T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

PANEL VERİLİ YARIPARAMETRİK REGRESYON MODELLERİ

Alper SİNAN

DOKTORA TEZİ

MATEMATİK ANABİLİM DALI

KONYA, 2010

T.C.
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

PANEL VERİLİ YARIPARAMETRİK REGRESYON MODELLERİ

Alper SİNAN

DOKTORA TEZİ

MATEMATİK ANABİLİM DALI

Bu tez 17.05.2010 tarihinde aşağıdaki jüri üyeleri tarafından oybirliği ile kabul edilmiştir.

……. ……. …….

Doç. Dr. Aşır GENÇ Prof. Dr. Fahrettin ARSLAN Doç. Dr. Galip OTURANÇ

……. …….

Doç. Dr. Coşkun KUŞ Yrd. Doç. Dr. Hasan KÖSE

ÖZET

DOKTORA TEZİ

PANEL VERİLİ YARIPARAMETRİK REGRESYON MODELLERİ

Alper Sinan

Selçuk Üniversitesi
Fen Bilimleri Enstitüsü
Matematik Anabilim Dalı

Danışman: Doç. Dr. Aşır GENÇ
2010, 136 sayfa

Jüri: Doç. Dr. Aşır GENÇ
Prof. Dr. Fahrettin ARSLAN
Doç. Dr. Galip OTURANÇ
Doç. Dr. Coşkun KUŞ
Yrd. Doç. Dr. Hasan KÖSE

Panel veri analizi sıklıkla ekonometri alanında kullanılmaktadır. Genellikle parametrik regresyon modelleri yardımıyla panel veri analizleri yapılmaktadır. Bu modeller yardımıyla tahmin yapılması ve tahminlerin yorumlanması kolaydır. Ancak parametrik modeller ile yapılan tahminler her zaman iyi sonuç vermemektedir. Bu problemin giderilmesi için panel veri analizlerinde parametrik olmayan regresyon modelleri kullanılmaktadır. Parametrik olmayan regresyon modelleri ile yapılan tahminler parametrik modellere göre daha iyi sonuç vermektedir. Parametrik olmayan regresyon modellerinde ise yorumlama zorluğu yaşanmaktadır.

Bu çalışmada panel veri analizleri için yarıparametrik regresyon modellerinin kullanımı incelenmiştir. Yarıparametrik regresyon modeli ile yapılan tahminler parametrik modele göre daha iyi sonuç vermekte ve parametrik olmayan modele göre yorumlaması kolay olmaktadır. Yarıparametrik model kullanarak 2009 Avrupa ekonomik krizi incelenmiş ve yorumlamalar yapılmıştır. Çalışmada ayrıca parametrik olmayan tahmin yöntemlerinden medyana dayalı yöntemler incelenmiş ve literatürde bulunan yöntemlere alternatif bir yöntem geliştirilerek simülasyonlar yardımıyla karşılaştırılmıştır. Geliştirilen yöntemin literatürde yer alan yöntemlere göre avantajları gösterilmiştir.

Anahtar Kelimeler: Yarıparametrik Regresyon, Panel Veri Modelleri, Parametrik Olmayan Regresyon

ABSTRACT

PhD Thesis

SEMIPARAMETRIC REGRESSION MODELS WITH PANEL DATA

Alper SİNAN

Selcuk University
Graduate School of Natural and Applied Sciences
Department of Mathematics

Supervisor: Assoc.Prof. Dr. Aşır GENÇ
2010, 136 page

Jury: Assoc. Prof. Dr. Aşır GENÇ
Prof. Dr. Fahrettin ARSLAN
Assoc. Prof. Dr. Galip OTURANÇ
Assoc. Prof. Dr. Coşkun KUŞ
Assoc. Prof. Dr. Hasan KÖSE

Panel data have been used frequently in the field of econometrics. Generally panel data analysis can be done by parametric regression models. Estimation and the interpretation of these estimates are very simple by these models. On the other hand estimates obtained by parametric models have'nt got always better results. For the remedy of this problem, nonparametric regression models are used in panel data analysis. The estimates obtained by nonparametric regression models give better results than parametric models. But there are some difficulties in interpreting the results in nonparametric regression models.

In this study the use of semiparametric regression models are investigated for panel data analysis. The estimates obtained by semiparametric regression models give better results than parametric models and the interpretation of the results are more simple than nonparametric model. By using semiparametric model, the economical crisis in Europa at 2009 is investigated and interpretations are made. Also in this study, the methods based on median from nonparametric estimation methods are investigated and by developing an alternative method, the results are compared with the methods given in the literature by simulations. The advantages of the developed method with respect to the methods given in the literature are shown.

Key Words: Semiparametric Regression, Panel Data Models, Nonparametric Regression.

İÇİNDEKİLER

ÖZET…….. i
ABSTRACT…… ii
İÇİNDEKİLER…… iii
ŞEKİLLER DİZİNİ…… v
ÇİZELGELER DİZİNİ…… vi
TEŞEKKÜR…… vii
1. GİRİŞ…… 1
1.1. Kaynak Araştırması…… 3
2. PARAMETRİK OLMAYAN REGRESYON…… 6
2.1. Medyana Dayalı Parametrik Olmayan Tahmin Yöntemi…… 8
2.1.1. Brown- Mood yöntemi…… 9
2.1.2. Theil yöntemi…… 12
2.1.3. Doğrusal ilişkiye sahip olmayan veriler için medyan tahmini……..... 14
2.2. Yoğunluk Tahmin Edicileri…… 17
2.2.1. Histogram tahmin edicisi…… 17
2.2.2. En yakın komşuluk tahmin edicisi…… 20
2.2.3. Kernel yoğunluk fonksiyonu tahmin edicisi ve özellikleri…… 21
2.2.4. Kernel çekirdek fonksiyonunun seçimi…… 27
2.2.5. Kernel tahmininde “h” bant genişliği seçimi…… 34
2.2.6. Çoklu kernel yoğunluk fonksiyonu tahmini…… 38
2.3. Parametrik Olmayan Regresyonda Yoğunluk Tahmini…… 40
2.3.1. Parametrik olmayan basit regresyon modelinde tahmin…… 42
2.3.1.1. Nadaraya-Watson kernel tahmin edicisi…… 42
2.3.2. Çoklu parametrik olmayan regresyon modeli…… 44
3. YARIPARAMETRİK REGRESYON…… 49
3.1. Yarıparametrik Regresyon Modelinin Tanımı…… 49
3.2. Yarıparametrik Regresyonda Model Belirleme…… 52
3.3. Yarıparametrik Regresyonda tahmin…… 53
3.3.1. Yarıparametrik regresyonda parametrik kısmın tahmini…… 53
3.3.2. Yarıparametrik regresyonda parametrik olmayan kısmın tahmini…… 57
4. PANEL VERİ ANALİZİ…… 59
4.1. Parametrik Regresyonda Panel Verili Modeller…….. 64

4.1.1. Parametrik sabit etkili panel veri modeli…… 65
4.1.2. Parametrik sabit etkili panel veri modelinin tahmini…… 67
4.1.3. Parametrik rasgele etkili panel veri modeli…… 68
4.1.4. Parametrik rasgele etkili panel veri modelinin tahmini…… 70
4.2. Parametrik Olmayan Regresyonda Panel Verili Modeller…… 74
4.2.1. Parametrik olmayan panel verili modellerde Nadaraya-Watson kernel tahmini …… 74
4.2.2. Lokal lineer kernel tahmini…… 76
4.2.3. Parametrik olmayan sabit etkili panel veri modeli…… 78
4.2.4. Parametrik olmayan sabit etkili panel veri modelinde tahmin yöntemi 79
4.2.5. Parametrik olmayan rasgele etkili panel veri modeli…… 81
4.2.6. Parametrik olmayan rasgele etkili panel veri modelinde tahmin…… 82
4.3. Yarıparametrik Regresyonda Panel Verili Modeller…… 85
4.3.1. Yarıparametrik sabit etkili panel veri modeli…… 88
4.3.2. Yarıparametrik sabit etkili panel veri modelinde tahmin…… 90
4.3.3. Yarıparametrik rasgele etkili panel veri modeli…… 92
4.3.4. Yarıparametrik rasgele etkili panel veri modelinde tahmin…… 93
5. UYGULAMA…… 99
5.1. Medyana Dayalı Parametrik Olmayan Tahmin Yöntemlerinin Simülasyon Yardımıyla Karşılaştırılması…… 99
5.2. Panel Veri Analizi İle 2009 Avrupa Ekonomik Krizinin Modellenmesi… 106
6. SONUÇ VE ÖNERİLER…… 129
KAYNAKLAR……..... 134

ŞEKİLLER DİZİNİ

Şekil 1. Brown-Mood yönteminin uygulanışı

Şekil 2. Doğrusal olmayan ilişki için medyan tahmini

Şekil 3. Düzgün kernel fonksiyonunun grafiği

Şekil 4. Epanechnikov kernel fonksiyonunun grafiği

Şekil 5. Üçgensel kernel fonksiyonunun grafiği

Şekil 6. Üçlü ağırlık kernel fonksiyonunun grafiği

Şekil 7. İkili ağırlık kernel fonksiyonunun grafiği

Şekil 8. Kosinüs kernel fonksiyonu grafiği

Şekil 9. Standart Normal kernel fonksiyonunun grafiği

Şekil 10. Kernel fonksiyonlarının grafiği

Şekil 11. Farklı ilişki yapılarına sahip veri setlerinin serpilme diyagramları

Şekil 12. Parametrik sabit etkili panel veri model grafiği

Şekil 13. Parametrik rasgele etkili panel veri modeli grafiği

Şekil 14. Parametrik olmayan sabit etkili panel veri modeli grafiği

Şekil 15. Parametrik olmayan rasgele etkili panel veri modeli grafiği

Şekil 16. Yarıparametrik sabit etkili panel veri modeli grafiği

Şekil 17. Yarıparametrik rasgele etkili panel veri modeli grafiği

ÇİZELGELER DİZİNİ

Çizelge 1. Panel veri tablosu

Çizelge 2. Medyan yöntemi için farklı biçimlerdeki veri setleri

Çizelge 3. Ekonomik göstergeler için panel veri tablosu

Çizelge 4. Parametrik sabit etkili modelin parametre tahminleri

Çizelge 5. Parametrik sabit etkili panel veri modeli $\hat{Y}_{it}$ tahminleri

Çizelge 6. Parametrik rasgele etkili panel veri modeline ilişkin değerler

Çizelge 7. h bant genişlikleri

Çizelge 8. Parametrik olmayan sabit etkili panel veri modeli hesaplamaları

Çizelge 9. Parametrik olmayan rasgele etkili panel veri modeline ilişkin tahminler

Çizelge 10. Model belirlemeye ilişkin değerler

Çizelge 11. Yarıparametrik sabit etkili panel veri modeline ilişkin hesaplamalar

Çizelge 12. Yarıparametrik rasgele etkili panel veri modeline ilişkin hesaplamalar

TEŞEKKÜR

Bu konunun seçiminde ve çalışma süresince, yardımlarını aldığım bütün çalışma arkadaşlarıma, benden desteğini esirgemeyen danışman hocam Sayın Doç. Dr. Aşır GENÇ'e saygı ve teşekkürlerimi sunarım. Ayrıca çalışma boyunca bana verdikleri manevi desteklerden dolayı aileme teşekkürü bir borç bilirim.

1. GİRİŞ

İstatistiksel araştırmalarda kullanılan yöntemler genel olarak parametrik ve parametrik olmayan yöntemler olmak üzere iki gruba ayrılmaktadır. Fakat özellikle modelleme yapılırken bağımlı değişken ile bağımsız değişkenler arasındaki ilişki her zaman parametrik veya parametrik olmayan modeller yardımıyla açıklanamamaktadır. Bu ilişki kısmen parametrik, kısmen de parametrik olmayan modellerle her iki tür modelden de iyi açıklanabilmektedir. Bir kısmı parametrik, bir kısmı ise parametrik olmayan regresyon modellerine yarıparametrik regresyon modelleri denilmektedir. İki ayrı kısımdan oluşan bu modelin parametrelerini tahmin etmek için, her iki kısmı da göz önüne alarak, çok sayıda farklı yöntemler kullanılmakta bu da parametre tahminini zorlaştırmaktadır. Bununla beraber, istatistiksel araştırmalarda kullanılan farklı veri türlerinin kendi yapılarına uygun modellerle incelenmesi gerektiğinden, modelin kompleks yapısı göz önüne alındığında yarıparametrik regresyon son yıllarda çok sık kullanılan aynı zamanda üzerinde çok çalışılan bir yöntem halini almıştır.

Bu çalışmanın temel amacı, panel veri analizi için yarıparametrik regresyon modelinin tanımlanması, parametrik kısımdaki bilinmeyen parametrelerin tahmin edilme süreci ve yarıparametrik model kullanımı ile bağımlı değişkenin tahmin edilmesidir.

Çalışma, son yıllarda çok fazla ilgi duyulan yarıparametrik regresyon yöntemini yine istatistiksel veri analizinde sıklıkla kullanılan panel verilerin analizi için kullanmayı amaçladığından, veri analizi alanında önemli bilgiler elde etmemizi sağlayacaktır. Özellikle ekonomi alanında kullanılan veri tiplerinin bilgi kaybını en aza indirmeyi amaçlayan panel veriler ile yapılmasının yaygınlaştığı bir ortamda veri analizlerinde bilgi kaybını en aza indirmek için geliştirilebilecek yeni yöntemlere kaynak olacaktır. Uygulama kısmında günümüz makro ekonomisi için çok önemli bir süreç olan Avrupa krizinin incelenmesi sebebi ekonomik modelleme ve panel veri analizi çalışmaları önem kazanmıştır.

Çalışmada ayrıca parametrik olmayan regresyon yöntemlerinden medyana dayalı yöntemlere alternatif bir tahmin yöntemi geliştirilmiştir. Bu yöntem de uygulama alanında farklı veri setleri için daha iyi sonuçlar elde etmeyi sağlayacak kolay uygulanabilir bir yöntemdir. Yöntemin algoritmasının basitliği nedeni ile ilerleyen zamanlarda daha fazla geliştirilebileceği düşünülebilir.

Çalışma altı ana bölümden oluşmaktadır. Çalışmanın ilk bölümünde tezin amacı açıklanmış, konu ile ilgili daha önceden yapılan çalışmalar incelenmiş, konunun zaman içerisinde gelişimi literatür özeti olarak verilmiştir.

Çalışmanın ikinci bölümünde parametrik olmayan regresyon kavramı ele alınmış ilk olarak medyana dayalı parametrik olmayan yöntemler incelenmiştir. Medyana dayalı parametrik olmayan yöntemlerden literatürde mevcut olan ve en çok kullanılan Brown-Mood yöntemi ile Theil yöntemi açıklanmıştır. Bu iki yöntemin iyi sonuç vermediği veri yapıları için yeni bir tahmin yöntemi öne sürülmüştür. Ardından parametrik olmayan regresyonun temeli olan yoğunluk tahmini konusu incelenmiştir. Ve son olarak parametrik olmayan regresyonda tahmin yöntemleri gösterilmiştir.

Üçüncü bölümde yarıparametrik regresyon modeli gösterilmiş, modelin kurulma aşamaları incelenmiştir. Yarıparametrik modelin parametrik kısmında yer alan parametrelerin tahmini için kullanılan tahmin yöntemi gösterilmiş ve ardından parametrik olmayan kısmın tahminleri yapılmıştır.

Çalışmanın dördüncü bölümünde ilk olarak panel veri kavramı açıklanmış ve panel veri analizinde oluşturulacak modeller tanıtılmıştır. Panel veri analizi için kullanılan sabit etkili panel veri modelleri ile rasgele etkili panel veri modellerinin regresyon modelleri yardımıyla tahmini anlatılmıştır. Bu amaçla iki tür model için de parametrik, parametrik olmayan ve yarıparametrik regresyon kullanımı gösterilmiş her bir model için tahminlerin nasıl yapılacağı ayrıntılı olarak incelenmiştir.

Çalışmanın beşinci bölümü olan uygulama bölümü iki farklı kısımdan oluşmaktadır. İlk kısımda medyana dayalı parametrik olmayan regresyon

yöntemlerinden literatürde yer alan bir yöntem ile bu çalışmada önerilen tahmin yöntemini karşılaştırmak amacı ile altı farklı durumda simülasyon yardımıyla oluşturulan veri setleri için tahminler elde edilmiş ve grafikler yardımıyla önerilen tahmin yönteminin literatürdeki yöntemlere göre uygunluğu gösterilmiştir. İkinci kısımda 2004 yılından 2009 yılına ilerleyen süreçte Avrupa kıtasını etkileyen ekonomik krizin modellenmesi amacı ile toplanan ekonomik veriler yardımıyla panel veri modelleri oluşturulmuştur. Her bir model için tahminler Maple 13, Spss13 ve Selcuk STAT paket programları ve Delphi programlama dili yardımıyla hesaplanmıştır. Sonuçlar tablo ve grafikler yardımıyla gösterilmiş ve elde edilen tahminler yorumlanmıştır.

Çalışmanın son bölümü elde edilen bilgilerin yorumlandığı ve değerlendirildiği sonuç bölümüdür.

1.1. Kaynak Araştırması

İstatistiksel tahmin teorisi, istatistiğin temel alanlarından biridir. Bu konu genel olarak ikiye ayrılabilir. Bunlardan en eski ve çok kullanılanı parametrik tahmin yöntemleridir. Parametrik tahmin yöntemlerinin yetersiz kaldığı noktadan sonra ortaya parametrik olmayan tahmin yöntemleri öne sürülmüş ve tahmin probleminin önemli bir dalı olmuştur. En son olarak ise bu iki tahmin yöntemine ek olarak yarıparametrik tahmin yöntemleri de geliştirilmiştir.

Parametrik olmayan yoğunluk tahmini, 19. yüzyılda istatistiksel araştırma alanı içerisine girmiştir.

Parametrik olmayan yoğunluk tahmini için çeşitli yöntemler kullanılmaktadır. Bunlardan en çok kullanılanı Kernel yoğunluk tahmin edicileridir. Kernel yoğunluk tahminin temelini oluşturan fikirler Fix ve Hodges (1951) tarafından teknik bir raporla ortaya koyulmuştur.

Kernel yoğunluk tahmin edicileri Rosenblatt (1956) ve Parzen tarafından önerilmiştir. Bu tahmin ediciler için ortalama hata kare ve ortalama hata kareler toplamı hesaplamalarını ilk olarak Rosenblatt (1956) ve Parzen (1962) tarafından hesaplanmıştır.

Rosenblatt (1971), bağımsız gözlemler ile olasılık yoğunluk fonksiyonunun tahminini ve bağımlı diziler için olasılık yoğunluk fonksiyonunun tahminini çalışmıştır.

k. en yakın komşuluk tahmini Loftsgaardfen ve Quesenberry (1965) tarafından önerilmiştir.

Nadaraya-Watson Kernel regresyon tahmin edicisi, Nadaraya (1964) ve Watson (1964) tarafından birbirlerinden bağımsız olarak önerilmişlerdir. Bu ilk çalışmadan sonra parametrik olmayan regresyonda en sık kullanılan yöntem Nadaraya-Watson Kernel tahmin edicisi olmuştur.

Priestley ve Chao (1972) kendi adlarını verdikleri tahmin ediciyi alternatif bir Kernel regresyon tahmini olarak önermişlerdir.

Miller (1976), En küçük kareler yönteminde kullanılan hata kareler toplamını ele alarak yarıparametrik regresyon modeline ilişkin parametre tahmini yapmıştır.

Pagan (1984) ve (1986), ardışık olarak yarıparametrik regresyon modelleri için yeni tahmin ediciler önermiştir. Robinson (1988), yarıparametrik modelde parametrik kısmın tahmini için $\sqrt{n}$ yoğunluğuna sahip bir tahmin edici göstermiştir.

Robinson (1987) ve Carroll (1982) yarıparametrik regresyon tahmin edicilerinin etkinlikleri ve varyansları üzerinde çalışmalar yapmışlardır. Eubank (1988) parametrik olmayan regresyonun parametrik olmayan yoğunluk tahminine dayandığını belirtmiştir.

Gassser ve Müler (1989) Kernel regresyon tahmini için farklı bir tahmin edici önermiştir. Newey (1990) minimum varyansı elde edebilmek için çalışmalar yapmıştır. Jhones ve ark. (1994) Kernel regresyon tahmin edicilerinin arasındaki ilişkiyi incelemişlerdir.

Yatchew (1997) yarıparametrik modelde tahmin yöntemleri üzerine çalışmış ve temel bir tahmin edici önermiştir. Bu yıllarda Eubank (1998), Reese (1998), Hardle (1998), Li ve Ullah (1998), Ullah ve Roy (1998) yarıparametrik regresyon üzerinde çalışmalar yapmışlar ve yarıparametrik regresyonda farklı modeller geliştirmişlerdir.

Sabit ve rasgele etkili modellerin yarıparametrik ve parametrik olmayan regresyonda kullanılmaya başlaması Robinson (1988) tarafından ele alınmış, Hardle (1990), Pagan ve Ullah (1999) yıllarındaki çalışmalarda detaylı olarak incelenmiştir.

Cula (1998), parametrik olmayan yöntemler için olasılık yoğunluk fonksiyonu tahmininin önemine değinmiş ve parametrik olmayan yaklaşımların en kolay şekilde yoğunluk tahmini ile tanımlanabileceğini belirtmiştir. Ayrıca parametrik olmayan yoğunluk tahminlerinden biri olan Kernel tahmin edicilerinin, parametrik olmayan regresyon fonksiyonunda tahmin için kullanılabildiğini belirtmiştir.

Wei ve Cheng (2000) panel verili regresyon modellerinde yarıparametrik regresyonun kullanımına değinmiş tekrarlı ölçümler ile model oluşturmuşlardır.

Schimek (2000) yarıparametrik modellerde düzgünleştirme yöntemi ile parametre tahminlerine çalışmış, Lee (2003) düzgünleştirme yönteminde h parametresinin seçimi için deneme çalışmaları yapmıştır.

Mukherjee (2002) tez çalışmasında yarıparametrik regresyonda sabit etkili ve rasgele etkili modeller için genelleştirilmiş tahmin edicileri kullanmış ve ekonomi üzerine bir uygulama yapmıştır.

Daha önceleri Yatchew (1997) tarafından öne sürülen yarıparametrik modelde farka dayalı tahmin yöntemlerini Kliple ve Eubank (2007) çalışmalarında ele almış ve varyans tahminleri için farka dayalı yöntemler geliştirmişlerdir.

Wasserman (2006) kitabında parametrik olmayan yöntemler başlığı altında parametrik olmayan ve yarıparametrik yöntemlerin uygulanmasına ilişkin temel noktaları incelenmiş ve temel bir kaynak kitap oluşturmuştur.

Racine ve Qui (2007) tarafından yarıparametrik modellerde kategorik veri analizi ile ilgili bir çalışma yapmıştır. Son yıllarda yarıparametrik regresyon üzerinde çok sayıda çalışmalar yapılmaktadır ve farklı veri türleri için tahmin ediciler geliştirilmektedir.

2. PARAMETRİK OLMAYAN REGRESYON

Regresyon analizi, istatistiksel veri analizinin en çok çalışılan konularından biridir. Regresyon analizi, uygulama alanında birçok yöntem ile kullanılan ve hala üzerinde çok fazla çalışılan bir konudur. Analiz edilmesi düşünülen veri setinin biçimine göre, değişken sayısına göre, tahmin yöntemlerine göre, model üzerinde yapılan değişikliklere göre birçok farklı regresyon modeli geliştirilmiştir. Regresyonda temel amaç, bağımlı değişken olarak kabul edilen Y değişkenini bağımsız değişken veya değişkenler $(x_1,...,x_q)$ yardımıyla açıklayacak bir model bulmak ve değişkenler arasındaki fonksiyonel bağlantıyı modelleyip Y bağımlı değişkenini en az hata ile tahmin etmektir. Regresyon modeli genel anlamda

$$Y_i = f(x_i) + \varepsilon_i, \qquad i = 1,2,...,N \tag{2.1}$$

biçiminde ifade edilir (Müller 2000). Modelde x_i ile gösterilen bağımsız değişkeni, Y_i bağımlı değişkeni ve i, gözlem numarasını göstermektedir. $f(\cdot)$ fonksiyonu ise bağımlı değişken ile bağımsız değişkenler arasındaki bağlantıyı veren ve x_i bağımsız değişkenlerine bağlı olan fonksiyondur. ε_i ' ler ise bağımsız ve $\varepsilon_i \sim N(0,\sigma^2)$ dağılıma sahip, hata terimi olarak adlandırılan rasgele değişkenlerdir. Eşitlik (2.1) ile verilen modeli ele aldığımızda eğer $f(\cdot)$ fonksiyonunun fonksiyonel biçimi önceden bilindiği kabul ediliyorsa ve parametrelerin sonlu bir kümesi tarafından tanımlandığı varsayılıyorsa model parametrik model olarak isimlendirilir. Bu durumda modeli

$$Y_i = f(x_i, \beta) + \varepsilon_i \qquad i = 1,2,...,N \tag{2.2}$$

biçiminde yazılabilir. Bu durumda $f(x_i, \beta)$ bilinen bir formdaki, x_i ' lere ve β parametrelerine bağlı bir fonksiyonu göstermektedir. Parametrik modeller parametrelerine göre doğrusal veya doğrusal olmayan biçimde olabilirler ancak önceden belirtildiği gibi $f(x_i, \beta)$ fonksiyonunun biçiminin fonksiyonel olarak bilindiği kabul edilmektedir. Parametrik modellerde hatalar ve alınan örneklem üzerinde bazı katı varsayımlar mevcuttur (Montgomery ve Peck 1992). Uygulama açısından basit yöntemler kullanılabildiği için yaygın olan parametrik yöntemler, bu varsayımların sağlanamadığı durumlarda kötü sonuçlar vermektedir (Karahasan 1999). Ayrıca parametrik regresyon modelleri etkili ve aykırı gözlemlerin varlığında kötü sonuçlar verebilmektedir. Bu durumlarda parametrik olmayan yöntemlerin kullanılması daha iyi sonuçlar verebilir. Eşitlik (2.1) ile verilen modelde $f(x_i)$ fonksiyonunun biçimi bilinmiyorsa modele parametrik olmayan regresyon modeli adı verilir (Ullah ve Pagan 1999). Parametrik olmayan regresyon modeli

$$Y_i = m(x_i) + \varepsilon_i, \qquad i = 1,2,...,N \tag{2.3}$$

eşitliği ile gösterilir. Parametrik olmayan regresyon modelinin gösteriminde karışıklık olmaması için $f(x_i)$ bilinmeyen fonksiyonu bundan sonra $m(x_i)$ olarak ifade edilecektir. Parametrik olmayan regresyonda da hatalar üzerinde varsayımlar mevcuttur. ε_i hatalarının 0 ortalamalı σ^2 sabit varyanslı, bağımsız ve aynı dağılımlı olduğu varsayılır. Parametrik olmayan regresyon, $m(x_i)$ fonksiyonu esnek bir fonksiyonel biçime sahip olduğu için verilerdeki hızlı değişimden, parametrik yöntemin aksine, çok fazla etkilenmeyecektir. Parametrik olmayan regresyonda önceden belirlenen bir fonksiyon ve parametre olmadığından amaç bilinmeyen $m(x_i)$ fonksiyonunu tahmin etmektir. Bu fonksiyon X bağımsız değişkeni ile Y bağımlı değişkeni arasındaki ilişkiyi tanımladığından $m(x_i)$ ' in tahmin edilmesi demek $E(Y|X)$ beklenen değerinin bulunması demektir (Hardle 1994). Tahmin probleminin temeli yoğunluk tahmini yaklaşımına göre kurulur. Bu tahmin genellikle x ' in bir komşuluğundaki Y bağımlı değişkeninin ağırlıklandırılması ile yapılmaktadır.

Yoğunluk tahmini, ağırlıklandırma ve regresyon modelinin tahmin yöntemleri ile ilgili detaylı bilgi ilerleyen bölümlerde verilecektir.

Parametrik olmayan yaklaşım yalnızca $f(x_i)$ fonksiyonunun bilinmeyen formda olması durumunda değil, doğrusal modelde varsayımların sağlanmaması durumunda da kullanılan tekniklere verilen ortak bir isimdir. Parametrik regresyonun dayandığı temel varsayımlar sağlanmadığı zaman, en yaygın tahmin yöntemi olan en küçük kareler yöntemi iyi sonuç vermemektedir. Bu varsayımlar her veri seti için tahmin yapmayı zorlaştırmaktadır. Bu gibi durumlarda popülasyon dağılımı ile ilgili herhangi bir varsayıma ihtiyaç duymayacak parametrik olmayan yaklaşımlar geliştirilmiştir(Pagan ve Ullah 1999). Bu yaklaşımlardan biri medyana dayalı basit doğrusal regresyon için tahmin yöntemidir.

2.1. Medyana Dayalı Parametrik Olmayan Tahmin Yöntemi

Parametrik olmayan tahmin yöntemlerinden biri olan medyana dayalı tahmin yöntemi bağımlı değişken ile bağımsız değişken arasındaki bağlantıyı doğrusal bir model yardımıyla açıklamayı amaçlar. Basit doğrusal regresyon modelinde parametrik olmayan tahmin yöntemlerinden medyana dayalı tahmin yöntemi ilk olarak Brown ve Mood tarafından uygulanmış. Ardından Theil tarafından farklı bir yöntem geliştirilmiştir. Her iki yöntemde basit bir algoritmaya dayalıdır. Görsel yöntemler olduğu için bağımsız değişken sayısı birden fazla olduğunda kullanılamazlar. Yani basit doğrusal model için tasarlanmışlardır ve bağımlı değişken ile bağımsız değişken arasında doğrusal bir ilişki olması durumunda kullanılmaktadır. Bu çalışmada değişkenler arasında doğrusal bir ilişki olmaması durumunda, bağımlı değişken ile bağımsız değişken arasındaki ilişkinin daha çok 2. dereceden bir polinom eğrisi yardımıyla ifade edildiği durumlarda uygulanabilecek bir tahmin yöntemi geliştirilmiştir ve ilişkinin daha yüksek dereceden polinomlar ile ifade edileceği durumlar için genelleştirilmiştir.

2.1.1. Brown- Mood yöntemi

Brown ve Mood tarafından geliştirilen bu yöntem basit doğrusal regresyon modelini ele alır. Basit doğrusal regresyon modeli

$$Y_i = \mu + \beta x_i + \varepsilon_i, \qquad i = 1,2,...,N \tag{2.4}$$

biçiminde olsun. Y_i ; bağımlı değişkenler, x_i ; bağımsız değişken, μ ve β parametrelerdir. N adet gözlem çiftlerinden oluşan örneklem $(x_{1,}y_1),...,(x_{N,}y_N)$ çiftleri ile gösterilsin. Amaç grafik üstünde bu gözlem çiftlerini göstererek elde edilen noktalara en yakın doğruyu geçirmektir. Bu yöntemde önce doğrunun geçtiği noktalar tespit edilecek ardından noktalar yardımıyla doğru denklemi ve buna bağlı olarak μ ve β parametre değerleri bulunacaktır. Brown ve Mood yönteminde uygulanacak işlemler bir algoritma yardımıyla gösterilirse

Adım 1. İlk olarak $(x_{1,}y_1),...,(x_{N,}y_N)$ gözlem çiftlerine ait serpilme diyagramı çizilir.

Adım 2. x_i değerlerine ilişkin medyan belirlenir ve serpilme diyagramının yatay ekseni üzerindeki x_i değerlerinin medyan noktasından dikey bir çizgi çizilir. Bu işlemin amacı veri setini iki eşit parçaya bölmektir. Bu yüzden çizginin üzerinde bulunan noktalar varsa çizgi veri setini mümkün olduğunca iki eşit guruba ayıracak şekilde sağa veya sola ötelenir.

Adım 3. Diyagramın üzerindeki iki gruptan birine ait gözlem çiftleri (x_{1j} , y_{1j}), diğer gruba ait gözlem çiftleri (x_{2k} , y_{2k}) ile gösterilirse ($j = 1,2,...,J$, J birinci gruptaki gözlem çifti sayısı ; $k = 1,2,...,K$, K ikinci gruptaki gözlem çifti sayısı) sırasıyla x_{1j} , y_{1j} , x_{2k} , y_{2k} değerlerine ait medyanlar hesaplanır.

Adım 4. İlk grup için x_{1j} gözlemlerinin medyan değerinden yatay eksene dik bir doğru çizilir ve y_{1j} gözlemlerinin medyan değerinden düşey eksene çizilen dikme

ile kesiştiği c_1 noktasının koordinatları kaydedilir. Aynı işlem ikinci gruptaki x_{2k} ve y_{2k} gözlemleri içinde yapılarak c_2 noktası elde edilir.

Adım 5. Kaydedilen c_1 ve c_2 noktaları birleştirilerek serpilme diyagramı üzerinde elde edilen doğru, regresyon doğrusunun tahmini için yapılan ilk yaklaşımdır.

Adım 6. Her iki gruptaki gözlem çiftlerinden bu doğruya olan dik uzaklıkların medyanı sıfıra eşit çıkıyorsa son tahmin olarak elde edilen doğru kabul edilir. Daha uygun bir doğrunun varlığını araştırmak için, dik uzaklıkların medyanı sıfır çıkmıyorsa her iki grup içinde uzaklıkların medyanı sıfır olacak şekilde doğru ötelenebilir. Doğrunun nihai pozisyonu için her iki grupta doğrunun geçtiği c_1^* ve c_2^* noktalarının koordinatları kaydedilir.

Adım 7. Son işlem olarak c_1^* ve c_2^* noktalarını birleştiren doğru denklemi hesaplanır ve $Y_i = \mu + \beta x_i + \varepsilon_i$ modeli için μ ve β değerleri bulunmuş olur(Okur 2009).

