLMAYAN REGRESYON YÖNTEMİ İLE TAHMİNİ: TÜRKİYE ÖRNEĞİ

AKMERCAN, Taner
Yüksek Lisans Tezi, Ekonometri Ana Bilim Dalı
Tez Danışmanı: Yrd. Doç. Dr. Noyan AYDIN
Ocak, 2016, 67 sayfa

Değişkenler arasındaki ilişkiyi incelemek için sıklıkla kullanılan geleneksel parametrik regresyon yöntemlerinde, bu ilişkinin fonksiyonel formu önceden belirlenmekte ve bu fonksiyonel formun parametrelerinin tahmin edilmesi amaçlanmaktadır. Parametrik olmayan regresyon yöntemlerinde ise, ilişkinin fonksiyonel formu açık bir şekilde belirlenmeden doğrudan regresyon fonksiyonunun tahmin edilmesi amaçlanmaktadır. Bu nedenle parametrik olmayan regresyon yöntemleri araştırmacılara değişkenler arasındaki ilişkiyi ortaya çıkartmada esneklik sağlamaktadır.

Bu çalışmada, hanehalkı zaruri tüketim harcamaları, hanehalkı kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü arasındaki ilişkiyi ortaya çıkartmak için parametrik olmayan bir regresyon yöntemi olan LOESS regresyon yöntemi ve 2012 yılı TÜİK Hanehalkı Bütçe Anketinden alınan veriler kullanılmıştır. Elde edilen sonuçlar göstermektedir ki, düşük gelir grubundaki hanelerin birey sayısı artarken, hanehalkı zaruri harcamaları önce hızlı bir şekilde artmakta ardından bu artış bir miktar azalmaktadır. Yüksek gelir grubunda ise hanehalkı birey sayısı artarken hanehalkı tüketim harcamalarında bir miktar azalma gözlemlenmektedir.

Anahtar Kelimleler: Parametrik Olmayan Regresyon, LOESS Regresyon Yöntemi, Hanehalkı, Harcama ve Gelir Arasındaki İlişki

ABSTRACT

ESTIMATION OF HOUSEHOLD CONSUMPTION EXPENDITURE BY NONPARAMETRIC REGRESSION METHOD: THE CASE OF TURKEY

AKMERCAN, Taner
M.A. Thesis, Department of Econometrics
Supervisor: Assistant Professor Noyan AYDIN
January, 2016, 67 pages

In traditional parametric regression methods which are often used to investigate relationship between variables, the functional form of the relationship is specified in advance and the aim is to estimate the parameters of this functional form. In nonparametric regression methods, in contrast, the aim is to estimate regression function directly without specified its functional form clearly. For this reason nonparametric regression methods provide to more flexibility for researchers in revealing relationship between the variables.

In this study, for purpose to reveal the relationship between household expenditure, households' income and size of household defined by OECD is used LOESS regression method which is a nonparametric regression method and data from TUIK Household Budget Survey of 2012. The results are obtained show that in low income group of household, initially household expenditure rapidly increase and then this increase decreases while increase size of household. İn high income group of hosehold, in conrast, household expenditure decreases a little bit while increase size of household.

Keywords: Nonparametric Regression, LOESS Regression Method, Household, The Relationship Between Expanditure and İncome

İÇİNDEKİLER

Sayfa

TABLOLAR LİSTESİ

Sayfa

ŞEKİLLER LİSTESİ

KISALTMALAR

CV	Cross Validation (Çapraz Geçerlilik)
EKK	En Küçük Kareler
EYO	En Yüksek Olabilirlik
K-NN	K-Nearest Neighborhood (K-En Yakın Komşuluk)
LF	Likelihood Function (Olabilirlik Fonksiyonu)
LOESS	Locally Weighted Scatterplot Smoothing (Lokal Olarak Ağırlıklandırılmış Dağılım Grafiği Düzleştirmesi)
LOWESS	Locally Weighted Scatterplot Smoothing (Lokal Olarak Ağırlıklandırılmış Dağılım Grafiği Düzleştirmesi)
LS	Least Squares (En Küçük Kareler)
ML	Maximum Likelihood (En Yüksek Olabilirlik)
OECD	Organisation for Economic Co-Operation and Development (Ekonomik Kalkınma ve İşbirliği Örgütü)
TÜİK	Türkiye İstatistik Kurumu

TEZ METNİ

GİRİŞ

Regresyon analizi değişkenler arasındaki ilişkiyi inceleyen ve sıklıkla kullanılan yöntemlerden biridir. Geleneksel regresyon yöntemi olarak bilinen parametrik regresyon yönteminde değişkenler arasındaki ilişkinin fonksiyonel formu önceden belirlenmektedir. Ardından belirlenen bu modelin katsayıları hesaplanarak değişkenler arasındaki ilişki incelenmektedir. Fakat her zaman değişkenler arasındaki ilişkinin nasıl bir fonksiyonel forma sahip olduğunu tespit etmek mümkün değildir. Ayrıca parametrik regresyon yönteminde önceden belirlenen ilişkinin fonksiyonel formu ile birlikte bir takım varsayımlarında sağlanması gerekmektedir. Varsayımların sağlanmadığı durumlarda değişkenler arasındaki ilişki incelenememektedir.

Parametrik olmayan regresyon yönteminde ise değişkenler arasındaki ilişkinin fonksiyonel formu hakkında herhangi bir varsayım yapılmadan doğrudan değişkenler arasındaki ilişki ortaya çıkarılmaya çalışılmaktadır. Bu nedenle parametrik olmayan regresyon yöntemi değişkenler arasındaki ilişkiyi incelemede araştırmacılara esneklik sağlamaktadır. Ayrıca parametrik olmayan regresyon yöntemleri parametrik regresyon yöntemleri ile elde edilen sonuçların geçerliliğini test etmek amacıyla da kullanılabilmektedirler.

Hanehalkı harcamaları ile geliri arasındaki ilişki, hanelerin sosyo-ekonomik yapılarının ve tüketim kalıplarının anlaşılması ve bu doğrultuda uygulanan politikaların test edilmesi veya uygulanacak politikaların belirlenmesi açısından oldukça önemlidir. Bu nedenle bu ilişkinin nasıl oluğunun anlaşılması ihtiyacı her zaman güncelliğini korumaktadır. Bununla birlikte hanehalkı harcamaları ve geliri üzerinde önemli bir etkiye sahip olan bazı sosyo-demografik özelliklerin bu ilişkiyi incelerken dikkate alınması hanelerin sosyo-ekonomik yapıları ve tüketim kalıpları hakkında daha geniş bilgiler sağlayacaktır.

Bu kapsamda yapılan çalışmalar incelendiğinde genellikle sıkı ön varsayımlar gerektiren parametrik yöntemlerin kullanıldığı görülmektedir. Bu çalışmada ise bu durumun dışına çıkılarak hanehalkı değişkenleri arasındaki ilişki parametrik olmayan bir yöntemle incelenmiştir. Böylelikle hem hanehalkı çalışmalarına farklı bir bakış açısı kazandırılmaya çalışılmış hem de değişkenler arasındaki ilişkinin fonksiyonel formu hakkında herhangi bir varsayımda bulunmadan doğrudan ilişkinin yapısı ortaya çıkarılmaya çalışılmıştır.

İlk bölümde regresyon analizinin önemine değinilmiş ve buna bağlı olarak parametrik ve parametrik olmayan yöntemlerden bahsedilerek aralarındaki farklılıklar kısaca ifade edilmiştir.

İkinci bölümde parametrik olmayan regresyon yöntemlerinden, kernel (çekirdek) regresyon yöntemi, en yakın komşuluk yöntemi (k-NN) ve LOESS regresyon yöntemi gibi yöntemler ele alınmıştır. Ayrıca parametrik olmayan regresyon yöntemlerinin dayandığı düzleştirme kavramı açıklanmış ve sık kullanılan düzleştirme yöntemleri ile birlikte düzleştirme parametresi seçimi hakkında ayrıntılı bilgiler verilmiştir. Ardından çalışmada kullanılacak olan LOESS regresyon yöntemi, span değeri seçimi ve uyum iyiliği ayrıntılı olarak tanıtılmıştır.

Üçüncü ve son bölümde ise hanehalkı zaruri tüketim harcamaları ile hanehalkı kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü arasındaki ilişki parametrik olmayan bir yöntem olan LOESS regresyon yöntemiyle incelenmiştir. Ardından aynı değişkenler kullanılarak değişkenler arasındaki ilişki parametrik regresyon yöntemiyle incelenmiştir. Yapılan incelemelerin sonuçları karşılaştırılmış ve elde edilen bulgular sunulmuştur.

BİRİNCİ BÖLÜM

REGRESYON ANALİZİ İLE PARAMETRİK VE PARAMETRİK OLMAYAN YÖNTEMLERE İLİŞKİN GENEL KAVRAMLAR

1.1. REGRESYON ANALİZİ

Regresyon terimi ilk olarak 19. Yüzyılın sonlarına doğru İngiliz bilim adamı Francis Galton tarafından kullanılmıştır. Galton (1886) çalışmasında ebeveynlerin boyları ile çocuklarının boyları arasındaki ilişkiyi incelemiş ve kısa boylu ailelerin çocukların genel ortalamaya göre kısa boylu, uzun boylu ailelerin çocuklarının ise genel ortalamaya göre uzun boylu olduklarını gözlemlemiştir. Bununla birlikte, çocukların boylarının ortalamalarının genel nüfusun boy ortalamasına doğru yaklaşma (regress) eğiliminde olduğunu saptamış ve bu eğilimi ortalamaya doğru yaklaşma (regression to modiocrity) olarak adlandırmıştır. Galton böylelikle iki nicel değişken arasında bir ilişkiyi tanımlamak için matematiksel bir fonksiyonun geliştirilebileceğini ortaya koymuştur.

Günümüzde değişkenler arasındaki ilişkiyi inceleyen regresyon analizinin temeli Galton' un yaptığı çalışmalara dayanmaktadır. Regresyon analizi; herhangi bir değişkenin (bağımlı değişken) bir veya daha fazla değişkenle (bağımsız değişken) arasındaki ilişkinin matematiksel bir form şeklinde yazılmasıdır (Orhunbilge, 1996: 12). Regresyon analizi ile bir veya daha fazla bağımsız değişkenin değerleri kullanılarak bağımlı değişkenin değeri tahmin edilmeye çalışılmaktadır. Regresyon analizinin başlıca üç amacı vardır (Maddala, 1992: 61);

1. Bağımsız değişkenlerdeki bireysel değişimlerin bağımlı değişken üzerinde yaptığı etkiyi analiz etmek
2. Bağımsız değişkenlerin aldığı değerlere göre, bağımlı değişkenin gelecekte alacağı değerleri tahmin etmek
3. Bağımsız değişken veya değişkenlerin bağımlı değişken üzerinde önemli bir etkiye sahip olup olmadığını incelemektir.

Regresyon analizi değişkenler arasındaki ilişkiyi incelerken sağladığı faydalardan ve kullanımda getirdiği kolaylıklardan dolayı bugün birçok alanda sıklıkla kullanılan bir analizdir.

1.1.1. Parametrik Yöntemler

Bir bağımlı değişken ve k tane bağımsız değişkenden oluşan standart çok değişkenli regresyon modeli aşağıdaki gibi ifade edilebilir;

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

Burada y bağımlı değişkeni, α sabit terimi, $x_1, x_2, \dots, x_k$ bağımsız değişkenleri, ε hata terimini ve $\beta_1, \beta_2, \dots, \beta_k$ regresyon katsayılarını ifade etmektedir. Regresyon katsayıları ait oldukları bağımsız değişkendeki bir birimlik değişmenin diğer değişkenler sabit tutulduğunda bağımlı değişkeni ne ölçüde değiştirdiğini göstermektedir. Regresyon katsayılarının işaretleri pozitif veya negatif olabilir. Pozitif regresyon katsayısı ait olduğu bağımsız değişken ile bağımlı değişken arasında aynı yönlü bir ilişki (biri artarken diğeri de artıyor) olduğunu, negatif regresyon katsayısı ise ait olduğu bağımsız değişken ile bağımlı değişken arasında ters yönlü bir ilişki (biri artarken diğeri azalıyor) olduğunu göstermektedir. Ayrıca sabit terim ve regresyon katsayılarına bu regresyon modelinin parametreleri denilmektedir.

Regresyon analizine parametrik yaklaşımda ilişkinin fonksiyonel biçimi ve hataların dağılımı ile ilgili önceden bir takım varsayımlar yapılmaktadır. Ardından bu varsayımlar altında belirlenen regresyon denkleminin parametreleri tahmin edilmeye çalışılır. Bu parametreleri tahmin ederken "En Küçük Kareler (EKK) Yöntemi" veya "En Yüksek Olabilirlik (EYO) Yöntemi" kullanılmaktadır.

1.1.1.1. En Küçük Kareler Yöntemi

Gözlem değerleri ile tahmin edilen değerler arasındaki farka hata terimi denilmektedir. En Küçük Kareler (EKK – Least Squares (LS)) Yöntemi, bu hata terimlerinin kareleri toplamını minimize edecek şekilde parametreleri tahmin ederek en uygun regresyon denklemini tahmin etmeyi amaçlayan bir yöntemdir. Bu amaçla hata terimini ifade eden eşitlikte denklemin katsayılarına göre sırasıyla birinci türevleri alınarak yeni denklemler elde edilir ve sıfıra eşitlenir. Elde edilen bu denklemlere "Normal Denklemler" adı verilmektedir. Bu normal denklemlerde uygun değerler yerlerine konularak parametreler bulunmaktadır.

Basit doğrusal regresyon denklemi $\hat{y}_i = a + bx_i$' yi ele alırsak (Serper, 2004: 286-287; Baltagi, 2011: 50);

$$\sum_{i=1}^{n} e_i{}^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \tag{1.1}$$

(1.1) eşitliği minimum olmalıdır. Yani a ve b parametreleri (1.1) numaralı denklemi minimum yapacak şekilde seçilmelidir. Eğer $S = \sum_{i=1}^{n} e_i{}^2$ alınıp sırasıyla a ve b parametrelerine göre birinci türevleri alınır ve sıfıra eşitlenirse;

$$\frac{\partial S}{\partial a} = 2\sum_{i=1}^{n} (-1)(y_i - a - bx_i) = 0$$

$$\frac{\partial S}{\partial a} = 2\sum_{i=1}^{n} (-y_i + a + bx_i) = 0$$

$$\frac{\partial S}{\partial a} = -\sum_{i=1}^{n} y_i + na + b\sum_{i=1}^{n} x_i = 0$$

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \tag{1.2}$$

Ve

$$\frac{\partial S}{\partial b} = 2\sum_{i=1}^{n} (y_i - a - bx_i)(-x_i) = 0$$

$$\frac{\partial S}{\partial b} = 2\sum_{i=1}^{n} (-x_i y_i + x_i a + bx_i{}^2) = 0$$

$$\frac{\partial S}{\partial b} = -\sum_{i=1}^{n} x_i y_i + a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i{}^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i{}^2 \tag{1.3}$$

Normal denklemleri elde edilir. Elde edilen (1.2) ve (1.3) normal denklemleri çözümlenerek a ve b regresyon katsayıları bulunduktan sonra regresyon denkleminde yerine yazılır ve bağımsız değişkenlerin değerleri yardımıyla bağımlı değişken tahmin edilir.

EKK yöntemi, belli varsayımları sağlaması halinde bağımlı ve bağımsız değişkenler arasındaki ortalama ilişkiyi gerçeğe en yakın şekilde tahmin eden yöntemdir (Tarı, 2008: 24).

1.1.1.2. En Küçük Kareler Yönteminin Varsayımları

Regresyon denkleminde parametreleri tahmin ederken en küçük kareler yöntemini kullanabilmek için bazı varsayımların sağlanması gerekir. Bu varsayımlar kısaca aşağıdaki gibi ifade edilebilir (Gujarati, 2004: 66-75):

Varsayım 1: Doğrusal regresyon modeli aşağıdaki denklemde gösterildiği gibi doğrusaldır. Fakat bağımlı ve bağımsız değişkenler kendi aralarında doğrusal olmayabilirler.

$$Y_i = \alpha + \beta_1 X_i + e_i$$

Varsayım 2: Tekrarlanan örneklerde bağımsız değişken *X* değerleri sabittir. Yani, *X* değişkeninin tesadüfi (rassal) değişken olmadığı varsayılır.