Parametre değerleri hesaplandıktan sonra uygunluğunun test edilmesi için Brown ve Mood yine medyana dayalı bir test geliştirmiştir. Test edilmek istenen hipotezler

$$H_0 : \mu = \mu_0, \quad \beta = \beta_0$$
$$H_1 : \mu \neq \mu_0 \text{ ve/veya } \beta \neq \beta_0 \qquad (2.5)$$

biçimindeki hipotezlerdir. H_0 hipotezi reddedilemez ise μ_0 ve β_0 olarak tahmin edilen parametrelerin anlamlı olduğuna karar verilecektir. Hipotezin test edilmesi için kullanılacak test istatistiği, yine grafik üzerinde gözlem çiftleri ile hesaplanmaktadır (Daniel 1990). Test istatistiğinin hesaplanma algoritması adım adım verilecek olursa;

1. Parametre tahmininde olduğu gibi, $(x_{1,}y_1),...,(x_{N,}y_N)$ gözlem çiftlerinin serpilme diyagramı çizilir.
2. Serpilme diyagramı üzerinde tahmin edilen parametre değerleri $\mu = \mu_0, \quad \beta = \beta_0$ alınarak $y = \mu_0 + \beta_0 x$ doğrusu çizilir.

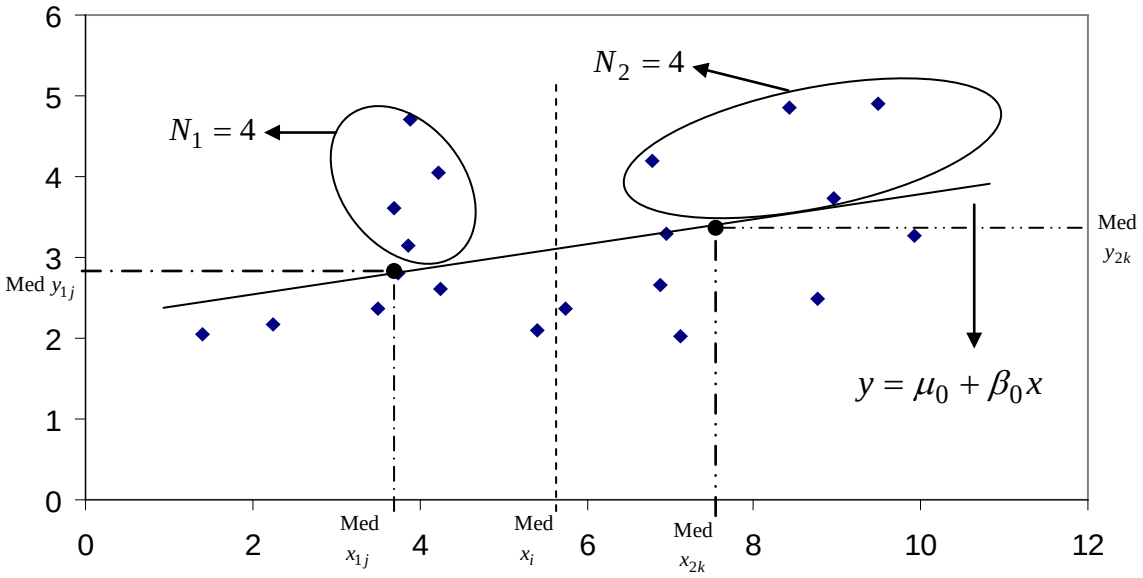
3. Yatay eksende bulunan x_i değerlerine ilişkin medyan belirlenir ve serpilme diyagramının yatay ekseni üzerindeki x_i değerlerinin medyan noktasından dikey bir çizgi çizilir. Bu adımlar parametre tahmini safhasında da yapıldığı için kolaylıkla elde edilir.

4. Diyagram üzerinde çizilen medyan dikmesinin solundaki yani x_i değerlerinin meydanından küçük olan x_i ' lere karşılık gelen gözlemlerden $y = \mu_0 + \beta_0 x$ regresyon doğrusunun üzerinde kalanlarının sayısı N_1 , medyan dikmesinin sağındaki yani x_i değerlerinin meydanından büyük olan x_i ' lere karşılık gelen gözlemlerden $y = \mu_0 + \beta_0 x$ regresyon doğrusunun üzerinde kalanlarının sayısı N_2 ,toplan gözlem sayısı da N olmak üzere test istatistiği;

$$\chi^2 = \left[\left(N_1 - \frac{N}{4} \right)^2 + \left(N_2 - \frac{N}{4} \right)^2 \right] \frac{8}{N} \tag{2.6}$$

eşitliği ile verilir. Mood N_1 ve N_2 değerlerinin 0.5 parametreli binom dağılımına sahip olduğunu ve buna bağlı olarak test istatistiğinin iki serbestlik dereceli χ^2 (ki-kare) dağılımına sahip olduğunu belirtmiştir. Yani $N_1 \sim binom(N^*, 1/2)$ olup burada N^*, veriler medyan doğrusu ile iki parçaya bölündüğü için N_1 değişkeninin alabileceği en büyük değer sayısı yani $\frac{N}{2}$ olmak üzere, $E(N_1) = \frac{N}{2}\frac{1}{2} = \frac{N}{4}$ ve $Var(N_1) = \frac{N}{8}$ ' dir. Temel kavramlar gereği bir X rasgele değişkeni için $\left(\frac{X - E(X)}{\sqrt{Var(X)}} \right) \sim N(0,1)$ olacağından $N_1 - \frac{N}{4} \Big/ \sqrt{N/8}$ ifadesi Standart Normal Dağılıma sahiptir. Bu ifadenin karesi de bir serbestlik dereceli χ^2 dağılımıdır. Aynı sonuç N_2 değişkeni için de elde edileceğinden toplam şeklindeki (2.6) eşitliği iki serbestlik dereceli χ^2 dağılımına sahip olacaktır. Bu durumda dikkat edilmesi gereken husus standart normal dağılım dönüşümü için N örneklem sayısının büyük olması gerektiğidir. Bazı araştırmacılar N sayısının 10 dan büyük olmasının test için yeterli

olacağını ifade etmişlerdir ancak örneklem miktarını daha büyük tutmakta testin geçerliliği açısından fayda vardır. Eşitlik (2.6) ile hesaplanan değer, χ^2 tablosundan bakılan iki serbestlik dereceli değerden büyük olursa H_0 hipotezi reddedilir. (Mood ve Brown 1951)



Şekil 1. Brown-Mood yönteminin uygulanışı

Şekil 1' de 20 adet gözlem değeri için elde edilen sonuçlara göre Brown-Mood yöntemi serpilme diyagramı üzerinde gösterilmiştir. Regresyon doğrusunun çizilmesi ve N_1 , N_2 değerlerinin bulunması şekil üzerinde net olarak görülmektedir.

2.1.2. Theil yöntemi

Brown ve Mood tarafından geliştirilen parametrik olmayan tahmin yöntemine farklı bir yaklaşım da Theil tarafından yapılmıştır. Theil gözlem değerlerinin medyanı yerine her bir gözlem çifti ile oluşturulacak modelden elde edilen tahmin değerlerinin medyanını kullanmayı daha uygun bulmuştur. Theil' e göre $Y_i = \mu + \beta x_i + \varepsilon_i$ modelindeki ε_i hata terimlerinin regresyon modeli varsayımına göre sıfır ortalamalı sabit σ^2 varyanslı olduğu düşünülürse, en iyi modeli bulmak için ε_i hata terimlerinin

medyanı sıfır olmalıdır. Bu amaçla parametre tahmini için uygulanacak yöntemler için bir algoritma hazırlanmıştır.

Adım 1. Bağımsız değişkenler $x_1 < x_2 < \cdots < x_N$ olacak biçimde sıralanır.

Adım 2. Bağımsız değişkenlerin sıralamasına göre N adet gözlem çifti arasından, $i, j = 1,2,...,N$ olmak üzere $i < j$ olacak biçimde elde edilebilecek $\binom{N}{2}$ sayısı kadar gözlem çifti için $b_{ij} = \dfrac{y_j - y_i}{x_j - x_i}$ eğimleri hesaplanır. Bu eğimler, aslında uydurulmak istenen $y = \mu + \beta x$ doğrusu için olası bütün $\hat{\beta}$ tahmin değerleridir.

Adım 3. Bulunan b_{ij} değerlerinin medyanı hesaplanır ve $\hat{\beta} = medyan(b_{ij})$ olarak tahmin edilir.

Adım 4. μ parametresini tahmin etmek içinde $\hat{\beta}$ tahmin değeri yardımıyla

$$\hat{\mu} = medyan(y_i) - medyan(x_i)\hat{\beta} \tag{2.7}$$

eşitliği kullanılır(Hussain ve Sprent 1983). Theil yöntemi kullanılarak elde edilen parametrelerin testleri için farklı test istatistikleri Theil tarafından geliştirilmiştir. Ancak elde edilen doğruyu Brown Mood Yönteminde kullanılan χ^2 test istatistiği yardımıyla test etmek mümkündür (Okur 2009).

Parametrik olmayan Brown-Mood ve Theil yöntemleri doğrusal ilişkiye sahip gözlemlerde tahmin yapmak için basit uygulanabilirliği ve varsayımlardan çok etkilenmemeleri sayesinde uygun olmakla beraber tahmin yapmadan önce verilerin serpilme diyagramı incelendiğinde gözlemler arasında ikinci dereceden bir polinom ifade edilecek şekilde bir ilişki varsa bu yöntemler uygun sonuç vermeyecektir. Buna karşılık medyan yöntemi aşağıda önerdiğimiz biçimde de kullanılabilir.

2.1.3. Doğrusal ilişkiye sahip olmayan veriler için medyan tahmini

Brown-Mood ve Theil yöntemleri değişkenler arasındaki bağlantıyı bir doğru yardımıyla açıklamaktadır. Bazı veri setleri arasındaki bağlantının polinom biçiminde bir eğri ile açıklanmasının daha uygun olacağı düşünülerek çalışmamızda yeni bir medyan tahmini yöntemi geliştirdik.

Veri setindeki gözlemlerin serpilme diyagramı çizildiğinde bağımlı değişken y_i' ler ile bağımsız değişken x_i' ler arasındaki ilişkinin doğrusal olmayan bir biçimde olduğu görülüyorsa yani matematiksel olarak ilişki $y = \mu + \beta_0 x + \beta_1 x^2$ şeklinde ikinci dereceden bir polinom ile ifade edilebilecek biçimde ise regresyon modeli

$$Y_i = \mu + \beta_0 x_i + \beta_1 x_i^2 + \varepsilon_i, \qquad i = 1,2,...,N \tag{2.8}$$

olarak kurulur. Bu model üzerinden parametre tahminini yapmak için önerdiğimiz yöntem Brown-Mood yöntemindeki gibi verilerin bölgelere ayrılması ile ilgilendiği gibi Theil yöntemindeki gibi olası tüm parametrelin hesaplanarak medyanının alınması fikrini de kullanmaktadır. Yöntemin uygulanma adımları aşağıdaki gibi sıralanabilir.

Adım 1. İlk olarak $(x_{1,}y_1),...,(x_{N,}y_N)$ gözlem çiftlerine ait serpilme diyagramı çizilir.

Adım 2. Bir parabolün biçimi hakkında kabaca bir fikir sahibi olmak için en azından üç noktaya ihtiyaç duyulacağını düşünerek x_i değerlerine ilişkin Q_{33} ile Q_{66} ile gösterilen %33' lük ve %66' lık kuantil değerleri hesaplanır ve serpilme diyagramının yatay ekseni üzerindeki x_i değerlerinin $Q_{.33}$ ve $Q_{.66}$ kuantil noktalarından dikey çizgiler çizilir. Bu sayede parabolün kollarını ve ortasını temsil edecek şekilde veri setinin 3 eşit parçaya bölünmesi hedeflenir. Dik çizgiler üzerinde bulunan noktalar her bir grubun mümkün olduğunca eşit sayıda gözlem çiftinden oluşması için sağa veya sola ötelenebilir.

Adım 3. Üç gruba ayrılan gözlem çiftlerinden ilk grupta olanlar (x_{1j} , y_{1j}), ikinci grupta olanlar (x_{2k} , y_{2k}) ve son grupta olanlar (x_{3t} , y_{3t}) ile gösterilsin. Burada $j = 1,...,J$, J birinci gruptaki gözlem çifti sayısı ; $k = 1,2...,K$, K ikinci gruptaki gözlem çifti sayısı; $t = 1,2...,T$, T üçüncü gruptaki gözlem çifti sayısını verir. $J + K + T = N$ olduğu görülmektedir. Theil yönteminde olduğu gibi olası bütün parametre tahminlerini bulmak amacıyla

$$\begin{aligned} y_{1j} &= \mu + \beta_0 x_{1j} + \beta_1 x_{1j}^2 \\ y_{2k} &= \mu + \beta_0 x_{2k} + \beta_1 x_{2k}^2 \quad , \quad j = 1,2,...,J\,;k = 1,2,...,K\,;t = 1,2,...,T \\ y_{3t} &= \mu + \beta_0 x_{3t} + \beta_1 x_{3t}^2 \end{aligned} \tag{2.9}$$

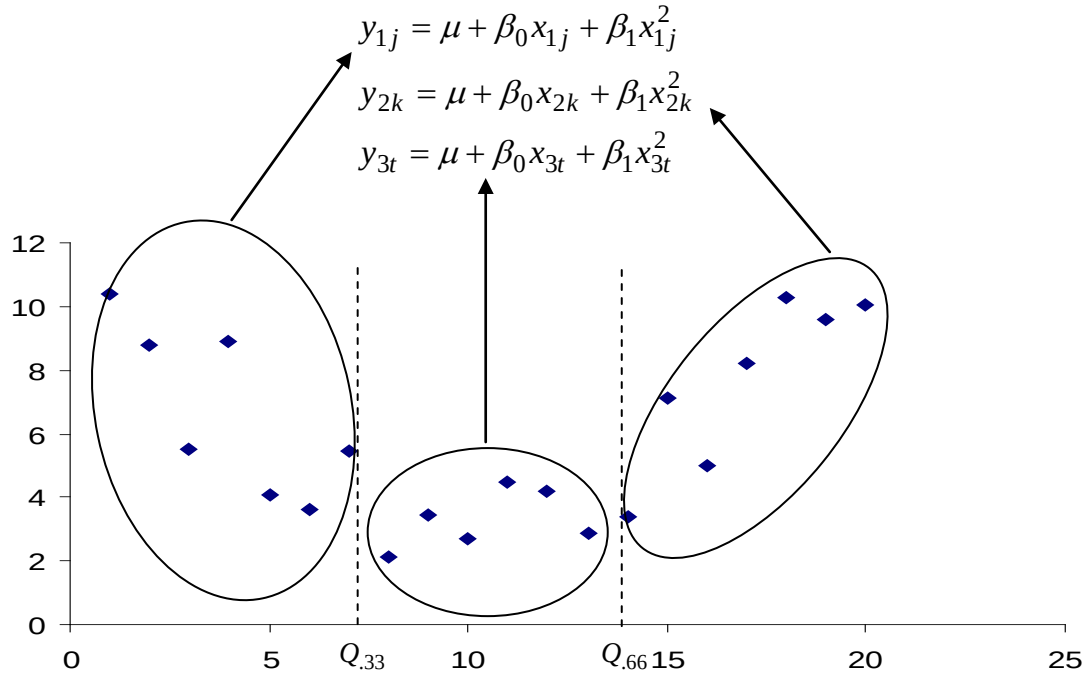
eşitliği ile verilen denklem sisteminin her bir satırına ilgili gruptan alınan gözlem çifti yerleştirilerek $J \times K \times T$ adet çözüm bulunur.

Adım 4. Bulunan $J \times K \times T$ adet çözümden elde edilen μ_{jkt}, $\beta_{0\,jkt}$ ve $\beta_{1\,jkt}$ parametreleri sıralanır ve nihai çözüm olarak kullanılacak tahmin değerleri

$$\begin{aligned} \hat{\beta}_0 &= medyan\left(\beta_{0\,jkt}\right) \\ \hat{\beta}_1 &= medyan\left(\beta_{1\,jkt}\right) \\ \hat{\mu} &= medyan\left(\mu_{jkt}\right) \end{aligned} \tag{2.10}$$

olacak biçimde hesaplanır. Böylece elde edilen regresyon eğrisi $\hat{\mu}$, $\hat{\beta}_0$ ve $\hat{\beta}_1$ lar tahmin değerleri olmak üzere $y = \hat{\mu} + \hat{\beta}_0 x + \hat{\beta}_1 x^2$ biçiminde elde edilir.

Şekil 2' de örnek olacak biçimde üretilen 20 adet gözlem çiftinin serpilme diyagramında çizilerek kuantiller yardımıyla 3 gruba ayrılması ve denklem sistemlerinin oluşturulması çizim yardımıyla gösterilmektedir.



Şekil 2. Doğrusal olmayan ilişki için medyan tahmini

Bu yöntem için daha farklı algoritmalarda düşünülmüştür. Değişkenler arasındaki ilişkiyi gösteren modelin derecesine göre algoritma genellenecek olursa polinomun derecesini p ile gösterilsin veri seti p adet eşit uzaklıktaki kuantil yardımıyla $p+1$ adet grup halinde parçalanabilir. Bu gruplar yardımıyla elde edilecek $p+1$ bileşenli denklem sistemleri çözülerek aynı algoritma uygulanabilir.

p. dereceden bir polinomun tahmini için oluşturulacak denklem sistemi

$$\begin{aligned} y_{1j} &= \mu + \beta_0 x_{1j} + \beta_1 x_{1j}^2 + \cdots + \beta_{p-1} x_{1j}^p \\ y_{2k} &= \mu + \beta_0 x_{2k} + \beta_1 x_{2k}^2 + \cdots + \beta_{p-1} x_{2k}^p \\ &\vdots \qquad\qquad \vdots \\ y_{(p+1)t} &= \mu + \beta_0 x_{(p+1)t} + \beta_1 x_{(p+1)t}^2 + \cdots + \beta_{p-1} x_{(p+1)t}^p \end{aligned} \tag{2.11}$$

biçimindedir. Her bir grubun eleman sayıları sırasıyla $j = 1,...,J$; $k = 1,...,K$; $t = 1,...,T$ olarak verilmektedir ve bu grupları oluştururken kullanılacak ilk kuantil

$Q_{\frac{1}{p+1}}$ biçiminde hesaplanabilir. Ardından gelecek kuantillerin derecesi sırayla ilk kuantilin yüzde miktarı eklenerek hesaplanır. Dikkat edilecek bir başka husus da hiçbir gözlem çiftini gözden kaçırmamaktır. Bunun kontrolü için $p+1$ gruptaki eleman sayıları toplamının N olup olmadığı kontrol edilebilir. Genelleştirilmiş çözüm sayesinde herhangi bir dereceden polinom için uygun çözüm değerlerine ulaşılabilir. Bu sayede bağımlı değişken ile bağımsız değişken arasındaki ilişkinin yapısı 2 veya daha fazla dereceden bir polinom ile açıklanabiliyorsa çözüm için bu çalışmada önerilen yöntem kullanılabilir.

2.2. Yoğunluk Tahmin Edicileri

Parametrik olmayan regresyon analizinin temeli yoğunluk fonksiyonu tahminine dayanmaktadır. Parametrik olmayan regresyon analizinde kitle dağılımının sürekli olduğu göz önüne alınırsa regresyon fonksiyonunun tahmini de yoğunluk fonksiyonu tahmini ile bulunmaktadır. Bir X rasgele değişkeni hakkında geniş bilgi elde etmek istenirse ilk olarak X değişkeninin yoğunluk fonksiyonuna bakılır (Wasserman 2006).

Yoğunluk tahmini, gözlenmiş verilerden yola çıkarak, o verilerin ait olabileceği bir yoğunluk fonksiyonunun elde edilmesi olarak adlandırılabilir. Yoğunluk tahmini parametrik tahmin ediciler ve parametrik olmayan tahmin ediciler olarak iki gurupta incelenebilir (Pagan ve Ullah 1999). Parametrik olmayan yoğunluk tahmini için farklı yöntemler vardır. Bunların arasında en çok kullanılanlar, histogram tahmin edicisi (Saf tahmin edici), en yakın komşuluk tahmin edicisi ve kernel tahmin edicisidir.

2.2.1. Histogram tahmin edicisi

Yoğunluk tahmini için kullanılan yöntemlerden en eski ve basit tahmin edici Histogram tahmin edicisidir. Tahmin yönteminin temelinde yoğunluk fonksiyonunu

şekline göre belirlemeye çalışmaktadır. Tahmin edici ilk olarak Fix ve Hodges (1951) tarafından öne sürülmüştür.

$X_1, X_2, ..., X_n$ n birimlik örneklem ve bu örneklemden elde edilen yoğunluk fonksiyonu tahmin edicisi de $\hat{f}_n(x)$ olsun. Aralık genişliği h olarak verildiğinde $\left[x - h/2, x + h/2\right]$ aralığına da histogram aralıkları denir (Hardle 1994). Dağılım fonksiyonu, $F(x) = P(X \le x)$ olsun, olasılık yoğunluk fonksiyonunun tanımı gereği

$$f(x) = \frac{d}{dx} F(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$

biçiminde ifade edilir. Bu tanım, artış miktarı $\frac{h}{2}$ alınır ve basit bir dönüşümle;

$$f(x) = \frac{d}{dx} F(x) = \lim_{h \to 0} \frac{F\left(x + \frac{h}{2}\right) - F\left(x - \frac{h}{2}\right)}{h}$$

$$= \lim_{h \to 0} \frac{P\left(x - \frac{h}{2} < X \le x + \frac{h}{2}\right)}{h} \qquad (2.12)$$

biçiminde olasılık yoğunluk fonksiyonu tanımlanır(Pagan ve Ullah 1999). Olasılık yoğunluk fonksiyonu $f(x)$'i tahmin etmek için h bant genişliğini n in pozitif bir fonksiyonu ve $n \to \infty$ iken $h \to 0$ olarak düşünülürse ve aralık genişlikleri eşit olarak alınırsa olasılık yoğunluk fonksiyonunun tahmin edicisi için, $P(x - \frac{h}{2} < X < x + \frac{h}{2})$ olasılığı $(x - \frac{h}{2}, x + \frac{h}{2})$ aralığına düşen örneklem sayısı ile tahmin edilirse

$$\hat{f}(x) = \frac{1}{nh}\left[\left(x - \frac{h}{2}, x + \frac{h}{2}\right) \text{ aralığındaki } X_1, X_2, ..., X_n \text{ lerin sayısı } \right] \qquad (2.13)$$

veya $P\left(x-\frac{h}{2}<X\le x+\frac{h}{2}\right)$ ifadesinin $P\left(-\frac{1}{2}<\frac{X-x}{h}\le\frac{1}{2}\right)$ biçiminde yazılması ile

$$\hat{f}(x)=\frac{1}{nh}\left[\left(\frac{-1}{2},\frac{1}{2}\right] \text{ aralığındaki } \frac{X_1-x}{h},...,\frac{X_n-x}{h} \text{ lerin sayısı } \right] \qquad (2.14)$$

olarak hesaplanır. $\hat{f}(x)$ saf tahmin edici olarak ta bilinmektedir. (Pagan ve Ullah 1999). Aralık genişliklerinin eşit olmadığı durumda ise,

$$\hat{f}(x)=\frac{1}{nh_x}\left[\left(x-\frac{h_x}{2},x+\frac{h_x}{2}\right] \text{ aralığındaki } X_1,X_2,\ldots,X_n \text{ lerin sayısı } \right] \qquad (2.15)$$

olarak verilebilir. Burada h_x , x 'in içinde bulunduğu aralık genişliği olarak alınmaktadır. Tahmin edici olasılık yoğunluk fonksiyonu olarak

$$\hat{f}(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}I\left(-1/2<\psi<1/2\right) \qquad (2.16)$$

biçiminde gösterilebilir. Burada $\psi=\frac{X_i-x}{h}$, h ; bant genişliğini göstermektedir ve

$$I(x)=\begin{cases}1, & -1/2<x<1/2\\ 0, & dy\end{cases} \qquad (2.17)$$

olarak ifade edilen indikatör fonksiyonudur. Saf tahmin edici $x_i\pm h/2$ noktalarında sıçramalara sahiptir ve sürekli bir fonksiyon olmadığından uygulamada sorunlar

çıkarabilir ve yanlış yorumlamalara sebep olabilir. Bu yüzden saf tahmin edici, çok sık tercih edilmemektedir (Müller 2000).

Histogram tahmini basit bir yöntem olmasına karşın uygulamalarda bilgi kaybına sebep olabilmektedir. Aralık genişliklerinin seçimi, seçilen aralıkların orta noktasının belirlenmesi sırasında tahmin edicinin davranışları değişebilmektedir (Wasserman 2006).

2.2.2. En yakın komşuluk tahmin edicisi

Histogram tahmin edicisinde $\left(x-\frac{h}{2}, x+\frac{h}{2}\right)$ aralığındaki $X_1, X_2, ..., X_n$' lerin oranına göre tahmin yapılmaktadır. Sabit bant genişliğinde işlem yapıldığından aralıkların bazılarında az düzleştirme bazı parçalarında ise çok düzleştirme yapılabilmektedir. Ancak yoğunluğun kuyruk tarafı tahmin edilmek istenildiğinde merkezinde olduğundan daha geniş aralık uzunluğuna ihtiyaç olur. Bu problemi gidermek amacıyla ilk olarak Fix ve Hodgesen yakın komşuluk tahminini önermiştir. En yakın komşuluk tahmin edicisi de saf tahmin edicinin özel bir halidir (Pagan ve Ullah 1999).

Tahmin edicide, d_{kx} ; x 'in $X_1, X_2, ..., X_n$ 'ler arasından k. en yakın komşuluğuna olan uzaklığını göstermek üzere $h = 2d_k(x)$ alınıp (2.13) eşitliğine uygulanırsa

$$\hat{f}(x) = \frac{1}{2nd_k(x)}\left[\left(x - d_k(x), x + d_k(x)\right) \text{ aralığındaki } X_1, X_2, ..., X_n \text{ lerin sayısı }\right]$$

$$= \frac{k}{2nd_k(x)}$$

$$= \frac{1}{2nd_k(x)} \sum_{i=1}^{n} I\left(\frac{x_i - x}{2d_k(x)} \right) \tag{2.18}$$

biçiminde tanımlanan tahmin ediciye *k.* en yakın komşuluk tahmin edicisi denir. Burada düzgünleştirme parametresinin derecesi olan h, k sabiti tarafından kontrol edilmekte ve h, dağılımın kuyruk tarafında $d_k(x)$ uzaklığı merkezde olduğundan daha geniş olmaktadır. Bu da kernel ve saf tahmin edicideki kuyruk yoğunluğu problemini gidermiş olur(Pagan ve Ullah 1999, Silverman 1990).

2.2.3. Kernel yoğunluk fonksiyonu tahmin edicisi ve özellikleri

$X_1, X_2, ..., X_n$ bağımsız ve aynı dağılımlı, f olasılık yoğunluk fonksiyonlu kitleden bir örneklem olmak üzere olasılık yoğunluk fonksiyonu f ' in kernel tahmin edicisi, ampirik dağılım fonksiyonu kullanılarak aşağıdaki gibi elde edilmektedir. f olasılık yoğunluk fonksiyonu olduğundan,

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} \tag{2.19}$$

şeklinde yazılabilir, burada F, f ' e karşılık gelen dağılım fonksiyonudur. Eşitlik (2.19) türevin tanımından

$$f(x) = \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h}$$

biçiminde de yazılabilir. Bu eşitlikten hareketle f ' in bir tahmin edicisi;

$$\hat{f}_h(x) = \lim_{h \to 0} \frac{\hat{F}_n(x+h) - \hat{F}_n(x-h)}{2h} \tag{2.20}$$

olarak yazılabilir. Burada dağılım fonksiyonunun bir tahmin edicisi olan $\hat{F}_n(x)$ ampirik dağılım fonksiyonudur ve

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{I}(X_i \le x) \tag{2.21}$$

şeklinde tanımlanır ve $I(\cdot)$ indikatör fonksiyonudur. Böylece (2.21) eşitliği (2.20) eşitliğinde yerine yazıldığında

$$\hat{f}_h(x) = \frac{\sum_{i=1}^{n} I(X_i \le x+h) - \sum_{i=1}^{n} I(X_i \le x-h)}{2nh}$$

$$= \frac{\sum_{i=1}^{n} I(x-h < X_i \le x+h)}{2nh} \tag{2.22}$$

elde edilir. Bu eşitlikte h yalnız bırakılarak mutlak değer yardımıyla

$$\hat{f}_h(x) = \frac{\sum_{i=1}^{n} I\left(|x - X_i| \le h\right)}{2nh}$$

$$= \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{2} I\left(\left|\frac{x - X_i}{h}\right| \le 1\right)$$

$$= \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{2.23}$$

biçiminde de yazılabilir (Silverman 1990), burada

$$K(u) = \frac{1}{2} I(|u| \leq 1)$$

eşitliği ile verilen uniform kernel fonksiyonudur. Bu fonksiyon *x'* in komşuluğundaki gözlemlere 0.5, komşuluğu dışındakilere ise 0 ağırlığını vermektedir. $K(u)$ fonksiyonu yerine 0 etrafında simetrik herhangi bir kernel olasılık fonksiyonu alınabilir (Wasserman 2006). $K(u)$ fonksiyonunun seçimine ileride yer verilmiştir. Burada $K(u)$ bir olasılık yoğunluk fonksiyonu olarak alındığında $\hat{f}_h(x)$ ' da bir olasılık yoğunluk fonksiyonu olduğunun gösterilmesi gerekmektedir. Olasılık yoğunluk fonksiyonu olması şartlarından biri $\hat{f}_h(x) \geq 0$ olmasıdır. (2.23) eşitliğinde $K(\cdot)$, düzgün dağılımın olasılık fonksiyonu olduğundan $K\left(\frac{x - X_i}{h}\right) \geq 0$ olacaktır ve $h \geq 0$ varsayımı altında $\hat{f}_h(x) \geq 0$ şartı sağlanacaktır. Bir diğer şart olasılık yoğunluk fonksiyonunun, değişkenin tanım aralığında integral sonucunun 1 olması gerekmektedir. Bu şart için,

$$\int_{-\infty}^{\infty} \hat{f}_h(x)dx = \int_{-\infty}^{\infty} \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) dx$$

$$= \frac{1}{nh} \sum_{i=1}^{n} \int_{-\infty}^{\infty} K\left(\frac{x - X_i}{h}\right) dx$$

integrali göz önüne alınsın. İntegral operatörü lineer operatör olduğundan toplam dışındaki integral toplam içine alınabilir. Yukarıdaki integralde $u = \frac{x - X_i}{h}$ dönüşümü yapılırsa $du = \frac{1}{h} dx$ olur ve

$$\int_{-\infty}^{\infty} \hat{f}_h(x)dx = \frac{1}{nh}\sum_{i=1}^{n} \int_{-\infty}^{\infty} K(u)du$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int_{-\infty}^{\infty} K(u)du$$

elde edilir. $K(u)$ düzgün dağılımın olasılık yoğunluk fonksiyonu olduğundan integralin sonucu

$$\int_{-\infty}^{\infty} \hat{f}_h(x)dx = \frac{1}{n}\sum_{i=1}^{n} 1 = 1$$

olarak elde edilir ve buna bağlı olarak $\hat{f}_h(x)$ fonksiyonunun da bir olasılık yoğunluk fonksiyonu olduğu görülür. $\hat{f}_h(x)$ tahmin edicisinin yanı (Bias) Teorem1'de verildiği gibidir.

Teorem 1. $\hat{f}_h(x)$ tahmin edicisinin Yanı (Bias)

$$E\left\{\hat{f}_h(x)\right\} = f(x) + \frac{h^2}{2} f''(x)\mu_2(K) + o\left(h^2\right) \qquad h \to 0 \tag{2.24}$$

dır. Burada, $\mu_2(K) = \int_{-\infty}^{\infty} u^2 K(u)du$ şeklindedir. Görüldüğü gibi Bias h ile doğru orantılı olup h küçüldükçe Bias azalır(Thompson ve Tapia 1990).

İspat .

$\hat{f}_h(x)$ tahmin edicisinin beklenen değeri,

$$E\left\{\hat{f}_h(x)\right\} = E\left\{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)\right\}$$

$$= \frac{1}{nh} E\left\{K\left(\frac{x - X_1}{h}\right) + \cdots + K\left(\frac{x - X_n}{h}\right)\right\}$$

olarak yazılır. Toplamın beklenen değeri X_i' lerin bağımsız ve aynı dağılımlı olması sebebiyle, ayrı ayrı beklenen değerlerin toplamı olarak yazılabileceğinden, sadece X_1 değişkeninin beklenen değeri olarak

$$E\left\{\hat{f}_h(x)\right\} = \frac{1}{nh} nE\left\{K\left(\frac{x - X_1}{h}\right)\right\}$$

biçiminde yazılır ve beklenen değerin tanımından

$$E\left\{\hat{f}_h(x)\right\} = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{X_1 - x}{h}\right) f(x_1)dx_1$$

elde edilir. İntegralde $v = \frac{x_1 - x}{h}$ dönüşümü yapılırsa $dv = \frac{1}{h}dx_1$ olur ve

$$E\left\{\hat{f}_h(x)\right\} = \int_{-\infty}^{\infty} K(v) f(x + vh)dv \qquad (2.25)$$

yazılabilir. $f(x + vh)$ ifadesinin x komşuluğunda Taylor serisine açılımı,

$$f(x+vh) = f(x) + f'(x)vh + \frac{1}{2} f''(x)v^2h^2 + o(h^2) \qquad (2.26)$$

olarak yazılır. eşitlik (2.26), eşitlik (2.25)'de yerine koyularak basit matematiksel işlemler yardımıyla

$$E\left\{\hat{f}_h(x)\right\} = \int K(v)\left\{ f(x) + f'(x)vh + \frac{1}{2} f''(x)v^2h^2 + o(h^2)\right\}dv$$

$$= f(x)\int K(v)dv + f'(x)h\int vK(v)dv + \frac{1}{2} f''(x)h^2\int v^2K(v)dv$$

$$+ o(h^2)\int K(v)dv$$

ifadesi elde edilir. $K(\cdot)$'nın 0 etrafında simetrik olasılık yoğunluk fonksiyonu olması sebebi ile $\int vK(v)dv = 0$ ve $\int K(v)dv = 1$ dır. Ayrıca $\int v^2K(v)dv = \mu_2(K)$ olduğundan

$$E\left\{\hat{f}_h(x)\right\} = f(x) + \frac{1}{2} f''(x)h^2\mu_2(K) + o\left(h^2\right) \quad , \quad h \to 0$$

elde edilir.

Teorem 2. $\hat{f}_h(x)$ tahmin edicisinin varyansı

$$Var\left\{\hat{f}_h(x)\right\} = \frac{1}{nh} R(K) f(x) + o\left(\frac{1}{nh}\right) \qquad nh \to \infty \qquad (2.27)$$

dir, burada $R(K) = \int K^2(u)du$ şeklindedir(Mukherjee 2002).

Buradan anlaşılıyor ki, varyans n ve h ile doğru orantılı olup n veya h büyük olursa varyans azalır. Ancak Teorem 1 de gösterildiği gibi h büyüdükçe de tahmin edicinin bias değeri büyüyeceğinden, h bant genişliğini dengeli seçmek gerekmektedir (Pagan ve Ullah 1999).

2.2.4. Kernel çekirdek fonksiyonunun seçimi

Parametrik olmayan kernel yoğunluk tahmini yönteminde kullanılacak kernel çekirdek fonksiyonu için bazı şartlar vardır. Bu şartları sağladığı sürece herhangi bir fonksiyon kernel çekirdek fonksiyonu olarak kullanılabilir. İlk temel şart $K(x)$ fonksiyonunun sıfır etrafında simetrik ve $K(x) \geq 0$ olması gerekmektedir. Kernel fonksiyonunun sağlaması gereken diğer şartlar

1. $\int_{-\infty}^{\infty} K(x)\, dx = 1$

2. $\int_{-\infty}^{\infty} x\, K(x)\, dx = 0$

3. $\int_{-\infty}^{\infty} x^2\, K(x)\, dx \neq 0$

4. $\int_{-\infty}^{\infty} K^2(x)\, dx < \infty$

olarak sıralanabilir. Özelliklerden de anlaşılacağı gibi kernel çekirdek fonksiyonlarının yerine bu şatları sağlayan olasılık yoğunluk fonksiyonları kullanılabilir. Bu özellileri sağlayan ve yoğunluk tahmini uygulamalarında sıkça kullanılan kernel çekirdek fonksiyonlarından bazıları derlenmiş ve grafikleri ile beraber gösterilmiştir (Wasserman 2006).

Düzgün (Uniform) Kernel Fonksiyonu

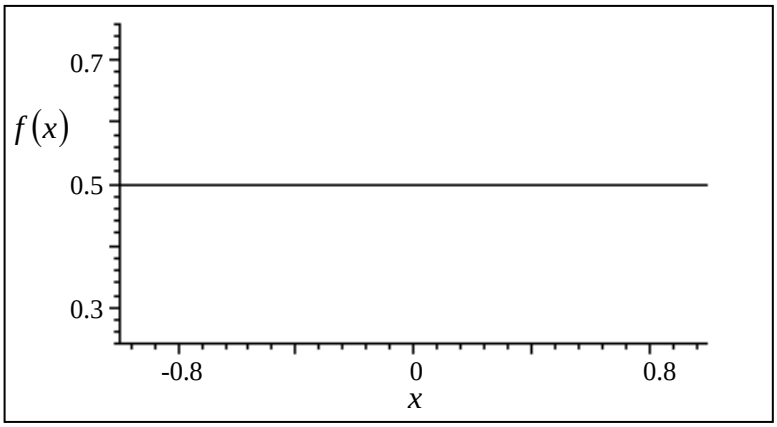
Düzgün kernel fonksiyonu,

$$K(x) = \begin{cases} \frac{1}{2} & -1 \le x \le 1 \\ 0 & d.y. \end{cases}$$

şeklinde tanımlanır, ayrıca

$$\int_{x=-1}^{1} x^2 K(x)\, dx = \sigma_K^2 = 1/3 \quad , \qquad \int_{x=-1}^{1} K^2(x)\, dx = 1/2$$

olup düzgün kernel fonksiyonu kernel çekirdek fonksiyonu olma şartlarını sağlar. Düzgün kernel fonksiyonunun grafiği Şekil 3' deki gibidir.



Şekil 3. Düzgün kernel fonksiyonunun grafiği

Epanechnikov Kernel Fonksiyonu

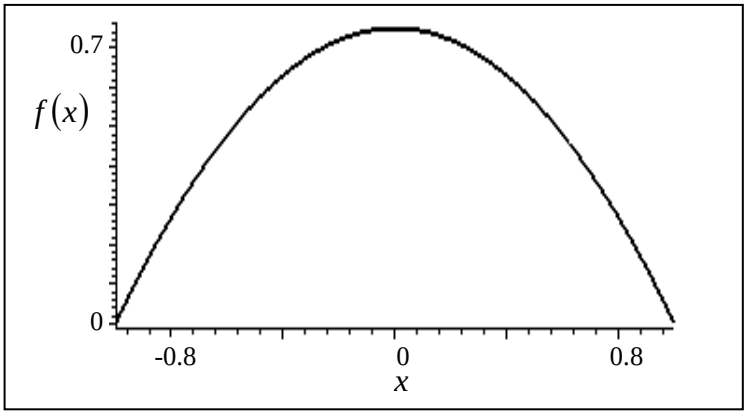
Epanechnikov kernel fonksiyonu

$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \le x \le 1 \\ 0 & d.y. \end{cases}$$

biçiminde tanımlanır, ayrıca

$$\int_{x=-1}^{1} x^2 K(x)\,dx = \sigma_K^2 = 1/5 \quad , \quad \int_{x=-1}^{1} K^2(x)\,dx = 3/5$$

olup, Epanechnikov kernel fonksiyonu kernel çekirdek fonksiyonu olma şartlarını sağlar. Epanechnikov kernel fonksiyonunun grafiği Şekil 4' de görüldüğü gibidir.



Şekil 4. Epanechnikov kernel fonksiyonunun grafiği

Üçgensel (Triangular) Kernel Fonksiyonu

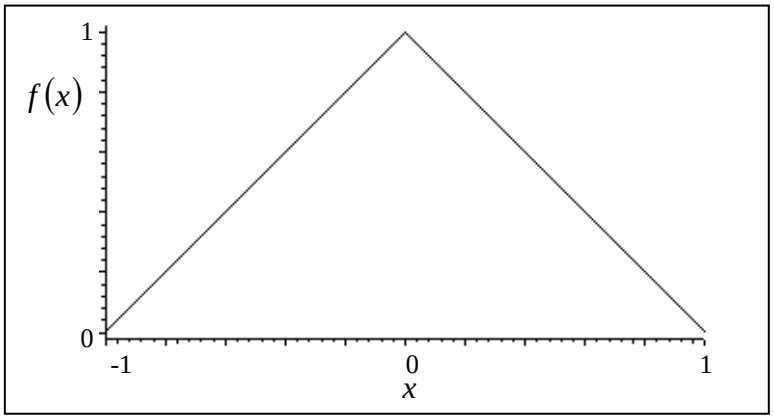
Üçgensel kernel fonksiyonu

$$K(x) = \begin{cases} 1-|x| & -1 \le x \le 1 \\ 0 & d.y. \end{cases}$$

biçiminde verilir ve

$$\int_{x=-1}^{1} x^2 K(x)\,dx = \sigma_K^2 = 1/6 \quad , \quad \int_{x=-1}^{1} K^2(x)\,dx = 2/3$$

olup, kernel fonksiyonu özelliklerinin sağlandığı görülür. Üçgensel kernel fonksiyonunun grafiği Şekil 5' deki gibidir.



Şekil 5. Üçgensel kernel fonksiyonunun grafiği

Üçlü ağırlık (Triweight) Kernel Fonksiyonu

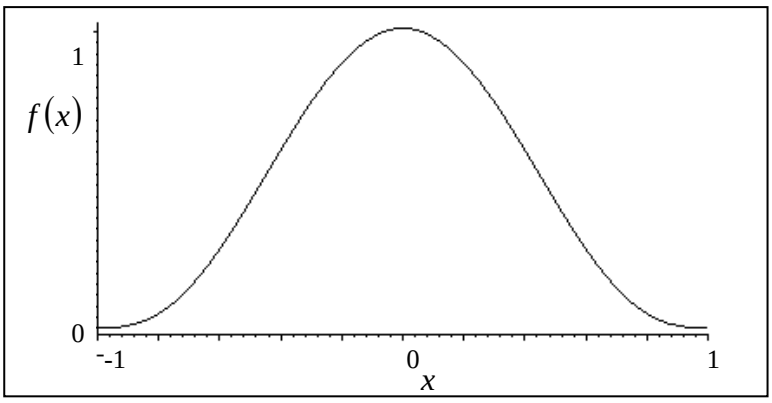
Üçlü ağırlık Kernel Fonksiyonu

$$K(x)=\begin{cases}\frac{35}{32}\left(1-x^2\right)^3 & -1\le x\le 1\\ 0 & d.y.\end{cases}$$

biçiminde ifade edilir. Üçlü ağırlık kernel fonksiyonu

$$\int_{x=-1}^{1} x^2\,K(x)\,dx \;=\; \sigma_K^2 = 1/9 \quad , \quad \int_{x=-1}^{1} K^2(x)\,dx \;= 350/429$$

özelliklerini sağladığından kernel fonksiyonudur. Grafiği Şekil 6' daki gibidir.



Şekil 6. Üçlü ağırlık kernel fonksiyonunun grafiği

İkili Ağırlık (Biweight) Kernel Fonksiyonu

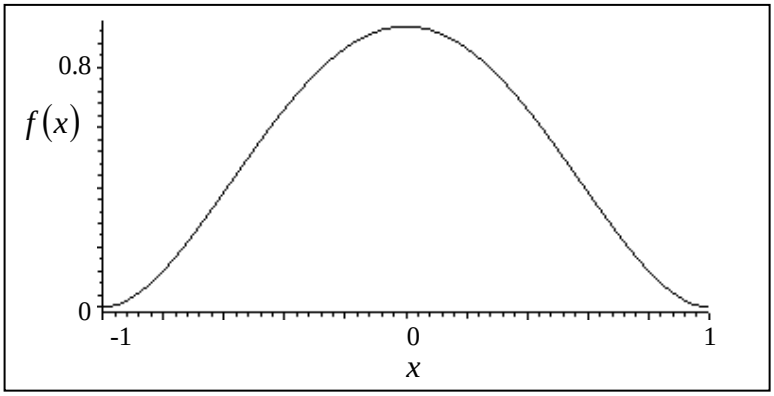
İkili ağırlık kernel fonksiyonu

$$K(x)=\begin{cases}\frac{15}{16}\left(1-x^2\right)^2 & -1\le x\le 1\\ 0 & d.y.\end{cases}$$

şeklinde tanımlanır ve ayrıca

$$\int_{x=-1}^{1} x^2\,K(x)\,dx \;=\; \sigma_K^2 = 1/7 \quad , \quad \int_{x=-1}^{1} K^2(x)\,dx \;= 5/7$$

olup ikili ağırlık kernel fonksiyonu kernel fonksiyonu olma şartlarını sağlamaktadır. İkili ağırlık kernel fonksiyonunun grafiği Şekil 7'de görüldüğü gibidir.



Şekil 7. İkili ağırlık kernel fonksiyonunun grafiği

Kosinüs (Cosine) Kernel Fonksiyonu

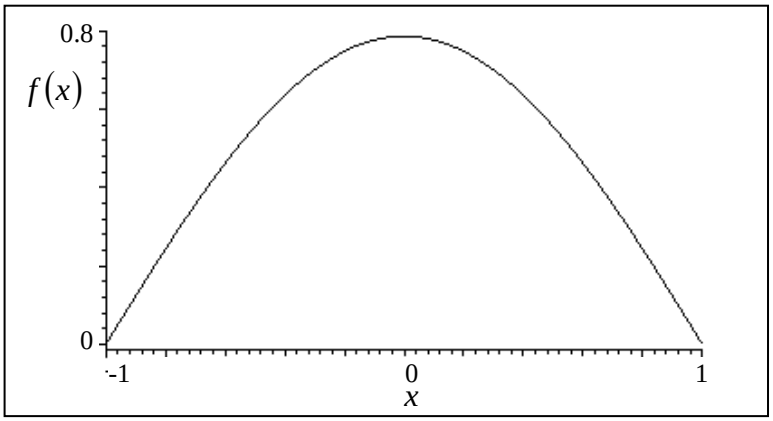
Kosinüs kernel fonksiyonu

$$K(x)=\begin{cases}\frac{\pi}{4}\cos\left(\frac{\pi x}{2}\right) & -1\le x\le 1\\ 0 & d.y.\end{cases}$$

olarak ifade edilir. Kosinüs kernel fonksiyonu

$$\int_{x=-1}^{1} x^2\,K(x)\,dx \;=\; \sigma_K^2 = 1-8/\pi^2 \quad , \quad \int_{x=-1}^{1} K^2(x)\,dx \;= \pi^2/16$$

özelliklerini sağladığı için kernel fonksiyonudur. Kosinüs kernel fonksiyonunun grafiği Şekil 8'deki gibidir.



Şekil 8. Kosinüs kernel fonksiyonu grafiği

Standart Normal (Gaussian) Kernel Fonksiyonu

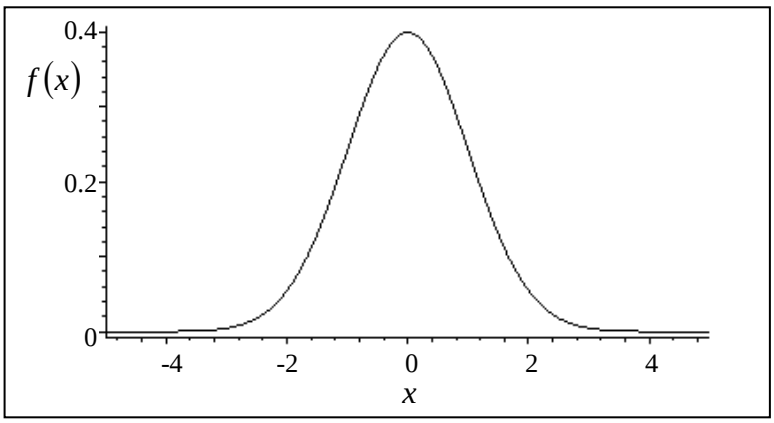
Standart normal dağılım yardımıyla verilen kernel fonksiyonu

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad , \qquad -\infty < x < \infty$$

biçimindedir. Standart normal kernel fonksiyonu,

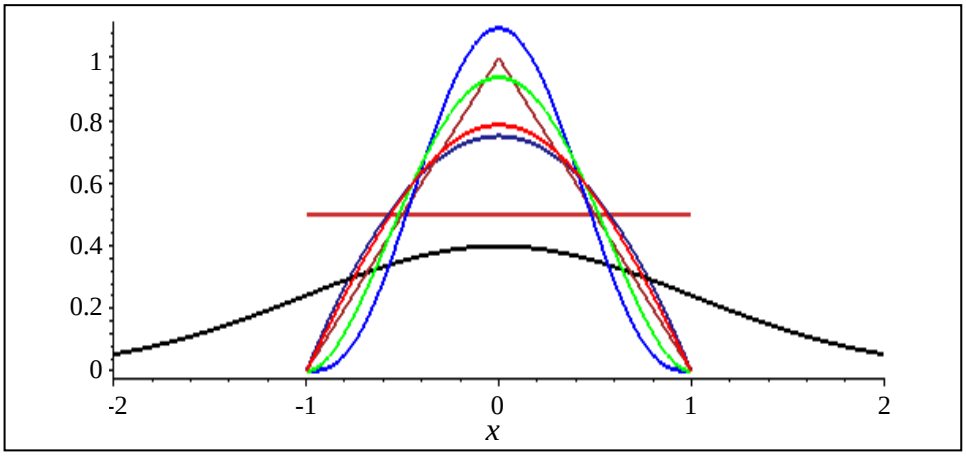
$$\int_{x=-1}^{1} x^2 K(x)\, dx = \sigma_K^2 = 1 \quad , \quad \int_{x=-1}^{1} K^2(x)\, dx = 1/\sqrt{2\pi}$$

özelliklerini sağlar ve grafiği Şekil 9' da görüldüğü gibidir. Diğer kernel fonksiyonlarının hepsi $-1 \le x \le 1$ aralığında tanımlanmış olmasına rağmen Standart Normal kernel fonksiyonu diğer fonksiyonlardan farklı olarak $-\infty < x < \infty$ aralığında tanımlanmıştır.



Şekil 9. Standart Normal kernel fonksiyonunun grafiği

Bu fonksiyonların tahmin sonucunda genellikle yakın değerler verdiği söylenmektedir. En çok kullanılan kernel çekirdek fonksiyonları ise Standart Normal ve Epanechnikov kernel çekirdek fonksiyonlarıdır. Bütün fonksiyonlar Şekil 10'daki gibi aynı grafik üstünde incelenirse hepsinin yakın değerlerde ağırlıklandırma yapacağı daha net görülür.



Şekil 10. Kernel fonksiyonlarının grafiği

2.2.5. Kernel tahmininde "*h*" bant genişliği seçimi

Kernel çekirdek fonksiyonunun seçimi yapıldıktan sonra uygulamalardaki ikinci problem *h* bant genişliğinin seçimidir. Daha önce gösterildiği gibi *h* bant genişliği büyüdükçe $\hat{m}(x)$ tahmin edicisinin varyansı azalır ancak buna karşılık tahmin edicinin yanı (bias) artar. Bu durumda oluşan çelişkiden kurtulmak amacıyla *h* bant genişliğini seçerken dikkatli olunması gerekmektedir. Temel düşünce verilerin az olduğu bir bölgede yeterli veriyi içine alan bir aralık bulmak için *h* bant genişliği daha büyük seçilecek, verilerin yoğun olarak bulunduğu bir bölgede küçük bir *h* bant genişliği seçilecektir. Bu durumda gözlem sayısı *n* arttıkça belli bir alandaki veri yoğunluğu da artacağından *h* bant genişliği daha küçük seçilecektir. *h* bant genişliği düzgünleştirme parametresi olarak da adlandırılır. Düzgünleştirme parametresinin seçimi birçok farklı yöntem yardımıyla yapılabilir. Bu yöntemlerden birine örnek olarak ASE(Averaged squared error) kriterine göre yapılan seçim verilebilir. Bu seçimde;

$$ASE(\text{h}) = n^{-1}\sum_{i=1}^{n}(\hat{m}(x_i) - m(x_i))^2 \tag{2.28}$$

olarak verilen hata kareler ortalaması değerini minimum yapacak şekilde bir *h* seçimini öngörür(Silverman 1990).

Bu kriterin dışında en çok kullanılan yöntemlerden bir diğeri AMISE (Approximate Mean Integrated Squared Error) değerini minimum yapmaya dayalı yöntemdir (Pehlivan 2005).

$$AMISE\left(\hat{f}_h(x)\right) = \frac{1}{nh}R(K) + \frac{h^4}{4}\mu_2^2(K)R(f'') \tag{2.29}$$

Şeklinde verilir. Burada,

$$R(f'') = \int \{f''(x)\}^2 dx$$

şeklindedir ve AMISE değerini minimum yapan h bant genişliği

$$h_{AMISE} = \left\{ \frac{R(K)}{\mu_2^2(K) R(f'') n} \right\}^{1/5} \tag{2.30}$$

biçimindedir. $X_1, X_2, ..., X_n$, f olasılık yoğunluk fonksiyonlu kitleden bir örneklem ve f , μ ortalamalı, σ^2 varyanslı normal dağılımın olasılık yoğunluk fonksiyonu olmak üzere

$$R(f'') = \sigma^{-5} \int \{\phi''(x)\}^2 dx$$

$$= \sigma^{-5} \frac{3}{8\sqrt{\pi}} = 0.212\sigma^{-5} \tag{2.31}$$

dir. Burada $\phi(x)$ standart normal dağılımın olasılık yoğunluk fonksiyonudur. Eşitlik (2.31)'de σ yerine

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}$$

tahmin edicisi yazılıp, eşitlik (2.30)'da yerine yazılırsa

$$\hat{h}_{opt1} = \left\{ \frac{R(K)}{\mu_2^2(K) R(f'') n} \right\}^{1/5}$$

$$= \left(\frac{4\hat{\sigma}^5}{3n} \right)^{1/5} \approx 1.06 \hat{\sigma} n^{-1/5} \qquad (2.32)$$

elde edilir. Silvermanın pratik (rule of thumb) $\hat{h}_{opt}$ tahmin edicisinin aykırı gözlemlere duyarlı olması uygulamada temel problemdir. Bir aykırı gözlem, σ parametresinin gerçeğinden çok daha büyük tahmin edilmesine yol açacağından $\hat{h}_{opt}$ ile seçilen h' ında gerçeğinden çok daha büyük tahmin edilmesine neden olur. Bu sebeple σ parametresinin daha sağlam(robust) bir tahmin edicisi olan kuantiller arası genişlik (Interquantil Range) kullanılması önerilir. Kuantiller arası genişlik tahmin edicisi

$$\hat{Q} = X_{\|0.75\times\|n:n} - X_{\|0.25\times\|n:n} \qquad (2.33)$$

eşitliğinden faydalanılarak elde edilebilir. Burada $X_{\|0.75\times\|n:n}$, %75 lik kuantil, $X_{\|0.25\times\|n:n}$, %25 lik kuantildir. $X_1, X_2, ..., X_n$' ler μ ortalamalı, σ^2 varyanslı Normal dağılıma sahip olduğundan

$$Z_i = \frac{X_i - \mu}{\sigma}, \qquad i = 1,2,...,n$$

şeklinde tanımlanan $i = 1,2,...,n$ rasgele değişkenleri Standart Normal dağılıma sahip olurlar ve

$$X_i = \mu + \sigma Z_i, \qquad i = 1,2,...,n$$

yazılabilir. Böylece

$$\begin{aligned}\hat{Q} &= X_{\|0.75\times\|n:n} - X_{\|0.25\times\|n:n} \\ &= \left(\mu + \sigma Z_{\|0.75\times\|n:n}\right) - \left(\mu + \sigma Z_{\|0.25\times\|n:n}\right) \\ &= \sigma\left(Z_{\|0.75\times\|n:n} - Z_{\|0.25\times\|n:n}\right) \\ &= 1.34\sigma\end{aligned}$$

elde edilir. Buradan σ' nın aykırı değerlere sağlam bir tahmin edicisi $\tilde{\sigma} = \dfrac{\hat{Q}}{1.34}$ olarak elde edilir. Buradan

$$h_{opt2} = 1.06\tilde{\sigma}n^{-1/5} = 1.06\frac{\hat{Q}}{1.34}n^{-1/5} = 0.79\hat{Q}n^{-1/5} \tag{2.34}$$

elde edilir. Bu işlemler sonucunda geliştirilen normal referans kuralı

$$h_n = \frac{1.06\hat{\sigma}}{n^{1/5}} \tag{2.35}$$

biçimindedir. Burada $\hat{\sigma} = \min\left\{S, \dfrac{\hat{Q}}{1.34}\right\}$ ile verilen eşitlikten hesaplanır. S, örneklem standart sapması, $\hat{Q}$: interquartile range($Q_{75} - Q_{25}$: %75 lik kuartil ile %25 lik kuartil arası fark) olarak tanımlanır. (Wasserman 2006)

2.2.6. Çoklu kernel yoğunluk fonksiyonu tahmini

Yoğunluk tahmini tek boyutlu fonksiyon için değil de çok boyutlu bir $\boldsymbol{X} = (X_1,\ldots,X_q)'$ vektörünün olasılık yoğunluk fonksiyonu için yapılıyorsa yoğunluk fonksiyonunun tahmini için çoklu kernel yoğunluk fonksiyonu kullanılır. Çoklu yoğunluk fonksiyonunda amaç $X_1,\ldots,X_q$ rasgele değişkenlerine ait $f(\boldsymbol{x}) = f(x_1,\ldots,x_q)$ yoğunluk fonksiyonunun tahminini bulmaktır. q tane değişkenin her birine ait n adet gözlem olduğu düşünülsün. Her bir değişkenin i. gözleminin alındığı $\boldsymbol{X}_i$ vektörü

$$\boldsymbol{X}_i = \left(X_{1i},\ldots,X_{qi}\right) \ , \qquad i = 1,2,\ldots,n \tag{2.36}$$

olur. q boyutlu bir rasgele vektör için yoğunluk fonksiyonu tahmini $K(\cdot)$ fonksiyonu, q boyutlu çoklu kernel fonksiyonu olmak üzere

$$\begin{aligned}\hat{f}_h(\boldsymbol{x}) &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^q}K\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^q}K\left(\frac{X_{i1}-x_1}{h},\ldots,\frac{X_{iq}-x_q}{h}\right)\end{aligned} \tag{2.37}$$

biçiminde tanımlanır (Hardle ve Müller 1997). Bu tahmin edicide her bir bağımsız değişken için h bant genişliğinin sabit olduğu durum gösterilmiştir. Her bir değişkenin farklı varyansa sahip olduğu düşünülürse bant genişliğinin hepsi için aynı olması tahminde sorun yaratacaktır. Uygun tahmin elde etmek için her bir değişkenin kendi yapısına uygun bant genişliği seçmek gerekir. Eğer her bir değişken için bant genişliği farklı olarak seçilirse tahmin edici

$$\hat{f}_h(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1 \times \cdots \times h_q} K\left(\frac{X_{i1}-x_1}{h_1},...,\frac{X_{iq}-x_q}{h_q}\right) \tag{2.38}$$

olarak yazılır. Tahmin edicinin kullanımındaki ilk önemli nokta çok boyutlu kernel fonksiyonu $K(\boldsymbol{x}) = K(x_1,\ldots,x_q)$ ' in nasıl alınacağıdır. Bu fonksiyon genellikle çarpımsal (*multiplicative*) kernel fonksiyonu olarak alınmaktadır (Wasserman 2006). $X_1,...,X_q$ değişkenlerinin bağımsız olması varsayımı altında çarpımsal kernel fonksiyonu

$$K(\boldsymbol{x}) = K(x_1)\times \cdots \times K(x_q)$$

olarak yazılan ifadenin kullanımı ile elde edilen tahmin edicidir. Daha basit olarak çarpım sembolü yardımıyla

$$\hat{f}_h(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\prod_{j=1}^{q} K\left(\frac{X_{ij}-x_j}{h_j}\right)\right\} \tag{2.39}$$

biçiminde verilir. Kullanılacak *h* bant genişliğinin seçimi için daha önce verilen normal referans kuralı çoklu durumlarda da uygulanmaktadır. Her bir değişkene ait bant genişliği belirlenirken

$$h_j = \left(\frac{4}{q+2}\right)^{1/(q+4)} n^{-1/(q+4)}\ \sigma_j \tag{2.40}$$

olarak belirlenir. Burada *q* değişken sayısı, *n* gözlem sayısını göstermektedir. σ_j standart sapmalarının yerine tahmin edicileri, $\hat{\sigma}_j = \min\left\{S_j, \frac{\hat{Q}}{1.34}\right\}$ ile verilen

eşitlikten hesaplanır. Burada S_j ; $j.$ değişkene ait örneklem standart sapması, $\hat{Q}$ daha öncede tanımlanan $j.$ değişken için %75 lik kuartil ile %25 lik kuartil arası fark olarak tanımlanır (Wasserman 2006).

2.3. Parametrik Olmayan Regresyonda Yoğunluk Tahmini

Regresyon analizinde en çok başvurulan yöntem parametrik regresyon modelidir. Regresyon analizinde amaç, bağımlı değişkenin bağımsız değişkenler tarafından ifade edilmesini sağlayacak bir model tahmin etmektir. Parametrik regresyonda, tahmin edilecek modelin yapısı önceden bilindiği için tahmin daha kolay olmaktadır. Ancak parametrik regresyonun kullanımı için bazı varsayımların sağlanması gerektiğinden daha önce de bahsedilmişti. Bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin yapısını bilinen parametrik modeller yardımıyla öngörülemiyorsa parametrik olmayan regresyon modeli kullanılabilmektedir. Parametrik olmayan regresyon modeli, verilerin ağırlıklandırılması ile tahmin yaptığı için bir ön bilgiye ihtiyaç duymamaktadır. Tek değişkenli parametrik olmayan regresyon modeli

$$Y_i = m(x_i) + \varepsilon_i \ , \qquad i = 1,2,...,N \tag{2.41}$$

biçiminde tanımlanır. Burada Y_i bağımlı değişken, x_i bağımsız değişken $E(\varepsilon_i) = 0$ ve $Var(\varepsilon_i) = \sigma^2$ olarak varsayılan bağımsız ve aynı dağılımlı hata terimleridir. $m(x_i)$, bilinmeyen regresyon fonksiyonudur ve $m(x_i)$ fonksiyonunun düzgün bir fonksiyon olması gerekmektedir. İlgilenilen $m(x_i)$ fonksiyonunun tahminini elde etmektir. $X = (x_1, x_2,..., x_N)$ bağımsız değişken ve Y bağımlı değişken olmak üzere, $m(\cdot)$ fonksiyonu

$$m(x) = E\left[Y \middle| X\right] = \frac{\int_{-\infty}^{\infty} y f(x,y)}{f(x)} \tag{2.42}$$

biçiminde yazılır (Pehlivan 2005) . Burada Y değişkeninin $X = x$ koşulu altında beklenen değeri elde edilmek istenmektedir. Eşitlikteki $f(x,y)$ fonksiyonu (X,Y) değişkenlerinin ortak olasılık yoğunluk fonksiyonudur. $f(x)$ ise X in marjinal olasılık yoğunluk fonksiyonudur. $m(x)$ in tahmininde yapılan eğri uydurma işlemine "*düzleştirme (smoothing)"* adı verilmektedir. Parametrik olmayan yöntem ile tahmin yapılacaksa daha önceden bahsedilen yoğunluk tahminine benzer bir ağırlıklandırma yöntemi kullanılacaktır (Hardle 1994).

Parametrik olmayan regresyon tahmin edicilerinin hesaplanmasın da kullanılan ağırlıklandırma işleminin temeli herhangi bir x noktasının h birim komşuluğunda bağımsız değişken x_i değerlerine karşılık gelen bağımlı değişken Y_i değerlerinin ağırlıklı ortalamasını hesaplamaktır. Böylece tahmin edici için x noktasına uzak olan x_i değerleri için küçük, x noktasına yakın olan x_i değerleri için büyük ağırlıklar verilmektedir (Fox 2002). Ağırlıklandırma fonksiyonu kullanarak $\hat{m}(x)$ fonksiyonunu

$$\hat{m}(x) = \sum_{i=1}^{N} W_h(x) Y_i \tag{2.43}$$

eşitliği ile elde edilir. Burada $W_h(.)$; bant genişliği h olan ve x_i' lere bağlı bir ağırlık fonksiyonudur. Yani x_i değişkenine bağlı bir fonksiyon yardımıyla Y_i bağımlı değişkeninin bir ağırlıklı ortalaması hesaplanmaktadır. Ağırlık fonksiyonu daha önce gösterilen yoğunluk tahmin edicilerinden biri olarak kullanılabilir ve buna bağlı olarak da h bant genişliği seçilir. Eşitlik (2.43) ayrıca,

$$\hat{m}(x) = \sum_{i=1}^{N} W_{hxi} Y_i \tag{2.44}$$

biçiminde de yazılabilir. Burada W_{hxi} fonksiyonunun x' e ve h' a bağlı bir ağırlık fonksiyonu olduğu ifade edilmektedir (Hardle 1994).

2.3.1. Parametrik olmayan basit regresyon modelinde tahmin

Parametrik olmayan regresyonda tahmin yapılması için ilk aşama (2.44) eşitliğinde verilen W_{hxi} ağırlık fonksiyonunu belirlemektir. Bunun için farklı ağırlık fonksiyonları önerilmişse de en yaygın olarak kullanılan yöntem kernel çekirdek fonksiyonlarının kullanımıdır. Seçilen kernel çekirdek fonksiyonu yardımıyla regresyon modelinin tahmin edilmesi için en yaygın olarak kullanılan yöntem Nadaraya-Watson kernel tahmin yöntemidir (Çağlayan 2002).