Varsayım 3: Hata terimleri sıfır ortalamaya sahiptir.

$$E(e_i|X_i) = 0$$

Varsayım 4: Tüm *X* gözlem değerleri için hata terimlerinin varyansı sabittir.

$$var(e_i|X_i) = E[e_i - E(e_i|X_i)]^2$$

$$var(e_i|X_i) = E({e_i}^2|X_i)$$

$$var(e_i|X_i) = \sigma^2$$

Bu varsayıma aynı zamanda eş varyans (homoscedastisity) denilmektedir. Bu varsayımla regresyon çizgisi etrafındaki değişimin X değerleri boyunca sabit olduğunu, yani X değerlerine göre değişmeyip sabit kaldığını ifade etmektedir.

Bu varsayım sağlanmadığında değişen varyans (heteroscedasticity) durumu ortaya çıkmaktadır. Böyle durumlarda EKK yöntemiyle elde edilen regresyon katsayıları tahminlerde kullanılamaz.

Varsayım 5: Hata terimleri arasında otokorelasyon yoktur. Yani; iki farklı X_i ve X_j $(i \neq j)$ değeri için, bu değerlere karşılık gelen e_i ve e_j $(i \neq j)$ hata terimleri arasında korelasyon yoktur.

$$cov(e_i, e_j | X_i, X_j) = E\{[e_i - E(e_i)] | X_i\}\{[e_j - E(e_j)] | X_j\}$$

$$cov(e_i, e_j | X_i, X_j) = E(e_i | X_i)(e_j | X_j)$$

$$cov(e_i, e_j | X_i, X_j) = 0$$

Varsayım 6: Hata terimi ve bağımsız değişken arasında ilişki yoktur.

$$cov(e_i, X_i) = E[e_i - E(e_i)][X_i - E(X_i)]$$

$$cov(e_i, X_i) = E\big[e_i\big(X_i - E(X_i)\big)\big]$$

$$cov(e_i, X_i) = E(e_i X_i) - E(X_i)E(e_i)$$

$$cov(e_i, X_i) = E(e_i X_i) = X_i E(e_i) = 0$$

Bu varsayım sağlanmazsa bağımsız değişkenin tek başına bağımlı değişken üzerindeki etkisini belirlenemez. Çünkü bağımsız değişken ve hata terimi arasında pozitif veya negatif bir ilişki olursa bağımsız değişken artarken veya azalırken hata terimi de artacak veya azalacaktır.

Varsayım 7: Gözlem sayısı tahmin edilecek parametre sayısından daha fazla olmalıdır.

Varsayım 8: Verilen bir örneklemde ki *X* değerleri aynı olamaz. Diğer bir ifadeyle, *X* değerlerinin varyansı

$$var(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Sınırlı pozitif bir sayıdır.

Varsayım 9: Regresyon modeli doğru bir şekilde belirlenmiştir. Diğer bir ifadeyle, ampirik analizde kullanılan modelde spesifikasyon yanlılığı veya hatası yoktur.

Varsayım 10: Çoklu doğrusal bağlantı yani, bağımsız değişkenler arasında güçlü doğrusal bir ilişki yoktur.

1.1.1.3. En Yüksek Olabilirlik Yöntemi

Regresyon denkleminin parametrelerini tahmin etmek için kullanılan diğer bir yöntemde En Yüksek Olabilirlik (EYO – Maximum Likelihood (ML)) Yöntemidir. Bu yöntem, hata terimlerinin bağımsız ve normal dağılımı varsayımı altında türetilen Olabilirlik Fonksiyonunu (Likelihood Function) maksimize etmeyi amaçlamaktadır.

Basit doğrusal regresyon denklemi; $Y_i = \alpha + \beta X_i + e_i$ olsun. Bu denklemde hataların olasılık yoğunluk fonksiyonu;

$$f(e_1, e_2, \dots, e_n; \alpha, \beta, \sigma^2) = \left(1/_{2\pi\sigma^2}\right)^{n/2} \exp\left(-\sum_{i=1}^{n} e_i^2/_{2\sigma^2}\right) \quad (1.4)$$

Şeklinde yazılabilir ve (1.4) denkleminde hata terimleri yerine, $e_i = Y_i - \alpha - \beta X_i$ yazılırsa;

$$f(Y_1, Y_2, \dots, Y_n; \alpha, \beta, \sigma^2) = \left(1/_{2\pi\sigma^2}\right)^{n/2} \exp\left(-\sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2/_{2\sigma^2}\right) \quad (1.5)$$

Fonksiyonu elde edilir (Baltagi, 2011: 57). (1.5) ile belirtilen bu fonksiyona "Olabilirlik Fonksiyonu (Likelihood Function - LF)" denir ve $LF(\alpha, \beta, \sigma^2)$ ile gösterilir (Gujarati, 2004: 114). Ardından (1.5) eşitliğinde her iki tarafın doğal logaritması alınır ve sırasıyla her bir parametreye göre birinci türevler alınıp sıfıra eşitlenirse aşağıdaki ifadeler elde edilir.

$$\ln LF = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\frac{(Y_i - \alpha - \beta X_i)^2}{\sigma^2}$$

$$\frac{\partial \ln LF}{\partial \alpha} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2(-1) = 0$$

$$\frac{\partial \ln LF}{\partial \beta} = -\frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2(-X_i) = 0$$

$$\frac{\partial \ln LF}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2 = 0$$

Elde edilen bu eşitlikler de gerekli çözümlemeler yapıldığında;

$$\sum_{i=1}^{n} X_i Y_i = \alpha \sum_{i=1}^{n} X_i + \beta \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} Y_i = n\alpha + \beta \sum_{i=1}^{n} X_i$$

$$\hat{\sigma}^2 = \frac{1}{n} e_i^2$$

Eşitlikleri elde edilir (Gujarati, 2004: 115-116). Burada dikkat edildiğinde En Yüksek Olabilirlik (EYO) Yöntemiyle elde edilen parametrelerin En Küçük Kareler (EKK) Yöntemiyle elde edilen parametreler ile aynı olduğu görülmektedir. Sadece tahmin edilen varyanslarda bir farklılık bulunmaktadır. EYO yöntemiyle tahmin edilen varyans yansız bir tahmin değildir (Baltagi, 2011: 57). Fakat bu yanlılık örneklem büyüklüğü *n* sonsuza giderken ortadan kalkmakta ve tahmin yansız olmaktadır (Tezcan, 2009: 28).

1.1.2. Parametrik Olmayan Yöntemler

Regresyon analizine parametrik yaklaşımda bağımlı değişken ile bağımsız değişken veya değişkenler arasında ki ilişkinin, ya analizden önce nasıl bir fonksiyonel forma sahip olduğu bilinmekte ya da ilişkinin belirlenen fonksiyonel bir forma uygun olduğu varsayılarak analiz yapılmakta ve bu fonksiyonel formun parametreleri tahmin edilmektedir. Regresyon analizine parametrik olmayan yaklaşımda ise bağımlı değişken ile bağımsız değişken veya değişkenler arasında ilişkinin fonksiyonel formu ile ilgili analizden önce herhangi bir varsayım yapılmamakta ve dolayısıyla parametre tahmini yapılmamaktadır. Doğrudan ilişkinin fonksiyonel formu tahmin edilmektedir. Bu nedenle parametre tahmini yapılmayan bu yöntemlere *"parametrik olmayan regresyon yöntemi"* denilmektedir. Daha teorik bir şekilde ifade edecek olursak, genel bir regresyon modeli;

$$y_i = m(x_i) + e_i \ , \qquad i = 1, 2, \dots , n \tag{1.6}$$

Şeklinde ifade edilebilir. Burada y_i bağımlı değişkeni, e_i hata terimlerini ve $m(x_i)$ bağımlı değişken ile bağımsız değişken veya değişkenler arasındaki ilişkiyi gösteren fonksiyonu belirtmektedir. Eğer bu $m(x_i)$ fonksiyonu biliniyor ise (1.6) eşitliği *parametrik regresyon modeli*, eğer $m(x_i)$ fonksiyonu bilinmiyor ise (1.6) eşitliği, *parametrik olmayan regresyon modeli* adını almaktadır (Hardle, 1992: 3-4).

Parametrik olmayan regresyon yönteminde amaç $m(x_i)$ fonksiyonunu tahmin etmektir ve bu tahmin yöntemine düzleştirici (smoother) adı verilmektedir. Bu nedenle düzleştirme kavramı parametrik olmayan regresyon yönteminde oldukça önemlidir. Düzleştirici (smoother); bir veya daha fazla bağımsız değişkenin fonksiyonu olan bağımlı değişkenin sahip olduğu trendi ifade etmek için kullanılan bir araçtır ve bağımlı değişkenin kendisinden daha az değişken bir trend tahmini yapmayı amaçlamaktadır (Tezcan, 2009: 14).

Parametrik olmayan regresyon yöntemi ilişkinin fonksiyonel formu ile ilgili belli ön varsayımlarda bulunmadığından genellikle veri seti içerisinde var olan yapısal modellerin türleri hakkında daha detaylı bilgi vermektedir (Jacoby, 2000: 578). Ayrıca parametrik regresyonda olduğu gibi sıkı varsayımlar söz konusu değildir. Sadece hata terimlerinin ortalamasının sıfır ve varyansının sonlu bir sayı olduğu varsayılmaktadır (Hart, 1997: 4).

1.1.3. Parametrik Olmayan Yöntemlerin Avantajları

Parametrik olmayan istatistiksel yöntemlerin son 75 yılda sürekli ve hızlı gelişimleri, parametrik olmayan yöntemlerin sahip oldukları avantajlardan kaynaklanmaktadır ve bu avantajlar aşağıdaki gibi özetlenebilir (Hollander vd., 2014: 1).

- Parametrik olmayan yöntemler, verilerin elde edildiği popülasyonun normal dağılımdan geldiği gibi güçlü varsayımlar gerektirmez.
- Parametrik olmayan yöntemlerin uygulanması her zaman olmasa da sıklıkla parametrik karşılıklarının uygulanmasından daha kolaydır.
- Parametrik olmayan yöntemler aykırı değerlerden daha az etkilenirler.
- Parametrik olmayan yöntemlerin anlaşılması genellikle oldukça kolaydır.
- Parametrik olmayan yöntemlerde ilk bakışta örneklemde bulunan temel bilginin çoğu kayboluyor gibi görülse de teorik olarak verimliliği incelendiğinde bunun böyle olmadığı görülmektedir. Genellikle bu parametrik olmayan yöntemler normal dağılıma uyan parametrik karşılıklarına nazaran daha az (slightly less) etkindirler. Buna karşılık normal dağılıma uymayan popülasyonlarda ise parametrik karşılıklarına göre çok daha fazla etkindirler.
- Parametrik olmayan yöntemler, parametrik yöntemlerin aksine gözlemlerin büyüklükleri yerine bunların sırasıyla ilgilenmektedir. Bu nedenle parametrik yöntemlerin kullanılamadığı birçok durumda parametrik olmayan yöntemler kullanılabilir.

Ayrıca, literatürde aykırı gözlemlerin etkilerini farklı biçimlerde ele alan güçlü (robust) parametrik yöntemler vardır. Bununla birlikte aşırı uç gözlemlerden dolayı parametreler bozulduğu için bu güçlü yöntemler bile uygun çözümler üretemeyebilir ve verinin gerçek yapısı modele yansıtılamaz. Bu gibi durumlarda parametrik olmayan yöntemlerin kullanılması daha etkili sonuçlar üretebilir (Hardle, 1992: 12).

Parametrik olmayan yöntemler sağladıkları bu avantajların yanında parametrik yöntemlere göre bazı dezavantajları da bulunmaktadır. Parametrik regresyona göre artan işlem kalabalığı ile işlemlerin karmaşıklığı ve bazı durumlarda elde edilen sonuçların açıklanmasında ki zorluk bunlardan bazılarıdır (Tezcan, 2009: 12).

İKİNCİ BÖLÜM

PARAMETRİK OLMAYAN REGRESYON YÖNTEMLERİ

2.1. PARAMETRİK OLMAYAN REGRESYON

Parametrik regresyon yönteminde bağımlı değişken ile bağımsız değişken veya değişkenler arasında ilişkinin fonksiyonel formu önceden belirlenmekte ve buna bağlı olarak bir takım ön varsayımların sağlanması gerekmektedir. Parametrik olmayan regresyon yönteminde ise bağımlı değişken ile bağımsız değişken veya değişkenler arasındaki ilişkiyle ilgili önceden bir fonksiyonel form belirlenmemekte ve herhangi bir varsayımda bulunulmamaktadır. Sadece hata terimlerinin ortalamasının sıfır ve varyansının sonlu bir sayı olduğu varsayılmaktadır (Hart, 1997: 4). Bu nedenle varsayımların sağlanmadığı veya değişkenler arasındaki ilişki hakkında bir ön bilginin bulunmadığı durumlarda parametrik olmayan regresyon yöntemi kullanılabilmektedir. Ayrıca parametrik bir model kurmaya yardımcı olması ve başta kernel (çekirdek) regresyon olmak üzere diğer tüm tahmin yöntemlerinin tutarlı tahminler sunmasından dolayı parametrik regresyon analizine alternatif olarak tercih edilmektedir (Gölveren, 2012: 7).

Parametrik olmayan regresyon yönteminin dört temel amacı vardır. Bunlar (Hardle, 1992: 7-8):

1. İki değişken arasındaki ilişkiyi ortaya çıkarmada çok yönlü bir metot sağlamak
2. Uygun bir parametrik modele ihtiyaç duymaksızın tahminler üretmek
3. Aykırı noktaların etkisini inceleyerek sahte gözlemleri ortaya çıkartmak
4. Bitişik X değerleri arasında bir interpolasyon yapmak veya kayıp değerlerin yerini doldurmak için esnek bir metot oluşturmaktır.

Parametrik olmayan regresyon yöntemi bu nedenle hem doğrusal olmayan ilişkilerin tespitinde hem de bu doğrusal olmayan ilişkilerin modellenmesinde oldukça yararlı bir metot ortaya koymaktadır.

2.1.1. Düzleştirme Kavramı

Düzleştirme teknikleri 20. yüzyılın sonlarına doğru büyük bir gelişme kaydetmesine rağmen temeli; 1857 yılında Saksonyalı ekonomist Engel’ in günümüzde regressogram olarak adlandırılan ünlü “Engelsches Gesetz” eğrisini bulmasıyla parametrik olmayan yaklaşımların ampirik analizler için önemli bir araç olarak kullanılmasına dayanmaktadır (Hardle, 1992: 17). Bu tarihten itibaren parametrik olmayan düzleştirme yaklaşımı uzun süre ihmal edilmiş ve buna bağlı olarak 20. yüzyılın ilk yarısında istatistik teorisinin matematiksel gelişimi, hesaplamalardaki kolaylık, model varsayımlarına uygunluk ve matematiksel uygunlukları nedeniyle parametrik yaklaşımlara ağırlık verilmiştir (Tezcan, 2009: 16).

Regresyon modeli $y_i = m(x_i) + e_i$, $i = 1, 2, \dots, n$ şeklinde olsun. Birinci bölümde de ifade edildiği gibi; parametrik olmayan regresyon yönteminin amacı değişkenler arasındaki ilişkiyi gösteren bilinmeyen $m(x_i)$ fonksiyonunu tahmin etmektir ve bu tahmin için kullanılan yönteme de düzleştirici (smoother) adı verilmektedir. Bir düzleştiricinin en önemli özelliği, değişkenler arasındaki ilişkinin biçimini kesin bir biçimde varsaymamasıdır. Bu özelliğinden dolayı parametrik olmayan regresyonda sıklıkla kullanılan bir araçtır (Hastie ve Tibshirani, 1990: 9). Düzleştirme fikrinin temelinde ise, verileri bir eğriye uydurmak ve daha basit fonksiyonların birleşimi olabilen esnek fonksiyonları kullanmak yatmaktadır (Yıldız, 2013: 12). Bu eğri tahmin süreci genel olarak düzeltme (smoothing) olarak adlandırılmaktadır (Hardle, 1992: 4). Ayrıca bir düzleştirici tarafından tahmin edilen eğriye düz (ya da düzgün) eğri (smooth curve) denilmektedir.