2.3.1.1. Nadaraya-Watson kernel tahmin edicisi

Kernel tahmin edicilerinin adı (2.44) eşitliğinde verilen, $\sum_{i=1}^{N} W_{hxi} Y_i$ şeklindeki parametrik olmayan regresyon tahmin edicisinde W_{hxi} ağırlıklarının kernel fonksiyonu kullanılarak elde edilmesinden dolayıdır. Nadaraya-Watson tahmin edicisinde, *i.* gözlemin *w* ağırlığı $x_i - x$ uzaklığının bir fonksiyonu olarak tanımlanır. Yani W_{hxi}, Y_i gözlemine atanan ağırlık olarak düşünülür. Bu durumda W_{hxi} ağırlıkları

$$W_{hxi} = \frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)} \qquad (2.45)$$

biçiminde ifade edilir ve görüleceği gibi $\sum_{i=1}^{N} W_{hxi} = 1$ olur. $i.$gözlem değeri için Y_i' ye atanan ağırlık bu şekilde hesaplandıktan sonra (2.44) eşitliğinde $\hat{m}(x)$' in tahmininde W_{hxi} ağırlığı yerine koyulduğunda

$$\hat{m}(x) = W_{hxi} Y_i = \sum_{i=1}^{N} \left[\frac{K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{N} K\left(\frac{x_i - x}{h}\right)} \right] Y_i \tag{2.46}$$

tahmin edicisi fonksiyonların bir toplamı olarak elde edilir(Pagan ve Ullah 1999). Daha düzgün bir ifade ile Nadaraya-Watson kernel tahmin edicisi

$$\hat{m}(x) = \frac{\sum_{i=1}^{N} K\left(\frac{x_i - x}{h}\right) Y_i}{\sum_{i=1}^{N} K\left(\frac{x_i - x}{h}\right)} \tag{2.47}$$

şeklinde tanımlanır. $\hat{m}(x)$' in tahmini elde edildikten sonra Y_i bağımlı değişkeninin tahmini olan $\hat{Y}_i$ ler, $\hat{m}(x)$ fonksiyonunda x_i değerlerinin yerine koyulması ile elde edilir. $\hat{m}(x)$ fonksiyonunun tahminini matris çarpımı ile ifade etmekte mümkündür. $\boldsymbol{1}_N$ ifadesi tüm elemanları 1 olan $N \times 1$ boyutlu vektör, $\boldsymbol{K}(x)$ köşegen elemanları $K\left(\frac{x_i - x}{h}\right)$ olan köşegen matris $\boldsymbol{Y}$ de bağımlı değişkenlere ait $N \times 1$ boyutlu vektör olmak üzere,

$$\hat{m}(x) = \left(\boldsymbol{1}_N' \boldsymbol{K}(x) \boldsymbol{1}_N\right)^{-1} \boldsymbol{1}_N' \boldsymbol{K}(x) \boldsymbol{Y} \tag{2.48}$$

eşitliği ile ifade edilebilir (Roy 1997). Kernel fonksiyonu ile ağırlıklandırma işlemini matematiksel bazı dönüşümler yardımıyla elde etmek mümkündür. Parametrik olmayan regresyon modeli $Y_i = m(x_i) + \varepsilon_i$ biçimindedir. $m(x_i)$ fonksiyonunun parametrik olmayan tahmini, $x_i - x$ gibi küçük bir aralıkta Y_i değerlerinin ağırlıklandırılması ile bulunur. $m(x_i)$ fonksiyonu x noktası komşuluğunda taylor serisine açılırsa

$$m(x_i) = m(x) + m'(x)(x_i - x) + \frac{1}{2!}m''(x)(x_i - x)^2 + \cdots \tag{2.49}$$

ifadesi elde edilir. Bu seride $m(x_i)$ fonksiyonunun x noktasındaki türevinden sonraki bileşenlerini O fonksiyonu yardımıyla gösterirsek $m(x_i) = m(x) + O(x_i - x)$, biçiminde olur ve $O(x_i - x)$ terimi özellikle gözlem sayısı fazla olduğu zaman hata terimlerinin içine dahil edilebilir. Sonuçta model $Y_i = m(x) + \varepsilon_i$ şeklinde yazılacaktır. $m(x)$, $m(x_i)$ fonksiyonunun x noktasındaki değeri olacaktır. Görüldüğü gibi parametrik olmayan model artık sabit bir fonksiyon biçiminde yazılmış ve tahmin kolaylaşmıştır. Modelin tahmini için, ağırlıklandırılmış en küçük kareler yöntemi uygulanır ve ağırlıklandırma işlemi için de $K(x)$ kernel yoğunluk fonksiyonu kullanılırsa, $\boldsymbol{\varepsilon}'\boldsymbol{K}(x)\boldsymbol{\varepsilon}$ ifadesinin minimum yapılması sonucunda (2.48) eşitliğindeki tahmin edicinin elde edileceği görülür (Mukherjee 2002).

2.3.2. Çoklu parametrik olmayan regresyon modeli

Regresyon uygulamalarında genellikle açıklayıcı değişken sayısının birden fazla olduğu durumlar ile karşılaşılmakta ve çoklu modeller için tahmin yöntemleri kullanılmaktadır. Parametrik olmayan regresyon modeli için de bağımsız değişken sayısının birden fazla olması durumunda çoklu parametrik olmayan modeller geliştirilmiş ve tahmin yöntemleri birden fazla bağımsız değişkeni kullanabilecek yapıya getirilmiştir. Eşitlik (2.41) ile verilen modelde bağımsız değişken sayısının birden fazla olduğunu düşünülürse çoklu parametrik olmayan regresyon modeli

$$Y_i = m(x_{1i}, x_{21},..., x_{qi}) + \varepsilon_i \ , \qquad i = 1,2,..., N \tag{2.50}$$

eşitliği ile verilir. Burada $m(\cdot)$ fonksiyonu q adet değişkeni barındıran bir fonksiyondur. Eşitlik (2.50) ile verilen model için tahmin yöntemleri geliştirilmiş ve değişken sayısının fazla olduğu durumlarda hesaplama zorlukları yaşanmıştır. Bu zorlukları yok edebilmek amacıyla çoklu parametrik olmayan regresyon modeli alternatif olarak toplamsal model biçimine de getirilebilmektedir. Toplamsal model olarak adlandırılan model

$$Y_i = \mu + m_1(x_{1i}) + m_2(x_{2i}) + \cdots + m_q(x_{qi}) + \varepsilon_i \ , \quad i = 1,2,..., N \tag{2.51}$$

biçiminde yazılmaktadır. μ anakütle sabit parametresidir. Toplamsal model toplam sembolü ile

$$Y_i = \mu + \sum_{k=1}^{q} m_k(x_{ki}) + \varepsilon_i \ , \qquad i = 1,2,..., N \tag{2.52}$$

biçiminde gösterilir. Eşitlik (2.51)'de verilen modelin tahmini için çok değişkenli Kernel yoğunluk fonksiyonu kullanılır. Basit parametrik olmayan regresyon tahmininde (2.47) eşitliği ile verilen tahmin edici, çoklu model için

$$\hat{m}(x_1,...x_q) = \frac{\sum_{i=1}^{N} K_1\left(\frac{x_{1i} - x_1}{h_1}\right) K_2\left(\frac{x_{2i} - x_2}{h_2}\right) \times \cdots \times K_q\left(\frac{x_{qi} - x_q}{h_q}\right) Y_i}{\sum_{i=1}^{N} K_1\left(\frac{x_{1i} - x_1}{h_1}\right) K_2\left(\frac{x_{2i} - x_2}{h_2}\right) \times \cdots \times K_q\left(\frac{x_{qi} - x_q}{h_q}\right)} \tag{2.53}$$

biçiminde yazılabilir. Sonuçta elde edilecek tahmin fonksiyonu $k=1,\ldots,q$' ya kadar x_k bağımsız değişkenlerine ve her bir bağımsız değişken için h_k bant genişliğine sahip bir fonksiyon olacaktır. Eşitlik (2.53) çarpım sembolü yardımıyla

$$\hat{m}(x_1,\ldots,x_q)=\frac{\sum_{i=1}^{N}\left(\prod_{k=1}^{q}K_k\left(\frac{x_{ki}-x_k}{h_k}\right)\right)Y_i}{\sum_{i=1}^{N}\left(\prod_{k=1}^{q}K_k\left(\frac{x_{ki}-x_k}{h_k}\right)\right)} \qquad (2.54)$$

biçiminde yazılabilir (Liu 1998). Görüldüğü gibi (2.50) eşitliğindeki modelin tahmini özellikle değişken sayısı arttıkça hesaplanması zor bir hal alacaktır. Buna karşın (2.51) veya (2.52) eşitliğinde verilen toplamsal model daha basit işlemlerle tahmin edilebilecek ancak toplamsal modelde bilgi kaybı daha fazla olacaktır(Çağlayan 2002). Toplamsal modelin tahmini için geriye uyum adında bir algoritma geliştirilmiştir. Eşitlik (2.52) ile verilen modelin anlamlı bir şekilde tahmin edilebilmesi için algoritma ilk olarak μ sabit parametresini yok etmeyi amaçlar ve ardından parametrik model yardımıyla bulunan en küçük kareler tahminlerini ilk tahmin yaklaşımı olarak alır. Bu algoritmanın yürütülmesi için uygulanacak işlemler sırasıyla aşağıdaki gibi verilir.

1. $\hat{\mu}=\bar{Y}$ olarak alınır.
2. $m_k(x_{ki})$, fonksiyonlarını ortalamadan sapmalar olarak düşünerek oluşturulan $Y_i-\bar{Y}=m_1^*(x_{1i}-\bar{x}_1)+\ldots+m_q^*(x_{qi}-\bar{x}_q)+\varepsilon_i$, $i=1,\ldots,N$ modeli parametrik gibi düşünülerek doğrusal regresyondaki klasik en küçük kareler tahmini yardımıyla $\hat{m}_k^*$ tahminleri hesaplanır.
3. Bulunan tahmin değerleri algoritmada kullanılacak ilk tahminler olarak düşünülür yani $\hat{m}_k^{(1)}$ olarak gösterilen ilk tahminler hesaplanan $\hat{m}_k^*$ tahminleri yardımıyla;

   $$\hat{m}_k^{(1)}(x_{ki})=\hat{m}_k^*(x_{ki}-\bar{x}_k), \quad k=1,\ldots,q$$

   olacak biçimde her bir değişken için elde edilir.

4. $m_1(x_{1i})$' in tahmini için, $Y_i - \overline{Y} = m_1^*(x_{1i} - \overline{x}_1) + ... + m_q^*(x_{qi} - \overline{x}_q) + \varepsilon_i$ eşitliğinde ε_i çekilirse;

$$\varepsilon_i = Y_i - \overline{Y} - m_1^*(x_{1i} - \overline{x}_1) + ... + m_q^*(x_{qi} - \overline{x}_q)$$

eşitliği elde edilir ve $m_1(x_{1i})$ tahmin edileceğinden $m_1^*(x_{1i} - \overline{x}_1)$ ifadesi sol taraf geçirilerek eşitlik,

$$\varepsilon_i + m_1^*(x_{1i} - \overline{x}_1) = Y_i - \overline{Y} - m_2^*(x_{2i} - \overline{x}_2) + ... + m_q^*(x_{qi} - \overline{x}_q)$$

olarak yazılır. Bu işlemi her bir $m_k(x_{ki})$ için yazarsak;

$$\begin{aligned}
e_{1i} &= \varepsilon_i + m_1^*(x_{1i} - \overline{x}_1) = Y_i - \overline{Y} - m_2^*(x_{2i} - \overline{x}_2) + ... + m_q^*(x_{qi} - \overline{x}_q) \\
e_{2i} &= \varepsilon_i + m_2^*(x_{2i} - \overline{x}_2) = Y_i - \overline{Y} - m_1^*(x_{1i} - \overline{x}_1) + ... + m_q^*(x_{qi} - \overline{x}_q) \\
&\vdots \\
e_{qi} &= \varepsilon_i + m_q^*(x_{qi} - \overline{x}_q) = Y_i - \overline{Y} - m_1^*(x_{1i} - \overline{x}_1) + ... + m_{q-1}^*(x_{(q-1)i} - \overline{x}_{q-1})
\end{aligned}$$

biçiminde q adet model elde edilir. yani hata terimlerine tahmin edilecek fonksiyona ait $\hat{m}_k^*(x_{ki} - \overline{x}_k)$ ilk tahmin değerleri eklenerek yeni e_{ki} değerleri elde edilmektedir.

5. $\hat{m}_k^{(2)}(x_{ki})$ tahminlerini elde etmek için, bulunan e_{ki} değerlerine karşılık x_{ki} değişkenlerinin yazılması ile elde edilecek model düzgünleştirme işlemi uygulanarak $\hat{e}_{ki}$, tahminleri bulunur. Bulunan tahmin değerleri $\hat{m}_k^{(2)}(x_{ki})$ tahmini olarak alınır. Model üzerinde gösterilecek olursa

$$\begin{aligned}
e_{1i} &= m(x_{1i}) + \dot{\varepsilon}_i \\
e_{2i} &= m(x_{2i}) + \dot{\varepsilon}_i \\
&\vdots \\
e_{ki} &= m(x_{ki}) + \dot{\varepsilon}_i
\end{aligned}$$

biçiminde e_{ki} değerlerinin bağımlı değişken olarak alındığı modellerden basit parametrik olmayan regresyon yardımıyla tahmininden elde edilecek $\hat{e}_{ki}$ tahminleri $\hat{m}_k^{(2)}(x_{ki})$ tahminleri olacaktır.

6. Son olarak elde edilen tahminler yardımıyla hesaplanacak $\dot{\varepsilon}_i$ hataları kullanılarak 4. adımda yapılan işleme geri dönülür ve aynı adımlar tekrarlanarak $\hat{m}_k^{(3)}(x_{ki})$ tahminleri elde edilir.

Görüldüğü gibi her bir adımda basit işlemler yapılarak sonuca ulaşılmasına rağmen toplam da çok fazla model kurulmakta ve çok fazla işlem yapılmaktadır. Ayrıca her bir fonksiyon ayrı ayrı tahmin edilmektedir. Bunun sonucunda hata artmaktadır. Bu yüzden (2.53) eşitliği ile verilen çoklu modelin tahminini kullanmak daha iyi sonuç verecektir. Toplamsal modelin avantajı ise basit işlemlerle sonuca ulaşılmasıdır (Çağlayan 2002).

Bu tahmin yöntemlerinin dışında lokal doğrusal kernel tahmin yöntemi adı altında başka bir tahmin yöntemi daha mevcuttur. Lokal doğrusal kernel tahmini panel verili parametrik olmayan model başlığı altında incelenecektir.

3. YARIPARAMETRİK REGRESYON

3.1. Yarıparametrik Regresyon Modelinin Tanımı

Parametrik regresyon analizinin temel varsayımlarından biri, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin fonksiyonel şeklinin biliniyor olmasıdır. Bu varsayım geçerli olmadığı zamanlarda parametrik regresyon analizi ile istenen tahminlere ulaşılamamaktadır. Bu durumlarda kullanılması önerilen yöntemler parametrik olmayan yöntemlerdir (Okur 2009). Parametrik olmayan yöntemler kullanıldığı zaman ise karşılaşılan problem ilk olarak uygulama zorluğudur. Tahmin yapıldıktan sonra karşılaşılan bir diğer sorun ise yorumlama problemidir. Parametrik olmayan regresyon sonucunda herhangi bir bağımsız değişken ile bağımlı değişken arasındaki ilişkinin biçimi kolayca görülememektedir. Bu zorlukların giderilmesi amacıyla son yıllarda yarıparametrik regresyon teknikleri geliştirilmiştir. Yarıparametrik regresyon modelleri yardımıyla tamamıyla parametrik olarak açıklanamayan modellerin en azından bir kısmını parametrik olarak kalan kısmını ise parametrik olmayan yöntemler yardımıyla tahmin edilmesi ve yorumlama imkanı getirilmesi amaçlanmaktadır(Çağlayan 2002). Bu amaçla oluşturulan, bir kısmı parametrik, bir kısmı ise parametrik olmayan regresyon modellerine yarıparametrik regresyon modelleri denilmektedir. Oluşturulan yarıparametrik model sayesinde parametrik kısmın tahmini kolaylaşacak ve tahmin sonucunda yorumlama kolaylığı sağlanacaktır. Parametrik olarak ilişkisi açıklanamayan değişkenler ise parametrik olmayan kısım yardımıyla modele dahil edilebilecek ve bilgi kaybı azaltılacaktır. Yarıparametrik modeli daha kolay gösterebilmek amacıyla ilk olarak iki kısmında da tek değişken bulunan modelin kurulması gösterilecek. İki bileşeni bulunan yarıparametrik modeli yazabilmek için öncelikle bileşenleri ayrı ayrı gösterilirse, tek değişkenli parametrik model,

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ , \qquad i = 1,2,...,N \tag{3.1}$$

şeklinde, parametrik olmayan model ise bağımsız değişkeni z_i olarak düşünerek,

$$Y_i = m(z_i) + \varepsilon_i, \qquad i = 1,2,...,N \tag{3.2}$$

şeklinde yazılır. Bu iki model birlikte ele alındığında tek değişkenli yarıparametrik model iki bileşenin toplamı biçiminde

$$Y_i = \beta_0 + \beta_1 x_i + m(z_i) + \varepsilon_i, \qquad i = 1,2,...,N \tag{3.3}$$

olarak gösterilir (Çağlayan 2002). Modelde β_0 anakütle sabit parametresini, β_1 parametrik bileşene ait bilinmeyen parametreyi, Y_i bağımlı değişkene ait gözlem değerlerini, x_i parametrik bileşene ilişkin bağımsız değişkenin gözlem değerlerini, z_i parametrik olmayan bileşene ilişkin bağımsız değişkenin gözlem değerlerini $m(.)$ fonksiyonu z_i' lere bağlı bilinmeyen formdaki fonksiyonu ve ε_i sıfır ortalamalı sabit σ^2 varyanslı hata terimlerini göstermektedir. Yarıparametrik regresyon modelinin farklı bir varsayımı parametrik kısma ait x_i bağımsız değişkeni ile parametrik olmayan kısma ait z_i bağımsız değişkenlerinin düzgün bir regresyon ilişkisine sahip olduğu varsayımıdır. Bu varsayımın gerekliliği tahmin konusunda görülecektir. Modelde parametrik kısım için konuşulursa bu kısma ait ε teriminin doğrusal regresyon modelinin varsayımlarına sahip olması gerektiğinden yarıparametrik modelin hata teriminin de klasik doğrusal regresyon modelinin varsayımlarına sahip olması gerekir (Çağlayan 2002). Modelde parametrik ve parametrik olmayan kısma ait bağımsız değişken sayısı birden fazla olduğu zaman model çoklu yarıparametrik model olarak adlandırılır. Genel gösterimi ile

$$Y_i = \beta_0 + \beta_1 x_{1i} + ... + \beta_q x_{qi} + m(z_{1i}, z_{2i}, ..., z_{Pi}) + \varepsilon_i, \quad i = 1,2,...,N \tag{3.4}$$

olarak ifade edilir. Modelde q adet parametrik kısma ait bağımsız değişken, P adet parametrik olmayan kısma ait bağımsız değişken bulunmaktadır. $\boldsymbol{X}$ matrisi

parametrik kısmın değişkenlerine ait $\boldsymbol{X} = \left[1\ x_{1i} \ldots x_{qi}\right]$ biçimindeki $N \times (q+1)$ boyutlu matris, $\boldsymbol{Y}$ vektörü bağımlı değişkenlere ait $N \times 1$ boyutlu vektör, $\boldsymbol{\beta} = \left[\beta_0\ \beta_1 \ldots \beta_q\right]'$ olacak şekilde $(q+1)\times 1$ boyutlu parametre vektörü, $m(\boldsymbol{z})$ içinde parametrik olmayan kısma ait $(z_{1i},\ldots,z_{Pi})$ bağımsız değişkenlerini bulunduran bilinmeyen formdaki çok değişkenli fonksiyon ve $\boldsymbol{\varepsilon}$ hata terimlerinin $N \times 1$ boyutlu vektörü olmak üzere (3.4) eşitliği ile verilen model

$$\boldsymbol{Y} = \boldsymbol{X\beta} + m(\boldsymbol{z}) + \boldsymbol{\varepsilon} \tag{3.5}$$

olarak matris gösterimi ile yazılır (Roy 1997). Parametrik olmayan regresyon modelinde bahsedildiği gibi $m(\boldsymbol{z})$ fonksiyonu tahmin edilirken karşılaşılan karmaşık işlemleri daha basite indirmek için çoklu parametrik olmayan modeli toplamsal model biçiminde ifade etmek mümkündür. Aynı işlem yarıparametrik model için de geçerli olup (3.4) eşitliğinde verilen çoklu yarıparametrik model toplamsal model olarak yazılabilir. Toplamsal yarıparametrik regresyon modelini

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_q x_{qi} + m_1(z_{1i}) + \cdots + m_P(z_{Pi}) + \varepsilon_i\,, \quad i = 1,2,\ldots,N \tag{3.6}$$

biçiminde ifade etmek mümkündür. Yazılan bu model sayesinde $m(.)$ bilinmeyen formdaki fonksiyonu P değişkenli bir fonksiyon yerine tek değişkenli P adet fonksiyonun toplamı biçiminde yazılmıştır. Eşitlik (3.6) toplam formülü yardımıyla

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \sum_{p=1}^{P} m_p(z_p) + \boldsymbol{\varepsilon} \tag{3.7}$$

biçiminde ifade edilebilir. İki ayrı kısımdan oluşan bu modelin parametrelerini tahmin etmek için, her iki kısmı da göz önüne alarak, çok sayıda farklı yöntemler kullanılmakta bu da parametre tahminini zorlaştırmaktadır**.** Bununla beraber, istatistiksel araştırmalarda kullanılan farklı veri türlerinin kendi yapılarına uygun modellerle incelenmesi gerektiğinden, modelin kompleks yapısı göz önüne alındığında

yarıparametrik regresyon son yıllarda çok sık kullanılan aynı zamanda üzerinde çok çalışılan bir yöntem halini almıştır.

3.2. Yarıparametrik Regresyonda Model Belirleme

Yarıparametrik regresyonda en önemli problemlerden biri regresyon modelinin belirlenmesidir. Çoklu regresyonda modelde birden fazla bağımsız değişkenin yer aldığı bilinmektedir. Çoklu yarıparametrik regresyon modeli;

$$Y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_q x_{qi} + m_1(z_{1i}) + \cdots + m_P(z_{Pi}) + \varepsilon_i, \quad i = 1,2...,N \quad (3.8)$$

biçiminde ifade edilir. Bu modelde q adet parametrik modele ait bağımsız değişken (x_i), P adet de parametrik olmayan modele ait bağımsız değişken (z_i) söz konusudur. Bir veri setinde elimizde bulunan bağımsız değişkenlerden hangilerinin parametrik, hangilerinin parametrik olmayan model kısmında yer alması gerektiğini belirleme sürecine yarıparametrik regresyonda model belirleme süreci denilmektedir.

Model belirleme sürecinde, çalışmanın konusuna göre daha önceden yapısı incelenmiş ve parametrik veya parametrik olmayan modelde yer alması gerektiği düşünülen değişkenler varsa model, çalışmacının belirlediği şekilde, herhangi bir test yapmadan kurulabilir. Modelde yer alacak değişkenler hakkında herhangi bir önbilgi bulunmuyor ise bağımlı değişken ile parametrik bir model ile ilişkilendirilebilecek bağımsız değişkenleri ve parametrik olmayan model ile ilişkilendirilebilecek bağımsız değişkenleri belirlemek için çeşitli yöntemler kullanılır. Bu yöntemlerden en uygunu parametrik regresyon modelinde parametrelerin anlamlılığının test edilmesidir. Bu yöntem için bütün değişkenler tek tek modele eklenerek modelin anlamlılığındaki değişme incelenir. Bu yöntemin uygulanma şekli algoritma olarak verilmiştir;

Veri setinde T adet bağımsız değişken olduğu durumda

1. Adım: Bağımlı değişken (y_i) ile her bir bağımsız değişkene (x_i) ait tekli parametrik regresyon modelleri kurulur ve parametre tahminleri yapılır.

2. Adım: Bağımlı değişken (y_i) ile bağımsız değişkenler(x_i) arasında sırası ile ikili, üçlü ,..., T 'li kurulabilecek bütün parametrik regresyon modelleri kurulur ve parametre tahminleri yapılır.

3. Adım: Elde edilen bütün parametrik regresyon modeller için bireysel parametrelerin anlamlılık testi(t testi) yapılır.

4. Adım: t testinin sonuçlarına göre, her bir modelde anlamlı çıkan bağımsız değişkenler modelin parametrik kısmına, genelde anlamsız çıkan değişkenlerde parametrik olmayan modele alınarak model belirlenir.

Bu seçim yapılırken genel bir bakış açısı ile yaklaşıldığından kesin bir seçim kriteri bulunmamaktadır. Modelde parametrelerin anlamlılığının yanında çoklu belirtme katsayısı olarak adlandırılan R^2 gibi kriterler de göz önüne alınabilir.

3.3. Yarıparametrik Regresyonda Tahmin

Yarıparametrik regresyon modelinde tahmin yapılırken modelde parametrik ve parametrik olmayan iki kısmın olduğu göz önüne alınarak kısımlar ayrı ayrı tahmin edilir. Ancak bu tahmin yapılırken kısımların birbirine etkileri göz önüne alınır. Bu amaçla önce parametrik kısım, parametrik olmayan kısma göre koşullu beklenen değer kullanımı tahmin edilmekte ardından elde edilen $\hat{\boldsymbol{\beta}}$ parametre tahminleri yerine konularak parametrik olmayan modeldeki bilinmeyen formdaki çok değişkenli fonksiyon olan $m(\boldsymbol{z})$ fonksiyonu tahmin edilir (Pagan ve Ullah 1999).

3.3.1. Yarıparametrik regresyonda parametrik kısmın tahmini

Tahmin probleminde yapılacak işlemlerin daha kolay anlaşılması için her bir kısmı tek değişkenli yarıparametrik regresyon modeli üzerinde gösterilirse model;

$$Y_i = x_i\beta + m(z_i) + \varepsilon_i \ , \qquad i = 1,2,...,N \tag{3.9}$$

şeklinde yazılır. Bu modelin parametrik olmayan kısımdaki değişken z_i' ye göre koşullu beklenen değeri alınır. ε_i hatalarının beklenen değerinin sıfır olduğu varsayımı altında model

$$E(Y_i|Z_i) = E(X_i|Z_i)\beta + m(z_i) \tag{3.10}$$

şeklini alır. Bilinmeyen formdaki $m(z_i)$ fonksiyonunu yok edebilmek için (3.9) eşitliği ile (3.10) eşitliğinin farkı alınırsa

$$\begin{aligned} Y_i - E(Y_i|Z_i) &= (x_i\beta + m(z_i) + \varepsilon_i) - (E(X_i|Z_i)\beta + m(z_i)) \\ &= x_i\beta + \varepsilon_i - E(X_i|Z_i)\beta \\ &= (x_i - E(X_i|Z_i))\beta + \varepsilon_i \end{aligned} \tag{3.11}$$

biçiminde olacaktır. Burada $E(X_i|Z_i)$ ve $E(Y_i|Z_i)$ yerine parametrik olmayan regresyondan elde edilen kernel tahmin edicileri kullanılacaktır. Yarıparametrik regresyonun varsayımı olan x_i değişkeni ile z_i değişkeni arasında düzgün bir regresyon ilişkisi olması durumunun gerekliliği bu aşamada $E(X_i|Z_i)$ beklenen değerinin hesaplanması açısından ortaya konulmuştur. $E(X_i|Z_i)$ ve $E(Y_i|Z_i)$ değerlerinin hesaplanması parametrik olmayan regresyon modeli üzerinde gösterilirse, $X_i = m(z_i) + \varepsilon_i$ modelinden elde edilecek $\hat{X}_i$ değerleri $E(X_i|Z_i)$ yerine, $Y_i = m(z_i) + \varepsilon_i$ modelinden elde edilecek $\hat{Y}_i$' ler $E(Y_i|Z_i)$ yerine kullanılacaktır. Eşitlik (3.11) de $Y_i^* = Y_i - E(Y_i|Z_i)$ ve $x_i^* = x_i - E(X_i|Z_i)$ olacak biçimde alınırsa dönüştürülmüş model

$$Y_i^* = x_i^* \beta + \varepsilon_i , \qquad i = 1,2,...,N \tag{3.12}$$

olarak elde edilir. Parametrik kısımdaki x_i bağımsız değişkenine ait β parametresinin tahmini (3.12) eşitliğine en küçük kareler tahmin yöntemi uygulanması ile $\boldsymbol{X}^*$ vektörü x_i^* değerlerine ait $N \times 1$ boyutlu vektör, $\boldsymbol{Y}^*$ ise Y_i^* değerlerine ait $N \times 1$ boyutlu vektör olmak üzere

$$\widehat{\boldsymbol{\beta}} = \left(\left(\boldsymbol{X}^* \right)' \boldsymbol{X}^* \right)^{-1} \left(\boldsymbol{X}^* \right)' \boldsymbol{Y}^* \tag{3.13}$$

eşitliği ile elde edilir (Pagan ve Ullah 1999). Modelde parametrik ve parametrik olmayan kısma ait birer değişken olduğu için $\hat{\boldsymbol{\beta}}$ tahmininin hesaplanması kolay anlaşılır biçimde görülmektedir. Modelde parametrik kısma ait birden fazla x_i bağımsız değişkeni olduğunda $\hat{\boldsymbol{\beta}}$ tahmininin hesaplanması aynı dönüşüm yardımıyla yine basit olarak görülebilir. Yarıparametrik modelin

$$Y_i = \beta_1 x_{1i} + \cdots + \beta_q x_{qi} + m(z_i) + \varepsilon_i , \qquad i = 1,2,...,N \tag{3.14}$$

eşitliği ile verildiği göz önüne alınsın. (3.10) ve (3.11) eşitliklerinde verilen dönüşümler (3.14) eşitliğindeki modele uygulanırsa

$$Y_i - E(Y_i | Z_i) = (x_{1i} - E(X_{1i} | Z_i))\beta_1 + \cdots + (x_{qi} - E(X_{qi} | Z_i))\beta_q + \varepsilon_i \tag{3.15}$$

eşitliği elde edilir. $E(X_{ki} | Z_i)$ ve $E(Y_i | Z_i)$ değerleri parametrik olmayan tahmin değerleri olmak üzere $\hat{\beta}_k$ tahminleri en küçük kareler tahmin yöntemi yardımıyla

(3.13) eşitliğinde verildiği gibi hesaplanır. Tek fark $\boldsymbol{X}^*$ matrisi elemanları $x_{ki} - E(X_{ki}|Z_i)$' ler olan $N \times q$ boyutlu tasarım matrisi olacaktır.

Modeldeki parametrik kısma ait değişken sayısının birden fazla olabileceği gibi parametrik olmayan kısma ait değişken sayısının da birden fazla olacağı düşünülürse modele ait $\hat{\beta}_k$ tahminlerini yapmak için aynı dönüşümler tekrar edilir. Yalnız bu durumda koşullu beklenen değer hesaplamasında tek değişkenin bilindiği durumda değil birden fazla değişkenin bilindiği durumda hesaplanan koşullu beklenen değerle karşılaşılacaktır. Bunun tahmini için de çoklu parametrik olmayan model kullanılacaktır. Bahsedilen yarıparametrik model,

$$Y_i = \beta_1 x_{1i} + \cdots + \beta_q x_{qi} + m(z_{1i}, z_{2i}, \ldots, z_{Pi}) + \varepsilon_i, \quad i = 1,2,..., N \tag{3.16}$$

eşitliği ile verilsin. Bu modelin parametrik olmayan kısımdaki değişkenler z_{pi}' lere göre koşullu beklenen değer alınırsa

$$\begin{aligned} E(Y_i|Z_{1i},..., Z_{Pi}) = {} & E(X_{1i}|Z_{1i},..., Z_{Pi})\beta_1 + ... \\ & + E(X_{qi}|Z_{1i},..., Z_{Pi})\beta_q + m(z_{1i},..., z_{Pi}) \end{aligned} \tag{3.17}$$

elde edilir. Eşitlik (3.16) ile (3.17) nin farkı alınarak gerekli düzenlemeler yapılırsa

$$\begin{aligned} Y_i - E(Y_i|Z_{1i},..., Z_{Pi}) = {} & (x_{1i} - E(X_{1i}|Z_{1i},..., Z_{Pi}))\beta_1 + \quad ... \\ & + (x_{qi} - E(X_{qi}|Z_{1i},..., Z_{Pi}))\beta_q + \varepsilon_i \end{aligned} \tag{3.18}$$

modeli elde edilir. Modelde verilen $E(Y_i|Z_{1i},..., Z_{Pi})$ ve $E(X_{1i}|Z_{1i},..., Z_{Pi})$ koşullu beklenen değerleri yerine yazılacak değerler, $Y_i = m(z_{1i}, z_{21},..., x_{Pi}) + \varepsilon_i$ ve $x_i = m(z_{1i}, z_{21},..., x_{Pi}) + \varepsilon_i$ çoklu parametrik olmayan regresyon modelinin tahmin

değerleri olacaktır. Çoklu parametrik olmayan regresyon modelinin tahmin edicisi (2.42) eşitliğinde verildiği gibidir (Müler 2000). Tahmin değerleri elde edildikten sonra yerlerine yazılarak elde edilecek parametrik model en küçük kareler tahmin yöntemi ile çözülür ve yarıparametrik modelin parametrik kısmına ait $\hat{\beta}_k$ $(k = 1,..., q)$ tahminleri elde edilir.

3.3.2. Yarıparametrik regresyonda parametrik olmayan kısmın tahmini

Yarıparametrik regresyonda parametrik kısma ait $\hat{\beta}_k$ tahminleri elde edildikten sonra parametrik olmayan kısma ait $m(z_{1i}, z_{2i},..., z_{Pi})$ bilinmeyen formdaki fonksiyonun tahmin işlemi kolaylıkla elde edilebilir. (3.17) ile verilen eşitlikte $m(z_{1i}, z_{2i},..., z_{Pi})$ fonksiyonunu çekersek

$$m(z_{1i},..., z_{Pi}) = E(Y_i | Z_{1i},..., Z_{Pi}) - \big(E(X_{1i} | Z_{1i},..., Z_{Pi})\beta_1 + ... + E(x_{qi} | Z_{1i},..., Z_{Pi})\beta_q\big) \quad (3.19)$$

elde edilir. Eşitlikte koşullu beklenen değerler parametre tahminleri yapılırken hesaplanmıştı. β_k ' lar yerine de bulunan parametre tahminleri $\hat{\beta}_k$ ' ları yazarsak $\hat{m}(z_{1i}, z_{2i},..., z_{Pi})$ tahmini elde edilir. $\hat{m}(z_{1i}, z_{2i},..., z_{Pi})$ tahminini elde etmenin bir başka yolu ise (3.16) modelinde parametrik kısmı eşitliğin sol tarafına geçirerek $\hat{\beta}_k$ tahmin değerlerinin yerine yazılmasıyla elde edilecek parametrik olmayan modelin tahminini bulmaktır. Yani elde edilecek parametrik olmayan model

$$Y_i - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_q x_{qi} = m(z_{1i}, z_{2i},..., z_{Pi}) + \varepsilon_i, \qquad i = 1,2,..., N \quad (3.20)$$

biçiminde olup $Y_i^* = Y_i - \hat{\beta}_1 x_{1i} - \cdots - \hat{\beta}_q x_{qi}$ yazılırsa tahmin edilecek çoklu parametrik olmayan model

$$Y_i^* = m(z_{1i}, z_{2i}, ..., z_{Pi}) + \varepsilon_i \quad , \qquad i = 1,2,..., N \tag{3.21}$$

olacak biçimde gösterilebilir. Ve $\hat{m}(z_{1i}, z_{2i}, ..., z_{Pi})$ tahmini parametrik olmayan modelden hesaplanır (Pagan ve Ullah 1999). Bu modelin tahmini çoklu parametrik olmayan regresyonda tahmin konusunda incelenmişti. İstenilirse model

$$Y_i^* = m_1(z_{1i}) + m_2(z_{2i}) + \cdots + m_P(z_{Pi}) + \varepsilon_i \quad , \qquad i = 1,2,..., N \tag{3.22}$$

olarak toplamsal model biçiminde de yazılıp $\hat{m}_p(z_{pi})$, $p = 1,..., P$ tahminleri, toplamsal modelin tahmini konusunda gösterilen algoritma yardımıyla elde edilir.