2.1.1.1. Spline Düzleştirme Yöntemi

Bir $m(\cdot)$ fonksiyonunun veriye uygunluğunun yaygın bir ölçütü artık kareleri toplamıdır (Hardle, 1992: 70). Bu toplam

$$\sum_{i=1}^{n} (Y_i - m(x_i))^2 \tag{2.1}$$

Şeklinde ifade edilir. Eğer $m(\cdot)$ herhangi bir fonksiyonel kısıtlamanın olmadığı bir eğri ise (2.1) ifadesi verilere interpolasyon uygulanarak $m(\cdot)$ yoluyla sıfıra indirgenebilir. Fakat bu durum değişkenler arasındaki ilişkinin yorumlanmasını güçleştirdiğinden, böyle bir eğri kabul edilebilir değildir (Çatalbaş, 2006: 160). Spline düzleştirme yaklaşımı, veriye uygun bir tahmin üretme amacı ile çok hızlı lokal değişimler olmaksızın bir eğri üretme amacı arasındaki rekabeti ölçerek veriye mantıksız interpolasyon uygulamayı engellemektedir (Hardle, 1992: 70). Spline düzleştirme yöntemi bunu hızlı lokal değişimlerden dolayı eğride meydana gelen pürüzlülüğü (roughness) dengeleyecek $\int (m''(x))^2\,dx$ pürüzlülük cezası (roughness penalty) ile yapmaktadır. Bu dengeleyici ile

$$S_\lambda(m) = \sum_{i=1}^{n} (Y_i - m(x_i))^2 + \lambda \int (m''(x))^2\,dx \qquad (2.2)$$

Eşitliği elde edilir ve bu eşitlik minimize edilerek $m(\cdot)$ tahmin edilir. Burada λ düzleştirme parametresini göstermekte ve $m(\cdot)$ eğrisinin pürüzlülüğü ve artıkların hatası arasındaki değişim oranını temsil etmektedir (Hardle, 1992: 71). Ayrıca $\widehat{m}(\cdot)$ fonksiyonu $S_\lambda(m)$' yi minimize eden fonksiyon olsun, bu durumda $\lambda \to \infty$ olduğunda $\widehat{m}(\cdot)$ fonksiyonu doğrusal forma yaklaşırken, $\lambda \to 0$ olduğunda ise $\widehat{m}(\cdot)$ fonksiyonu sadece Y gözlemlerinin interpolasyonu olacaktır. Yani, Y değerleri interpolasyon fonksiyonuna yaklaşacak, dolayısıyla (2.2)' deki ilk terim herhangi bir $m(x)$ fonksiyonu için sıfır olacaktır (Gölveren, 2012: 24). Bu nedenle spline düzleştirme yöntemi, uyum iyiliği ile yerel değişim arasındaki değiş tokuşu araştırma işlemi olarak kabul edilir (Çatalbaş, 2006: 161).

$[a,b] = [X_{(1)}, X_{(n)}]$ aralığında iki defa türevlenebilen tüm fonksiyonların sınıfı üzerinde $S_\lambda(m)$' nin minimizasyon problemi kübik spline $\widehat{m}_\lambda(x)$ olarak tanımlanan tek bir çözüme sahiptir. Bununla birlikte tahmin edilen $\widehat{m}_\lambda(x)$ eğrisi aşağıdaki özelliklere sahiptir (Hardle, 1992: 71):

- $\widehat{m}_\lambda(x)$ ardışık X değerleri arasında kübik polinomiyaldir.
- X_i gözlem noktalarındaki $\widehat{m}_\lambda(x)$ eğrisi ve onun ilk iki türevi süreklidir. Fakat üçüncü türevi sürekli olmayabilir.
- $X_{(1)}$ ve $X_{(n)}$ sınır noktalarında $\widehat{m}_\lambda(x)$' in ikinci türevi sıfırdır.

$\widehat{m}_\lambda(x)$ fonksiyonu bu özelliklerden ve ilk ikisini sağlarsa "kübik spline", üç özelliği bir arada sağlarsa "doğal kübik spline" olarak adlandırılır.

Spline düzleştirme yöntemi ile yapılan eğri tahmininde en büyük problem, veri değerlerini içeren kesin bir formülden ziyade çözüm olarak üstü kapalı bir şekilde bir minimizasyon probleminin ortaya konmasıdır. Bu durumda tahminin eğilimini ve tahmin edilen eğrinin gerçekte nasıl modellendiğini anlamak zordur (Tezcan, 2009: 77).

2.1.1.2. Ortogonal Seriler Düzleştirme Yöntemi

Bilinmeyen regresyon fonksiyonu $m(\cdot)$ fourier serileri açılımı kullanılarak

$$m(x) = \sum_{i=0}^{\infty} \beta_i \varphi_i(x)$$

Şeklinde yazılabilir. Burada $\{\varphi_i(x)\}_{i=0}^{\infty}$ bilinen temel fonksiyonları, $\{\beta_i\}_{i=0}^{\infty}$ ise bilinmeyen fourier katsayılarını ifade etmektedir ve amaç bilinmeyen fourier katsayılarını tahmin etmektir. Fakat bu tahmin yapılırken birtakım sorunlarla karşılaşılmaktadır. Bunlardan ilki, sonsuz sayıda ve sıfırdan farklı β_i olduğunda toplam sonsuz olacaktır. İkincisi ise sonlu sayıda gözlemden sonsuz sayıda katsayı tahmininin yapılamayacak olmasıdır. Bu yüzden fourier serilerini temsil edecek şekilde, örneklem büyüklüğü n' nin bir fonksiyonu olan N terimleri sayısı belirlenmelidir. Böylelikle seri tahmin yöntemleri üç adımda incelenir (Hardle vd., 2004: 105):

- Temel fonksiyonun seçilmesi
- Örneklem büyüklüğü n' den daha küçük bir tam sayı olan N' nin seçilmesi
- N tane bilinmeyen katsayının uygun bir metot ile tahmin edilmesi.

N bu tahmin yöntemlerinin düzleştirme parametresidir. N büyüdükçe fourier serilerindeki terim sayısı da artacak ve tahmin veriyi interpole etme yönünde gelişecektir. N' nin küçük değerleri için ise nispeten düz tahminler üretilecektir. Ayrıca β_i katsayıları hızlı fourier dönüşümü kullanılarak tahmin edilebilir (Gölveren, 2012: 26).

2.1.1.3. Medyan Düzleştirme Yöntemi

Medyan düzleştirme yöntemi, tahmin probleminin çözümü için koşullu olasılık fonksiyonu yerine koşullu medyan fonksiyonunu kullanan en yakın komşuluk yöntemi olarak tanımlanabilir. Koşulla medyan fonksiyonu uç değerlere karşı koşullu olasılık fonksiyonundan daha dirençlidir. Ayrıca medyan düzleştirme yöntemi, regresyon eğrisindeki süreksizlikleri modellemeye imkân sağlamaktadır (Hardle vd., 2004: 101). Bir medyan düzleştirici aşağıdaki gibi tanımlanır.

$$\widehat{m}(x) = med\{Y_i : i \in J_x\}$$

Burada J_x ={i: X_i, x' in en yakın komşuluğundaki k tane gözlemden biri} dir. Yani x' in en yakın komşuluğundaki k tane gözlemden biri olan X_i değerine karşılık gelen Y_i değerlerinin medyanı hesaplanmaktadır.

2.2. KERNEL REGRESYON YÖNTEMİ

Parametrik olmayan regresyon analizinde bir tahmin yöntemi olarak kernel regresyon yöntemi sıklıkla kullanılan ve bilinen bir yöntemdir. Bu regresyon yönteminde bağımlı değişkene kernel (çekirdek) adı verilen fonksiyonlarla lokal ağırlıklar verilerek eğri tahmini yapılmaktadır. Bu nedenle böyle parametrik olmayan tahmin yöntemleri kernel (çekirdek) regresyon yöntemi olarak anılmaktadır. Ayrıca kernel regresyon yönteminde kullanılan kernel (çekirdek) fonksiyonları $K(\cdot)$ ile gösterilmekte ve aşağıdaki özellikleri sağlamaktadır:

- $\forall\, u$ için, $K(u) \geq 0$
- $\int_{-\infty}^{+\infty} K(u)du = 1$
- $K(-u) = K(u)$, yani simetriktir.

Bir kernel fonksiyonuna bağlı olarak elde edilen ağırlık,

$$W_i(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)}$$

Şeklinde gösterilir. Burada $W_i(x) = W(x, x_i)$, $|x - x_i|$ uzaklığına bağlı olarak i. gözleme verilen ağırlığı ifade etmektedir. Bu ağırlıklar, $|x - x_i|$ uzaklığı arttıkça küçülmekte, $|x - x_i|$ uzaklığı azaldıkça da büyümektedir. Ayrıca $K(\cdot)$ kullanılan kernel (çekirdek) fonksiyonunu, h ise pencere (bant) genişliğini (bandwidth) ifade etmektedir ve $\sum_{i=1}^{n} W_i(x) = 1$' dir.

Burada en önemli husus kernel (çekirdek) fonksiyonunun ve h pencere genişliğinin seçimidir. Fakat pencere genişliği seçimiyle karşılaştırıldığında kernel fonksiyonunun seçimi daha az önem arz etmektedir. Kernel fonksiyonunun şeklinde meydana gelen değişikliğin yapılan tahmine olan etkisi, pencere genişliğinde meydana gelen değişikliğin neden olacağı etkiden daha azdır (Gölveren, 2012: 13).

Uygulamada çeşitli kernel fonksiyonları bulunmaktadır. Bu fonksiyonlardan en sık kullanılanları Tablo 2.1' de verilmiştir (Hardle vd., 2004: 41).

Tablo 2.1: Kernel Fonksiyonları

Kernel Fonksiyonu	$\boldsymbol{K(u)}$
Uniform	$\frac{1}{2}, \quad \lvert u\rvert \leq 1$
Triangle	$1-\lvert u\rvert, \quad \lvert u\rvert \leq 1$
Epanechnikov	$\frac{3}{4}(1-u^2), \quad \lvert u\rvert \leq 1$
Quartic (Biweight)	$\frac{15}{16}(1-u^2)^2, \quad \lvert u\rvert \leq 1$
Triweight	$\frac{35}{32}(1-u^2)^3, \quad \lvert u\rvert \leq 1$
Cosine	$\frac{\pi}{4}\cos(\frac{\pi}{2}u), \quad \lvert u\rvert \leq 1$
Gaussian	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$

2.2.1. Nadaraya-Watson Tahmin Yöntemi

Nadaraya-Watson tahmin yöntemi Nadaraya (1964) ve Watson (1964) tarafından rassal tasarım durumu için ileri sürülen bir kernel regresyon tahmin yöntemidir. Bu tahmin yönteminde temel amaç, bağımlı değişkenin aldığı değerlerin lokal ağırlıklı ortalamasını (locally weighted averaging) alarak bilinmeyen regresyon fonksiyonu $m(\cdot)$'i tahmin etmektir. Nadaraya ve Watson tarafından ileri sürülen bu tahmin aşağıdaki gibidir (Hardle vd., 2004: 89).

$$\widehat{m}(x) = \frac{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right)}$$

Pencere genişliği h, $\widehat{m}(x)$ tahmininin düzgünlüğünün derecesini ifade etmektedir. Bu durum h sırasıyla sıfıra ve sonsuza giderken limitleri alındığında kolaylıkla görülebilir. Bir X_i gözlem noktasında $h \to 0$ iken $\widehat{m}(X_i) \to Y_i$ olmakta, keyfi bir x noktasında ise $h \to \infty$ iken $\widehat{m}(x) \to \bar{Y}$ olmaktadır. Bu iki limit durumu açıkça göstermektedir ki h düzleştirme parametresi sıfıra ne çok hızlı ne de çok yavaş yakınsamalıdır (Hardle ve Linton: 1994: 14).

2.2.2. Gasser-Müller Tahmin Yöntemi

İntegral tahmin yöntemi olarak ta bilinen bu tahmin yöntemi Gasser ve Müller (1984) tarafından ileri sürülmüştür. $[a, b]$ aralığından elde edilen $x_{(i)},\ i = 1, \dots, n$ sıralı tasarım noktaları için, $K_h(\cdot) = \frac{K(\cdot/h)}{h}$ olmak üzere, Gasser ve Müller tarafından önerilen ağırlık fonksiyonu;

$$W_{hi}^{GM}(x) = n \int_{s_{i-1}}^{s_i} K_h(x-u)du$$

Şeklinde gösterilmektedir. Burada $s_i = \left(x_{(i)} + x_{(i+1)}\right)/2\,, s_0 = a, s_{n+1} = b$ olarak ifade edilir. Böyle bir ağırlık fonksiyonuna sahip Gasser-Müller tahmin yöntemi,

$$\widehat{m}(x) = \frac{1}{n}\sum_{i=1}^{n} W_{hi}^{GM}\,(x)Y_i$$

Şeklinde ifade edilmektedir (Hardle vd., 2004: 90-91).

Nadaraya-Watson ve Gasser-Müller tahmin yöntemleri karşılaştırıldığında, Nadaraya-Watson yöntemi özellikle sınır bölgelerinde önemli ölçüde bir yanlılığa sahiptir. Gasser-Müller yöntemi ise bu yanlılığı azaltmakta fakat buna karşılık varyansı arttırmaktadır (Tezcan, 2009: 56).

2.3. KERNEL REGRESYONDA PENCERE GENİŞLİĞİ SEÇİMİ

Kernel regresyon yöntemiyle tahmin yaparken en çok dikkat edilmesi gereken husus, yapılan tahminin düzgünlüğünü kontrol eden h pencere genişliğinin (düzleştirme parametresi) seçimidir. Literatürde optimum pencere genişliğini elde edebilmek için çeşitli yöntemler geliştirilmiş olsa da bu yöntemlerle elde edilen pencere genişliği değerlerinden hangisinin optimum pencere genişliği olduğu hakkında kesin olarak bir şey söylenemememektedir.

Pencere genişliği seçiminde esasında elde edilen tahminin yanlılığı ve varyansı arasında bir denge (trade-off) bulunmaktadır. Geniş bir pencere genişliği seçildiğinde tahmin aşırı düzleştirilmiş (oversmooth) olacaktır. Benzer şekilde dar bir pencere genişliği seçildiğinde bu kez tahmin az düzleştirilmiş (undersmooth) olacaktır (Hardle, 1992: 24). Ayrıca, geniş bir pencere genişliği tahminin varyansını azaltırken yanlılığını arttırmaktadır. Benzer şekilde dar bir pencere genişliği de tahminin yanlılığını azaltırken varyansını arttırmaktadır. Bu nedenle tahminin yanlılığı ve varyansı arasındaki bu dengeyi en iyi şekilde sağlayacak bir pencere genişliği seçilmelidir.

2.3.1. Çapraz Geçerlilik Yöntemi

Optimal pencere genişliği seçimi için sıklıkla kullanılan bir yöntem olan çapraz geçerlilik (CV- Cross Validation) yöntemi Wahba ve Wold (1975) tarafından önerilmiştir. Bu yöntem, bir gözlemi dışarıda bırakma (leave one out) olarak ifade edilen ve bağımlı değişkenin her bir değerinin geri kalan veri tarafından tahmin edilmesi düşüncesine dayanmaktadır. Diğer bir ifadeyle bu yöntemde; örnek büyüklüğü n ve gözlem noktaları $\{x_i, y_i\}_{i=1}^{n}$ olmak üzere, bu (x_i, y_i) gözlemlerinden biri veri setinden çıkartılır ve kalan *(n-1)* adet gözleme bağlı olarak bilinmeyen regresyon fonksiyonu $m(\cdot)$ için kareli artıkların toplamını minimum yapacak düzeltme parametresi (pencere genişliği) seçilmektedir. Bu pencere genişliği seçim yöntemi aşağıdaki gibi ifade edilmektedir.

$$CV(h) = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \widehat{m}_{h,-i}(X_i)\right\}^2 w(X_i) \tag{2.3}$$

Burada amaç (2.3) eşitliğini minimum yapacak h parametresini seçmektir. Ayrıca (2.3) eşitliğinde *"-i"* i. gözlemin veriden çıkartıldığını, $w(X_i)$ ağırlıkları, $\widehat{m}_{h,-i}(X_i)$ ise bir gözlem dışarıda bırakılarak tahmin edilen eğriyi göstermekte ve aşağıdaki gibi ifade edilmektedir (Hardle vd., 2004: 114).

$$\widehat{m}_{h,-i}(X_i) = \frac{\sum_{j\neq i} K_h\left(X_i - X_j\right)Y_j}{\sum_{j\neq i} K_h\left(X_i - X_j\right)}$$

2.3.2. Plug-in Yöntemi

Plug-in yöntemi, optimal pencere genişliği seçiminde kullanılan hata kareleri ortalamasının asimptotik bir genişlemesine dayanmaktadır (Hardle, 1992: 189). Burada bilinmeyen fonksiyonların yerine onların tutarlı tahmincileri olan $\widehat{m}_h''(x)$, $\hat{\sigma}_h^2(x)$, $\hat{f}_h'(x)$ ve $\hat{f}_h(x)$ kullanılmaktadır. Böylece $\widehat{m}_{\hat{h}_{opt}}(x)$ kullanılarak $m(x)$ tahmin edilir (Hardle ve Linton, 1994: 25).