4. PANEL VERİ ANALİZİ

Ekonometrik veriler genelde iki biçimde derlenmektedir. Üzerine geniş çalışmalar yapılan veri tiplerinden ilki zaman serileri, diğeri ise yatay kesit verileridir. Zaman serisi analizi temelde bir zaman aralığındaki değerleri ele alır ve zamana göre modelleme yapmayı amaçlar. Ekonometrik verilerin yine sık kullanılan bir diğer biçimi yatay kesit verileridir. Yatay kesit verilerinde ise tek bir zaman noktasında farklı değişkenlerin değerleri ele alınır ve değişkenlerin zamandan bağımsız modelleme yapılır. t_1 ile t_2 zamanları arasında X ve Y değişkenleri arasındaki ilişkinin bir tek birey için incelenmesinde zaman serisi analizi, t zamanında N adet birey için X ve Y değişkenleri arasındaki ilişki incelendiğinde yatay kesit analizi kullanılır. Zaman serisi ve yatay kesit verileri ile ayrı ayrı analizler yapılabilmektedir (Sayan 2000).

Ancak her iki durumunda birlikte incelenme ihtiyacı duyulduğunda zaman boyutuna sahip kesit serilerini kullanarak ekonomik ilişkilerin tahmin edilmesi yöntemine panel veri analizi adı verilmektedir. Bu analizde zaman serileri ile kesit serileri bir araya getirilerek, hem zaman hem de kesit boyutuna sahip veri seti oluşturulmaktadır. Zaman serileri ve yatay kesit verilerinin bir arada kullanılması ile panel veri seti oluşmaktadır. Daha açık ifade edilecek olursa t_1 ile t_2 zamanları arasında N adet birey için X ve Y değişkenlerinin arasındaki ilişki inceleniyorsa veri seti Panel Veri seti olacaktır. Bu veri setinin analizinde panel veri için geliştirilen yöntemler kullanılacaktır.

Panel veri kullanımının diğer veri setlerine göre bazı üstünlükleri vardır. Bunların başında gözlem hacmi ile ilgili avantajlar yer almaktadır. Panel veri yöntemi kesit ve zaman serisi gözlemlerini birleştirdiğinden gözlem sayısı daha fazladır. Örneklemde gözlem sayısı arttıkça tahminler daha iyi sonuçlar verir (Baltagi 1995). Buna rağmen gözlemin az yapılabildiği durumlarda da analiz yapılmasına imkan verir. Panel veri, zaman içinde farklı bireyler ile ilgili değerler aldığından bu bireylerin arasındaki heterojen yapıdan etkilenmeyecek biçimde bireylere özgü bazı değişkenleri kullanabilir. Bunun sayesinde bireylerin etkilerini ayrı ayrı değerlendirme imkanı verir.

Bir diğer avantajı ise panel veri setinin değişkenler arasında daha az çoklu doğrusal bağlantı sorunu oluşturmasıdır. Bunun sebebi modele eklenen bilinmeyen formdaki fonksiyon sayesinde değişkenlerin hepsini lineer yapıda gösterme zorunluluğunun kalkmasıdır. Ayrıca panel veri modelleri tamamıyla kesit ya da zaman serilerinden daha karmaşık modellerin kurulmasına ve test edilmesine olanak tanır. Panel verinin en önemli avantajı değişkenler arasındaki ilişkiyi hem zaman hem de bireyler açısından ele alabilmesidir (Gujarati 2003).

Bu avantajlar göz önüne alındığında panel verilerin özellikle ekonomi ve sağlık alanında çok yaygın uygulama alanı bulabileceği görülmektedir. Ekonomi alanında yıllara göre farklı ülkelerin iktisadi göstergeleri veya sağlık alanında aynı türden farklı ilaçların zaman içerisinde vücutta meydana getirdiği değişimler incelendiğinde kullanılacak veri seti panel veri yapısında olmaktadır. Panel veri setine ait tablo genel olarak aşağıdaki biçimde gösterilir.

Çizelge 1. Panel veri tablosu

i	t	Y_{it}	k		
			1	...	q
1 ⋮ 1	1 ⋮ T	Y_{11} ⋮ Y_{1T}	x_{111} ⋮ x_{11T}	...	x_{q11} ⋮ x_{q1T}
⋮	⋮	⋮	⋮	...	⋮
N ⋮ N	1 ⋮ T	Y_{N1} ⋮ Y_{NT}	x_{1N1} ⋮ x_{1NT}	...	x_{qN1} ⋮ x_{qNT}

Tabloda; $i = 1,\ldots,N$ yatay kesit indisidir yani i. bireye ait gözlem değerlerini gösterir, $t = 1,...,T$ zaman indisidir ve t. zamandaki gözlem değerini verir. $k = 1,...,q$ bağımsız değişkenlerin indisidir ve q adet bağımsız değişkenin varlığını belirtir. Y_{it} bağımlı değişken, x_{kit} bağımsız değişkenlerdir.

Genel olarak panel veri analizleri için kurulan modeller iki farklı grup altında incelenir. Verilerde meydana gelen farklılıklar tahmin edilecek modelin katsayılarında değişime yol açacak kadar önemli değilse katsayıları sabit modeller adı altında incelenir. Ancak birimlere veya zamana göre meydana gelebilecek değişimler katsayıları etkileyecek boyutta ise panel veriler değişken katsayılı modeller adı altında incelenmektedir. Basit parametrik regresyon modeli üzerinde panel veri ile kurulabilecek farklı model tiplerini vardır. Panel veri modeli en genel olarak sabit parametrelerin ve eğim parametrelerinin hem zamana hem de bireylere göre farklılıklardan etkilendiği düşünülerek değişken katsayılı modelin en genel hali

$$Y_{it} = \alpha_{it} + x_{it}\beta_{it} + u_{it}, \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.1}$$

biçiminde kurulur. Burada y_{it} bağımlı değişkene ait *i*. bireyin *t*. zamanda aldığı değeri, x_{it} bağımsız değişkene ait *i*. bireyin *t*. zamanda aldığı değeri, α_{it} her bir birimin her bir zamanına göre farklı olduğu düşünülen sabit parametreleri, β_{it} her birim ve zamana göre farklı olan eğim parametresini ve u_{it} hata terimlerini gösterir. Görüldüğü gibi bütün bireyler ve bireylerin her zamanı için farklı sabit parametreler ve eğim parametreleri modelde yer almaktadır. Böyle bir modelin tahmini, parametre sayısının gözlem sayısından fazla olması nedeniyle mümkün değildir. Ancak sabitlenen parametrelerin ne anlama geldiğinin görülmesi açısından genel model yazılmıştır. Panel veri analizinde zamana bağlı farklılıklardan kaynaklanan değişim göz ardı edilerek yalnızca bireylerin farklılıklarından kaynaklanan değişim incelenmek istenirse panel verili model ,

$$Y_{it} = \alpha_{i} + x_{it}\beta_{i} + u_{it}, \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.2}$$

olarak yazılır. α_i yalnızca bireylerin farklılıklardan kaynaklanan sabit parametreyi, β_i ise sadece bireylere göre farklılık gösteren eğim parametresidir. Eşitlik (4.2) ile verilen modelde sabit parametrelerin ve eğim parametrelerinin bireylere göre meydana gelen farklılıklardan da etkilenmediği düşünülürse model

$$Y_{it} = \alpha + x_{it}\beta + u_{it}, \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.3}$$

olarak yazılabilir. Bu modele kısıtlı model veya katsayıları sabit model adı verilir. Panel veri analizi için yazılabilecek en basit model bu şekilde verilen klasik parametrik regresyon modelidir. Katsayıları sabit modelde hata terimleri ile ilgili farklı varsayımlara göre farklı modeller düşünülebilir ancak modeldeki katsayıların herhangi bir değişimden hiçbir zaman etkilenmediği düşünülmektedir. Bu model bütün gözlemleri birlikte kullanarak hiçbir etkiyi düşünmeden tahmin yaptığından verilerin homojen yapıda olduğunu varsayar (Sayyan 2000). Bu homojenliğin test edilmesi için u_{it} hatalarının bağımsız ve aynı dağılımlı $\left(u_{it} \sim N(0,\sigma_u^2)\right)$ olduğu varsayımı altında

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_N, \;\; \beta_1 = \beta_2 = \cdots = \beta_N \tag{4.4}$$
$$H_1 : \alpha_1 \neq \alpha_2 \neq \cdots \neq \alpha_N \;\; \text{ve/veya} \;\; \beta_1 = \beta_2 = \cdots = \beta_N$$

hipotezinin α anlam düzeyinde test edilmesi gereklidir. Hipotezin test edilmesi için F test istatistiği

$$F = \frac{(HKT_2 - HKT_1)(Nq + N - q - 1)^{-1}}{HKT_1(NT - Nq + N)^{-1}} \tag{4.5}$$

biçiminde olup, F test istatistiği $(Nq + N - q - 1)$ ve $(NT - Nq + N)$ serbestlik dereceli F dağılımına sahiptir. Eşitlik (4.5) ile hesaplanan test istatistiği, $(Nq + N - q - 1)$ ve $(NT - Nq + N)$ serbestlik dereceli F tablo değerinden büyükse H_0 reddedilir. Test istatistiğinin hesaplanmasında kullanılan HKT_1 değeri (4.2) eşitliğindeki modelden elde edilen hata kareler toplamı, HKT_2 değeri ise (4.3) eşitliğinde verilen modelden elde edilen hata kareler toplamıdır. Test istatistiğinin hesaplanmasında verilen q değeri ise modelde yer alan bağımsız değişken sayısıdır. Modeller yazılırken gösterim kolaylığı açısından tek değişkenli model biçiminde yazılmıştır, ancak test istatistiği değişken sayısının birden fazla olduğu durumda da geçerlidir. Test sonucunda H_0 hipotezi reddedilemez ise (4.3) ile verilen katsayıları

sabit model kullanılabilir. Ancak H_0 hipotezi reddedilir ise heterojen yapının sabit parametrelerden mi yoksa eğim parametrelerinden mi kaynaklandığını test etmek gerekir. Bu amaçla bireylere ait farklılıklardan kaynaklanan değişimin α_i sabit parametresi ile gösterildiği ve eğim parametrelerinin bu değişimden etkilenmediği düşünülerek $\beta_1 = \beta_2 = ... = \beta_N = \beta$ olarak kabul edilir ve değişken katsayılı model olan

$$Y_{it} = \alpha_i + x_{it}\beta + u_{it}, \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.6}$$

modeli kurulur. β_i parametrelerinin farklı olmadığı kabul edilen bu modelin kullanılıp kullanılmayacağını test etmek için α_i sabit parametrelerinin eşitliğinin test edilmesi gerekmektedir. Bu amaçla

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_N$$
$$H_1 : \alpha_1 \neq \alpha_2 \neq \cdots \neq \alpha_N$$

hipotezi kurulur ve hipotezin test edilmesi için kullanılacak test istatistiği

$$F = \frac{(HKT_2 - HKT_3)(N-1)^{-1}}{HKT_3(NT - N + q)^{-1}} \tag{4.7}$$

eşitliği ile verilir. Burada verilen HKT_3 değeri (4.6) modeli yardımıyla elde edilen hata kareler toplamıdır. Eşitlik (4.7) ile hesaplanan test istatistiği $(N-1)$ ve $(NT - N + q)$ serbestlik dereceli F tablo değerinden büyükse H_0 red edilir. Test sonucunda H_0 hipotezi reddedilirse (4.6) eşitliğinde verilen değişken katsayılı kısıtsız modelin kullanımı uygun olacaktır. Hipotez testleri sonucunda değişken katsayılı bir modelin seçildiği varsayılırsa karşılaşılacak iki durum vardır. Bu durumlar; bireylerin farklılıklarından kaynaklanan değişimi gösteren α_i sabit parametrelerinin modele dahil edilme yöntemine göre belirlenmektedir. Değişimin modele eklenme yöntemine göre, "sabit etki" ve "rasgele etki" olmak üzere iki tür

model ortaya çıkarmaktadır. Bireylere göre meydana gelen farklılıklardan kaynaklanan etkilerin sabit olduğu düşünülerek modele sabit bir terim ile ilave ediliyorsa modelin adı sabit etkili panel veri modeli, etkilerin rasgele bir değişken olduğu varsayılan durumda ise modele rasgele etkili panel veri modeli adı verilmektedir (Roy 1997).

Elde edilen iki tür model içinde literatürde yer alan çeşitli regresyon yöntemleri kullanılmaktadır. En yaygın regresyon yöntemi olan parametrik lineer regresyon yöntemi panel verili modeller için de sıklıkla kullanılmaktadır(Balestra ve Sevestre 1996). Ancak son yıllarda parametrik regresyondan başka parametrik olmayan ve yarıparametrik modellerde panel veri analizleri için kullanılmaya başlanmıştır. Bu bölümde bahsedilen bu regresyon yöntemlerinin panel veri analizi için kullanımı gösterilecek ve klasik modelleri panel veri analizine uygun olarak biçimlendirip tahmin yöntemlerinin belirlenmesi gösterilecektir. Daha öncede belirttiğimiz gibi bu çalışmada (1.6) ile verilen değişken katsayılı model üzerinde durulacaktır. Yani α_i yalnızca bireylerin farklılıklardan etkilenen sabit parametreyi gösterdiği durumlar için analizler yapılacaktır.

4.1. Parametrik Regresyonda Panel Verili Modeller

Panel verili modeller için en çok kullanılan regresyon yöntemi parametrik regresyondur. Bunun sebebi parametrik regresyonun kullanım alanının geniş olması ve daha basit işlemler ile sonuca ulaşılmasıdır. Parametrik regresyonda panel veri kullanımı ile oluşacak parametrik model

$$Y_{it} = \alpha_i + x_{kit}\beta_k + u_{it}, \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \; ; \; k = 1,2,...,q \qquad (4.8)$$

biçimindedir (Mukherjee 2002). Burada N birey sayısı, T her bir birey için gözlenen zamanı göstermektedir. Y_{it} değişkeni, *i.* bireyin *t*. zamanda aldığı değeri gösteren bağımlı bir değişken olduğundan, kesit boyutu için i, zaman için t olmak üzere iki alt indisle ifade edilmektedir. x_{kit} rasgele değişkeninde ise bağımlı değişkene ilave olarak birden fazla bağımsız değişken olabileceği için değişken

sayısını gösteren k indisi vardır. Parametrik regresyonda panel veri kullanımı α_i sabit parametresinin sabit etkili ya da rasgele etkili olması durumuna göre iki farklı biçimde yapılmaktadır. Her iki durumda da tahmin yöntemleri bu bölümde incelenecektir.

4.1.1. Parametrik sabit etkili panel veri modeli

Panel verili regresyon modellerinde birimler arasında meydana gelen farklılıklardan kaynaklanan değişmeyi ele almak için bu değişimin regresyon modelinin katsayılarından bazılarında veya tümü üzerinde değişmeye yol açtığı varsayılır (Baltagi 1995). Bu durumda Panel veri için kullanılacak regresyon modeli,

$$Y_{it} = \alpha_i + x_{kit}\beta_k + u_{it}\ , \qquad i = 1,2,\ldots,N\ ;\ t = 1,2,\ldots,T\ ;\ k = 1,2,...,q \tag{4.9}$$

eşitliğindeki gibi verilsin. Burada $i = 1,...,N$ bireylerin indisini, $t = 1,...,T$ zamanın indisini, k da bağımsız değişken indisini göstermek üzere; Y_{it} : bağımlı değişkenleri, x_{kit} : bağımsız değişkenleri, α_i : birimler arası veya zaman içinde meydana gelen farklılıklardan kaynaklanan değişmeyi, β_k : $q \times 1$ boyutlu $\boldsymbol{\beta}$ parametre vektörünün $k.$ bileşenini, u_{it} ise bağımsız ve aynı dağılımlı olduğu varsayılan hata terimlerini göstermektedir. u_{it} hata teriminin beklenen değeri sıfır, varyansı ise σ_u^2 olarak varsayılmaktadır. Modelde yer alan ve sabit parametre olarak varsayılan α_i terimi sebebi ile bu modele parametrik sabit etkili model adı verilmektedir. Sabit etki katsayısının her bir bireyin t anındaki aldığı değerler için sabit olduğu bu modele aynı zamanda Kovaryans modeli adı verilir.

Parametrik sabit etkili modeli $i.$ birey için matris gösterimi ile yazılırsa

$$\boldsymbol{Y}_i = \alpha_i \boldsymbol{1}_T + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{u}_i\,, \qquad i = 1,2,...,N \tag{4.10}$$

şeklinde olur. Burada $\boldsymbol{Y_i}$; $T \times 1$ boyutlu bağımlı değişkene ait vektör, $\boldsymbol{X_i}$; $T \times q$ boyutlu bağımsız değişkenler matrisi, $\boldsymbol{\beta}$: $q \times 1$ boyutlu parametre vektörünü, $\boldsymbol{1}_T$; $T \times 1$

boyutlu bir vektörü ve $\boldsymbol{u_i}$; hata terimine ait $T \times 1$ boyutlu vektördür. Hatalar için, $E(u_i) = 0$, $V(u_i) = \sigma_u^2 I_T$ olarak varsayılmaktadır. (4.10) eşitliğinde verilen model hem i. birey hem de t. zaman için matris gösterimi ile yazılırsa

$$\boldsymbol{Y} = \boldsymbol{D\alpha} + \boldsymbol{X\beta} + \boldsymbol{u} \tag{4.11}$$

elde edilir. Eşitlik (4.11) de $\boldsymbol{Y}$; $NT \times 1$ boyutlu bağımlı değişken vektörü, $\boldsymbol{X}$; $NT \times q$ boyutlu bağımsız değişkenler matrisi, $\otimes$ sembolü kroneker çarpımı ifade etmek üzere, $\boldsymbol{D} = \boldsymbol{I}_N \otimes \boldsymbol{1}_T$ biçiminde tanımlanan katsayı matrisi, $\boldsymbol{\alpha}$; $N \times 1$ boyutlu sabit parametreler vektörü, $\boldsymbol{\beta}$; $q \times 1$ boyutlu parametre vektörü ve $\boldsymbol{u}$; $NT \times 1$ boyutlu hata terimleri vektörüdür. Matris gösterimi daha açık olarak yazılacak olursa,

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & \\ 0 & 1 & \vdots \\ \vdots & \vdots & 0 \\ & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \\ \vdots \\ \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} x_{111} & x_{211} & \cdots & x_{q11} \\ x_{112} & x_{212} & \cdots & x_{q12} \\ \vdots & \vdots & \vdots & \vdots \\ x_{11T} & x_{21T} & \cdots & x_{q1T} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1N1} & x_{2N1} & \vdots & x_{qN1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1NT} & x_{2NT} & \cdots & x_{qNT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \\ \vdots \\ \\ \vdots \\ \beta_q \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T} \\ \vdots \\ u_{N1} \\ \vdots \\ u_{NT} \end{bmatrix}$$

biçimindedir. Burada $\boldsymbol{Z} = [\boldsymbol{D}\ \boldsymbol{X}]$ ve $\boldsymbol{\delta} = [\boldsymbol{\alpha}\ \boldsymbol{\beta}]'$ biçimindeki matrisler olmak üzere model,

$$\boldsymbol{Y} = \boldsymbol{Z\delta} + \boldsymbol{u} \tag{4.12}$$

şeklinde yeniden ifade edilir (Sayyan 2000). Klasik regresyon modeline benzer bir şekilde yazılan bu modelde $\hat{\boldsymbol{\delta}}$ tahminini en küçük kareler yöntemine bağlı olarak $\boldsymbol{u'u}$ ifadesini minimize etmek yoluyla bulmak mümkündür. Ancak tasarım matrisi $\boldsymbol{Z}$ biçiminde oluşturulduğu zaman boyutu büyüyecek ve en küçük kareler yöntemi gereği $(N+q)\times(N+q)$ boyutlu $\boldsymbol{Z'Z}$ matrisinin tersini bulma işlemiyle karşılaşılacaktır.

İşlem zorluğunu yok etmek için öncelikle (4.11) eşitliğindeki modelde $\boldsymbol{\beta}$ parametre vektörü tahmin edilecek ardından $\boldsymbol{\alpha}$ ' nın tahmini bulunacaktır.

4.1.2. Parametrik sabit etkili panel veri modelinin tahmini

Eşitlik (4.11) deki model ele alınırsa, $\boldsymbol{\beta}$ parametre vektörünü tahmin etmek için en küçük kareler yöntemi kullanılabilir. En küçük kareler yöntemine bağlı olarak $\boldsymbol{u'u}$ ifadesini minimize ederek $\boldsymbol{\beta}$ parametre vektörünün en iyi yansız bir tahminini bulmak mümkündür. Minimum yapma işlemi sonucunda $\boldsymbol{\beta}$ parametre vektörünün tahmin edicisi

$$\hat{\boldsymbol{\beta}}_{SE} = \left(\boldsymbol{X'M}_D\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X'M}_D\boldsymbol{Y}\right) \tag{4.13}$$

olarak elde edilir. Eşitlik (4.13) de

$$\boldsymbol{M}_D = \boldsymbol{I}_{NT} - \boldsymbol{D}\left(\boldsymbol{D'D}\right)^{-1}\boldsymbol{D'} \tag{4.14}$$

biçimindeki matrisi gösterir. Aynı zamanda $\boldsymbol{M}_D$ matrisini

$$\boldsymbol{Q}_T = \boldsymbol{I}_T - \overline{\boldsymbol{Q}}_T \qquad \text{ve} \qquad \overline{\boldsymbol{Q}}_T = \frac{\boldsymbol{1}_T\boldsymbol{1}_T'}{T}$$

olmak üzere

$$\boldsymbol{M}_D = \boldsymbol{I}_N \otimes \boldsymbol{Q}_T \tag{4.15}$$

şeklinde yazmak mümkündür. $\boldsymbol{\beta}$ parametre vektörünü tahmin ettikten sonra modelde yer alan α_i sabit parametresinin tahmini ise

$$\hat{\alpha}_i = \overline{Y}_{i.} - \sum_{k=1}^{q} \overline{x}_{ki.}\hat{\beta}_{kSE} \tag{4.16}$$

biçiminde yapılır (Roy 1997). Burada $\bar{Y}_{i.}$ bağımlı değişkene ait gözlemlerde her bir i bireyine ait T adet zamandaki gözlemlerin ortalaması, $\bar{x}_{ki.}$ ise k. bağımsız değişkenin i bireyine ait T adet zamandaki gözlemlerin ortalamasıdır. Matematiksel olarak ifade edilecek olursa

$$\bar{Y}_{i.} = \frac{1}{T}\sum_{t=1}^{T} y_{it}\,, \qquad \bar{x}_{ki.} = \frac{1}{T}\sum_{t=1}^{T} x_{kit}$$

dir. α sabit parametresinin tahmini matris çarpımı yolu ile hesaplanmak istenirse

$$\hat{\boldsymbol{\alpha}}_{SE} = \left(\boldsymbol{D}'\boldsymbol{D}\right)^{-1}\boldsymbol{D}'\left(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{SE}\right) \tag{4.17}$$

eşitliği kullanılır. Burada $\boldsymbol{D}$ matrisi daha önceden $\boldsymbol{D} = \boldsymbol{I}_N \otimes \boldsymbol{1}_T$ biçiminde tanımlanan katsayı matrisidir.

4.1.3. Parametrik rasgele etkili panel veri modeli

Panel verili regresyon analizinde kullanılan regresyon modellerinden biri de Rasgele Etkili Modeldir. Sabit etkili panel veri modelinde birimlere göre meydana gelen farklılıklardan kaynaklanan değişmeyi gösterebilmek için modele sabit katsayı adı verilen her bir bireyin T zamanda aldığı değer için sabit olan bir α_i sabit etki katsayısı eklenmişti. Rasgele etkili modelde ise yine birimlere göre meydana gelen değişimi göstermek mümkündür. Modeldeki hata terimi olan u_{it} ' lerin birçok gözlenememiş etkileri, zamana ve bireylere göre değişimden kaynaklanan bilgi kaybını içine dahil ettiği göz önüne alınırsa α_i' lerin de u_{it} gibi rasgele olabileceği düşünülebilir. Yine aynı düşünce tarzı ile bireylerin ve zamanın göstereceği farklılıklardan meydana gelen α_i' ler, hata terimi olan u_{it}' lerin bir bileşeni olarak kabul edilebilir. Bu işlem sonucunda yeni oluşacak olan hata terimini

$$v_{it} = \alpha_i + u_{it} \tag{4.18}$$

biçiminde ifade etmek mümkündür. Eşitlik (4.18)'de rasgele olduğu kabul edilen α_i teriminin bağımsız ve aynı dağılımlı , 0 ortalamalı ve σ_α^2 varyanslı olduğu varsayımı geçerlidir. Bu şekilde yeni oluşturulan hata terimi sayesinde sabit etkili modelde oluşan serbestlik derecesi kaybı önlenmiş olur. Ancak buna karşın rasgele etkili modelde hata teriminin varyans-kovaryans matrisinin oluşturulması zorlaşacaktır.

Yeni hata teriminin kullanımı ile oluşturulacak rasgele etkili panel veri modeli

$$Y_{it} = \mu + x_{kit}\beta_k + v_{it}, \quad i = 1,2,\ldots,N \; ; t = 1,2,\ldots,T \; ; \; k = 1,2,...,q \tag{4.19}$$

olarak elde edilir. Rasgele etkili panel veri modelinde, sabit etkili modelde olduğu gibi sabit katsayı kullanmak mümkündür. μ katsayısı burada anakütle ortalama sabit parametresini temsil etmektedir ve herhangi bir tanımlama problemine yol açmamaktadır. μ parametresinin modele dahil edilip edilmemesine araştırıcı tarafından karar verilir. Eşitlik (4.19) ile verilen modelde, bireylere veya zamana göre meydana gelebilecek değişme hata terimine dahil edildiği için modelin bir diğer adı da hata bileşenler modelidir.

Rasgele etkili model, $\boldsymbol{\beta} = \left[\beta_1 \cdots \beta_q\right]'$ şeklinde $q \times 1$ boyutlu vektör, $\boldsymbol{X}$ $NT \times q$ boyutlu tasarım matrisi, $\boldsymbol{v}$' de $NT \times 1$ boyutlu hataların vektörü olmak üzere matris gösterimi ile yazılırsa

$$\boldsymbol{Y} = \mu + \boldsymbol{X\beta} + \boldsymbol{v} \tag{4.20}$$

veya model $i.$ birey için matris gösterimi ile yazılırsa

$$\boldsymbol{Y}_i = \tilde{\boldsymbol{X}}_i\boldsymbol{\delta} + \boldsymbol{v}_i \tag{4.21}$$

biçiminde ifade edilir. Eşitlik (4.21)' de $\tilde{\boldsymbol{X}}_i = \left[\boldsymbol{1}_T \quad \boldsymbol{X}_i\right]$ olacak şekilde $T \times (q+1)$ boyutlu matris, $\boldsymbol{\delta} = \left[\mu \quad \boldsymbol{\beta}'\right]'$ biçiminde $(q+1)\times 1$ boyutlu parametre vektörü, $\boldsymbol{v_i} = \left[v_{i1} \quad \cdots \quad v_{iT}\right]'$ olacak şekilde hata terimleri vektörüdür.

4.1.4. Parametrik rasgele etkili panel veri modelinin tahmini

Rasgele etkili panel regresyon modelinde parametre tahmini yaparken dikkat edilecek husus hata teriminin iki ayrı bileşenden oluşmuş olmasıdır. Rasgele etkili modelde parametre tahmini yapmak için sıradan en küçük kareler tahmin yöntemi kullanılabilir. Bu yöntemi uygulamak için (4.21) eşitliğinde verilen *i.* birey için rasgele etkili panel regresyon modelini daha önce tanımlanan $\boldsymbol{Q}_T$ matrisi ile çarpıp daha önce sabit etkili model için verilen (4.13) eşitliği yardımıyla $\hat{\boldsymbol{\beta}}_{RE}$ parametre vektörünün tahmininin elde edilmesi mümkündür. Elde edilen

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{Y}\right) \tag{4.22}$$

şeklindeki rasgele etkili model tahmin edicisi görüldüğü gibi sabit etkili model için elde edilen $\hat{\boldsymbol{\beta}}_{SE}$ parametre vektörünün aynısıdır. Bu işlemi yaparken dikkat edilecek husus sabit etkili modelde $\hat{\boldsymbol{\beta}}_{SE}$ parametre vektörünü bulduktan sonra N adet model için α_i sabit parametrelerinin tahmini (4.17) eşitliği yardımıyla hesaplanıyordu. Rasgele etkili panel regresyon modeli olarak eşitlik (4.20) ile verilen modelde ise her bir birey ve zaman için tek bir μ anakütle ortalama sabit parametresi mevcut olduğundan anakütle ortalama sabit parametresinin tahmini

$$\hat{\mu} = \bar{Y}_{..} - \sum_{k=1}^{q} \hat{\beta}_{kRE}\bar{x}_{k..} \tag{4.23}$$

eşitliği yardımıyla hesaplanır. Eşitlik (4.23)' de $\bar{Y}_{..}$; Y_{it} gözlemlerinin ortalamasını, $\bar{x}_{k..}$ ise x_{kit} gözlemlerinin ortalamasını göstermektedir yani

$$\overline{Y}_{..} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it} \ , \quad \overline{x}_{k..} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{kit} \tag{4.24}$$

eşitlikleri ile matematiksel olarak gösterilir. Eşitlik (4.22) ile elde edilen $\hat{\boldsymbol{\beta}}_{RE}$ tahmini yansız ve tutarlı bir tahmin olmasına rağmen v_{it} hata teriminin birleşik terimlerden oluşmasını ve hataların varyans-kovaryans yapısını göz önüne almadığından gerçek veri seti için etkin bir tahmin edici değildir. $\boldsymbol{\beta}$ parametre vektörünün en iyi yansız tahmin edicisi için Genelleştirilmiş en küçük kareler tahmin yöntemi kullanılması önerilir. Genelleştirilmiş en küçük kareler tahmini için (4.20) eşitliğinde verilen model göz önüne alınırsa modeldeki $\hat{\boldsymbol{\beta}}_{RE}$ parametre vektörünün en iyi yansız tahmin edicisi olan $\hat{\boldsymbol{\beta}}_{RE}$ nin bulunması için $\boldsymbol{X}$ bağımsız değişkenlerin matrisi ile $\boldsymbol{Y}$ bağımlı değişken vektörü olmak üzere, genelleştirilmiş en küçük kareler tahmin edicisi

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{RE}} = \left(\boldsymbol{X}'\boldsymbol{M}_V{}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{M}_V{}^{-1}\boldsymbol{Y}\right) \tag{4.25}$$

eşitliği ile hesaplanabilir. Burada $\boldsymbol{M}_V$ matrisinin hesaplanması için iki bileşenden oluşan hatalara ait varyans kovaryans matrisinin tahmin edilmesi gerekmektedir. Modelde i. birey için bütün T zamandaki hata terimleri ele alınırsa, v_i 'nin varyans kovaryans matrisi

$$E(v_i v_i') = \boldsymbol{V} = \sigma_u^2 \boldsymbol{I}_T + \sigma_\alpha^2 \boldsymbol{1}_T \boldsymbol{1}_T' \tag{4.26}$$

biçiminde elde edilir (Darnell 1994). Her bir i birey için gözlemler birbirinden bağımsız olduğundan birleşik hata terimi v_{it} ' nin NT gözlemi için elde edilecek varyans-kovaryans matrisi yani (4.25) eşitliğinde kullanılan $\boldsymbol{M}_V$ matrisi

$$\boldsymbol{M}_V = \boldsymbol{I}_N \otimes \boldsymbol{V} \tag{4.27}$$

eşitliği yardımıyla elde edilir. Eşitlik (4.25) de kullanılan $\boldsymbol{M}_V^{-1}$ ters matrisi açık olarak

$$\boldsymbol{M}_V^{-1} = \frac{1}{\sigma_u^2}\left[\boldsymbol{I}_{T\times N} - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2}\boldsymbol{1}_{T\times N}\boldsymbol{1}'_{T\times N}\right] \tag{4.28}$$

şeklinde hesaplanır.