Pencere genişliği seçiminde çapraz geçerlilik yöntemine göre daha düşük değişkenlik göstermesi ve iyi bir yakınsama oranına sahip olması plug-in yönteminin en önemli özelliğidir (Tezcan, 2009: 83). Bunun yanında diğer pencere genişliği seçim yöntemleri kadar etkin sonuçlar vermesine karşın ilk adımda nasıl bir pencere genişliği seçileceğiyle ilgili büyük bir belirsizliğe sahiptir. Bir diğer dezavantajı ise optimal pencere genişliği seçiminin belli düzgünlük sınıflarına (iki defa türevlenebilir regresyon fonksiyonu gibi) sınırlandırılmış olmasıdır (Hardle, 1992: 189).

2.4. EN YAKIN KOMŞULUK YÖNTEMİ (K-NN)

Kernel regresyon yöntemi ile bilinmeyen regresyon fonksiyonu $m(\cdot)$' nin tahmini yapılırken, x' in h pencere genişliği tarafından belirlenen sabit bir komşuluğundaki y değerlerinin ağırlıklı ortalamaları alınmaktaydı. K-en yakın komşuluk (K-Nearest Neighborhood) yöntemi ile tahmin yaparken de benzer şekilde x' in komşuluğundaki y değerlerinin ağırlıklı ortalaması alınmakta fakat buradaki en önemli fark kernel regresyonda sabit olan pencere genişliğinin k-en yakın komşuluk yönteminde değişken olmasıdır. K-en yakın komşuluk tahmini,

$$\widehat{m}(x) = \frac{1}{n}\sum_{i=1}^{n} W_{ki}(x)Y_i$$

Şeklinde gösterilir. Burada $\{W_{ki}(x)\}_{i=1}^{n}$ ağırlıkları göstermekte ve

J_x ={i : X_i, x' in en yakın komşuluğundaki k tane gözlemden biridir.} olmak üzere,

$$W_{ki} = \begin{cases} \frac{n}{k}, & i \in J_x \ ise \\ 0\,, & diğer\ durumlarda \end{cases}$$

Şeklinde tanımlanmaktadır. Ayrıca bu tahmin yönteminde k, düzleştirme parametresi olarak ifade edilmektedir (Hardle vd., 2004: 99).

Bu tahmin yönteminde en önemli husus k düzleştirme parametresinin seçimidir. Eğer k=*1* seçilirse veriler tekrardan üretilmiş olacaktır. Eğer k örneklem büyüklüğüne eşit veya ondan daha büyük seçilirse bu durumda da bağımlı değişkenin aldığı değerlerin ortalaması elde edilmiş olacaktır. Bu nedenle k düzleştirme parametresinin hem tahmin edilen eğrideki pürüzlülüğü azaltıp düzgün bir tahmin elde edecek kadar büyük, hem de yanlılığı azaltacak kadar küçük seçilmesi gerekmektedir.

2.5. LOESS REGRESYON YÖNTEMİ

Lokal polinomial regresyon yöntemi veya lokal olarak ağırlıklandırılmış regresyon yöntemi olarak ta bilinen bu yöntem ilk olarak Cleveland (1979) tarafından önerilmiştir. Dağılım grafiği düzleştirmesi esasına dayanan bu yöntem "LOcally WEighted Scatterplot Smoothing" ifadesinin kısaltılmış biçimi olan "LOWESS" adı ile anılmaktadır. Daha sonra Cleveland ve Devlin (1988) tarafından LOWESS yönteminin metodolojisi geliştirilerek çok değişkenli verilere uyarlanmış ve LOESS adını almıştır.

Dağılım grafiği düzleştirmesinin ilk örneği Ezekiel tarafından 1941 yılında verilmiştir. Bu yöntemde, noktalar x_i değerlerine göre gruplandırılmakta ve her bir grup için x_i ve y_i değerlerinin ortalaması alınarak bulunan yeni $(x_{i_{ort}}, y_{i_{ort}})$ değerleri dağılım grafiği üzerinde işaretlenmektedir. Böylelikle düzleştirme işlemi yapılmaktadır. Daha sonra Clark (1977) tarafından önerilen dağılım grafiği düzleştirme yönteminde ise, ardışık noktalar düz bir hat boyunca interpole edilmekte ve ardından bu noktaların bir ağırlık fonksiyonuyla konvolüsyonu yapılarak düzleştirilmektedir (Cleveland, 1979: 829). Bu yöntemlerde yapılan düzleştirmeler sapan değerlere karşı oldukça hassas davranmaktadır. LOESS yönteminde yapılan düzleştirme sürecinde ise iterasyonlar kullanılarak düzleştirme sapan değerlere karşı oldukça dirençli hale getirilmektedir. Bu yöntemde yapılan düzleştirme süreci aşağıdaki gibidir.

Düzleştirme sürecine geçmeden önce bazı kavramların tanımların yapılaması gerekmektedir. Bir dağılım grafiği üzerinde bulunan noktalar $(x_i, y_i),\ i = 1, 2, \dots, n$ olsun. Bu dağılım grafiği üzerinde yapılan düzleştirme sonucu elde edilen noktalar da $(x_i, \hat{y}_i)$ olsun. Bu $(x_i, \hat{y}_i)$ noktalarından oluşan yeni düzen dağılım grafiği düzleştirmesi olarak ifade edilmektedir. $(x_i, \hat{y}_i)$ noktaları x_i noktasında düzleştirilmiş nokta, $\hat{y}_i$ değerleri de x_i noktasında uydurulmuş değer (fitted value) olarak adlandırılmaktadır.

W bir ağırlık fonksiyonu olsun ve aşağıdaki özellikleri sağlasın;

- $|x| < 1$ için, $W(x) > 0$
- $|x| \geq 1$ için, $W(x) = 0$
- $x \geq 0$ için, $W(x)$ artmayan bir fonksiyondur
- $W(x) = W(-x)$

Her bir i için h_i, x_i' nin x_j' den r. en yakın komşuluğuna olan uzaklığı olsun. Diğer bir ifadeyle h_i, $|x_i - x_j|$ $(j = 1, \dots, n)$ arasındaki r. en küçük sayı olsun. $k = 1, \dots, n$ için,

$$w_k(x_i) = W\left(\frac{x_k - x_i}{h_i}\right)$$

olsun. Düzleştirme sürecinde aşağıdaki adımlar izlenir (Cleveland, 1979: 830-831).

1. (x_k, y_k) için, $w_k(x_i)$ ağırlıkları ile ağırlıklandırılmış en küçük kareler yardımıyla x_k üzerinde y_k' nın d. dereceden polinomial regresyonundaki parametrelerin $\hat{\beta}_j(x_i),\ i = 1, \dots, d$ tahminleri her bir i değeri için hesaplanır. Yani;

$$\min_{\beta_0, \beta_1, \dots, \beta_n} \sum_{k=1}^{n} w_k(x_i)(y_k - \beta_0 - \beta_1 x_k - \cdots - \beta_d x_k{}^d)^2$$

minimizasyonunu sağlayacak β_j değerleri elde edilir. Böylece,

$$\hat{y}_i = \sum_{j=0}^{d} \hat{\beta}_j(x_i) x_i{}^j = \sum_{k=1}^{n} r_k(x_i) y_k$$

elde edilir. Burada $r_k(x_i)$, y_i' ye $(i = 1, \dots , n)$ bağlı değildir, sadece regresyondan elde edilen y_k noktalarına ait katsayıları ifade etmektedir.

2. W aşağıdaki gibi tanımlanan çift kare (bisquare) ağırlık fonksiyonu olsun.

$$W(x) = \begin{cases} (1 - x^2)^2, & |x| < 1 \\ 0 \quad , & |x| \geq 1 \end{cases}$$

Uydurulmuş değerlerden elde edilen artıklar $e_i = y_i - \hat{y}_i$ ve artıkların mutlak değerinin $|e_i|$ medyanı s olsun. Bu durumda dirençlilik ağırlıkları (robustness weights);

$$\delta_k = W\left({}^{e_k}/_{6s}\right)$$

şeklinde ifade edilir.

3. (x_k, y_k) noktasında $\delta_k w_k(x_i)$ ağırlığı ile ağırlıklandırılmış en küçük kareler yöntemi kullanılarak $d.$ dereceden polinomial tahminiyle her bir i için yeni $\hat{y}_i$ hesaplanır.

4. İkinci ve üçüncü adımlar toplam t kere tekrarlanır. Son olarak elde edilen $\hat{y}_i$ değerleri dirençli lokal olarak ağırlıklandırılmış regresyonun (robust locally weighted regression) uydurulmuş (fitted) değerleridir.

İkinci, üçüncü ve dördüncü adımdaki iteratif süreç, aykırı gözlemlere karşı daha dirençli tahminler elde etmek içi yapılmaktadır. Çünkü büyük artıklar her defasında küçük dirençlilik ağırlıkları δ_k ile ağırlıklandırılacak ve düzleştirilen noktaların belirlenmesinde diğer gözlemler kadar etkili olamayacaktır.

LOESS regresyon yöntemi aykırı gözlemlere karşı olan bu güçlü düzleştirme özelliğinden dolayı kernel regresyon tahmin yöntemlerinde karşılaşılan yanlılığı azaltarak üstünlük sağlamaktadır.

Ayrıca LOESS regresyon yöntemi bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi grafiksel olarak özetlemektedir. Bu nedenle bu yöntemi diğerlerinden ayıran en önemli özellik, "verinin kendi kendisine konuşmasına izin vermesi" olarak ifade edilmektedir. Bu yöntemle tahmin edilen eğri, veri içerisinde olabilecek herhangi bir yapı (fonksiyonel ilişki) hakkında sıkı ön varsayımlarda bulunmaktan ziyade ampirik olarak elde edilmektedir. Bu nedenle geleneksel istatistiksel yöntemlerle incelendiğinde kolaylıkla gözden kaçırılabilecek karmaşık ilişkiler, LOESS yöntemi ile incelendiğinde sıklıkla ortaya çıkarılmaktadır (Jacoby, 2000: 578).

LOESS regresyon yönteminde uygulamaya geçmeden önce belirlenmesi gereken dört temel öğe bulunmaktadır. Bunlar; ağırlık fonksiyonu W, polinomialin derecesi d, iterasyon sayısı t ve düzleştirme parametresi olan span değeridir. W ağırlık fonksiyonu için tavsiye edilen fonksiyon,

$$W(x) = \begin{cases} (1 - x^3)^3, & |x| < 1 \\ 0 \quad , & |x| \geq 1 \end{cases}$$

şeklinde tanımlanan Tricube ağırlık fonksiyonudur. Çünkü bu ağırlık fonksiyonu hata varyansının tahmini için χ^2 dağılımı yaklaşımını iyileştirmektedir. İterasyon sayısı t için ise, gerçek ve yapay veri setleri ile yapılan çok sayıda çalışma iki iterasyonun neredeyse tüm durumlar için yeterli olduğunu göstermektedir (Cleveland, 1979: 833). Polinomialin derecesinin seçiminde yapılan çalışmalarda genellikle $d = 1$ (doğrusal polinomial) ve $d = 2$ (karesel polinomial) kullanılmaktadır. Üç ve üzeri polinomial derecesi ise tahmine (fitting) fazla katkı sağlamamaktadır. Ayrıca veri setinde uç değerler varsa doğrusal polinomial $(d = 1)$ tahminde yapay bir düzlük meydana getirecektir. Bu nedenle karesel polinomial $(d = 2)$ hemen her durum için yeterli sonuçlar vereceği için tercih edilmektedir (Keele, 2008: 36; Jacoby, 2000: 586-587). Düzleştirme parametresi olan span değeri ise ayrı bir başlık altında incelenecektir.

Ayrıca LOESS regresyon yönteminde serbestlik derecesi, parametrik regresyondan farklı olarak bir tam sayı olmak zorunda değildir (Cleveland, 1979: 835). Eşdeğer parametre sayısı (equivalent number of parameters) olarak ifade edilen serbestlik derecesi negatif olmayan bir sayı olmakla birlikte elde edilen LOESS eğrisine benzer bir eğri parametrik bir yöntemle elde edilmek istendiğinde gerekli olan parametre sayısını ifade etmektedir (Jacoby, 2000: 595).

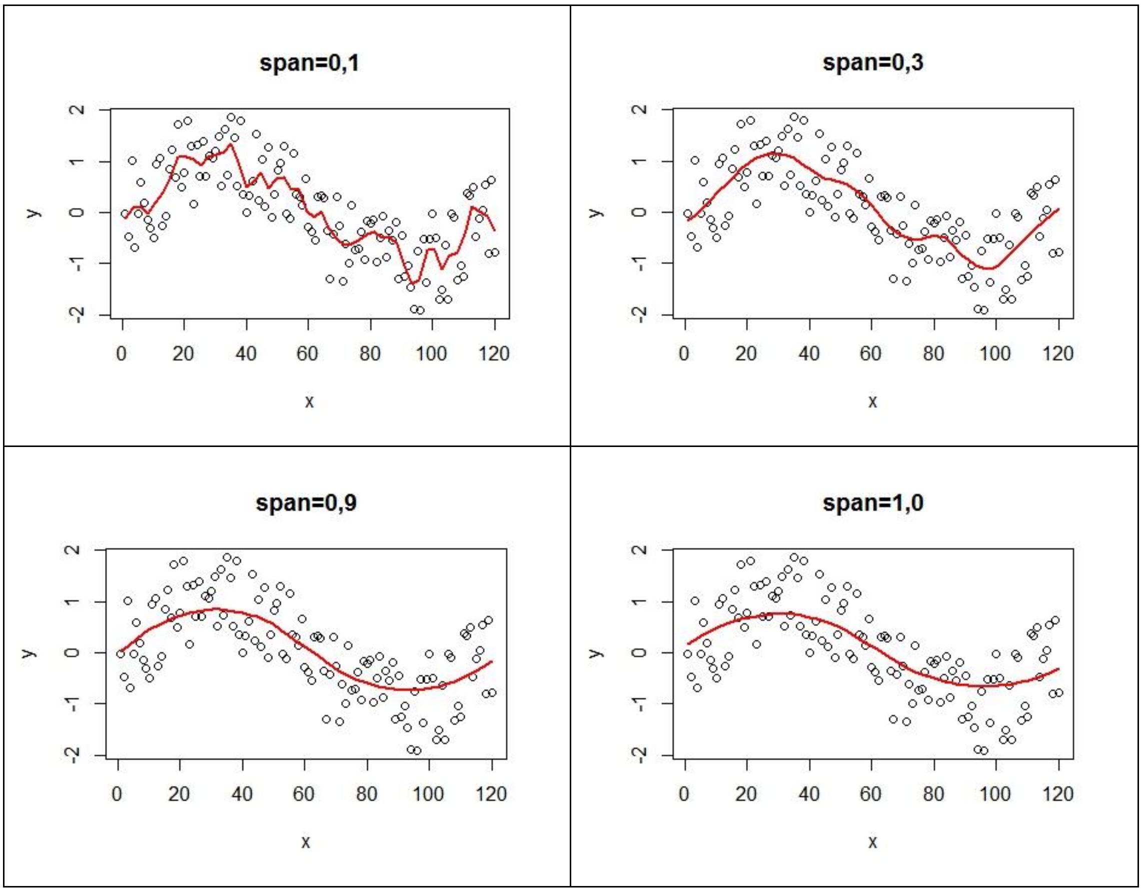
2.5.1. Span Değeri Seçimi

Parametrik olmayan regresyon yöntemlerinde en önemli konulardan biri düzleştirme parametresinin seçimidir. LOESS regresyon yönteminde de düzleştirme parametresi olan span değerinin seçimi büyük önem taşımaktadır. Burada span değeri diğer parametrik olmayan yöntemlerde seçilen düzleştirme parametresine benzer bir rol oynamaktadır. Fakat diğer parametrik olmayan yöntemlerden farklı olarak her bir pencereye kaç gözlem düşeceğinden ziyade her bir pencereye düşecek gözlemler tüm gözlemlerin bir oranı olarak belirlemektedir.

Span 0-1 aralığında değer alan bir düzleştirme parametresidir. Bu değer belirlenirken varyans ve yanlılık arasında en iyi dengeyi sağlayacak şekilde seçilmelidir. Çünkü çok büyük span değeri varyansı küçültmekle birlikte yanlılığı artıracaktır. Benzer şekilde çok küçük span değeri de yanlılığı azaltırken varyansın şişmesine neden olacaktır. Span değeri seçiminde kullanılan en kolay ve en yaygın yöntem görsel deneme yanılma yöntemidir. Bu yöntem de span değeri başlangıçta 0,5 olarak belirlenir ve elde edilen eğri eğer pürüzlü (rough) ise span değeri 0,1 arttırılır. Bu şekilde devam ederek en düz eğri elde edilmeye çalışılır (Keele, 2008: 31-34). Ayrıca span değerinin 0,5 olması her bir lokal regresyon bölgesinde eğriyi tahmin etmek için toplam verinin %50' sinin kullanıldığını ifade etmektedir (Jacoby, 2000: 584).