Parametrik rasgele etkili panel veri modelinin tahmini doğrudan veriler üzerine uygulanacak bir dönüşüm yardımıyla da hesaplanabilmektedir. Eşitlik (4.25) ile verilen tahmin edicide kullanılan $\boldsymbol{M}_V$ matrisini ortadan kaldıracak bu dönüşüm için öncelikle bir θ değerinin hesaplanması gerekmektedir. θ değeri,

$$\theta = 1 - \sqrt{\frac{\sigma_u^2}{T\sigma_\alpha^2 + \sigma_u^2}}$$

eşitliği yardımıyla hesaplanır. Hesaplanan θ yardımıyla yeni dönüştürülmüş değişkenler $Y_{it}^{*} = Y_{it} - \theta\overline{Y}_{i.}$ ve $x_{kit}^{*} = x_{kit} - \theta\overline{x}_{ki.}$ olarak bulunur. Dönüştürülmüş değişkenler yardımıyla kurulan

$$Y_{it}^{*} = \mu + x_{kit}^{*}\beta_k + v_{it} \tag{4.29}$$

modeline klasik en küçük kareler tahmini uygulanırsa parametrik rasgele etkili panel veri modeline ait $\hat{\boldsymbol{\beta}}_{RE}$ tahmin edicisi,

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^{*}\right)^{-1}\left(\boldsymbol{X}^{*\prime}\boldsymbol{Y}^{*}\right) \tag{4.30}$$

olacak biçimde elde edilir. Parametre tahmini için yapılması gereken son adım ise σ_α^2 σ_u^2 varyans değerlerinin tahmin edilmesidir. Klasik regresyonda σ^2' nin bir tahmin edicisi olarak öne sürülen s^2 tahmin edicisi, *n* gözlem, *k* değişken sayısı, $\hat{\beta}$ da β parametrelerinin en küçük kareler tahmin edicisi olmak üzere

$$s^2 = \frac{1}{n-k-1}\sum_{k=1}^{q}\sum_{i=1}^{n}\left(Y_i - x_{ki}\hat{\beta}_k\right)^2$$

biçiminde idi(Montgomery ve Peck 1992). Panel regresyon modelinde bireylerin zamanlara göre aldıkları değerlerin varlığından dolayı varyans-kovaryans yapısını daha iyi açıklayabilmek için σ_α^2 ve σ_u^2 için Swamy ve Arora(1972) tarafından aşağıda tanımlanan tahmin ediciler önerilmiştir. Buna göre

$$\hat{\sigma}_u^2 = \frac{\sum_{k=1}^{q}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[Y_{it} - \overline{Y}_{i.} - \left(x_{kit} - \overline{x}_{ki.}\right)\hat{\beta}_{k\,SE}\right]^2}{N(T-1)-q}$$

$$\hat{\sigma}_\alpha^2 = \frac{\sum_{k=1}^{q}\sum_{i=1}^{N}\left(\overline{Y}_{i.} - \hat{\mu}_G - \overline{x}_{ki.}\hat{\beta}_{k\,G}\right)^2}{N-(q+1)} - \frac{1}{T}\hat{\sigma}_u^2 \qquad (4.31)$$

olarak verilen tahmin edicilerdir. Eşitlik (4.31)' de $\boldsymbol{\beta}_G$ gruplar arası tahmin edici olarak adlandırılır ve $\overline{\boldsymbol{X}}$ elemanları $\left(\overline{x}_{ki.} - \overline{x}_{k..}\right)$ biçiminde olan $N \times q$ boyutlu matris ve $\overline{\boldsymbol{Y}}$ 'de elemanları $\left(\overline{Y}_{i.} - \overline{Y}_{..}\right)$ şeklinde olan $N \times 1$ boyutlu vektör olmak üzere

$$\boldsymbol{\beta}_G = \left(\overline{\boldsymbol{X}}'\overline{\boldsymbol{X}}\right)^{-1}\overline{\boldsymbol{X}}\,\overline{\boldsymbol{Y}} \qquad (4.32)$$

olarak tahmin edilir. $\boldsymbol{\beta}_G$' nin tahmin edilmesinin ardından

$$\hat{\mu}_G = \overline{Y}_{..} - \sum_{k=1}^{q} \hat{\beta}_{kG} \overline{x}_{k..} \tag{4.33}$$

eşitliği ile $\hat{\mu}_G$ tahmin edilir. $\hat{\boldsymbol{\beta}}_{RE}$ tahmininin hesaplanmasından sonra (4.20) modelinde yer alan μ sabit parametresinin tahmini ise basit olarak, $\hat{\mu} = \overline{Y}_{..} - \hat{\boldsymbol{\beta}}'_{RE} \overline{\boldsymbol{X}}_{..}$ olacak biçimde hesaplanır.

4.2. Parametrik Olmayan Regresyonda Panel Verili Modeller

Panel veri analizleri için kullanılan bir diğer yöntem parametrik olmayan regresyon modeli kullanımıdır. Parametrik olmayan regresyon modelinin tahmin yöntemleri ve yoğunluk tahmini konusu 2. bölümde ayrıntılı olarak anlatılmıştı. Parametrik olmayan regresyonda tahmin için kullanılan Nadaraya-Watson kernel yöntemi panel verilere de uyarlanabilmektedir. Bu yöntemden farklı olarak lokal lineer kernel yöntemi de geliştirilmiştir. Öncelikle bu yöntemlerin panel veri için kullanımı hakkında bilgi verdikten sonra parametrik olmayan regresyonda sabit ve rasgele etkili panel veri modellerinin tahmini anlatılacaktır. Nadaraya Watson kernel tahmin yönteminin matrisler ile uygulanması konusu da panel veri analizleri için gerekli olduğundan bu bölümde incelenecektir.

4.2.1. Parametrik olmayan panel verili modellerde Nadaraya-Watson kernel tahmini

Parametrik olmayan regresyonda panel veri kullanımı için yazılacak tek değişkenli model

$$Y_{it} = m(x_{it}) + u_{it} \ , \qquad i = 1,2,\ldots,N \ ; \ t = 1,2,\ldots,T \tag{4.34}$$

eşitliği ile verilsin. Bu modelde $m(x_{it})$ bağımsız değişkene ait bilinmeyen biçimdeki fonksiyonu temsil etmektedir. Y_{it} bağımlı değişkenleri, u_{it} ise hata terimlerini göstermektedir. Hata terimi üzerinde yapılan varsayımlar (4.34)'deki model içinde

geçerli olup, u_{it} hata terimlerinin bağımsız ve aynı dağılımlı, beklenen değeri sıfır ve sabit bir σ_u^2 varyansına sahip olduğu varsayılır. $m(x)$ fonksiyonunun parametrik olmayan tahmini yani bir x noktasındaki koşullu beklenen değer, $x_{it} - x$ gibi küçük bir aralıkta y değerlerinin ağırlıklandırılması ile bulunur. $m(x_{it})$ fonksiyonunu ağırlıklandırılacak forma getirmek amacıyla $m(x_{it})$ fonksiyonu taylor açılımı yardımıyla x komşuluğunda taylor serisine açılırsa

$$m(x_{it}) = m(x) + m'(x)(x_{it} - x) + m''(x)(x_{it} - x)^2 / 2 + \cdots \tag{4.35}$$

elde edilir. Taylor serisine açılmış haldeki (4.35) eşitliği O fonksiyonu yardımıyla

$$m(x_{it}) = m(x) + O(x_{it} - x), \tag{4.36}$$

şeklinde yazılır (Pagan ve Ullah 1999). $O(x_{it} - x)$ terimi özellikle N sayısı büyük olduğu zaman u_{it}' ye dahil edilebilir. Bu işlem iki bileşenden oluşan hata teriminin beklenen değerini değiştirmeyecektir ve sonuç olarak parametrik olmayan panel verili regresyon modeli

$$Y_{it} = m(x) + u_{it} \tag{4.37}$$

yada daha açık bir ifade olarak matris gösterimi ile

$$\boldsymbol{Y} = \boldsymbol{1}_{NT}\, m(x) + \boldsymbol{u} \tag{4.38}$$

eşitliği ile ifade edilebilir. Eşitlik (4.38) de, $\boldsymbol{Y}$: $NT \times 1$ boyutlu bağımlı değişkenler vektörü, $m(x)$ bilinmeyen biçimdeki fonksiyon ve $\boldsymbol{u}$' da $NT \times 1$ boyutlu hata terimini göstermektedir. Eşitlik (4.38) de verilen modelinin tahmini için, ağırlıklandırılmış en küçük kareler yöntemine benzer olarak, $\boldsymbol{K}(x)$; elemanları kernel yoğunluk fonksiyonu

olan matris olmak üzere $\boldsymbol{u}'\boldsymbol{K}(x)\boldsymbol{u}$ ifadesinin minimum yapılması gerekmektedir. $\boldsymbol{u}'\boldsymbol{K}(x)\boldsymbol{u}$ ifadesi toplam olarak yazılırsa

$$\boldsymbol{u}'\boldsymbol{K}(x)\boldsymbol{u} = \sum_i \sum_t (Y_{it} - m(x))^2 K\left(\frac{x_{it} - x}{h}\right) \tag{4.39}$$

biçimindedir. Bu ifadenin minimum yapılma işlemi ile karşılaşılır. Eşitlik (4.39)' da $\boldsymbol{K}(x)$, köşegen elemanları $K\left(\frac{x_{it} - x}{h}\right)$' lerden oluşan köşegen matris, $K\left(\frac{x_{it} - x}{h}\right)$' de kernel ağırlık fonksiyonu olarak adlandırılır ve $\boldsymbol{K}_{it}$ ile de gösterilebilir. $\boldsymbol{K}_{it}$ fonksiyonu x_{it} değeri x' den uzak iken küçük değerler, x' e yaklaştıkça büyük değerler alan bir ağırlık fonksiyonudur. Kernel ağırlık fonksiyonunun özellikleri detaylı olarak yoğunluk tahmini başlığı altında incelenmişti. Eşitlik (4.39)'un minimum yapılması ile elde edilecek $\hat{m}(x)$ tahmini matris gösterimi ile

$$\hat{m}(x) = (\boldsymbol{1}'_{NT}\, \boldsymbol{K}(x)\boldsymbol{1}_{NT})^{-1} \boldsymbol{1}'_{NT}\, \boldsymbol{K}(x)\boldsymbol{Y} \tag{4.40}$$

eşitliğinde görüldüğü gibidir (Roy 1997). Bu tahmin yöntemi parametrik olmayan regresyon modelinde tahmin konusunda gösterilen Nadaraya-Watson kernel regresyon tahmini olarak adlandırılır. Değişken sayısı 1' den fazla olduğunda çarpımsal kernel fonksiyonu kullanılmaktadır(Wasserman 2006).

4.2.2. Lokal lineer kernel tahmini

Önceki başlık altında, (4.34) eşitliği ile verilen parametrik olmayan model Nadaraya-Watson kernel tahmini yapabilmek amacıyla Taylor serisine açılmış ilk terim olan $m(x)$ alınarak kalan kısım u_{it} hata terimine dahil edilmişti. Lokal lineer kernel tahmin yöntemi ise parametrik olmayan modeli, parametreye göre lineer model biçimine getirmeyi amaçlar. (4.34) eşitliğinde $m(x_{it})$ fonksiyonunun $x_{it}\beta$ şeklinde yazılabileceğini düşünürsek, modelimiz $Y_{it} = x_{it}\beta + u_{it}$ biçimindeki lineer modele

dönüşecek ve parametre tahmin problemi bilinen basit yöntemler yardımıyla çözülebilecektir. Bu amaçla Taylor serisine açılan $m(x_{it})$ fonksiyonunun ilk iki terimi alınarak yeni parametrik olmayan model

$$Y_{it} = m(x) + (x_{it} - x)\psi(x) + u_{it} \ , \qquad i = 1,2,\ldots,N \ ; \ t = 1,2,\ldots,T \tag{4.41}$$

eşitliği ile ifade edilir (Pagan ve Ullah 1999). Burada $\psi(x)$ olarak verilen fonksiyon $m(x)$ fonksiyonunun birinci türevidir. (4.41) eşitliğini q değişken için daha basit bir matris gösterimi ile

$$\boldsymbol{Y} = \boldsymbol{Z\delta} + \boldsymbol{u} \tag{4.42}$$

biçiminde yazmak mümkündür. Burada $\boldsymbol{Z}$ elemanları $\begin{bmatrix} 1 & x_{1it} - x_1 \cdots x_{qit} - x_q \end{bmatrix}$ biçiminde olan $NT \times (q+1)$ boyutlu matris, $\boldsymbol{\delta}$ ' da elemanları $\begin{bmatrix} m(x) & \boldsymbol{\psi}'_k(x) \end{bmatrix}'$ biçiminde olan $(q+1)\times 1$ boyutlu parametre vektörüdür. (4.42) eşitliğindeki model ağırlıklandırılmış en küçük kareler yöntemi ile çözülürse, $\boldsymbol{u}'\boldsymbol{K}(x)\boldsymbol{u}$ ifadesinin minimum yapılması ile elde edeceğimiz tahmin edici

$$\hat{\boldsymbol{\delta}} = \left(\boldsymbol{Z}'\boldsymbol{K}(x)\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{K}(x)\boldsymbol{Y} \tag{4.43}$$

tahmin edicisinin elde edilmesinden sonra $m(x)$ ve $\boldsymbol{\psi}(x)$' in lokal en küçük kareler tahmin edicileri

$$\hat{m}(x) = \begin{pmatrix} 1 & \boldsymbol{0}'_q \end{pmatrix}\hat{\boldsymbol{\delta}} \tag{4.44}$$

ve

$$\hat{\boldsymbol{\psi}}(\boldsymbol{x}) = \begin{pmatrix} 0 & \boldsymbol{1}'_q \end{pmatrix}\hat{\boldsymbol{\delta}} \tag{4.45}$$

eşitlikleri yardımıyla elde edilir. $\boldsymbol{0_q}$ ile gösterilen $q\times 1$ boyutlu elemanları 0 olan vektör, $\boldsymbol{1_q}$ ile gösterilen $q\times 1$ boyutlu elemanları 1 olan vektördür.

4.2.3. Parametrik olmayan sabit etkili panel veri modeli

Parametrik olmayan panel verili regresyon modelinde bireylere göre meydana gelecek değişimi modele dahil etmenin iki farklı yolu vardır. Daha önce parametrik panel verili regresyon modelinde meydana gelen değişimi sabit veya rasgele etki olarak iki farklı şekilde modele dahil etmiş ve iki farklı model üzerinde tahmin yöntemlerini incelemiştik. Parametrik modelde olduğu gibi parametrik olmayan modelde de değişim modele iki farklı yöntemle eklenebilir. Birimlere göre meydana gelen değişimi sabit bir etki parametresi ile modele eklersek modelimizin adı parametrik olmayan sabit etkili model olarak adlandırılır. Sabit etki parametresini α_i, $i = 1,...,N$ olarak gösterelim. α_i bireylere ait özel sabit parametre olarak adlandırılır, ve parametrik olmayan sabit etkili panel veri modeli

$$Y_{it} = \alpha_i + m(x_{it}) + u_{it} \ , \qquad i = 1,2,\ldots,N \ ; \ t = 1,2,\ldots,T \tag{4.46}$$

eşitliği ile verildiği gibidir (Mukherjee 2002). Burada Y_{it} ; bağımlı değişkenleri, $m(x_{it})$ bağımsız değişkene ait bilinmeyen biçimdeki fonksiyonu ve u_{it} de hata terimlerini göstermektedir. Eşitlik (4.46) ile verilen modelde $m(x_{it})$ yerine $x_{it}\beta$ olduğu zaman $Y_{it} = \alpha_i + x_{it}\beta + u_{it}$ biçiminde olan parametrik sabit etkili model için $\hat{\beta}$ ve $\hat{\alpha}_i$ tahminleri herhangi bir dönüşüme ihtiyaç duyulmadan yapılabilmekte idi. Ancak parametrik olmayan sabit etkili modelde $m(x_{it})$ bilinmeyen formda bir fonksiyon olduğundan α_i sabit etki katsayısının tahmini zorlaşmaktadır. Bu problemi giderebilmek ve bireylerde *T* zaman için oluşan etkiyi sabitleyebilmek için modeli lokal kernel tahmininde olduğu gibi lineer bir forma getirmeye ihtiyaç duyulmaktadır. Bu amaçla (4.46) eşitliği ile verilen modeli, (4.41) eşitliğinin elde edilişine benzer olarak $m(x_{it})$ fonksiyonunun taylor serisinden elde edilen ilk iki terimi yardımıyla

$$Y_{it} = \alpha_i + m(x) + (x_{it} - x)\psi(x) + u_{it} \ , \qquad i = 1,2,\ldots,N \ ; \ t = 1,2,\ldots,T \tag{4.47}$$

biçiminde yazmak mümkündür. Aynı zaman da T zaman için bireylerin aldığı değerleri sabitleyebilmek amacı ile

$$\overline{Y}_{i.} = \alpha_i + m(x) + (\overline{x}_{i.} - x)\psi(x) + \overline{u}_{i.} \quad , \qquad i = 1,2,...,N \tag{4.48}$$

eşitliği yazılabilmektedir. Burada $\psi(x)$ fonksiyonu $m(x)$ fonksiyonunun birinci türevidir.

4.2.4. Parametrik olmayan sabit etkili panel veri modelinde tahmin yöntemi

Parametrik olmayan sabit etkili modelde tahmin yapabilmek için modelin lineer bir formda yazılacağı daha önce belirtilmişti. Aynı zamanda modelde yer alan $m(x)$ ve α_i terimlerinin de yok edilebilmesi için eşitlik (4.47)'den eşitlik (4.48) çıkartılarak gerekli düzenlemeler yapılırsa

$$Y_{it} - \overline{Y}_{i.} = (x_{it} - \overline{x}_{i.})\psi(x) + u_{it} - \overline{u}_{i.} \quad , \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.49}$$

modeli elde edilir. Bu model yardımıyla $m(x)$'in türevi olan $\psi(x)$ tahmin edilebilecek ve ardından basit matematiksel işlemler yardımıyla $\hat{m}(x)$ ve $\hat{\alpha}_i$ tahminleri de bulunabilecektir. Eşitlik (4.49) ile verilen modelde $\psi(x)$ fonksiyonunu tahmin etmek için lokal lineer kernel tahminden yola çıkarak

$$\sum_i \sum_t \left(Y_{it} - \overline{Y}_{i.} - (x_{it} - \overline{x}_{i.})\psi(x)\right)^2 K\left(\frac{x_{it} - x}{h}\right) \tag{4.50}$$

toplamını minimum yaparak, $\hat{\psi}_{SE}(x)$ tahminini değişken sayısının tek olduğu durumda yani $q = 1$ için

$$\hat{\psi}_{SE}(x)=\sum_i\sum_t \frac{K\left(\frac{x_{it}-x}{h}\right)(Y_{it}-\bar{Y}_{i.})(x_{it}-\bar{x}_{i.})}{\sum_i\sum_t K\left(\frac{x_{it}-x}{h}\right)(x_{it}-\bar{x}_{i.})^2} \tag{4.51}$$

eşitliği yardımıyla elde ederiz. Değişken sayısı birden fazla ise yani $q \geq 1$ için (4.50) ifadesini minimum yapan değer matris gösterimi ile

$$\hat{\boldsymbol{\psi}}_{SE}(x)=\left(\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{K}(x)\boldsymbol{M}_D\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{K}(x)\boldsymbol{M}_D\boldsymbol{Y}\right) \tag{4.52}$$

biçiminde elde edilir (Roy 1997). Burada $\boldsymbol{K}(x)$, köşegen elemanları $K\left(\frac{x_{it}-x}{h}\right)$, lerden oluşan $NT \times NT$ boyutlu köşegen matris, $\boldsymbol{M}_D$ bölüm (4.1.2.) de verilen $NT \times NT$ boyutlu matris, $\boldsymbol{X}$ ise $NT \times q$ boyutlu tasarım matrisidir. $\hat{\boldsymbol{\psi}}_{SE}(x)$ parametre vektörünün tahminidir ve $q \geq 1$ olduğundan $\hat{\boldsymbol{\psi}}_{SE}(x)$, $q \times 1$ boyutlu elemanları $\hat{\boldsymbol{\psi}}_{kSE}(x)$ ' lerden oluşan vektördür. $\hat{\boldsymbol{\psi}}_{SE}(x)$ tahminini yapmak için dönüştürülmüş değişkenlerde kullanılabilir. Eşitlik (4.49)'da verilen modeli yazarken kullanılan $(Y_{it}-\bar{Y}_{i.})$ dönüştürülmüş değişkenini Y_{it}^* , $(x_{it}-\bar{x}_{i.})$ dönüştürülmüş değişkenini x_{it}^* ve $(u_{it}-\bar{u}_{i.})$ dönüştürülmüş hataları da u_{it}^* ile gösterirsek modelimizi

$$Y_{it}^* = x_{it}^*\,\psi(x)+u_{it}^* \tag{4.53}$$

biçiminde yazabiliriz. Matris gösterimi ile bu ifade $\boldsymbol{\psi}(x)$ türev fonksiyonlarının vektörü, $\boldsymbol{X}^*$; dönüştürülmüş x_{kit}^* değişkenlerinin $NT \times q$ boyutlu tasarım matrisi, $\boldsymbol{Y}^*$; dönüştürülmüş bağımlı değişkenin $NT \times 1$ boyutlu vektörü, $\boldsymbol{u}^*$ hataların $NT \times 1$ boyutlu vektörü olmak üzere

$$\boldsymbol{Y}^* = \boldsymbol{X}^*\boldsymbol{\psi}(x)+\boldsymbol{u}^* \tag{4.54}$$

şeklinde yazılır. Bu modele lokal lineer kernel tahmininde kullandığımız ağırlıklandırılmış en küçük kareler tahmin yöntemi uygulanırsa,

$$\hat{\boldsymbol{\psi}}_{SE}(x) = \left(\boldsymbol{X}^{*'} \boldsymbol{K}(x) \boldsymbol{X}^{*} \right)^{-1} \left(\boldsymbol{X}^{*'} \boldsymbol{K}(x) \boldsymbol{Y}^{*} \right) \tag{4.55}$$

eşitliği ile sabit etkili model için $\hat{\boldsymbol{\psi}}_{SE}(x)$ vektörünün tahmini elde edilir. Tahmin sonucunda ham tasarım matrisinin kullanıldığı eşitlik (4.52) ile verilen tahmin edici ile aynı değerlere ulaşılır. $\hat{\boldsymbol{\psi}}_{SE}(x)$ türev fonksiyonu tahmin edildikten sonra sıra ile (4.47) eşitliğindeki $m(x)$ fonksiyonunun ve α_i sabit parametrelerinin tahmin edilmesi gerekmektedir. Bu tahminler sırası ile

$$\hat{m}(x) = \breve{Y}_{..} - \sum_{k=1}^{q} (\breve{x}_{k..} - x_k) \hat{\boldsymbol{\psi}}_{kSE}(x) \tag{4.56}$$

ve

$$\hat{\alpha}_i = \breve{Y}_{i.} - \hat{m}(x) - \sum_{k=1}^{q} (\breve{x}_{ki.} - x_k) \hat{\boldsymbol{\psi}}_{kSE}(x) \tag{4.57}$$

denklemleri kullanılarak elde edilir. Eşitliklerde ($\bar{x}_{ki.}, \bar{Y}_{i.}$) ile gösterilen ifade bireylere göre zamanların ortalamaları, ($\bar{x}_{k..}, \bar{Y}_{..}$) gösteren ifade ise gözlemlerin tümünün ortalamasıdır (Roy 1997). Parametrik panel verili regresyon modelinde ortalama değerlerinin hesaplanması gösterilmiştir.

4.2.5. Parametrik olmayan rasgele etkili panel veri modeli

Parametrik olmayan panel verili regresyon modelinde bireylere göre meydana gelecek değişimi modele dahil etmenin bir diğer yolu Rasgele etkili model kullanmaktır. Parametrik panel verili regresyon modelinde kullanıldığı gibi bireylere göre meydana gelen değişimi temsil eden α_i' lerin rasgele olduğu düşünülürse, α_i' ler hata terimi olan u_{it} 'lerin bir bileşeni olarak kabul edilebilir. Yeni hata terimleri

$$v_{it} = u_{it} + \alpha_i, \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.58}$$

olacak biçimde düşünülür. Parametrik olmayan panel verili regresyonda $u_{it} + \alpha_i$ biçiminde olduğu düşünülen yeni v_{it} hata terimlerinin kullanımı ile elde edilen parametrik olmayan rasgele etkili model

$$Y_{it} = m(x_{it}) + v_{it}, \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.59}$$

biçiminde ifade edilir. Burada Y_{it} ; bağımlı değişkenleri, $m(x_{it})$ bağımsız değişkene ait bilinmeyen biçimdeki fonksiyonu ve v_{it} ' de birleşik hata terimlerini göstermektedir. Modeli lokal lineer kernel tahmininde gösterildiği gibi taylor serisine açıp aynı matematiksel işlemler yardımıyla parametrik olmayan rasgele etkili model

$$Y_{it} = m(x) + (x_{it} - x)\psi(x) + v_{it} \; , \qquad i = 1,2,\ldots,N \; ; \; t = 1,2,\ldots,T \tag{4.60}$$

olarak elde edilir. $\psi(x)$ fonksiyonu $m(x_{it})$ fonksiyonunun birinci türevinin x noktasındaki değeridir. Rasgele etkili modele sabit bir μ katsayısı eklenebileceği gibi sabit katsayının $m(x)$ değeri içine dahil edildiği de düşünülebilir (Roy 1997). Parametrik rasgele etkili modelde μ sabit parametresi modele dahil edilmiş ve tahmini yapılmıştı. Parametrik olmayan rasgele etkili modelde ise sabit parametrenin $m(x)$ değeri içinde olduğunu varsayılarak işlemlere eşitlik (4.60) ile verilen model ile devam edilmektedir.

4.2.6. Parametrik olmayan rasgele etkili panel veri modelinde tahmin

Parametrik olmayan rasgele etkili panel veri modelinin tahminine (4.60) eşitliği ile elde edilen modeldeki türev fonksiyonunun tahmini ile başlanır. Türev fonksiyonu tahmin edildikten sonra $m(x_{it})$ fonksiyonunun tahmini (4.56) eşitliğinde olduğu gibi hesaplanır. Unutulmaması gereken husus, lokal lineer kernel tahmininde

olduğu gibi rasgele etkili model tahmin edicisinin de kernel fonksiyonu yardımıyla ağırlıklandırılması gerekmektedir.

Eşitlik (4.60) ile verilen model için türev fonksiyonu olan $\boldsymbol{\psi}(x)$' in parametrik olmayan rasgele etki tahmin edicisi bağımsız değişken sayısı yani $q \geq 1$ için

$$\hat{\boldsymbol{\psi}}_{RE}(x) = \left(\boldsymbol{X}'\boldsymbol{M}_V^{-1/2}\boldsymbol{K}(x)\boldsymbol{M}_V^{-1/2}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{M}_V^{-1/2}\boldsymbol{K}(x)\boldsymbol{M}_V^{-1/2}\boldsymbol{Y}\right) \qquad (4.61)$$

olarak hesaplanır. $\hat{\boldsymbol{\psi}}_{RE}(x)$ vektörü $q \times 1$ boyutlu $\hat{\boldsymbol{\psi}}_{kRE}(x)$' lerden oluşmaktadır. $\boldsymbol{V}$ matrisi birleşik hata terimlerine ait Varyans-Kovaryans matrisi olmak üzere $\boldsymbol{M}_V$ matrisinin $\boldsymbol{I}_N \otimes \boldsymbol{V}$ biçiminde hesaplandığını ve $\boldsymbol{V}$ matrisinin elemanlarını parametrik rasgele etkili modelin tahmini konusunda ele alınmıştı. $\boldsymbol{K(x)}$ matrisi ise köşegen elemanları $K\left(\frac{x_{it} - x}{h}\right)$ şeklindeki kernel ağırlık fonksiyonu olan köşegen matrisdir.

Bu tahmin yöntemi ile $\boldsymbol{\psi}(x)$ fonksiyonu tahmin edilebileceği gibi alternatif biçimlerde de tahmin yapmak mümkündür. Eşitlik (4.61) ile verilen tahmin edicide $\boldsymbol{M}_V{}^{-1/2}$ matrisi zamanlara göre bireylerin ayrı olarak yazabilmesi için kullanılmakta ve bu ifadenin hesaplanması tahmin işlemini zorlaştırmaktadır. Aynı zamanda $\boldsymbol{\psi}(x)$ tahmin edildikten sonra $m(x)$ in tahmini için parametrik olmayan sabit etkili modeldeki gibi ayrı bir işlem yapılması gerekmektedir. Bu zorlukları yok etmek amacıyla eşitlik (4.60) ile verilen modele farklı bir dönüşüm uygulayarak eşitlik (4.61)'deki tahmin edici ile aynı sonucu verecek bir tahmin edici kullanılır. Bu dönüşümdeki temel amaç (4.61) eşitliğinde kullanılan birleşik hata terimine ait varyans-kovaryans matrisini dönüştürülmüş değişkenlerin içine dahil ederek parametre tahmininde $\boldsymbol{M}_V^{-1/2}$ kullanımını kaldırmaktadır. Bu amaçla eşitlik (4.60) ile verilen model, *i.* bireyin T zaman için ortalamasına göre,

$$\overline{Y}_{i.} = m(x) + (\overline{x}_{i.} - x)\psi(x) + v_{i.}\,, \qquad i = 1,2,...,N \qquad (4.62)$$

şeklinde gösterilir. $\lambda = \hat{\sigma}_u^2 / \left(\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2\right)$ ve $\theta = 1 - \lambda^{1/2}$ olmak üzere eşitlik (4.62) ile verilen model her iki taraftan θ ile çarpılırsa

$$\theta\overline{Y}_{i.} = \theta m(x) + \theta(\overline{x}_{i.} - x)\psi(x) + \theta v_{i.} \tag{4.63}$$

eşitliği elde edilir. Eşitlik (4.60)' da verilen model ile (4.63)' de verilen modelin farkı alınıp gerekli düzenlemeler yapılırsa farkları alınmış yeni model

$$Y_{it} - \theta\overline{Y}_{i.} = (1-\theta)m(x) + (x_{it} - \theta\overline{x}_{i.} - (1-\theta)x)\psi(x) + v_{it} - \theta\overline{v}_{i.} \tag{4.64}$$

eşitliği ile elde edilir. $Y_{it}^* = Y_{it} - \theta\overline{Y}_{i.}$, $x_{kit}^* = x_{kit} - \theta\overline{x}_{ki.}$ ve $x_k{}^* = \lambda^{1/2} x_k$ olmak üzere, $\boldsymbol{Z}^*(x)$ matrisi $\left[\lambda^{1/2} \quad x_{1it}^* - x_1{}^* \cdots x_{qit}^* - x_q{}^*\right]$ biçiminde $NT \times (q+1)$ boyutlu matris olarak tanımlansın. $\boldsymbol{Y}^*$ elemanları Y_{it}^* ' lar olan $NT \times 1$ boyutlu vektör ve $\boldsymbol{v}^*$ elemanları v_{it}^* 'lar olan $NT \times 1$ boyutlu hatalar vektörü olmak üzere $q \geq 1$ için (4.64) eşitliği matris gösterimi ile

$$\boldsymbol{Y}^* = \boldsymbol{Z}^*(x)\boldsymbol{\delta}(x) + \boldsymbol{v}^* \tag{4.65}$$

olarak ifade edilebilir. Burada parametre vektörü olarak alınan $\boldsymbol{\delta}(\boldsymbol{x})$ vektörü açık olarak $\left[m(x) \quad \boldsymbol{\psi}(x)'\right]$ biçimindeki $(q+1)\times 1$ boyutlu vektördür. (4.65) eşitliğindeki model için $\boldsymbol{\delta}(x)$ parametre vektörünün lokal parametrik olmayan rasgele etki tahmin edicisi $\boldsymbol{v}^{*'}\boldsymbol{K}(x)\boldsymbol{v}^*$ veya bu ifade daha açık olarak gösterilirse $\left(\boldsymbol{Y}^* - \boldsymbol{Z}^*(x)\boldsymbol{\delta}(x)\right)' \boldsymbol{K}(x)\left(\boldsymbol{Y}^* - \boldsymbol{Z}^*(x)\boldsymbol{\delta}(x)\right)$ ifadesinin minimum edilmesi ile

$$\hat{\boldsymbol{\delta}}_{RE}(x) = \left(\boldsymbol{Z}^{*\prime}(x)\boldsymbol{K}(x)\boldsymbol{Z}^*(x)\right)^{-1}\left(\boldsymbol{Z}^{*\prime}(x)\boldsymbol{K}(x)\boldsymbol{Y}^*\right) \tag{4.66}$$

eşitliği yardımıyla elde edilir. $m(x)$ ve $\boldsymbol{\psi}(\boldsymbol{x})$ ' in tahminleri ise sırasıyla $\hat{m}_{RE}(x) = \begin{pmatrix} 1 & \mathbf{0}_q' \end{pmatrix} \hat{\boldsymbol{\delta}}_{RE}(x)$ ve $\hat{\boldsymbol{\psi}}_{RE}(x) = \begin{pmatrix} 0 & \mathbf{1}_q' \end{pmatrix} \hat{\boldsymbol{\delta}}_{RE}(x)$ işlemleri ile elde edilir.

Değişkenlere uygulanan bu dönüşümde kullanılan λ ifadesinin hesaplanması için $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ tahminlerinin yapılması gerekmektedir. Bu tahminlerin hesaplanması parametrik rasgele etkili model tahminindeki hesaplama biçimiyle temelde benzerlik göstermektedir. $\lambda = \hat{\sigma}_u^2 / \left(\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2 \right)$ olarak verilen ifadede

$$\hat{\sigma}_u^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{k=1}^{q} \left(Y_{it} - \overline{Y}_{i.} - \hat{\boldsymbol{\psi}}_{kSE}(x) \left(x_{kit} - \overline{x}_{ki.} \right) \right)^2 / NT \tag{4.67}$$

eşitliği yardımıyla $\hat{\sigma}_u^2$ tahmin edilir. Burada $\hat{\boldsymbol{\psi}}_{kSE}(x)$ değerleri (4.52) eşitliğinde verilen sabit etkili model için türev fonksiyonunun tahmin değerleridir. $\hat{\sigma}_\alpha^2$ tahmini için ise

$$\hat{\sigma}_\alpha^2 = \left(\sum_{i=1}^{N} \left(\overline{Y}_{i.} - \tilde{m}(\overline{x}_{i.}) \right)^2 / N \right) - \hat{\sigma}_u^2 / T \tag{4.68}$$

tahmin edicisi kullanılır. $\tilde{m}(\overline{x}_{i.})$ tahmininin yerine ise *i.* bireyler için *T* zamandaki ortalamaları alarak eşitlik (4.48) ile yazılan parametrik olmayan sabit etkili modeldeki $m(x)$ fonksiyonunun tahmini kullanılır.