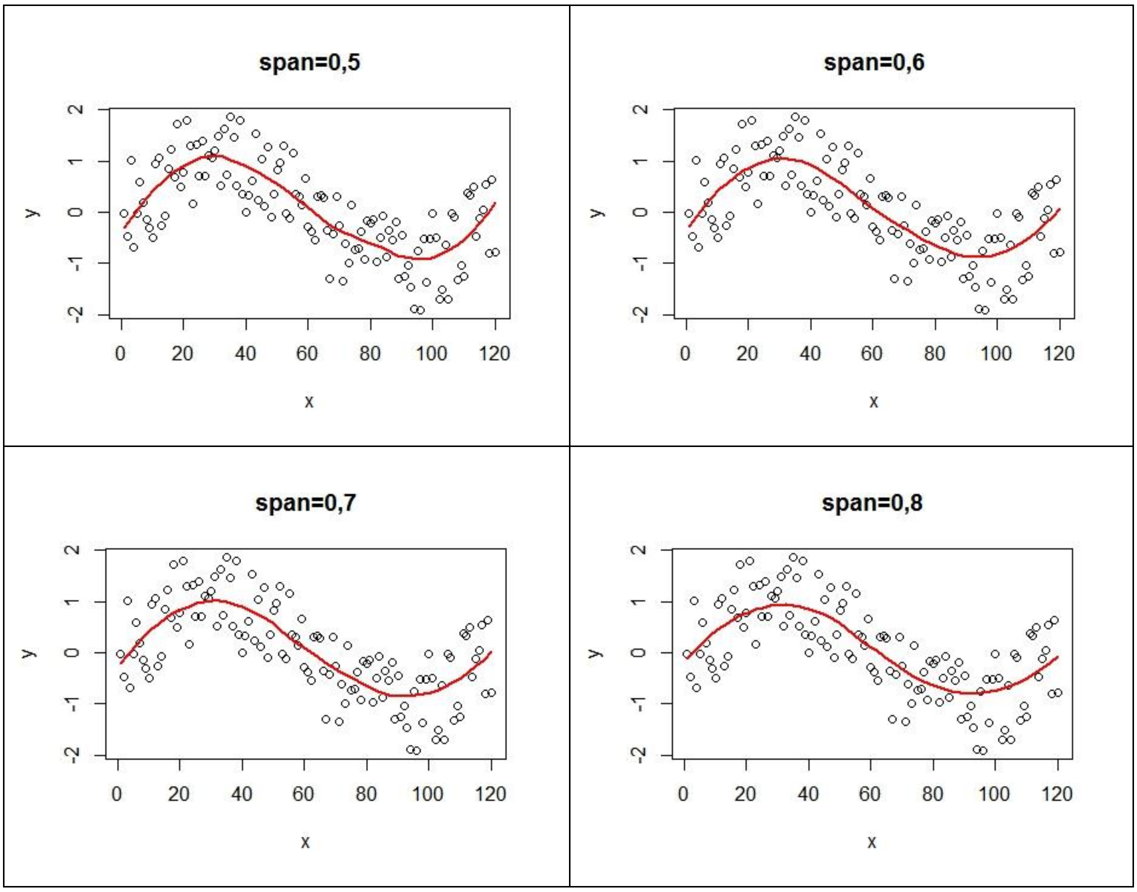
Çalışmanın üçüncü bölümünde kullanılacak veri setinde çok fazla veri olduğundan span değerindeki ufak değişimler elde edilen eğride kayda değer bir değişim meydana getirmemektedir. Bu nedenle bu bölümde, span değerinde meydana gelen değişimin tahmin edilen eğriyi nasıl etkilediğini incelemek ve daha iyi anlaşılmasını sağlamak amacıyla daha az verinin bulunduğu örnek bir veri setiyle uygulama yapılmıştır. Bu amaçla R paket programında yapay bir veri seti üretilmiş ve farklı span değerleri ile elde edilen LOESS eğrileri incelenmiştir. Bu yapay veri seti yardımıyla kurulan modelin uç span değerleri ile elde edilen düzleştirilmiş eğrileri Şekil 2.1' de verilmiştir.

Şekil 2.1: Uç Span Değerleri ile Elde Edilen LOESS Eğrileri



Şekil 2.1' de incelendiğinde, küçük span değerleri ile elde edilen eğrilerin oldukça pürüzlü (rough) diğer bir ifadeyle az düzleştirilmiş (undersmooth) eğriler olduğu, büyük span değerleri ile elde edilen eğrilerin ise aşırı derecede düz (oversmooth) eğriler olduğu görülmektedir. Ayrıca küçük span değerleri Şekil 2.1' de de görüldüğü gibi yanlılığı azaltmakta fakat eğride meydana gelen aşırı değişkenlik varyansı şişirmektedir. Benzer şekilde büyük span değerleri ile elde edilen eğrilerde bu değişkenlik azalarak varyansı küçültmekte fakat eğride ortaya çıkan aşırı düzlük tahminin yanlı olmasına neden olmaktadır. Uygun span değerini belirlemek için daha önce de değinildiği gibi 0,5 değerinden başlayıp görsel deneme yanılma yoluyla eğrinin durumuna göre 0,1 arttırılarak veya azaltılarak en uygun eğri elde edilmeye çalışılır. Bu yöntemle elde edilen LOESS eğrileri Şekil 2.2' de verilmiştir.

Şekil 2.2: Farklı Span Değerleri ile Elde Edilen LOESS Eğrileri



En uygun eğriyi belirlerken, eğrinin dağılım grafiği üzerinde bulunan veri bulutundaki noktaların olabildiğince ortasından geçecek şekilde seçilmesine özen gösterilmelidir. Fakat bunu yaparken eğrinin pürüzlü olmamasına dikkat edilmelidir. Şekil 2.2 incelendiğinde span değeri 0,5 olan eğri az da olsa pürüzlü olduğundan daha düz eğriler elde edilmeye çalışılmalıdır. Span değeri 0,6 ve 0,7 olan eğriler veri bulutundaki noktaları ortasından geçmeye daha çok meyilliler ve birbirlerine oldukça yakın sonuçlar vermektedirler. Fakat span değeri 0,6 olan eğri büküm noktalarında veri bulutuna daha iyi adapte olduğundan en uygun span değeri olarak 0,6 seçilmelidir. Span değeri 0,8 olan eğri ise diğerlerine nazaran veri bulutundaki noktaların ortasından geçmeye daha az meyilli olduğundan en uygun span değeri olarak ifade edilemez.

2.5.2. Uyum İyiliği

Parametrik regresyon analizinde bir uyum iyiliği ölçütü olan belirlilik katsayısı R^2, bağımlı değişkende meydana gelen değişimin ne kadarının bağımsız değişken veya değişkenler tarafından açıklandığını göstermektedir. Parametrik olmayan LOESS regresyon yönteminde bu durum tam olarak böyle olmamakla birlikte R^2 değeri aşağıdaki gibi hesaplanmaktadır.

$$R^2_{LOESS} = \frac{\sum_{i=1}^{n}\left(\hat{g}(x_i) - \overline{\hat{g}(x_i)}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{Y})^2}$$

Burada; $\hat{g}(x_i)$ *i*. gözlem için tahmin edilen değerini (fitted value), $\overline{\hat{g}(x_i)}$ tahmin edilen değerlerin (fitted values) ortalamasını, y_i *i*. gözlem için bağımlı değişkenin gerçek değerini, $\bar{Y}$ ise bağımlı değişken için örneklem ortalamasını ifade etmektedir.

LOESS regresyon yönteminde R^2_{LOESS} değeri parametrik regresyonda olduğu gibi açıklanan varyans olarak yorumlanamaz. Çünkü R^2_{LOESS}, bağımsız değişkenler tarafından açıklanan varyansı değil uydurulan değerler (fitted value) tarafından açıklanan varyansı ifade etmektedir. Bununla birlikte tahmin sürecinde dirençlilik ağırlıkları kullanıldığında yanlış veya anlamsız R^2_{LOESS} değerleri ortaya çıkabilir. Örneğin, veri setinde sapan değerler varsa bu değerler bağımlı değişkenin kareler toplamını şişirerek olduğundan daha büyük hatta bazı durumlarda 1' den büyük R^2_{LOESS} değerlerinin elde edilmesine neden olabilir. Ayrıca diğer parametrik olmayan yöntemlerde olduğu gibi LOESS regresyon yönteminde de veriye özel bir model uydurulmaya çalışılmadığından bağımlı değişkenin açıklanan varyansı daha az önem arz etmektedir. Tüm bu sebeplerden dolayı LOESS regresyon yöntemiyle yapılan çalışmalarda R^2_{LOESS} değeri nadiren rapor edilmektedir (Jacoby, 2000: 594).

ÜÇÜNCÜ BÖLÜM

TÜRKİYE HANEHALKI ZARURİ TÜKETİM HARCAMALARI İLE GELİRİ ARASINDAKİ İLİŞKİNİN LOESS REGRESYON YÖNTEMİ İLE İNCELENMESİ

3.1. ÇALIŞMANIN AMACI

Hanehalkı harcamalarının incelenmesi; hanelerin sosyo-ekonomik yapılarının anlaşılması, geleceğe dönük politika üretimi ve ekonomik dinamiklerin nasıl şekillendiğinin ortaya çıkarılmasında oldukça önem taşımaktadır. Bu anlamda harcama ve gelir arasındaki ilişkinin incelenmesi ekonomik dinamiklerin anlaşılması konusunda temel dayanak noktası olmaktadır.

Bu kapsamda yapılan çalışmalarda; Sarımeşeli (1999) yaptığı çalışmada harcama eğilimlerini belirlemek amacıyla harcama ve gelir arasındaki ilişkiyi yerleşim yeri sosyo-demografik değişkeni ile birlikte, farklı modeller kullanarak incelemiştir. Emeç (2001) yaptığı çalışmada bölgelere göre hanehalkı reisinin harcama ve geliri arasındaki ilişkiyi çeşitli sosyo-demografik özelliklerle birlikte incelemiştir. Beyaz (2007) yaptığı çalışmada hanehalkı zaruri harcamaları ve geliri arsındaki ilişkiyi, eğitim düzeyi ve yerleşim yeri sosyo-demografik değişkenleri ile incelemiştir. Zortuk (2009) yaptığı çalışmada Dumlupınar üniversitesi öğrencileri gelir ve harcamaları arasındaki ilişkiyi cinsiyet değişkeni ile birlikte incelemiştir. Avcı (2011) yaptığı çalışmada bayanların hanehalkı harcama ve geliri arasındaki ilişkiyi çeşitli sosyo-demografik değişenler ile birlikte incelemiştir.

Yine bu kapsamda yapılan fakat parametrik olmayan yöntemlerin kullanıldığı çalışmalarda; Bieren ve Pott-Butter (1990) ve Delgado ve Miles (1997) hanehalkı zaruri tüketim harcamaları ve hanehalkı kullanılabilir geliri arasındaki ilişkiyi, hanehalkı tipi sosyo-demografik değişkeni ile birlikte incelemiş ve parametrik olmayan kernel (çekirdek) regresyon yöntemini kullanmıştır. Gölveren (2012) yaptığı çalışmada ise hanehalkı zaruri tüketim harcamaları ve hanehalkı kullanılabilir geliri arasındaki ilişkiyi, çeşitli sosyo-demografik özelliklerle birlikte incelemiş ve parametrik olmayan kernel (çekirdek) regresyon yöntemini kullanmıştır.

Yapılan çalışmalar yöntemsel açıdan incelendiğinde, parametrik yöntemlerin baskınlığı dikkat çeken önemli bir durumdur. Her ne kadar parametrik yöntemlerin uygulanması bazı sıkı ön varsayımlar gerektirse de parametrik yöntemlerin uygulama sonrasında tahmin amaçlı kullanımının daha kolay olması ve araştırmacıların tahmin sürecinde kolaylıkla senaryo üretebilmeleri ve buna dayalı talep öngörülerini kolaylıkla hesaplayabilmeleri araştırmacıları bu yöntemleri kullanmaya teşvik etmiş olabilir.

Bu araştırmanın amacı ise; daha önce yapılan hanehalkı araştırmalarında çoğunlukla kullanılan ve sıkı ön varsayımların olduğu parametrik yöntemlerin aksine parametrik olmayan bir tekniği kullanarak hem hanehalkı çalışmalarına farklı bir bakış açısı kazandırmak hem de varsa veri yapısı içerisinde bulunan gizli yapıyı ortaya çıkartmaktır. Bu yapıyı ortaya çıkartabilmek için daha önce de değinildiği gibi “verinin kendi kendisine konuşmasına izin verme” olarak da ifade edilen ve parametrik olmayan bir yöntem olan LOESS regresyon yöntemi kullanılmıştır.

3.2. ÇALIŞMADA KULLANILAN VERİ SETİ

Hanehalkı Bütçe Anketleri, hanelerin sosyo-ekonomik yapıları, yaşam düzeyleri ve tüketim kalıpları hakkında bilgi veren, uygulanan sosyo-ekonomik politikaların test edilmesi amacıyla kullanılan en önemli kaynaklardan biridir (Gölveren, 2012: 30). Bu bağlamda, çalışmada kullanılacak olan veri seti 2012 Hanehalkı Bütçe Anketidir. Bu anket, 1 Ocak – 31 Aralık 2012 tarihleri arasında bir yıl süreyle ve her ay değişen 1104, yıllık toplam da ise 13248 örnek hanehalkına uygulanarak elde edilmiştir.

Çalışmada; *hanehalkı zaruri tüketim harcamaları, hanehalkı kullanılabilir geliri* ve *OECD tanımlı hanehalkı büyüklüğü* değişkenleri kullanılmıştır. Seçilen bu değişkenlerden *hanehalkı zaruri tüketim harcamaları* bağımlı değişken*, hanehalkı kullanılabilir geliri* ve *OECD tanımlı hanehalkı büyüklüğü* değişkenleri ise bağımsız değişken olarak ele alınmıştır. Ele alınan bu değişkenlerden *hanehalkı zaruri tüketim harcamaları* ve *hanehalkı kullanılabilir geliri* değerleri yıllık ortalamaları ifade etmektedir ve sayısal olarak büyük değerler aldığından logaritmaları alınarak kullanılmıştır. Bu değişkenlerin tanımları ise Tablo 3.1’ de verilmiştir.

Tablo 3.1: Çalışmada Kullanılan Hanehalkı Değişkenlerinin Tanımları

Hanehalkı Zaruri Tüketim Harcamaları	Hanehalkı zaruri tüketim harcama kalemlerinin tümü toplanarak elde edilmiştir.
Hanehalkı Kullanılabilir Geliri	Hanehalkı fertlerinden gelir getirenlerin çalıştıkları işlerden kazandıkları gelir, sermaye ve mülk (ücret, kâr, faiz, kira) geliri ile emekli maaşı, dul-yetim aylıkları ve yaşlılara yapılan ödemeler, karşılıksız burs vb. transfer gelirleri gibi parasal gelirleri ve ayni gelirlerin toplamı kişisel kullanılabilir gelir kapsamı içinde yer almaktadır. Hanede yer alan her bir ferdin kişisel yıllık kullanılabilir gelirinin toplamından hanehalkı kullanılabilir gelirine ulaşılmıştır.
OECD Tanımlı Hanehalkı Büyüklüğü	Hanedeki ilk yetişkin için 1, 14 ve daha yukarı yaştaki fertler için 0,5, 14 yaşından küçük fertler için 0,3 değerleri dikkate alınarak hesaplanan hanehalkı büyüklüğüdür.

3.3. LOESS REGRESYON YÖNTEMİNİN UYGULANMASI

Çalışmanın bu kısmında parametrik olmayan regresyon yöntemiyle tahmin yapmak amacıyla ikinci bölümde tanıtılan LOESS regresyon yöntemi kullanılmış ve kullanılan değişkenlerin istatistiksel olarak anlamlı olup olmadıkları test edilip, sonuçlar değerlendirilmiştir. Daha önce de ifade edildiği gibi LOESS yöntemi LOWESS yönteminin metodolojik olarak geliştirilmiş ve çok değişkenli verilere uyarlanabilir hale getirilmiş versiyonu olduğundan uygulamada LOESS yöntemi tercih edilmiştir.

Uygulamada ilk olarak tek değişkenli modeller kurulmuş, ardından diğer değişken modele eklenerek değişkenlerin istatistiksel olarak anlamlılıkları test edilmiştir. Elde edilen bu sonuçlarla birlikte değişkenlerin üç boyutlu grafiği çizilerek sonuçlar değerlendirilmiştir.

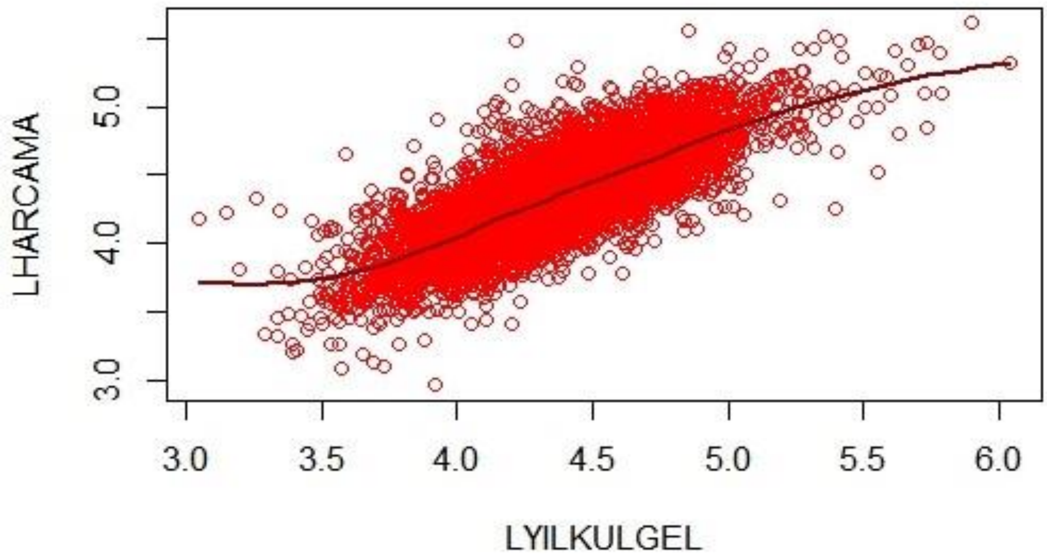
3.3.1. Tek Değişkenli Modellerin Kurulması

Çalışmanın bu aşamasında tek değişkenli modellerin dağılım grafikleri üzerine LOESS düzleştirme eğrileri (smooth curve) çizilmiş ve değişkenler arasındaki ikili ilişki hakkında bir ön görüş sağlanmıştır.

Burada en önemli konu span değerinin seçimidir. 0-1 aralığında değer alabilen span değeri için literatürde önerilen değer 0,5 ve civarındaki değerlerdir (Tezcan, 2009: 114). Bu bağlamda 0,5 ve civarındaki farklı değerler alınarak elde edilen LOESS düzleştirme eğrileri incelenmiştir. Fakat çalışmada kullanılan veri setinde çok fazla veri bulunduğundan, elde edilen dağılım grafiğinde veri bulutu belli bir yoğunluk sergilemektedir. Bu nedenle farklı span değerleri ile çizilen LOESS düzleştirme eğrileri arasında görsel olarak pek fark oluşmamaktadır. Bu yüzden bu çalışmada span değeri olarak 0,5 seçilmiştir. Farklı span değerleri ile çizilen LOESS düzleştirme eğrileri ve sonuçları Ek 1' de verilmiştir. İlk adımda, bağımlı değişken olarak hanehalkı zaruri tüketim harcamaları, bağımsız değişken olarak ise hanehalkı yıllık kullanılabilir geliri değişkeninin bulunduğu model 1 ve bağımlı değişken olarak hanehalkı zaruri tüketim harcamaları, bağımsız değişken olarak ise OECD tanımlı hanehalkı büyüklüğünün bulunduğu model 2 olmak üzere iki farklı model kurulmuştur.

Model 1' e ait LOESS düzleştirme eğirişi Şekil 3.1' de verilmiştir.

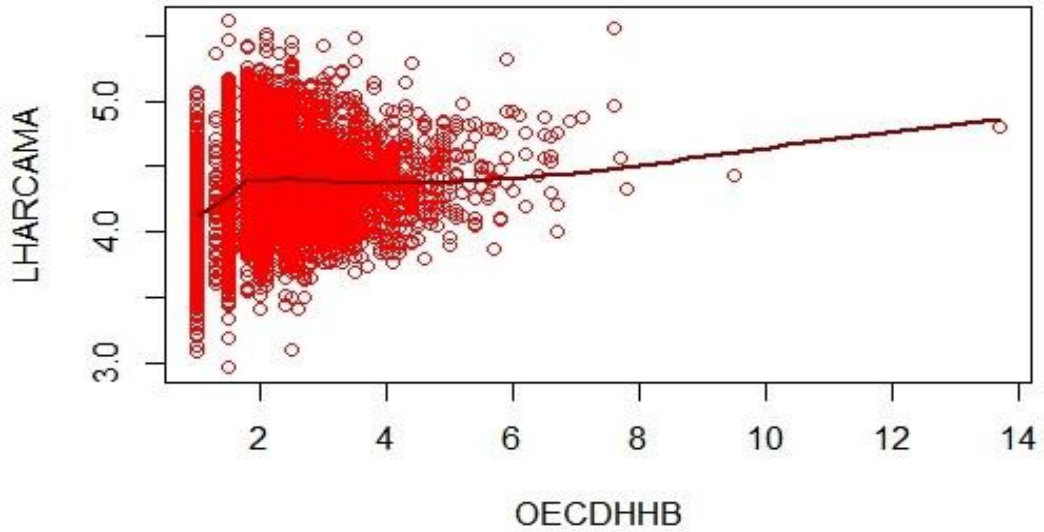
Şekil 3.1: Model 1'e Ait LOESS Düzleştirme Eğrisi



Burada LHARCAMA; hanehalkı zaruri tüketim harcamalarını, LYILKULGEL ise hanehalkı yıllık kullanılabilir gelirini göstermektedir. Şekil 3.1' de de görüldüğü gibi bu iki değişken arasındaki ilişki verilerin yoğunlaştığı bölümde neredeyse doğrusal bir ilişki olarak, uç kısımlarda ise bu doğrusallıktan saparak eğrisel bir ilişki olarak ortaya çıkmaktadır. Şekil 3.1' in sol tarafındaki eğrisellik incelendiğinde, hanehalkı kullanılabilir gelirinin 3,0-3,5 değerleri arasında hanehalkı zaruri tüketim harcamalarında önce çok az bir azalmanın meydana geldiği ardından durularak tekrar artmaya başladığı görülmektedir. Daha sonra bu artış gelir arttıkça artmakta ve hanehalkı kullanılabilir gelirinin yaklaşık olarak 5 değerinden sonra bu artışın hızının azaldığı görülmektedir.

Model 2’ ye ait LOESS düzleştirme eğrisi ise Şekil 3.2’ de verilmiştir.

Şekil 3.2: Model 2’ ye Ait LOESS Düzleştirme Eğrisi



Burada OECDHHB; OECD tanımlı hanehalkı büyüklüğü değişkenini ifade etmektedir. Şekil 3.2 incelendiğinde OECD tanımlı hanehalkı büyüklüğü yaklaşık olarak 2 değerine doğru artarken hanehalkı zaruri tüketim harcamalarında hızlı bir artış meydana gelmektedir. Ardından 2 ile yaklaşık olarak 4,5 değerleri arasında hanehalkı zaruri tüketim harcamaları nispi olarak azalmakta ve daha sonra tekrar artmaktadır. Fakat bu artış ilk kısımda olduğu gibi sert bir artış değildir.

Her iki modele ait LOESS regresyon sonuçları ise Tablo 3.2’ de verilmiştir.

Tablo 3.2: Tek Değişkenli Modellere Ait LOESS Regresyon Sonuçları

	Model 1	**Model 2**
Gözlem Sayısı	9980	9980
Artıkların Standart Hatası	0,1804	0,278
Eşdeğer Parametre Sayısı (Modelin Serbestlik Derecesi)	8,1	7,92
Span Değeri	0,5	0,5

3.3.2. Modellerdeki Değişkenlerin Anlamlılıklarının Test Edilmesi

Bu aşamada her iki modelde bulunan değişkenlerin istatistiksel olarak anlamlılıkları test edilip tüm değişkenlerin yer aldığı model kurulmuştur. Daha önce kurulan tek değişkenli modellerde, değişkenlerin istatistiksel olarak anlamlı olup olmadıklarını diğer bir ifadeyle değişkenler arasında istatistiksel bir ilişki olup olmadığını test etmek için F testi kullanılmıştır. Bu F test istatistiği;

$$F = \frac{(RSS_1 - RSS_2)/(df_2 - df_1)}{RSS_2/(n - df_2)}$$

şeklinde hesaplanmaktadır. Burada; RSS_1=Birinci modelin hata kareleri toplamı, RSS_2=İkinci modelin hata kareleri toplamı, df_1=Birinci modelin serbestlik derecesi (eşdeğer parametre sayısı), df_2=İkinci modelin serbestlik derecesi (eşdeğer parametre sayısı) ve n toplam gözlem sayısını göstermektedir (Jacoby, 2000: 595).

Bu testi uygularken test etmek istediğimiz her değişken için, o değişken içinde bulunduğu model ile o değişkenin içinde bulunmadığı model karşılaştırılacak ve modeller arasında istatistiksel olarak bir farklılık olup olmadığı incelenecektir. Diğer bir ifadeyle değişkenin modele istatistiksel olarak anlamlı bir katkı yapıp yapmadığı incelenecektir.

Hanehalkı kullanılabilir geliri değişkeni için yapılan F testi sonuçları Tablo 3.3' te verilmiştir.

Tablo 3.3: Hanehalkı Kullanılabilir Geliri Değişkeni İçin Yapılan F Testi Sonuçları

	Eşdeğer Parametre Sayısı	**Hata Kareleri Toplamı**	**F değeri**	**p**
LHARCAMA ~ OECD	7,92	770,54	1064	2,2e-16***
LHARCAMA ~ OECD + LYILKULGEL	16,66	319,12		

***0,001 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.

Tablo 3.3' te verilen sonuçlara göre hanehalkı kullanılabilir geliri değişkeninin modele eklenmesiyle hata kareleri toplamında meydana gelen değişikliğin istatistiksel olarak 0,001 anlamlılık düzeyinde farklı olduğu görülmektedir. Bu durumda hanehalkı kullanılabilir geliri değişkeninin hanehalkı zaruri tüketim harcamaları üzerinde istatistiksel olarak anlamlı bir etkiye sahip olduğu anlaşılmaktadır.

Benzer şekilde, OECD tanımlı hanehalkı büyüklüğü değişkeni için yapılan F testi sonuçları Tablo 3.4’ te verilmiştir.

Tablo 3.4: OECD Tanımlı Hanehalkı Büyüklüğü Değişkeni İçin Yapılan F Testi Sonuçları

	Eşdeğer Parametre Sayısı	**Hata Kareleri Toplamı**	**F değeri**	**p**
LHARCAMA ~ LYILKULGEL	8,10	324,38	12,604	2,2e-16***
LHARCAMA ~ LYILKULGEL + OECD	16,66	319,12		

***0,001 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.

Tablo 3.4 incelendiğinde, OECD tanımlı hanehalkı büyüklüğü değişkeni modele eklendiğinde hata kareleri toplamında meydana gelen değişikliğin istatistiksel olarak 0,001 anlamlılık düzeyinde farklı olduğu görülmektedir. Bu durumda OECD tanımlı hanehalkı büyüklüğü değişkeninin hanehalkı zaruri tüketim harcamaları üzerinde istatistiksel olarak anlamlı bir etkiye sahip olduğu anlaşılmaktadır.

3.3.3. Çok Değişkenli Modelin Kurulması ve Bu Modele İlişkin Tahminler

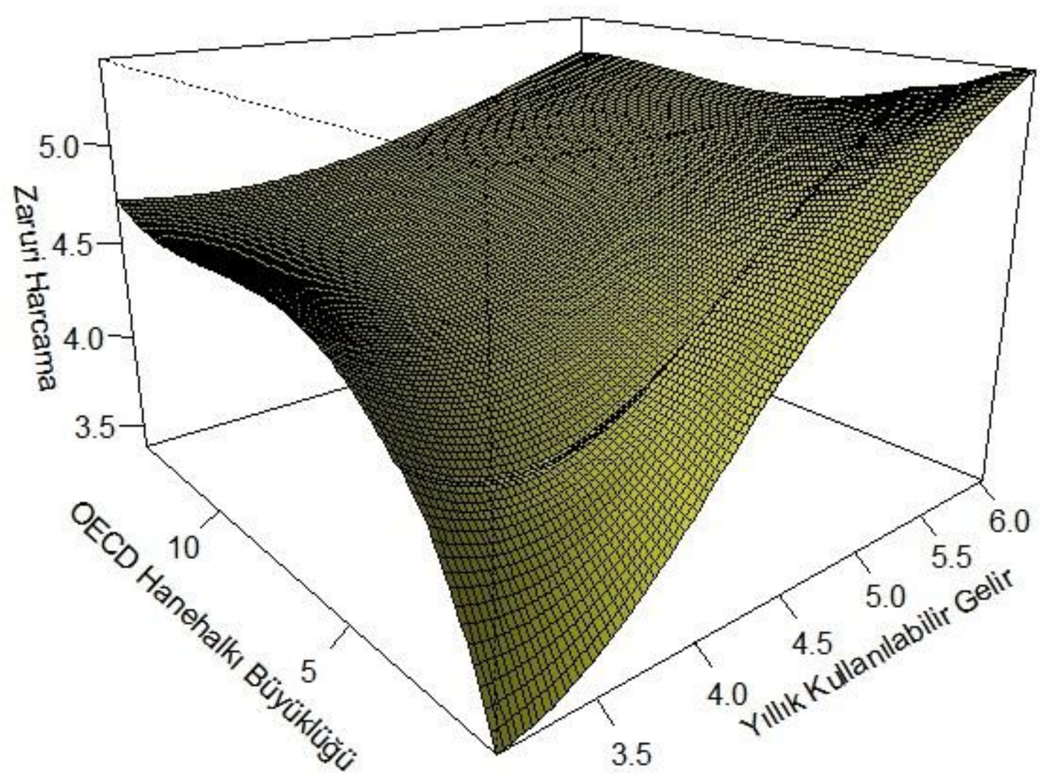
Yapılan testler sonucunda her iki bağımsız değişkeninde bağımlı değişken üzerinde istatistiksel olarak anlamlı bir etkiye sahip olduğu görülmüştür. Şu durumda her iki bağımsız değişken de modele eklenerek çok değişkenli model kurulmuş ve sonuçlar Tablo 3.5’ te verilmiştir.

Tablo 3.5: Çok Değişkenli Model Sonuçları

LHARCAMA ~ LYILKULGEL + OECD	
Gözlem Sayısı	9980
Artıkların Standart Hatası	0,179
Hata Kareleri Toplamı	319,12
Eşdeğer Parametre Sayısı (Modelin Serbestlik Derecesi)	16,66
Span Değeri	0,5

Elde edilen tüm bu sonuçlarla birlikte çok değişkenli modelle ilgili çıkarımda bulunmak amacıyla üç boyutlu grafiği Şekil 3.3' de verilmiş ve sonuçlar değerlendirilmiştir.

Şekil 3.3: Çok Değişkenli Modelin Üç Boyutlu Grafiği



Şekil 3.3 incelendiğinde hanehalkı yıllık kullanılabilir geliri ile hanehalkı zaruri tüketim harcamaları arasındaki ilişkinin doğrusala yakın fakat tam olarak doğrusal bir ilişki olmadığı görülmektedir. Benzer şekilde OECD tanımlı hanehalkı büyüklüğü ile hanehalkı zaruri tüketim harcamaları arasındaki ilişkininde doğrusal olmadığı görülmektedir. Genel olarak hanehalkı kullanılabilir geliri artarken hanehalkı zaruri tüketim harcamaları da artmakta fakat aynı durum OECD tanımlı hanehalkı büyüklüğü için geçerli değildir. İlişkinin daha iyi görülebilmesi için Şekil 3.3' de verilen çok değişkenli modelin üç boyutlu grafiğinin farklı açılardan elde edilen şekilleri Ek 2' de verilmiştir.