4.3. Yarıparametrik Regresyonda Panel Verili Modeller

Panel veri analizinde parametrik ve parametrik olmayan modeller kullanılabileceği gibi son yıllarda yaygın olarak zamana bağlı değişkenlerde yarıparametrik modellerin de kullanıldığı görülmüştür. Parametrik modelde zamana bağlı değişkenler kullanıldığında otokorelasyon problemi ortaya çıkabilmektedir. Parametrik olmayan modellerde ise değişken sayısının artması bazı yorum zorluklarına sebep olabilir. Yarıparametrik modelde ise hem parametrik hem de

parametrik olmayan kısım yer aldığından değişken sayısının fazla olması nedeniyle modelin yorumlanması da parametrik olmayan modellerden daha kolay olmaktadır. Aynı zamanda modelde zamana bağlı etkiler parametrik olmayan kısım ile modele dahil edilebilir buda otokorelasyon problemini parametrik modele oranla düşük seviyede tutar. Panel veri kullanımı için yazılabilecek genel yarıparametrik regresyon modeli

$$Y_{it} = x_{kit}\beta_k + \sum_{p=1}^{P} m_p\left(z_{pit}\right) + u_{it} \;, \quad i = 1,2,...,N \;;\; t = 1,2,...,T$$

$$k = 1,2,...,q \;;\; p = 1,2,...,P \qquad (4.69)$$

biçimdedir. Toplamsal olarak yazılan bu eşitliğin yerine z_{it} değişkenlerinin vektör olarak düşünüldüğü

$$Y_{it} = x_{kit}\beta_k + m\left(z_{1it},...,z_{Pit}\right) + u_{it} \qquad (4.70)$$

ifadesi de kullanılabilir. (4.69) eşitliğinde Y_{it} bağımlı değişkenleri, x_{kit} modelin parametrik kısmını oluşturan bağımsız değişkenleri, β_k parametresi x_{kit} değişkenlerine ait parametreleri, $m(\cdot)$ modelin parametrik olmayan kısmındaki bağımsız değişkenler olan z_{pit}' lere ait bilinmeyen formdaki fonksiyonlar ve u_{it}' de hata terimlerini göstermektedir. Yarıparametrik panel verili regresyon modeli için hata terimleri üzerinde temel varsayımlar geçerlidir. Buna göre u_{it} hatalarının sıfır ortalamalı ve sabit σ^2 varyanslı olduğu varsayılır. (4.69) eşitliği matris gösterimi ile

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \sum_{p=1}^{P} m_p(z_{pit}) + \boldsymbol{u} \qquad (4.71)$$

biçiminde yazılabilir. Bu ifade panel verili model için

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{211} & \cdots & x_{q11} \\ x_{112} & x_{212} & \cdots & x_{q12} \\ \vdots & \vdots & \vdots & \vdots \\ x_{11T} & x_{21T} & \cdots & x_{q1T} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1N1} & x_{2N1} & \vdots & x_{qN1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1NT} & x_{2NT} & \cdots & x_{qNT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \\ \vdots \\ \\ \vdots \\ \beta_q \end{bmatrix} + \begin{bmatrix} \sum m_p\left(z_{p11}\right) \\ \sum m_p\left(z_{p12}\right) \\ \vdots \\ \sum m_p\left(z_{p1T}\right) \\ \vdots \\ \sum m_p\left(z_{pN1}\right) \\ \vdots \\ \sum m_p\left(z_{pNT}\right) \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T} \\ \vdots \\ u_{N1} \\ \vdots \\ u_{NT} \end{bmatrix} \qquad (4.72)$$

biçiminde gösterilebilir. Yarıparametrik modelde parametrik ve parametrik olmayan kısımların ayrı olarak tahmin edildiği bilinmektedir. Panel verili yarıparametrik modelde bireylere veya zamana göre değişimler modele parametrik ve parametrik olmayan modeldeki gibi iki farklı biçimde dahil edilir. Bunun sonucunda sabit etkili yarıparametrik regresyon modeli ve rasgele etkili yarıparametrik regresyon modeli olarak iki farklı model elde edilir. Bu modelleri daha ayrıntılı olarak inceleyeceğiz. Ancak her iki modelde de parametrik olan kısmın tahmini için bir dönüşüm uygulanacağından eşitlik (4.69) ile verilen modelin koşullu beklenen değeri alınıp yine aynı şekilde fark alma işlemi uygulanırsa parametrik olmayan kısmında tek değişken olduğu varsayılan panel verili yarıparametrik modeli

$$Y_{it} - E\left(Y_{it}\middle|Z_{it}\right) = \left(x_{kit} - E\left(X_{kit}\middle|Z_{it}\right)\right)\beta_k + u_{it} \qquad (4.73)$$

olarak yazılır. Modelde birden fazla parametrik olmaya kısma ait değişken varsa z_{it} ifadesi vektör olarak düşünülür. $E\left(Y_{it}\middle|,Z_{it}\right)$ ve $E\left(X_{kit}\middle|Z_{it}\right)$ değerleri parametrik olmayan regresyondan elde edilen tahmin değerleridir. Bu tahmin değerleri Nadaraya-Watson kernel tahmini ile yapılabileceği gibi lokal lineer kernel tahmini yardımıyla da bulunabilir. Yani $E\left(Y_{it}\middle|Z_{it}\right)$ değerleri yerine $Y_{it} = m(z_{it}) + u_{it}$ biçiminde verilen modelden tahmin edilecek $\hat{Y}_{it}$ değerleri ve $E\left(X_{kit}\middle|Z_{it}\right)$değerleri yerine ise x_{kit}' lerin ayrı ayrı bağımlı değişken olduğu düşünülerek yazılan $X_{kit} = m(z_{it}) + u_{it}$ modellerinden elde edilen $\hat{X}_{kit}$ tahminleri kullanılır. $Y_{it} - E\left(Y_{it}\middle|Z_{it}\right)$ farkı R_{it}^{yz} , $x_{kit} - E\left(X_{kit}\middle|Z_{it}\right)$ farkı ise R_{kit}^{xz} ile gösterilirse (4.73) eşitliğinde verilen modeli

$$R_{it}^{yz} = R_{kit}^{xz}\beta_k + u_{it} \tag{4.74}$$

olacak şekilde lineer bir formda yazmak mümkündür (Pagan ve Ullah 1999). Bu modelde parametrik kısmın tahmin edicisi olan $\hat{\beta}_k$ değerleri klasik en küçük kareler yöntemi ile tahmin edilebilmektedir. Sabit etkili yarıparametrik panel regresyon modelleri için $\hat{\beta}_k$ değerlerinin hesaplanışı (4.74) eşitliğindeki gibi yapılır fakat rasgele etkili yarıparametrik panel regresyon modellerinde farklı bir tahmin edici kullanılır.

4.3.1. Yarıparametrik sabit etkili panel veri modeli

Birimlerin farklılıklarından meydana gelen değişimi modele sabit bir α_i katsayısı yardımıyla ilave edecek olursak yarıparametrik sabit etkili panel regresyon modeli

$$Y_{it} = \alpha_i + x_{kit}\beta_k + \sum_{p=1}^{P} m_p\left(z_{pit}\right) + u_{it}, \quad i = 1,2,...,N \;;\; t = 1,2,...,T$$
$$k = 1,2,...,q \;;\; p = 1,2,...,P \tag{4.75}$$

şeklinde verildiği gibi olur. Burada α_i sabit etki parametresidir. Modelde daha önceden z_{pit} olarak verdiğimiz parametrik olmayan kısma ait bağımsız değişkenlerin toplam biçiminde yazıldığı durum gösterilmektedir. Eşitlik (4.75) ile verilen model açık matris gösterimi ile

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & \\ 0 & 1 & \vdots \\ \vdots & \vdots & 0 \\ & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \\ \vdots \\ \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} x_{111} & x_{211} & \cdots & x_{q11} \\ x_{112} & x_{212} & \cdots & x_{q12} \\ \vdots & \vdots & \vdots & \vdots \\ x_{11T} & x_{21T} & \cdots & x_{q1T} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1N1} & x_{2N1} & \vdots & x_{qN1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1NT} & x_{2NT} & \cdots & x_{qNT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \\ \vdots \\ \\ \vdots \\ \beta_q \end{bmatrix} + \begin{bmatrix} \sum m_p\left(z_{p11}\right) \\ \sum m_p\left(z_{p12}\right) \\ \vdots \\ \sum m_p\left(z_{p1T}\right) \\ \vdots \\ \sum m_p\left(z_{pN1}\right) \\ \vdots \\ \sum m_p\left(z_{pNT}\right) \end{bmatrix} + \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T} \\ \vdots \\ u_{N1} \\ \vdots \\ u_{NT} \end{bmatrix}$$

biçiminde yazılır. Burada $\boldsymbol{Y}$; $NT \times 1$ boyutlu bağımlı değişken vektörü, $\boldsymbol{X}$; $NT \times q$ boyutlu bağımsız değişkenler matrisi, $\boldsymbol{D} = \boldsymbol{I}_N \otimes \boldsymbol{1}_T$ biçiminde tanımlanan katsayı matrisi, $\boldsymbol{\alpha}$; $N \times 1$ boyutlu sabit etki parametreleri vektörü, $\boldsymbol{\beta}$; $q \times 1$ boyutlu parametre vektörü, $m(\boldsymbol{z})$ parametrik olmayan kısma ait değişkenlerin bilinmeyen formdaki fonksiyonu ve $\boldsymbol{u}$; $NT \times 1$ boyutlu hata terimleri vektörü olarak belirtilirse eşitlik

$$\boldsymbol{Y} = \boldsymbol{D\alpha} + \boldsymbol{X\beta} + m(\boldsymbol{z}) + \boldsymbol{u} \tag{4.76}$$

şeklinde yazılır. Parametrik olmayan kısma ait değişkenler için bilinmeyen $m_p\left(z_{pit}\right)$ fonksiyonları toplam biçiminde yazılmayıp önceden verildiği gibi vektör olarak gösterilmiştir. (4.75) eşitliği

$$Y_{it} = \alpha_i + x_{kit}\beta_k + m\left(z_{it}\right) + u_{it}, \qquad \begin{matrix} i = 1,2...,N \;\; ; \;\; t = 1,2,...,T \;\; ; \\ k = 1,2,...,q \end{matrix} \tag{4.77}$$

olarak da yazılabilir. $m\left(z_{it}\right)$' nin z_{pit} değişkenlerine ait ortak olasılık fonksiyonu olduğu varsayılırsa, (4.76) eşitliğindeki $\boldsymbol{z}$ ' nin parametrik olmayan kısımdaki değişkenlerin bir vektörü olduğu söylenebilir.

4.3.2. Yarıparametrik sabit etkili panel veri modelinde tahmin

Yarıparametrik regresyon tahmininde modelin parametrik olan kısmının ve parametrik olmayan kısmının ayrı tahmin edildiğini söylemiştik. Sabit etkili modelde de öncelikle parametrik olan kısım tahmin edilir ve ardından parametrik kısmın tahmini yardımıyla parametrik olmayan kısım tahmin edilir. Modelin parametrik olan kısmının tahmini için daha önce yazılan (4.74)'deki modeli göz önüne alalım. α_i sabit etki parametresinin fark alma işlemi sırasında kaybolacağı düşünülürse α_i sabit etki parametresinin varlığı parametrik kısmın tahmininde bir değişikliğe yol açmayacaktır. (4.74) modelinin klasik en küçük kareler tahmin edicisi sabit etkili model için

$$\hat{\boldsymbol{\beta}}_{YPSE} = \left(\boldsymbol{R}'_{xz}\boldsymbol{R}_{xz}\right)^{-1}\left(\boldsymbol{R}'_{xz}\boldsymbol{R}_{yz}\right) \tag{4.78}$$

eşitliği ile elde edilir. burada $\boldsymbol{R}_{xz}$ ile gösterilen matris elemanları R_{it}^{xz} olan ve $x_{kit} - E\left(X_{kit}\middle|Z_{it}\right)$ farklarından oluşan $NT \times q$ boyutlu matrisdir. $\boldsymbol{R}_{yz}$ vektörü ise elemanları R_{it}^{yz} ile gösterilen $Y_{it} - E\left(Y_{it}\middle|Z_{it}\right)$ farklarından oluşan $NT \times 1$ boyutlu vektördür. Tahmin sonucunda elimizde $q \times 1$ boyutlu yarıparametrik regresyon modelinin parametrik kısmındaki x_{kit} ' lere ait $\hat{\boldsymbol{\beta}}_{YPSE}$ parametre tahmin vektörü olacaktır. Parametrik olmayan kısma ait z_{pit} bağımsız değişkenlerinin bilinmeyen bir fonksiyonu olan $m\left(z_{it}\right)$ fonksiyonunun tahmini elde edilen $\hat{\boldsymbol{\beta}}_{YPSE}$ tahminleri yardımıyla hesaplanır. Bu amaçla oluşturulan yeni dönüştürülmüş model

$$Y_{it}^{*} = \alpha_i + m\left(z_{it}\right) + u_{it}, \qquad i = 1,2,...,N \;\; ; \;\; t = 1,2,...,T \tag{4.79}$$

olarak yazılsın. Burada $Y_{it}^{*} = Y_{it} - \sum_{k=1}^{q} x_{kit}\hat{\beta}_{kYPSE}$ farkı ile elde edilen yeni dönüştürülmüş bağımlı değişkenlerdir (Kneisner ve Li 1996). Böylece (4.79) eşitliği (4.46)' da verilen parametrik olmayan sabit etkili model şekline dönüşmüş olur.

Parametrik olmayan sabit etkili modelde yapılan dönüşümler (4.79) eşitliğindeki modele de uygulanırsa

$$Y^{*}_{it} - \overline{Y}^{*}_{i.} = (z_{it} - \overline{z}_{i.})\psi(z) + u_{it} - \overline{u}_{i.} \tag{4.80}$$

olarak elde edilir. Burada $\psi(z)$, $m(z)$ fonksiyonunun birinci türevi olup parametrik olmayan kısımda birden fazla değişken olması durumunda $\boldsymbol{\psi}_p(z_{pit})$'lerden oluşan $\boldsymbol{\psi}(\boldsymbol{z})$ vektörünün tahmin edicisini (4.52) eşitliğine benzer şekilde

$$\hat{\boldsymbol{\psi}}_{YPSE}(z) = \left(\boldsymbol{Z}'\boldsymbol{M}_D\boldsymbol{K}(z)\boldsymbol{M}_D\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{M}_D\boldsymbol{K}(z)\boldsymbol{M}_D\boldsymbol{Y}^{*}\right) \tag{4.81}$$

olarak elde etmek mümkündür. Bu tahmin ediciyi elde etmenin diğer bir yolu parametrik olmayan sabit etkili model tahmininde anlatılmıştı. Bu bölümde detaylı olarak anlatılmayacaktır ancak eşitlik (4.80) ile verilen modeldeki farkları alınmış değişkenler ile yazılacak yeni modele lokal lineer kernel yöntemindeki ağırlıklandırılmış en küçük kareler yöntemi uygulanırsa (4.55) eşitliğinde verilen tahmin ediciye benzer olarak $\hat{\boldsymbol{\psi}}_{YPSE}$ tahmin edicisi elde edilebilir. Türev fonksiyonunun tahmini elde edildikten sonra $m(z)$ sabit fonksiyonunun tahmini ve α_i sabit etki parametrelerinin tahmini yine parametrik olmayan sabit etkili regresyondaki (4.56) ve (4.57) eşitliklerindeki tahminler yardımıyla hesaplanır. Bu tahmin ediciler eşitlik (4.79) ile verilen model için

$$\hat{m}(z) = \breve{Y}^{*}_{..} - \sum_{p=1}^{P}\left(\breve{z}_{p..} - z_p\right)\hat{\boldsymbol{\psi}}_{pYPSE}(z) \tag{4.82}$$

ve

$$\hat{\alpha}_i = \breve{Y}^{*}_{i.} - \hat{m}(z) - \sum_{p=1}^{P}\left(\breve{z}_{pi.} - z_p\right)\hat{\boldsymbol{\psi}}_{pYPSE}(z) \tag{4.83}$$

olarak yazılır ve $\hat{m}(z)$ ve $\hat{\alpha}_i$ tahminlerinde kullanılan $\breve{Y}^{*}_{..}$ değeri dönüştürülmüş bağımlı değişkenin tüm birey ve zamanlar için ortalaması, $\breve{Y}^{*}_{i.}$ değeri dönüştürülmüş

bağımlı değişkenin her bir $i.$ birey için ortalaması , $\breve{z}_{p..}$ değeri $p.$ bağımsız değişkenin tüm birey ve zamanlar için ortalaması, $\breve{z}_{pi.}$ değeri de $p.$ bağımsız değişkenin her bir $i.$ birey için ortalamasıdır (Roy 1997).

4.3.3. Yarıparametrik rasgele etkili panel veri modeli

Panel verilerin yarıparametrik regresyonda kullanılmasının diğer yolu daha önce parametrik ve parametrik olmayan modellerde gösterildiği gibi rasgele etkili model kullanılmasıdır. α_i etkilerinin sabit değil de rasgele olduğu düşünülürse yarıparametrik regresyon modelinin hata terimi olan u_{it} ' lere dahil edilmesiyle oluşan yeni birleşik hata terimi

$$v_{it} = u_{it} + \alpha_i \ , \qquad i = 1,2,...,N \ ; \ t = 1,2,...,T \tag{4.84}$$

olarak düşünülür ve yarıparametrik rasgele etkili panel regresyon modeli parametrik olmayan kısmının toplam biçiminde yazıldığı durumda

$$Y_{it} = x_{kit}\beta_k + \sum_{p=1}^{P} m_p\left(z_{pit}\right) + v_{it} \ , \qquad i = 1,2,...,N \ ; \ t = 1,2,...,T$$
$$k = 1,2,...,q \ ; \ p = 1,2,...,P \tag{4.85}$$

olarak elde edilir. $m(z_{it})$ ile gösterilen fonksiyonunun, z_{pit} değişkenlerinin ortak olasılık fonksiyonu olduğu varsayıldığı durumda model

$$Y_{it} = x_{kit}\beta_k + m(z_{it}) + v_{it} \tag{4.86}$$

şeklinde gösterilir. Modele sabit bir μ parametresi dahil etmek mümkündür. Modele dâhil edilecek μ sabit parametresi $\hat{\beta}_k$ ve $\hat{m}_p\left(z_{pit}\right)$ tahminlerinin bulunuşunu değiştirmeyecektir. (4.85) eşitliğindeki model açık olarak

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{211} & \cdots & x_{q11} \\ x_{112} & x_{212} & \cdots & x_{q12} \\ \vdots & \vdots & \vdots & \vdots \\ x_{11T} & x_{21T} & \cdots & x_{q1T} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1N1} & x_{2N1} & \vdots & x_{qN1} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1NT} & x_{2NT} & \cdots & x_{qNT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \\ \vdots \\ \\ \vdots \\ \beta_q \end{bmatrix} + \begin{bmatrix} \sum m_p\left(z_{p11}\right) \\ \sum m_p\left(z_{p12}\right) \\ \vdots \\ \sum m_p\left(z_{p1T}\right) \\ \vdots \\ \sum m_p\left(z_{pN1}\right) \\ \vdots \\ \sum m_p\left(z_{pNT}\right) \end{bmatrix} + \begin{bmatrix} v_{11} \\ v_{12} \\ \vdots \\ v_{1T} \\ \vdots \\ v_{N1} \\ \vdots \\ v_{NT} \end{bmatrix} \tag{4.87}$$

matris çarpımı ile ifade edilebilir ve

$$\boldsymbol{Y} = \boldsymbol{X\beta} + m(\boldsymbol{z}) + \boldsymbol{v} \tag{4.88}$$

olarak gösterilir. burada $\boldsymbol{Y}$; bağımlı değişkenlere ait $NT \times 1$ boyutlu vektör, $\boldsymbol{X}$;parametrik kısımdaki bağımsız değişkenlerin $NT \times q$ boyutlu matrisi, $\boldsymbol{\beta}$; $q \times 1$ boyutlu parametre vektörü, $m(\boldsymbol{z})$; parametrik olmayan kısma ait değişkenlerin ortak olasılık fonksiyonu ve $\boldsymbol{v}$; hata terimlerine ait $NT \times 1$ boyutlu vektördür.

4.3.4. Yarıparametrik rasgele etkili panel veri modelinde tahmin

Önceki rasgele etkili modellerde olduğu gibi yarıparametrik rasgele etkili panel veri modelinde de temel amaç iki bileşenden oluşan v_{it} hata terimlerinin varyans-kovaryans yapısını göz önüne alarak ağırlıklandırılmış tahmin edicinin bulunmasıdır. Yarıparametrik modelde parametrik kısım öncelikle tahmin edildiği için parametrik kısmın tahmininde de v_{it}' lerin varyans-kovaryans yapısına göre ağırlıklandırılmış en küçük kareler yöntemi kullanılmalıdır. Yarıparametrik panel verili regresyon başlığı altında yapılan dönüşüm sonucu elde edilen eşitlik (4.74)'deki modeli rasgele etkili modele uyarlarsak hata terimindeki değişimi göz önüne alarak rasgele etkili modeli

$$R_{it}^{yz} = R_{kit}^{xz}\beta_k + v_{it} \tag{4.89}$$

olarak yazmak mümkündür. Bu modelde R_{it}^{yz} gösterimi; elemanları $Y_{it} - E\left(Y_{it}\middle|Z_{it}\right)$ farkları olan dönüştürülmüş bağımlı değişkeni, R_{kit}^{xz} gösterimi ise elemanları $x_{kit} - E\left(X_{kit}\middle|Z_{it}\right)$ farkları olan dönüştürülmüş bağımsız değişkenleri temsil etmektedir. Eşitlik (4.74)'den farklı olan ise birleşik hata terimlerinin varlığıdır. Modelden $\hat{\beta}_k$ ların elde edilmesi için ağırlıklandırılmış en küçük kareler tahmin edicisi

$$\hat{\boldsymbol{\beta}}_{YPRE} = \left(\boldsymbol{R}_{it}^{xz*\prime} \boldsymbol{R}_{it}^{xz*} \right)^{-1} \left(\boldsymbol{R}_{it}^{xz*\prime} \boldsymbol{R}_{it}^{yz*} \right) \tag{4.90}$$

eşitliği ile elde edilir. $\boldsymbol{R}_{it}^{xz*}$, elemanları R_{kit}^{xz*}' lardan oluşan $NT \times q$ boyutlu matrisidir. $\hat{\boldsymbol{\beta}}_{YPRE}$ tahmini elemanları $\hat{\beta}_{kYPRE}$' lar olan vektördür. Eşitlikte kullanılan $\boldsymbol{R}^{*yz}_{it}$ ve $\boldsymbol{R}^{*xz}_{it}$ matrisleri ağırlıklandırma işlemini doğrudan veriler üzerinden yapılmasını sağlayan dönüşümle elde edilmiştir. Bu matrislerin elemanları sırasıyla

$$\begin{aligned} R_{it}^{yx*} &= R_{it}^{yz} - \left(1 - \lambda^{1/2}\right) \overline{R}_{i.}^{yz} \\ R_{kit}^{xz*} &= R_{kit}^{xz} - \left(1 - \lambda^{1/2}\right) \overline{R}_{ki.}^{xz} \end{aligned} \tag{4.91}$$

işlemleri ile bulunur (Roy 1997). $\overline{R}_{i.}^{yz}$ ve $\overline{R}_{ki.}^{xz}$ ifadeleri *i.* birey için T zamanın hepsine ait R_{it}^{yz} ve R_{kit}^{xz} dönüştürülmüş verilerinin ortalama değerleridir. λ ise parametrik olmayan rasgele etkili panel veri modelinde verilen

$$\lambda = \hat{\sigma}_u^2 / \left(\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2 \right) \tag{4.92}$$

eşitliği ile hesaplanır. Yarıparametrik rasgele etkili panel veri modelinde kullanılacak λ değerinin hesaplanması bölüm sonunda verilmiştir. Parametrik olan kısma ait $\hat{\boldsymbol{\beta}}_{YPRE}$ parametresi tahmin edildikten sonra eşitlik (4.86) ile verilen modeldeki $m\left(z_{it}\right)$

fonksiyonunun tahmin edilmesi işlemine geçilir. Bu amaçla yarıparametrik sabit etkili modelde yapılan dönüşüm (4.86) eşitliğindeki modele uygulanırsa model

$$Y_{it}^{*} = m(z_{it}) + v_{it} \tag{4.93}$$

şeklinde yazılır. Burada $Y_{it}^{*} = Y_{it} - \sum_{k=1}^{q} x_{kit}\hat{\beta}_{kYPRE}$ farkı ile elde edilen dönüştürülmüş bağımlı değişkenlerdir. Eşitlik (4.93)' da verilen modele parametrik olmayan rasgele etkili modelin tahmini konusunda açık olarak gösterilen dönüşüm uygulanırsa eşitlik (4.93) ile verilen model (4.64) eşitliğine benzer bir şekilde

$$Y_{it}^{*} - \theta\overline{Y}_{i.}^{*} = (1-\theta)m(z) + (z_{it} - \theta\overline{z}_{i.} - (1-\theta)z)\psi(z) + v_{it} - \theta\overline{v}_{i.} \tag{4.94}$$

veya $P \geq 1$ olması durumunda matris gösterimi ile

$$\boldsymbol{Y}^{**} = \boldsymbol{Z}^{*}(z)\boldsymbol{\delta}(z) + \boldsymbol{v}^{*} \tag{4.95}$$

olarak elde edilir. $\boldsymbol{Y}^{**}$; elemanları Y_{it}^{**}' lar olan $NT \times 1$ boyutlu vektör ve $Y_{it}^{**} = Y_{it}^{*} - \theta\overline{Y}_{i.}^{*}$ biçiminde tanımlanır. $\boldsymbol{Z}^{*}(z)$ matrisi; elemanları $\left[\lambda^{1/2} \quad z_{1it}^{*} - z_{1}^{*} \cdots \quad z_{Pit}^{*} - z_{P}^{*}\right]$ olan $NT \times (P+1)$ boyutlu matris olarak tanımlansın. $z_{it}^{*} = z_{it} - \theta\overline{z}_{i.}$ ve $z^{*} = \lambda^{1/2}z$ olacak şekilde yazılır. $\boldsymbol{v}^{*}$ ise elemanları v_{it}^{*}' lar olan $NT \times 1$ boyutlu hatalar vektörüdür. $\overline{z}_{pi.}$ ve $\overline{Y}_{i.}^{*}$ değerleri $i.$ birey için T zamanın ortalamasıdır. $\overline{z}_{p..}$ ve $\overline{Y}_{..}^{*}$ gösterimleri, tüm birey ve zamanlara ait ortalamalardır. Burada parametre vektörü olarak alınan $\boldsymbol{\delta}(z)$ vektörü açık olarak $\left[m(z) \quad \boldsymbol{\psi}(z)'\right]$ biçimindeki $(P+1)\times 1$ boyutlu vektördür. $\boldsymbol{\delta}(z)$ parametre vektörünün lokal parametrik olmayan rasgele etki tahmin edicisi hesaplamak için $\left(\boldsymbol{Y}^{**} - \boldsymbol{Z}^{*}(z)\boldsymbol{\delta}(z)\right)' \boldsymbol{K}(\boldsymbol{z})\left(\boldsymbol{Y}^{**} - \boldsymbol{Z}^{*}(z)\boldsymbol{\delta}(z)\right)$ ifadesini minimum yapan değer

$$\hat{\boldsymbol{\delta}}_{\boldsymbol{YPRE}}(\boldsymbol{z}) = \left(\boldsymbol{Z}^{*\prime}(z)\boldsymbol{K}(z)\boldsymbol{Z}^{*}(z)\right)^{-1}\left(\boldsymbol{Z}^{*\prime}(z)\boldsymbol{K}(z)\boldsymbol{Y}^{*}\right) \tag{4.96}$$

eşitliği yardımıyla elde edilir. $\boldsymbol{K}(z)$ matrisi köşegen elemanları $K\left(\frac{z_{it}-z}{h}\right)$ biçimindeki köşegen matristir. $m(z)$ ve $\boldsymbol{\psi}(z)$' in tahminleri ise sırasıyla $\hat{m}_{YPRE}(z) = \begin{pmatrix}1 & \boldsymbol{0}_p\end{pmatrix}\hat{\boldsymbol{\delta}}_{YPRE}(z)$ ve $\hat{\boldsymbol{\psi}}_{YPRE}(z) = \begin{pmatrix}0 & \boldsymbol{1}_p\end{pmatrix}\hat{\boldsymbol{\delta}}_{YPRE}(z)$ işlemleri ile elde edilir.

Değişken dönüşümleri yapılırken kullanılacak λ değerinin yerine tahmin edicisi olarak $\hat{\lambda}$ kullanılır. $\hat{\lambda}$ ise daha önce verildiği gibi $\hat{\lambda} = \hat{\sigma}_u^2 / \left(\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2\right)$ biçiminde hesaplanmaktadır. $\hat{\sigma}_u^2$ tahmini

$$\hat{\sigma}_u^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(R_{it}^{yz} - \overline{R}_{i.}^{yz} - \sum_{k=1}^{q}\hat{\beta}_{kSE}\left(R_{kit}^{xz} - \overline{R}_{ki.}^{xz}\right)\right)^2 / NT \tag{4.97}$$

eşitliği yardımıyla tahmin edilir. Burada $\hat{\beta}_{kSE}$ değeri $R_{it}^{yz} = \alpha_i + R_{kit}^{xz}\beta_k + u_{it}$ modelinden elde edilecek parametre tahmin değeridir. $\hat{\sigma}_\alpha^2$ değeri ise

$$\hat{\sigma}_\alpha^2 = \left(\sum_{i=1}^{N}\left(\overline{R}_{i.}^{yz} - \sum_{k=1}^{q}\hat{\beta}_{kG}R_{ki.}^{xz}\right)^2 \Big/ N\right) - \hat{\sigma}_u^2 / T \tag{4.98}$$

olacak şekilde bulunur. $\hat{\beta}_{kG}$ elemanlarının oluşturduğu $\hat{\boldsymbol{\beta}}_G$ vektörü ise $\overline{R}_{i.}^{yz} = \alpha_i + \overline{R}_{ki.}^{xz}\beta_{kG} + u_{it}$ modelinin en küçük kareler tahmini ile bulunan değerdir(Roy 1997).

Yarıparametrik rasgele etkili panel veri modelin tahmininde dönüştürülmüş değişkenleri kullanmadan ağırlıklandırma işlemini doğrudan tahmin ediciye uygulamak mümkündür. Parametrik olmayan rasgele etkili panel veri modelinde

dönüşüm yapılmadan kullanılan tahmin ediciye benzer olarak; $\boldsymbol{V}$ matrisi birleşik hata terimlerine ait Varyans-Kovaryans matrisi olmak üzere, $\boldsymbol{I}_N \otimes \boldsymbol{V}$ biçiminde hesaplanan $\boldsymbol{M}_V$ matrisi yardımıyla ilk olarak parametrik kısma ait $\hat{\boldsymbol{\beta}}_{YPRE}$ parametre vektörü tahmin edilir. $\boldsymbol{R}_{it}^{yz*} = \boldsymbol{M}_V^{-1/2}\,\boldsymbol{R}_{it}^{yz}$ ve $\boldsymbol{R}_{it}^{xz*} = \boldsymbol{M}_V^{-1/2}\,\boldsymbol{R}_{it}^{xz}$ olmak üzere $\hat{\boldsymbol{\beta}}_{YPRE}$ tahmin edicisi,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{YPRE}} = \left(\boldsymbol{R}_{it}^{xz*\prime}\, \boldsymbol{R}_{it}^{xz*} \right)^{-1} \left(\boldsymbol{R}_{it}^{xz*\prime}\, \boldsymbol{R}_{it}^{yz*} \right) \tag{4.99}$$

eşitliği yardımıyla tahmin edilir. $\boldsymbol{R}_{it}^{xz*}$ matrisi, elemanları R_{kit}^{xz*} değişkenlerinden oluşan $NT \times q$ boyutlu matrisdir. $Y_{it}^{*} = Y_{it} - \sum_{k=1}^{q} x_{kit}\hat{\beta}_{kYPRE}$ farkı ile elde edilen dönüştürülmüş bağımlı değişkenler kullanılarak yazılan $Y_{it}^{*} = m(z_{it}) + v_{it}$ modeline lokal lineer kernel tahmini uygulanır. $m(z_{it})$ fonksiyonunun birinci türevi olan $\boldsymbol{\psi}(z)$ fonksiyonunun tahmini değişken sayısının birden fazla olması durumunda yani $P \geq 1$ için

$$\hat{\boldsymbol{\psi}}_{YPRE}(z) = \left(\boldsymbol{Z}'\boldsymbol{M}_V^{-1/2}\boldsymbol{K}(z)\boldsymbol{M}_V^{-1/2}\boldsymbol{Z} \right)^{-1} \left(\boldsymbol{Z}'\boldsymbol{M}_V^{-1/2}\boldsymbol{K}(z)\boldsymbol{M}_V^{-1/2}\boldsymbol{Y}^{*} \right) \tag{4.100}$$

eşitliği ile hesaplanabilir. Burada $\boldsymbol{Z}$ parametrik olmayan kısma ait bağımsız değişkenlerin $NT \times P$ boyutlu tasarım matrisi, $\boldsymbol{Y}^{*}$ dönüştürülmüş bağımlı değişkenlerin vektörü, $\boldsymbol{K}(z)$ köşegen elemanları $K\left(\frac{z_{it} - z}{h} \right)$ şeklindeki kernel ağırlık fonksiyonu olan köşegen matristir. $\boldsymbol{\psi}(z)$ fonksiyonunun tahmini yapıldıktan sonra $m(z_{it})$ fonksiyonunun tahmini

$$\hat{m}(z) = \breve{Y}^{*}_{..} - \sum_{p=1}^{P} \left(\breve{z}_{p..} - z_p \right) \hat{\boldsymbol{\psi}}_{pYPRE}(z) \tag{4.101}$$

olarak yazılır. $\breve{Y}^{*}_{..}$ değeri dönüştürülmüş bağımlı değişkenin tüm birey ve zamanlar için ortalaması, $\breve{Y}^{*}_{i.}$ değeri dönüştürülmüş bağımlı değişkenin her bir $i.$ birey için ortalaması , $\breve{z}_{p..}$ değeri $p.$ bağımsız değişkenin tüm birey ve zamanlar için ortalaması, $\breve{z}_{pi.}$ değeri de $p.$ bağımsız değişkenin her bir $i.$ birey için ortalamasıdır.

5. UYGULAMA

Uygulama ana başlığı altında iki alt konu ele alınmıştır. Birinci kısımda medyana dayalı parametrik olmayan tahmin yöntemleri başlığında incelenen Brown-Mood tahmin yöntemi ile doğrusal ilişkiye sahip olmayan veri setleri için önerilen yeni tahmin yöntemi karşılaştırılmıştır. Karşılaştırma yapmak amacıyla bağımlı ve bağımsız değişken arasındaki ilişkinin farklı şekillerde olduğu durumlar ele alınmış ve her bir durum için tahmin yönteminin uyumu incelenmiştir.