Farklı açılardan elde edilen üç boyutlu grafikler de incelendiğinde, OECD tanımlı hanehalkı büyüklüğü yaklaşık olarak 5'e doğru artarken hanehalkı zaruri tüketim harcamalarındaki artış yüksek, daha sonra bu artış durulmakta ve yerini nispi bir artışa bıraktığı görülmektedir. Benzer ilişki ikili karşılaştırmalarda da ortaya çıkmıştı fakat orada bu artış miktarı OECD tanımlı hanehalkı büyüklüğü değeri yaklaşık olarak 2' ye doğru artarken, hanehalkı zaruri tüketim harcamalarında sert bir artış görülmekte ve daha sonra nispi bir azalma ardından nispi bir artış gerçekleşmekteydi. Bunun sebebi modele yeni bir değişkenin eklenmesiyle ilişkide meydana gelen değişimdir.

Ayrıca hanehalkı kullanılabilir geliri değeri yaklaşık olarak 3,5 değerine gelirken OECD tanımlı hanehalkı büyüklüğünün hanehalkı zaruri tüketim harcamasına etkisi, daha düşük hanehalkı kullanılabilir geliri düzeyine göre azalmaktadır. Benzer şekilde hanehalkı kullanılabilir geliri en üst seviyede iken, OECD tanımlı hane halkı büyüklüğü arttığında bunun hanehalkı zaruri tüketim harcamasına olan etkisi OECD tanımlı hanehalkı büyüklüğünün yaklaşık 6 değerine kadar olan kısımda azalmakta, daha sonra durulup nispi bir artış sergilemektedir.

İktisadi olarak bakıldığında gerçekten de düşük gelir grubunda bulunan ailelerde birey sayısındaki artışın hanehalkı zaruri tüketim harcamalarına dönüşme miktarı, yüksek gelir grubunda bulunan ailelere oranla daha yüksektir. Ayrıca birey sayısındaki artış da her zaman aynı oranda zaruri harcamaya dönüşmemektedir.

3.4. PARAMETRİK VE PARAMETRİK OLMAYAN REGRESYON TAHMİNLERİNİN KARŞILAŞTIRILMASI

Çalışmanın bu kısmında En Küçük Kareler (EKK) tahmincisine dayalı parametrik regresyon yöntemiyle, LOESS regresyon yönteminde ele alınan değişkenler kullanılarak regresyon modeli kurulmuş ve regresyon katsayılarının tahmini yapılıp sonuçlar değerlendirilmiştir.

3.4.1. Parametrik Regresyon Modelinin Kurulması

İlk olarak regresyon denklemi oluşturulmuş ve regresyon katsayıları tahmin edilmiştir. Kurulan parametrik regresyon modeli;

$$LHARCAMA = \beta_0 + \beta_1 LYILKULGEL + \beta_2 OCED + \varepsilon$$

şeklindedir. Bu parametrik regresyon modelinin sonuçları Tablo 3.6 ve Tablo 3.7’ de verilmiştir.

Tablo 3.6: Parametrik Regresyon Modelinin Katsayılarının Anlamlılığı ile İlgili Sonuçlar

		Bağımsız Değişkenler	
	Sabit Terim	**Yıllık Kullanılabilir Gelir**	**OECD Tanımlı Hanehalkı Büyüklüğü**
Regresyon Katsayıları	0,942	0,768	0,023
Standart Hata	0,027	0,006	0,002
t – İstatistiği	34,663	122,471	9,902
p – Değeri	0,000***	0,000***	0,000***

*** 0,001 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.

Tablo 3.6’ da elde edilen sonuçlar incelendiğinde regresyon katsayılarının 0,001 anlamlılık düzeyinde istatistiksel olarak anlamlı oldukları görülmektedir. Bu veriler ışığında parametrik regresyon modelimiz aşağıdaki gibidir;

$$LHARCAMA = 0{,}942 + 0{,}768\, LYILKULGEL + 0{,}023\, OECD + \varepsilon$$

Tablo 3.7: Parametrik Regresyon Modelinin Sonuçları

Bağımlı Değişken	Hanehalkı Zaruri Tüketim Harcaması
Bağımsız Değişkenler	Hanehalkı Yıllık Kullanılabilir Gelir ve OECD Tanımlı Hanehalkı Büyüklüğü
Gözlem Sayısı	9980
Tahminin Standart Hatası	0,18004
Durbin_Watson İstatistiği	1,781
R^2	0,613
F İstatistiği	7927,086
p Değeri (F İstatistiği)	0,000***

*** 0,001 anlamlılık düzeyinde istatistiksel olarak anlamlıdır.

Tablo 3.7 incelendiğinde kurulan regresyon modelimizin 0,001 anlamlılık düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Parametrik regresyon modelimiz ve regresyon katsayılarımız anlamlı olduğundan katsayılara bakarak istatistiksel çıkarımda bulunabiliriz. Burada OECD tanımlı hanehalkı büyüklüğü değişkeni ile hanehalkı yıllık kullanılabilir geliri karşılaştırıldığında, hanehalkı yıllık kullanılabilir değişkeni katsayısı daha büyük olduğundan hanehalkı zaruri tüketim harcamaları değişkenine daha çok etki etmektedir. Ayrıca hanehalkı yıllık kullanılabilir geliri değişkeninde ki her bir birimlik artış veya azalış, hanehalkı zaruri tüketim harcamalarını %76,8 arttırmakta veya azaltmaktadır. Benzer şekilde OECD tanımlı hanehalkı büyüklüğündeki her bir birimlik artış veya azalış, hanehalkı zaruri tüketim harcamalarını %2,3 arttırmakta veya azaltmaktadır.

3.4.2. Tahmin Sonuçlarının Karşılaştırılması

Hanehalkı zaruri tüketim harcamalarının hanehalkı kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü ile açıklanması amacıyla parametrik ve parametrik olmayan regresyon analizleri yapılmıştır. Yapılan bu analizler sonucunda hanehalkı zaruri tüketim harcamaları ile hanehalkı kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklükleri arasında istatistiksel olarak anlamlı bir ilişki bulunmuştur.

Değişkenler arasındaki ilişki her iki yöntemde de incelendiğinde; parametrik regresyon yönteminde özellikle OECD tanımlı hanehalkı büyüklüğündeki artışın hanehalkı zaruri tüketim harcamalarına olan etkisinin sabit bir şekilde ve %2,3 olduğu görülmektedir. Oysa parametrik olmayan regresyon yöntemi olan LOESS regresyonda ise bu etkinin OECD tanımlı hanehalkı büyüklüğü değerinin yaklaşık olarak 5' e kadar olan artışında hanehalkı zaruri tüketim harcamalarına olan etkisinin çok yüksek olduğu daha sonra bu etkinin azaldığı ve yerini nispi bir artışa bıraktığı gözlemlenmiştir. Benzer şekilde parametrik regresyon yönteminde hanehalkı yıllık kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü değişkenlerinin hanehalkı zaruri tüketim harcamalarına olan etkisinin sırasıyla sabit olarak %76,8 ve %2,3 olduğu görülmektedir. Oysa parametrik olmayan regresyonda bu etkinin doğru orantılı fakat sabit bir şekilde olmadığı ve hanehalkı kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü birlikte artarken ilk başlarda hanehalkı zaruri tüketim harcamalarına yüksek etki gösterirken bu etkinin daha sonra azaldığı görülmektedir.

SONUÇ VE ÖNERİLER

Hanehalkı harcamaları ile geliri arasındaki ilişki ekonomik dinamiklerin nasıl şekillendiğinin anlaşılmasında en temel unsurlardan biri olmaktadır. Bu nedenle bu ilişkinin incelenmesi hem ekonomik dinamiklerin nasıl şekillendiği hakkında bilgi vermekte hem de bu ekonomik dinamiklere bağlı olarak politika üretilmesine imkân sağlamaktadır.

Bu kapsamda yapılan bu çalışmada hanehalkı zaruri tüketim harcamaları ile hanehalkı yıllık kullanılabilir geliri arasındaki ilişki incelenmiştir. Literatürde bu ilişkiyi inceleyen ve parametrik yöntemlerin kullanıldığı birçok çalışma bulunmaktadır. Fakat parametrik olmayan yöntemlerin kullanıldığı çalışmalar oldukça azınlıktadır. Bu nedenle bu çalışmada, sıklıkla incelenen bu ilişkiye yöntemsel açıdan bir zenginlik katmak ve varsa ilişki içerisindeki bilinmeyen yapıyı ortaya koymak amacıyla parametrik olmayan bir yöntem kullanılmıştır.

Hanehalkı harcamaları ve geliri arasındaki ilişkiyi tek başına incelemek ekonomik dinamiklerin nasıl şekillendiğinin anlaşılması açısından yeterli değildir. Bu nedenle hanehalklarının gelirini üreten ve bu geliri harcamalara dönüştüren bireyler de bu ilişkiye dâhil edilmiştir. Bu bağlamda, ilişkiye OECD tanımlı hanehalkı büyüklüğü ilave edilerek ilişki hakkında daha kapsamlı bilgi edinmeye çalışılmıştır.

Yapılan ikili karşılaştırmalarda; hanehalkı zaruri tüketim harcamaları ve hanehalkı yıllık kullanılabilir geliri arasındaki ilişkinin, hanehalkı kullanılabilir gelirinin 3,0-3,5 değerleri için ilk aşamada ufak bir azalma ve ardından artış gösterdiği görülmektedir. Daha sonra bu artış devam etmekte fakat hanehalkı kullanılabilir gelirinin yaklaşık olarak 5,0 değerinden sonra artış oranında azalma meydana gelmektedir. Parametrik olmayan regresyon yöntemiyle elde edilen bu bulgu, parametrik regresyon yöntemiyle elde edilen ve ilişkinin doğru orantılı olduğu bulgusuyla karşılaştırıldığında bir miktar paralellik göstermesine karşın, ilişki her noktada sabit bir şekilde doğru orantılı olmadığından parametrik yöntemle elde edilen tahmin yetersiz olmaktadır.

Hanehalkı zaruri tüketim harcamaları ve OECD tanımlı hanehalkı büyüklüğü arasındaki ilişki de ise haneye giren her bir bireyin aynı oranda zaruri harcamaya etki etmediği, hatta belli bir seviyede bu etkinin nispi olarak azaldığı görülmektedir. Ayrıca bu nispi azalmadan hemen önce hanedeki birey sayısı artarken bunun zaruri harcamaya olan etkisinin oldukça yüksek olduğu gözlemlenmektedir.

Hanehalkı zaruri tüketim harcamaları ile hanehalkı yıllık kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü değişkenleri birlikte incelendiğinde ise dikkat çekici bulgular elde edilmiştir. Bunlardan en dikkat çekici olanı, hanehalkı yıllık kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü birlikte artarken bunun hanehalkı zaruri tüketim harcamalarına olan etkisinin ilk başlarda belli oranda arttığı fakat sonrasında bu artışın durulduğu hatta yüksek gelir grubuna geçildiğinde ise azaldığı gözlemlenmektedir. İlk kısımdaki bu artış, ikili karşılaştırmalarda da görüldüğü üzere küçük hanelerin zaruri harcamalara olan etkisinin yüksek olmasıyla açıklanabilir. Bu da bize küçük hanelerde gelirden bağımsız harcamaların yüksek olduğunu göstermektedir. Benzer şekilde ikinci kısımdaki durulma ikili karşılaştırmalarla birlikte incelendiğinde ise orta büyüklükteki hanelerin zaruri harcamalara olan etkisinin nispi olarak azalmasıyla açıklanabilir. Bu da orta gelir düzeyinde bulunan ve orta büyüklükteki hanelerde gelirden bağımsız harcamaların düşük gelir grubunda bulunan hanelere oranla daha düşük olduğunu göstermektedir. Üçüncü ve son kısım incelendiğinde ise, büyük hanelerin hanehalkı zaruri tüketim harcamalarına olan etkisi tek başına yüksek olmasına rağmen yüksek gelir düzeyinde bulunan hanelerde hanedeki birey sayısı arttıkça bunun hanehalkı zaruri tüketim harcamalarına olan etkisi azalmaktadır. Bu da yüksek gelir düzeyinde bulunan hanelerde birey sayısı arttıkça bunun zaruri harcamalara sınırlı ölçüde etki ettiğini göstermektedir.

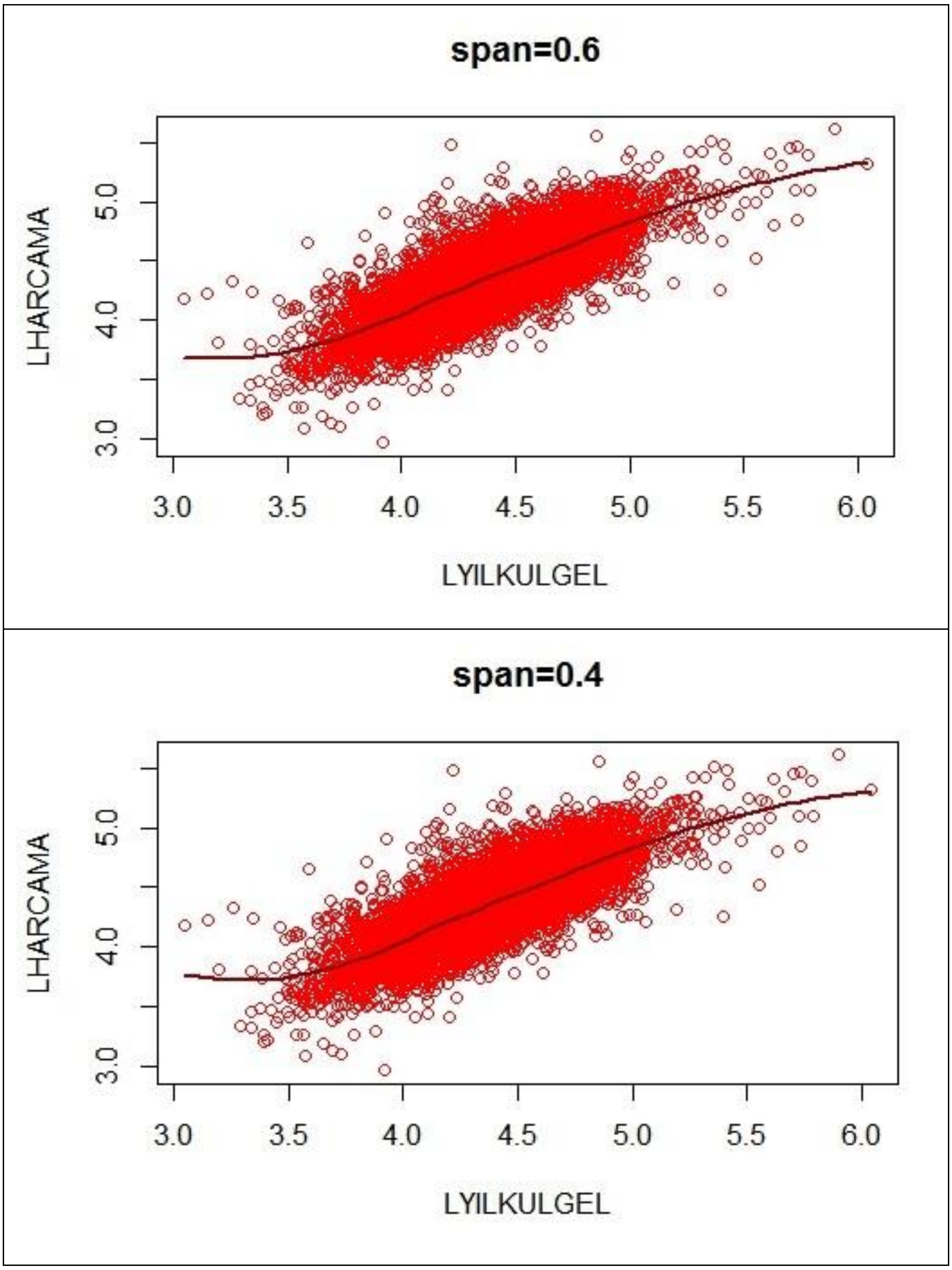
Buraya kadar özetlediğimiz bulgular parametrik yöntemle elde edilen bulgularla karşılaştırıldığında parametrik olmayan yöntemle elde edilen bulguların daha kapsamlı olduğu görülmektedir. Ayrıca parametrik yöntemlerde olduğu gibi ilişkinin fonksiyonel formu belli bir kalıba sokulmadığından elde edilen bulgular oldukça zengindir. Bu açıdan bakıldığında hanehalkı harcamaları ve geliri arasındaki ilişkinin parametrik olmayan yöntemlerle incelenmesi, bu ilişkinin daha iyi anlaşılmasına ve ekonomik dinamiklerin nasıl şekillendiğinin daha iyi ortaya koyularak daha etkili politikalar üretilmesine yardımcı olabilir.

Bunun dışında parametrik olmayan yöntemler incelenen ilişkinin fonksiyonel formunun kesin olarak bilinmediği durumlarda veya parametrik yöntemlerle elde edilen sonuçların güvenilirliğini test etmek amacıyla da kullanılabilir. Bunula birlikte her ne kadar iktisat teorisince harcama ile gelir ve hanehalkı birey sayısı arasında bir ilişki olduğu ortaya koyulsa da bu ilişkinin fonksiyonel formu hakkında önsel olarak kesin bir bilgi bulunmamaktadır.

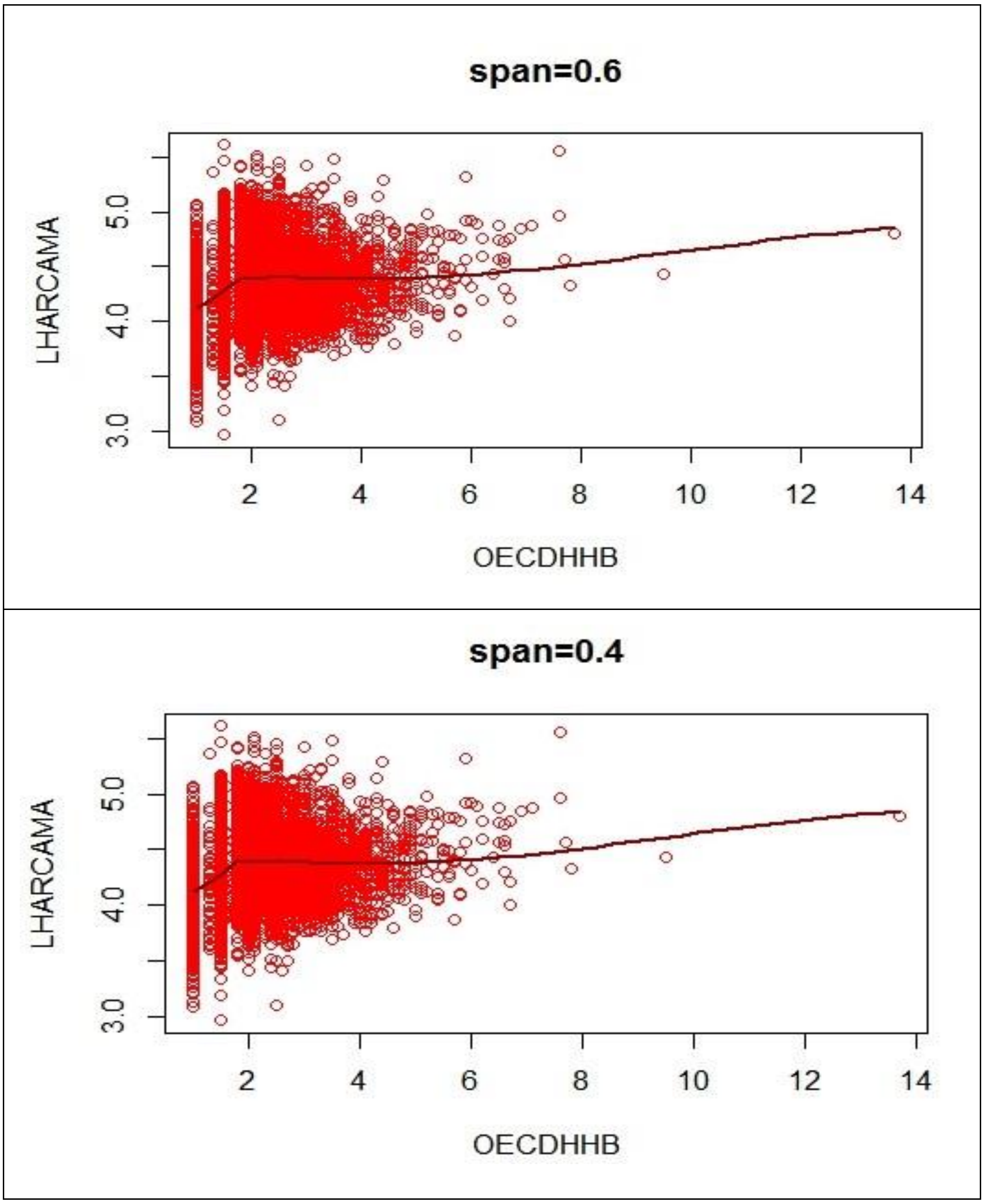
Bu kapsamda, yapılan bu çalışmada hanehalkı zaruri tüketim harcamaları ile hanehalkı yıllık kullanılabilir geliri ve OECD tanımlı hanehalkı büyüklüğü arasındaki ilişki parametrik olmayan bir yöntem olan LOESS regresyon yöntemiyle incelenmiştir. Böylece hem uygulanan parametrik yöntemin sonuçları parametrik olmayan yöntemle elde edilen sonuçlarla karşılaştırılmış hem de önsel olarak fonksiyonel formu kesin olarak bilinmeyen bu ilişki incelenmiştir. Bu sayede hem ilişkinin fonksiyonel formu ortaya konmuş hem de parametrik yöntemle elde edilen sonuçlar test edilmiştir.

EKLER

EK 1: Tek Değişkenli Modellerin Farklı Span Değerleri İle Elde Edilen LOESS Düzleştirme Eğrileri ve LOESS Regresyon Sonuçları



Model 1’ e Ait Farklı Span Değerleri İle Elde Edilen LOESS Düzleştirme Eğrileri

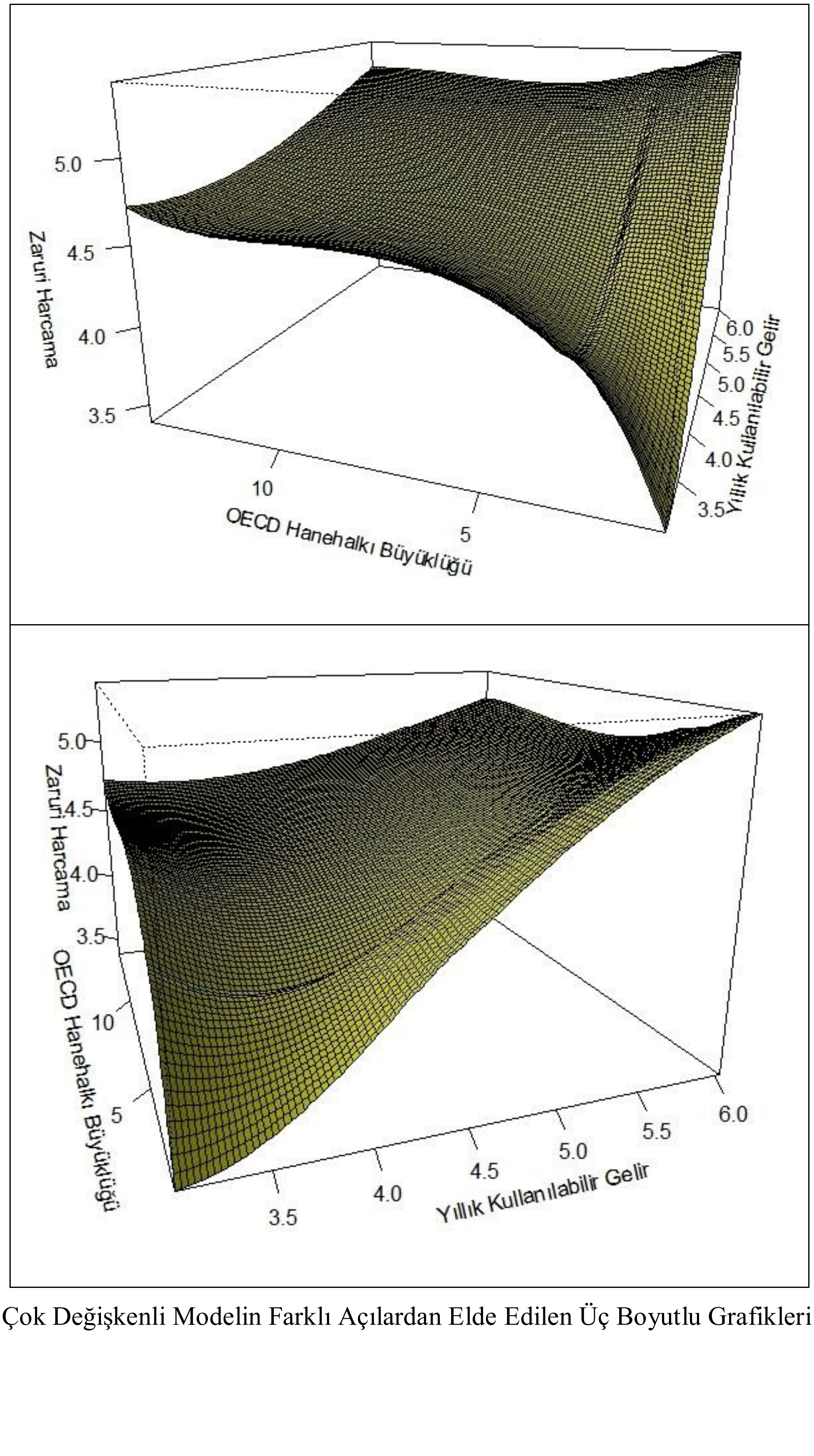


Model 2'ye Ait Farklı Span Değerleri İle Elde Edilen LOESS Düzleştirme Eğrileri

Farklı Span Değerleri İle Elde Edilen Tek Değişkenli Modellere Ait LOESS Regresyon Sonuçları

	Model 1	**Model 2**	**Model 1**	**Model 2**
Span Değeri	0,4		0,6	
Gözlem Sayısı	9980	9980	9980	9980
Artıkların Standart Hatası	0,1804	0,2781	0,1804	0,2781
Eşdeğer Parametre Sayısı (Modelin Serbestlik Derecesi)	9,51	9,05	7,1	6,49

EK 2: Çok Değişkenli Modelin Daha İyi İncelenebilmesi ve Anlaşılabilmesi İçin Farklı Açılardan Oluşturulan Üç Boyutlu Grafikleri



Çok Değişkenli Modelin Farklı Açılardan Elde Edilen Üç Boyutlu Grafikleri

KAYNAKÇA

AVCI, Gizem Mukiyen, (2011), "Türkiye' de Çalışan Kadınların Hanehalkı Harcama Örüntüleri Üzerine Etkisi", Yüksek Lisans Tezi, **Bülent Ecevit Üniversitesi Sosyal Bilimler Enstitüsü**, Zonguldak.

BALTAGİ, Badi H., (2011), **Econometrics**, Springer, New York.

BEYAZ, Fatma Banu, (2007), "Türkiye'de Hanehalkı Tüketim Harcamaları ve Talep Tahmini", Yüksek Lisans Tezi, **Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü**, Antalya.

BİERENS, H. J. ve POTT-BUTER, H. A., (1990), "Specification of Household Engel Curves by Nonparametric Regression". **Econometric Reviews**, Sayı: 9 (2), pp. 123-184.

CLARK, R. M., (1977), "Non-Parametric Estimation of A Smooth Regression Function", **Journal of The Royal Statistical Society**, Sayı: 39 (1), pp. 107-113.

CLEVELAND, William S., (1979), "Robust Locally Weighted Regression And Smoothing Scatterplots", **Journal of The American Statistical Association**, Sayı: 74, pp. 829-836.

CLEVELAND, W. S., S. J. DEVLİN, (1988), "Locally Weighted Regression: An Approach to Regresion Analysis by Local Fitting ", **Journal of The American Statistical Association**, Sayı: 83, pp. 596-610.

ÇATALBAŞ, Gaye Karpat, (2006), "Parametrik Olmayan Tahmin Teknikleri: Türkiye Verileri ile Bir Uygulama", Doktora Tezi, **Gazi Üniversitesi Sosyal Bilimler Enstitüsü**, Ankara.

DELGADO, Miguel A. ve MİLES, Daniel, (1997), "Household Characteristics And Consumption Behaviour: A Nonparametric Approach", **Empirical Economics**, Sayı: 22, pp. 409-429.

EMEÇ, H., (2001), "Türkiye'de Bölgelerarası Tüketim Harcamaları Tobit Model Yaklaşımı", **Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, Sayı: 16 (2), ss. 61-73, İzmir.

GASSER, H. G. ve MÜLLER, T., (1984), “Estimating Regression Functions and Their Derivatives by The Kernel Method”, **Scandinavian Journal of Statistics**, Sayı: 11 (3), pp. 171-185.

GÖLVEREN, Elif, (2012), “Parametrik Olmayan Regresyon Modellerinden Çekirdek Regresyon ve Türkiye’de Hanehalkı Harcamaları Üzerine Bir Uygulama”, Yüksek Lisans Tezi, **Atatürk Üniversitesi Sosyal Bilimler Enstitüsü**, Erzurum.

GUJARATİ, Domador N., (2004), **Basic Econometrics**, McGraw-HiII, New York.

HOLLANDER, M., WOLFE, D. ve CHİCKEN, E., (2014), **Nonparametric Statistical Methods**, John Wiley & Sons, Inc., New York.

HARDLE, W., (1992), **Applied Nonparametric Regression**, Cambridge University Press, Cambridge.

HARDLE, W., MÜLLER, M., SPERLİCH, S. ve WERWATZ, A., (2004), **Nonparametric and Semiparametric Models,** Springer, Berlin.

HARDLE, W., ve LİNTON, O., (1994), “Applied Nonparametric Methods”, **Handbook of Econometrics**, Sayı: 4 (38), pp.2295-2339.

HART, Jeffrey D., (1997), **Nonparametric Smoothing and Lack-Of-Fit Tests**, Springer, New York.

HASTİE, T. J. ve TİBSHİRANİ, R. J., (1990), **Generalized Additive Models**, Chapman and Hall, London.

JACOBY, William G., (2000), “Loess: A Nonparametric, Graphical Tool For Depicting Relationships Between Variables”, **Electoral Studies**, Sayı: 19, pp. 577-613.

KEELE, Luke, (2008), **Semiparametric Regression For The Social Sciences**, John Wiley & Sons Ltd., Ohio.

MADDALA, G. S., (1992), **İntroduction to Econometrics**, Macmillan Publishing Company, New York.

ORHUNBİLGE, Neyran, (1996), **Uygulamalı Regresyon ve Korelasyon Analizi**, Avcıol Basım Yayın, İstanbul.

SARIMEŞELİ, M., (1999), “Hanehalkları Harcama Eğilimleri” **Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi**, Sayı: 2, ss. 41-50, Ankara.

SERPER, Özer, (2004), **Uygulamalı İstatistik 2**, Ezgi Kitabevi, Bursa.

TARI, Recep, (2008), **Ekonometri**, Avcı Ofset, İstanbul.

TEZCAN, Nuray, (2009), "Tahmine Parametrik Olmayan Regresyon Yöntemiyle Yaklaşım", Doktora Tezi, **İstanbul Üniversitesi Sosyal Bilimler Enstitüsü**, İstanbul.

TÜİK, (2012), **Hanehalkı Bütçe Anketlerinin Uygulama Yöntemi, Tanım ve Kavramlar**, Ankara.

WAHBA, G. ve WOLD, S., (1975), "A Completely Automatic French Curve: Fitting Spline Function by Cross-Validation", **Communication in Statistics**, Sayı: 4, pp. 1-17.

YILDIZ, Münevvere, (2013), "Parametrik Olmayan Bulanık Regresyon Modelleri Analizi", Doktora Tezi, **Anadolu Üniversitesi Fen Bilimleri Enstitüsü**, Eskişehir.

ZORTUK, Mahmut, (2009), "Cinsiyet Değişkeni Bağlamında Harcama Alt Grupları ve Gelir İlişkisi: Dumlupınar Üniversitesi Öğrencileri Üzerine Bir Uygulama", **Dumlupınar Üniversitesi Sosyal Bilimler Dergisi**, Sayı: 23, Kütahya.

Elektronik Kaynaklar

GALTON, Sir Francis, (1886), "Regression Towards Mediocrity in Hereditary Stature", **Journal of Anthropological Institute of Great Britain and Ireland,** Sayı: 15, pp. 246–263, *http://www.jstor.org/stable/2841583* (12.10.2015).

NADARAYA, A. E., (1964), "On Estimating Regression", **Theory of Probability and its Applications** Sayı: 9 (1), *http://epubs.siam.org/doi/abs/10.1137/1109020*, (11.11.2015).

WATSON G. S. (1964), "Smooth Regression Analysis", **Sankhya, Series A**, Sayı: 26, pp. 359-372, *http://www.jstor.org/stable/25049340* (11.11.2015).

DİZİN