İkinci kısımda 4. bölümde bahsedilen Panel veri analizlerinden parametrik, parametrik olmayan ve yarıparametrik modellerin her biri için sabit etkili ve rasgele etkili modeller oluşturulmuş ve tahminler hesaplanmıştır. Modeller yardımıyla 2006 yılından başlayarak 2009 da zirve noktasına ulaşan Avrupa krizi incelenmiştir. Ülkelerin ekonomisinde etkili olan ve krizden en çok etkilenen temel göstergeler bağımsız değişken olarak alınmış ve bunlara karşılık ülkelerin büyüme yüzdeleri bağımlı değişken olarak tahmin edilmiştir. Modele krizden Avrupa'da en çok etkilenen ülkeler ile Türkiye'de alınmıştır.

Verilerin analizi için Selçuk Stat, SPSS 13, Maple 11 paket programları ve Delphi programlama dili kullanılmıştır.

5.1. Medyana Dayalı Parametrik Olmayan Tahmin Yöntemlerinin Simülasyon Yardımıyla Karşılaştırılması

Uygulamanın birinci kısmında basit regresyon modeli için uygulanan parametrik olmayan tahmin yöntemlerinden medyana dayalı yöntemlerin karşılaştırılması yapılmak istenmiştir. Parametrik olmayan regresyonda medyana dayalı tahmin yöntemlerinden en çok kullanılan yöntem olan Brown-Mood yöntemi ile bu çalışmada öne sürülen medyan yöntemini karşılaştırmak amacı ile 6 farklı veri seti simülasyon yardımıyla elde edilmiştir. Veri setleri 25 adet gözlem çiftinden oluşmaktadır. Veri seti oluşturulurken karşılaşılabilecek farklı durumlar düşünülerek her bir duruma uygun veri setleri oluşturulmuştur. Bağımlı değişken Y_i ler ile

bağımsız değişken x_i lerin arasındaki ilişkinin yapısı verilerden elde edilen serpilme diyagramlarında görülmektedir. Uygulamada kullanılan veriler tablo olarak verilmiştir.

Çizelge 2. Medyan yöntemi için farklı biçimlerdeki veri setleri

Durum 1		Durum 2		Durum 3		Durum 4		Durum 5		Durum 6	
Y	X	Y	X	Y	X	Y	X	Y	X	Y	X
3.094	1.764	5.264	1.219	0.765	1.379	2.345	1.069	4.377	2.616	2.159	1.454
4.314	1.800	7.189	1.222	1.064	1.464	3.138	1.724	3.913	2.950	2.944	1.773
3.666	2.153	9.195	1.631	2.857	1.740	3.664	2.716	4.946	3.275	3.529	2.441
3.682	2.293	7.753	1.941	3.567	2.016	3.278	3.236	4.075	3.850	3.308	2.936
4.820	2.553	8.873	2.127	3.576	2.539	4.917	3.936	3.201	3.876	3.233	3.086
3.368	2.665	8.279	2.296	2.087	2.522	7.297	4.062	3.251	4.930	3.905	3.186
4.317	3.350	7.497	3.049	4.757	3.070	5.934	5.147	3.392	5.325	4.117	3.218
4.430	3.402	8.256	3.167	3.834	3.319	8.938	6.347	3.549	6.394	4.480	4.583
2.988	3.910	9.573	3.362	4.692	3.654	9.491	7.137	2.868	6.554	5.060	4.812
3.539	4.011	10.561	3.467	8.015	4.106	9.362	7.793	2.875	6.200	5.101	5.257
3.319	4.322	9.625	3.752	6.905	4.235	12.184	6.651	2.390	7.177	5.600	5.416
4.340	4.881	8.836	3.820	6.848	4.508	12.388	8.759	1.894	7.345	5.678	5.907
3.027	4.945	9.517	4.192	7.556	4.936	11.433	8.971	1.927	7.922	6.535	6.002
4.178	4.959	8.841	4.321	8.425	5.361	10.880	9.107	2.080	8.338	5.374	6.871
2.715	5.767	11.591	4.388	8.826	5.226	9.827	10.896	2.188	9.181	5.474	7.958
4.245	6.048	12.482	4.488	11.252	5.800	12.017	11.314	2.319	9.233	5.437	8.133
4.932	6.082	12.995	4.561	10.065	6.086	12.543	12.602	2.320	9.709	4.748	8.369
3.783	6.583	11.612	5.092	12.713	6.002	12.658	13.615	2.894	10.174	3.875	8.934
4.588	6.890	12.874	5.140	15.718	6.107	14.342	13.643	2.463	10.300	3.827	9.446
4.227	7.408	10.887	5.268	14.447	6.326	13.349	14.620	4.382	10.555	3.695	10.108
2.558	7.451	12.359	5.327	14.760	6.197	12.380	17.663	4.187	10.717	3.196	10.322
4.099	7.550	13.226	5.760	15.306	6.337	14.347	18.039	3.431	11.153	3.620	10.462
2.850	7.619	11.633	6.094	16.834	6.462	13.250	19.867	3.406	12.368	4.280	10.499
3.758	8.284	11.897	6.392	16.613	6.601	12.647	21.596	4.033	13.116	2.199	11.479
4.608	9.073	13.677	6.723	17.270	6.207	14.385	16.014	5.085	13.157	2.997	12.857

Çizelge 2'de verilen 6 farklı veri setinin serpilme diyagramları çizilmiştir. Her birinin serpilme diyagramları sırasıyla verilmiştir. İlk iki durum için

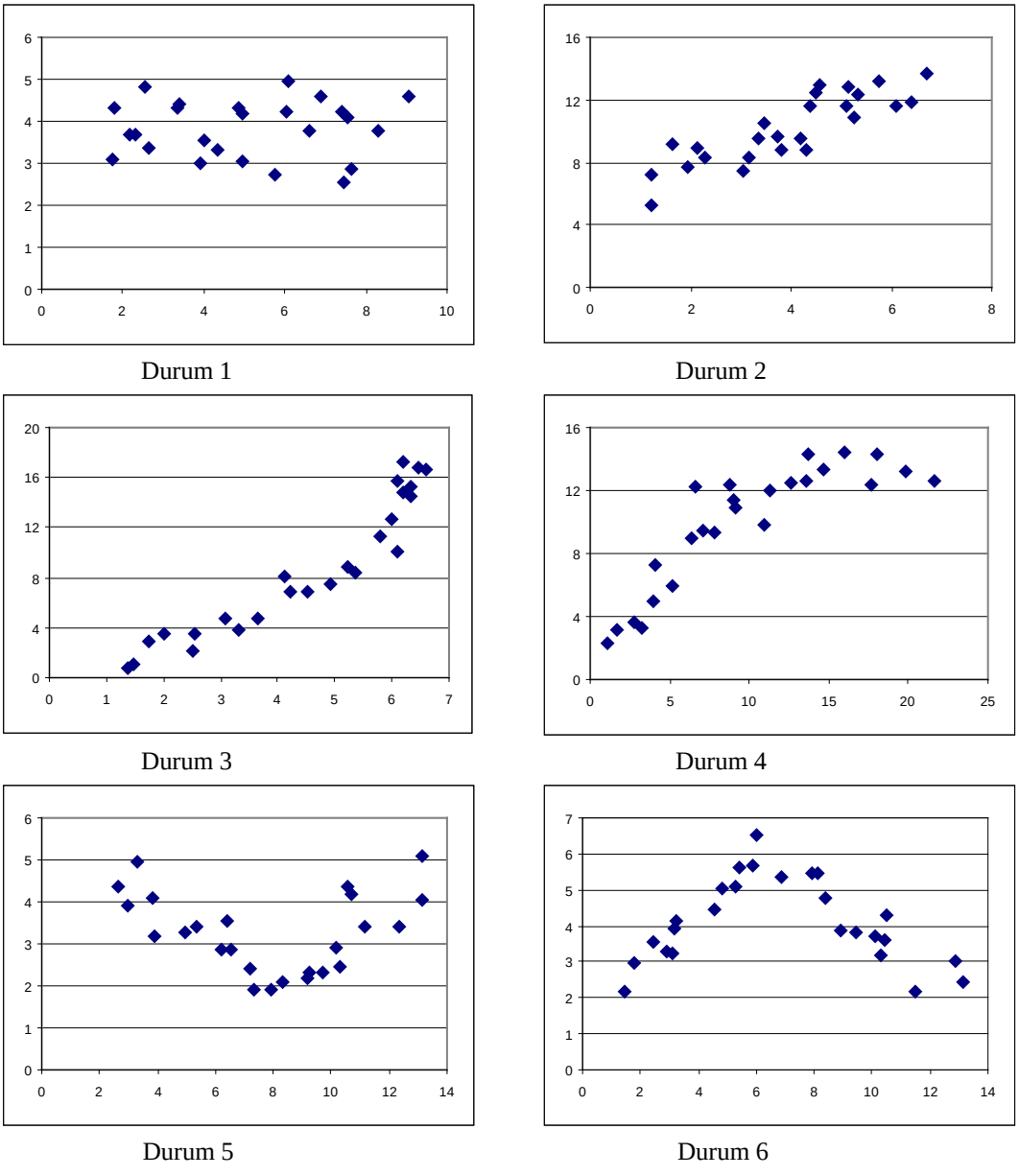
$$Y_i = \mu + \beta x_i + \varepsilon_i, \qquad i = 1,..., N \tag{5.1}$$

eşitliği ile verilen basit regresyon modeli kurulabilir. Y_i bağımlı değişkeni ile x_i bağımsız değişkeni arasında doğrusal bir ilişkinin olduğu görülmektedir. Bu durumda ilk iki model için literatürde yer alan Brown-Mood veya Theil yöntemi kullanılabilir.

Ancak geriye kalan 4 durum için aynı modelin kurulması çok uygun olmayacaktır. Bağımlı değişken ile bağımsız değişken arasındaki ilişkinin doğrusal olmadığı ve bu durumlar için

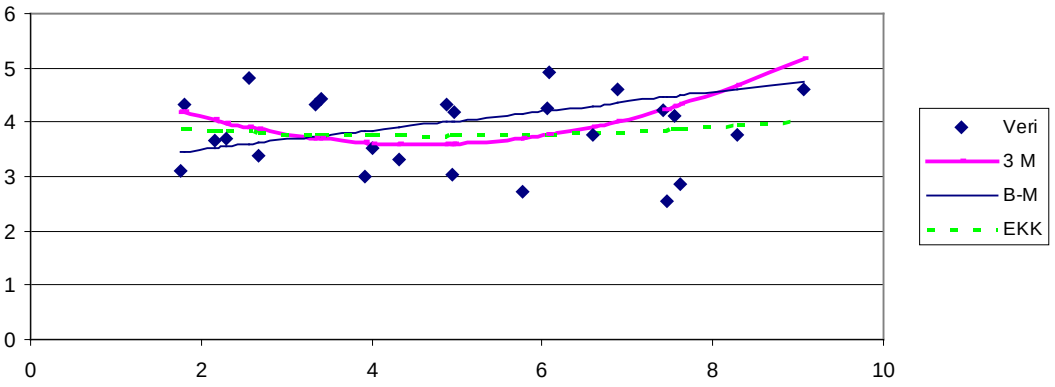
$$Y_i = \mu + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i \quad , \qquad i = 1,...,N \tag{5.2}$$

biçimindeki parabolik modelin kullanılmasının daha uygun olacağı görülmektedir. (5.2) eşitliğindeki modelin tahmini için Brown-Mood yöntemi yerine bu çalışmada önerilen medyan tahmini yönteminin daha iyi sonuç verdiği gösterilmiştir.



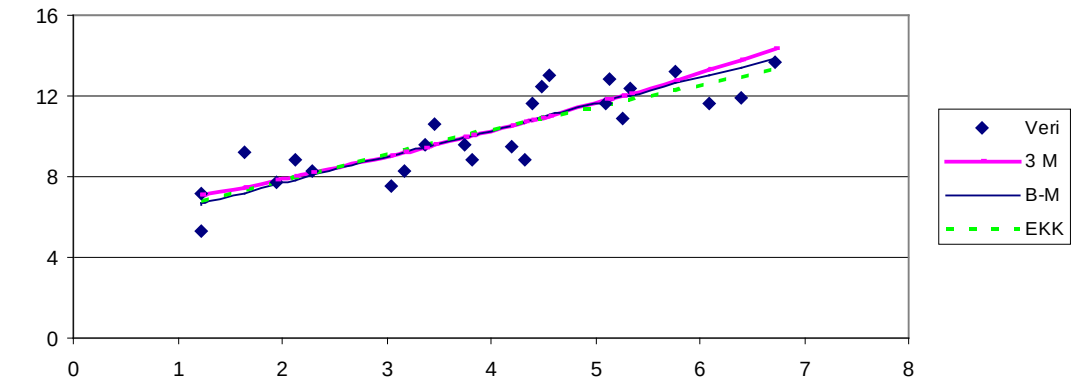
Şekil 11. Farklı ilişki yapılarına sahip veri setlerinin serpilme diyagramları

Şekil 11'de verilen her durum için (5.1) eşitliği ile verilen basit doğrusal modelin tahmini Brown-Mood yöntemi yardımıyla yapıldı. Bununla beraber yine her bir durum için (5.2) eşitliği ile verilen modelin tahmini bu çalışmada önerilen medyan yöntemi yardımıyla elde edildi. Her bir durum için (5.2) modelinin klasik en küçük kareler tahminleri karşılaştırma yapmak amacı ile elde edildi. Sırasıyla her bir durum için elde edilen parametre tahminleri, hata kareler ortalamaları ve tahmin edilen model serpilme diyagramı üzerinde gösterildi. (5.1) eşitliğindeki model için literatürde yer alan medyan tahmini yöntemlerinden Theil yöntemi serpilme diyagramlarının rahat görünebilmesi için hesaplanmamıştır. Theil yöntemi kullanılarak Brown-Mood yönteminde elde edilen doğrulara yakın sonuçlar elde edilmektedir. Bu yüzden karşılaştırma açısından bir faydası olmayacaktır. Her bir durum için elde edilen sonuçlar şekillerde görüldüğü gibidir.



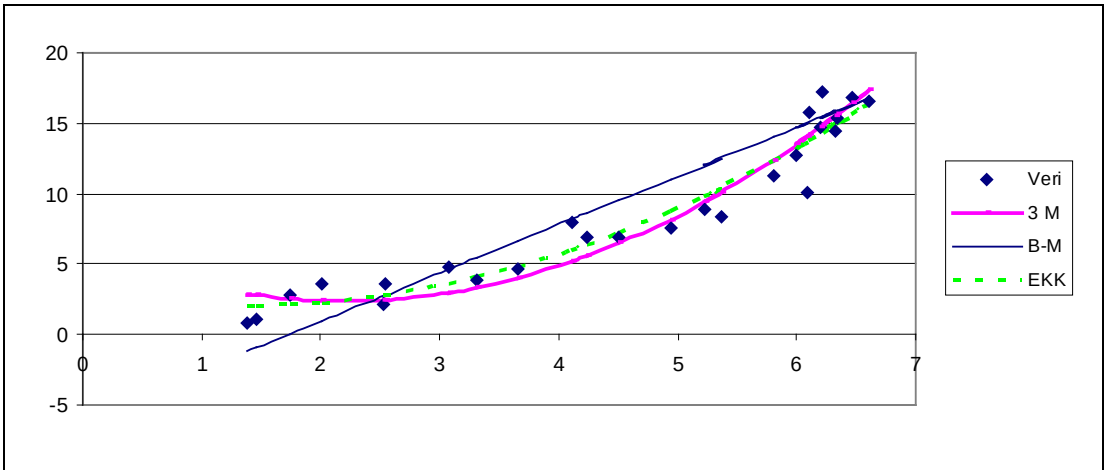
	3 parçalı medyan	**Brown-Mood**	**EKK**
Durum 1	$\hat{\mu}$ =5.197, $\hat{\beta}_1$ =-0.710, $\hat{\beta}_2$ =0.078 **HKO=** 0.570	$\hat{\mu}$ =3.141, $\hat{\beta}$ = 0.177 **HKO**= 0.629	$\hat{\mu}$ =4.078, $\hat{\beta}_1$ =-0.137, $\hat{\beta}_2$ =0.014 **HKO**= 0.449

İlk grafikte bağımlı değişken ve bağımsız değişken arasındaki ilişkinin doğrusal bir yapıda olduğu görülmektedir. Bu durumda kullanılan 3 metot hata kareler ortalaması baz alındığında birbirine yakın sonuçlar vermiştir.

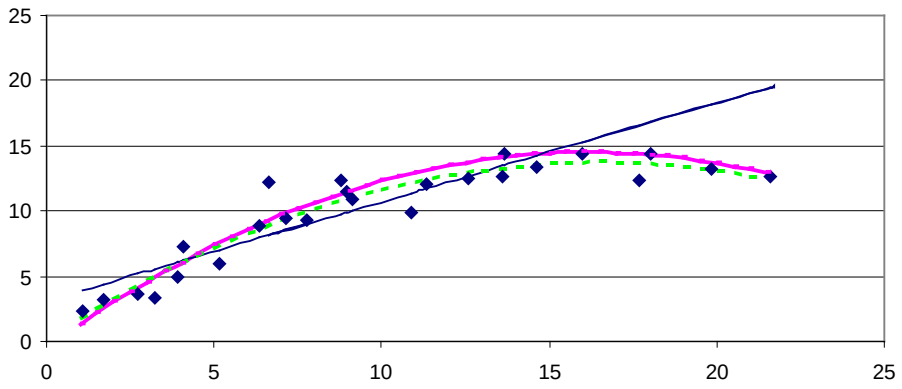


	3 parçalı medyan	Brown-Mood	EKK
Durum 2	$\hat{\mu}$ =5.998, $\hat{\beta}_1$ =0.802, $\hat{\beta}_2$ =0.065 **HKO=** 1.349	$\hat{\mu}$ =5.051, $\hat{\beta}$ = 1.312 **HKO**= 1.276	$\hat{\mu}$ =5.118, $\hat{\beta}_1$ =1.417, $\hat{\beta}_2$ =-0.029 **HKO**= 1.228

İkinci durumda değişkenler arasında doğrusal ve aynı yönde bir ilişki söz konusudur. Bu durumda da her üç metot hata kereler ortalamasına göre yakın sonuçlar vermiştir. Grafikte de görüldüğü gibi tahminler neredeyse çakışmaktadır.

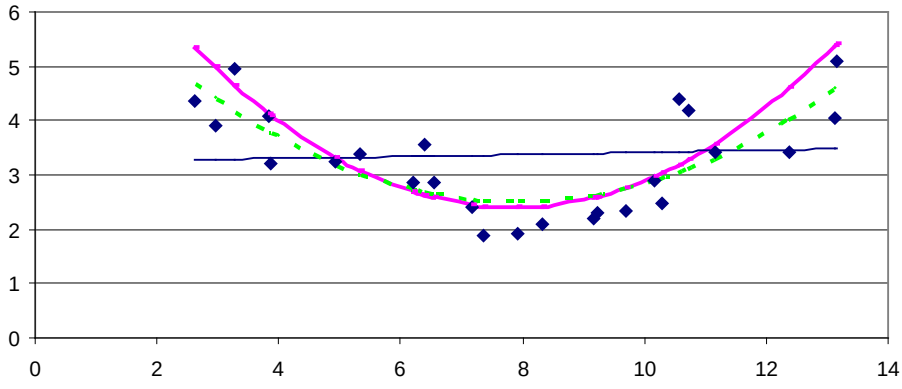


	3 parçalı medyan	Brown-Mood	EKK
Durum 3	$\hat{\mu}$ =6.031, $\hat{\beta}_1$ =-3.369, $\hat{\beta}_2$ =0.770 **HKO=** 2.242	$\hat{\mu}$ =-5.959, $\hat{\beta}$ = 3.446 **HKO**= 4.907	$\hat{\mu}$ =3.017, $\hat{\beta}_1$ =-1.413, $\hat{\beta}_2$ =0.520 **HKO**= 1.901

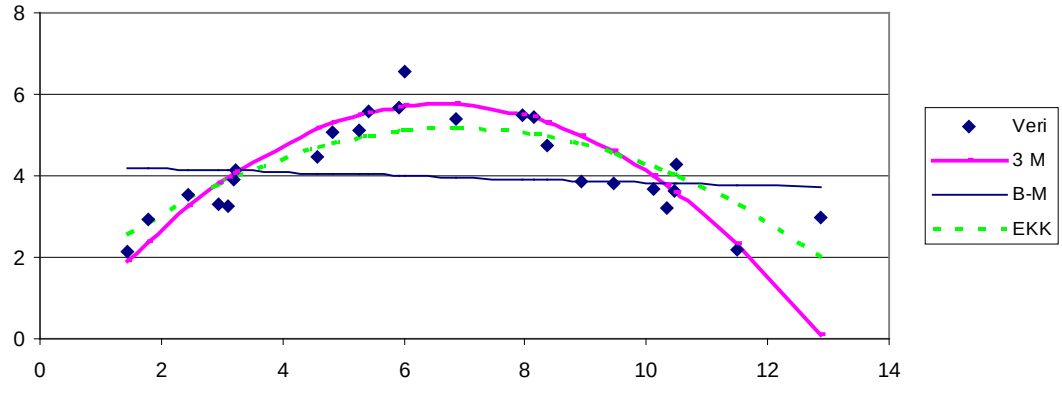


	3 parçalı medyan	Brown-Mood	EKK
Durum 4	$\hat{\mu}$ =-0.448, $\hat{\beta}_1$ =1.844, $\hat{\beta}_2$ =-0.056 **HKO=** 1.595	$\hat{\mu}$ =3.120, $\hat{\beta}$ = 0.758 **HKO**= 5.754	$\hat{\mu}$ =0.148, $\hat{\beta}_1$ =1.649, $\hat{\beta}_2$ =-0.050 **HKO**= 1.287

3. ve 4. durumlarda veriler daha geniş bir aralıkta ve az eğime sahip olarak üretilmiştir. Her üç yönteme ait tahminler ve hata kareler ortalamaları grafiklerde verilmiştir. Farklı olarak 5. ve 6. durumlarda ise daha küçük aralıkta ve eğimi fazla olarak üretilmiştir.



	3 parçalı medyan	Brown-Mood	EKK
Durum 5	$\hat{\mu}$ =9.028, $\hat{\beta}_1$ =-1.690, $\hat{\beta}_2$ =0.107 **HKO=** 0.441	$\hat{\mu}$ =3.223, $\hat{\beta}$ = 0.019 **HKO**= 0.871	$\hat{\mu}$ =7.320, $\hat{\beta}_1$ =-1.216, $\hat{\beta}_2$ =0.076 **HKO**= 0.314



	3 parçalı medyan	Brown-Mood	EKK
Durum 6	$\hat{\mu}$ =-0.578, $\hat{\beta}_1$ =1.916, $\hat{\beta}_2$ =-0.145 **HKO=** 0.568	$\hat{\mu}$ =4.252, $\hat{\beta}$ = -0.042 **HKO**= 1.308	$\hat{\mu}$ =0.962, $\hat{\beta}_1$ =1.223, $\hat{\beta}_2$ =-0.089 **HKO**= 0.368

3,4,5 ve 6. durumlar için tahminlere bakılırsa Brown-Mood yöntemi ile elde doğrusal modelin hata kareler ortalamasının çalışmada önerilen medyan yöntemine göre çok yüksek olduğu görülmektedir. Veriler serpilme diyagramında doğrusal olmayan bir ilişki içinde olduğundan Brown-Mood yöntemi ile tahmin edilecek doğru uygun sonuç vermemiştir. Bu durumda önerilen medyan yöntemi verilerin yapısına uygun bir eğri tahmin ettiğinden daha iyi sonuç vermiştir. 3 parçalı medyan yöntemi her bir durum için parametrik EKK yöntemine yakın sonuçlar vermiştir. Medyan yönteminin uygulanması için herhangi bir varsayım bulunmaması parametrik olmayan bu tahmin yönteminin EKK tahmin yöntemine göre bir avantajı olarak kabul edilir. Bağımlı değişken ile bağımsız değişken arasındaki ilişkinin doğrusal olmadığı durumlarda, parametrik tahmin yöntemlerinin varsayımları sağlanmadığı zaman Brown-Mood yönteminden daha iyi sonuç elde etmek için önerilen medyan yöntemi kullanılabilir.

5.2. Panel Veri Analizi İle 2009 Avrupa Ekonomik Krizinin Modellenmesi

Bu bölümde parametrik, parametrik olmayan ve yarıparametrik modeller yardımıyla panel veri analizleri yapılması amaçlanmıştır. Uygulamada gerçek veri seti kullanılmıştır. Veriler Uluslararası Para Fonu (IMF) tarafından yayınlanan World Economic Outlook (WEO) veri tabanından alınmıştır. Panel veri seti, Avrupa ekonomik krizinden en çok etkilenen beş Avrupa ülkesi ile birlikte Türkiye' ye ait 2004 - 2009 yılları arasındaki ekonomik göstergeler alınarak oluşturulmuştur. Alınan ülkeler; Portekiz, İspanya, İtalya, Yunanistan, İrlanda ve Türkiye'dir. Bu ülkelere ait büyüme yüzdeleri bağımlı değişken olarak alınıp tahmin edilmeye çalışılmıştır. Büyümeyi etkileyen faktörler olarak da sırasıyla Enflasyon, İşsizlik ve Cari İşlemler Dengesinin Gayri Safi Yurt İçi Hasıla'ya (GSYİH) oranı alınmıştır. Büyüme yüzdesini etkileyen en önemli faktörlerden enflasyon ve işsizlik, yıllık yüzde değeri olarak alınmıştır. Basit olarak, bir ülkede zaman içerisinde üretilen mal ve hizmet miktarının artmasına iktisadi büyüme denilmektedir. Büyümeyi etkileyen bir çok gösterge mevcuttur. Enflasyonun büyüme üzerini etkisi ekonomi alanında 1950 yıllarından günümüze çok tartışılmaktadır. Literatürde enflasyonun büyümeyi kısa dönemde olumlu etkilediğini savunan görüşlerle birlikte hem kısa hem de uzun vadede enflasyonun büyümeye olumsuz yansıdığını savunan görüşlerde mevcuttur. Bu durumda büyüme için önemli bir ekonomik gösterge olarak kabul edilmektedir(Artan 2008). Bir diğer ekonomik gösterge olan işsizlik, iktisat teorisinde kurulan bütün modellerin içine dolaylı veya direk olarak dahil edilmiş ve analizlerde kullanılmıştır (Yılmaz 2005). Büyüme kişilerin reel gelirlerinin sürekli artması anlamına geldiğinden işsizlik göstergesi büyüme hızını dolaylı veya direk olarak etkilemektedir. Bir diğer gösterge olan cari işlemler dengesi bir ülkenin döviz gelirleri ile döviz giderleri arasındaki fark olarak açıklanır ve bu dengenin GSYİH ye oranı önemli bir ekonomik gösterge olarak kabul edilir. Bu oranın -5,-6 düzeyinde olması ekonomistler tarafından kritik eşik kabul edilir ve üzerine çıkması istenmez. Bu orandaki değişimler iktisadi büyüme verilerini yeterince etkilemektedir. Bu amaçla büyüme üzerinde etkileri olan bu üç değişken modele alınmıştır. Değişkenlerin aldığı değerler Panel veri tablosu biçiminde hazırlanmıştır.

Çizelge 3. Ekonomik göstergeler için panel veri tablosu

Ülkeler ($i.$)	Yıllar ($t.$)	Değişkenler			
		Büyüme(%)	Enflasyon(%)	İşsizlik(%)	Cari İşlemler Dengesi/GSYİH
		Y	X1	X2	X3
Portekiz	2004	1.5	2.5	7	-7.6
	2005	0.9	2.1	7.6	-9.5
	2006	1.4	3	7.7	-10
	2007	1.9	2.4	8	-9.4
	2008	0	2.7	7.6	-12.1
	2009	-3	-0.6	9.5	-9.9
İspanya	2004	3.3	3.1	10.9	-5.3
	2005	3.6	3.4	9.2	-7.4
	2006	4	3.6	8.6	-9
	2007	3.6	2.8	8.3	-10
	2008	0.9	4.1	11.3	-9.6
	2009	-3.8	-0.3	18.2	-6
İtalya	2004	1.5	2.3	9.1	-0.9
	2005	0.7	2.2	8.1	-1.7
	2006	2	2.2	7.8	-2.6
	2007	1.6	2	6.1	-2.4
	2008	-1	3.5	6.8	-3.4
	2009	-5.1	0.7	9.1	-2.5
Yunanistan	2004	4.9	3	9.7	-5.8
	2005	2.9	3.5	9.9	-7.3
	2006	4.5	3.3	9.5	-11.1
	2007	4	3	8.3	-14.2
	2008	2.9	4.2	7.6	-14.4
	2009	-0.8	1.1	9.5	-10
İrlanda	2004	4.6	2.3	5.3	-0.6
	2005	6.2	2.2	4.3	-3.5
	2006	5.4	2.7	4.1	-3.6
	2007	6	2.9	4.5	-5.3
	2008	-3	3.1	6.1	-5.2
	2009	-7.5	-1.6	12	-1.7
Türkiye	2004	9.4	8.6	10.3	-3.7
	2005	8.4	8.2	10.2	-4.6
	2006	6.9	9.6	9.9	-6
	2007	4.7	8.8	10.3	-5.8
	2008	0.9	10.4	11	-5.7
	2009	-6.5	6.2	14.8	-1.9

Panel veri tablosunda görüldüğü gibi $i = 1,...,6$ olmak üzere *i.* indisi 6 adet ülkeyi, $t = 1,...,6$ olmak üzere *t.* indisi de 6 farklı zamanı göstermektedir.

Parametrik Panel Verili Regresyon Modeli

İlk olarak panel verilerde en çok kullanılan parametrik model incelenecektir. Parametrik model sabit etkili ve rasgele etkili model olmak üzere iki farklı biçimde oluşturulmakta idi. Panel veri analizi için parametrik sabit etkili panel veri modeli

$$Y_{it} = \alpha_i + x_{1it}\beta_1 + x_{2it}\beta_2 + x_{3it}\beta_3 + u_{it} \ , \qquad i = 1,...,6 \ ; \ t = 1,...,6$$

biçiminde kurulur. Modelin tahmini için gerekli olan M_D matrisi $M_D = I_6 \otimes Q_6$ biçiminde hesaplanır. Q_6 ile ifade edilen 6×6 boyutlu matris

$$Q_6 = \begin{bmatrix} 0.833 & -0.166 & -0.166 & -0.166 & -0.166 & -0.166 \\ -0.166 & 0.833 & -0.166 & -0.166 & -0.166 & -0.166 \\ -0.166 & -0.166 & 0.833 & -0.166 & -0.166 & -0.166 \\ -0.166 & -0.166 & -0.166 & 0.833 & -0.166 & -0.166 \\ -0.166 & -0.166 & -0.166 & -0.166 & 0.833 & -0.166 \\ -0.166 & -0.166 & -0.166 & -0.166 & -0.166 & 0.833 \end{bmatrix}$$

olarak hesaplanmıştır. M_D matrisi ise 36×36 boyutlu köşegen elemanları Q_6 matrisinden oluşan

$$M_D = \begin{bmatrix} Q_6 & 0 & \cdots & 0 \\ 0 & Q_6 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & Q_6 \end{bmatrix}_{36\times 36}$$

biçimindeki matristir. Tahminler yapıldıktan sonra her bir ülke için hesaplanan $\hat{\alpha}_i$ değerleri ve modelin $\hat{\beta}_k$ tahminleri tabloda verilmiştir.

Çizelge 4. Parametrik sabit etkili modelin parametre tahminleri

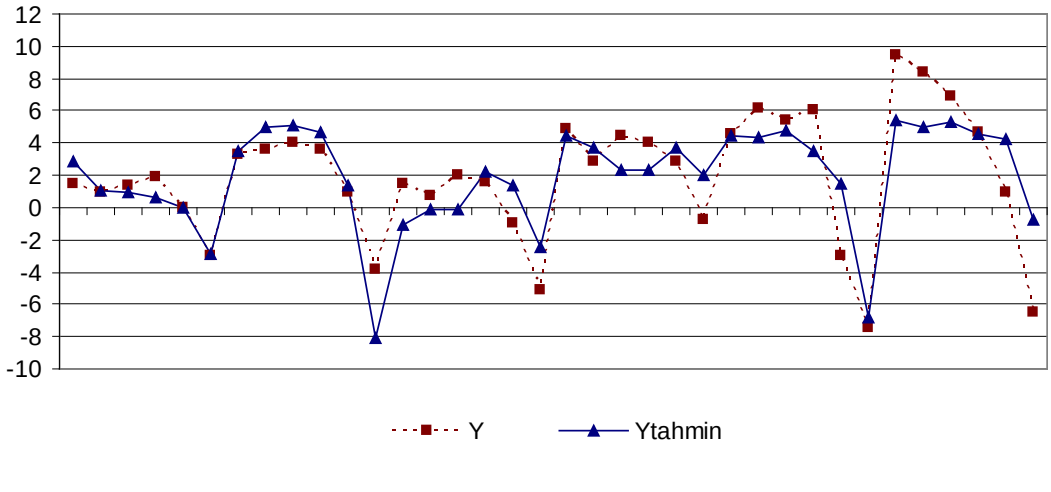
i	$\hat{\alpha}_i$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	15.082	0.395	-1.360	0.481
2	19.694			
3	10.834			
4	19.258			
5	11.007			
6	17.848			

Çizelgede özellikle $i = 2$ ve $i = 4$ için sabit etki parametresinin değer olarak diğerlerinden yüksek çıktığı ve $i = 3$ için düşük çıktığı görülmektedir. 2 ve 4 indisine sahip ülkeler sırasıyla İspanya ve Yunanistan'dır. Bu çalışmada büyümeyi etkileyen faktörler olarak işsizlik, enflasyon ve cari işlem dengesinin GSYİH ye oranı seçilmişti. Bu sonuca göre bu ilki ülke için diğerlerine oranla bu faktörlerin yanında büyüme oranını etkileyen daha farklı faktörlerin olduğu söylenebilir. 3 indisine sahip ülke İtalya'dır. İtalya'nın diğer ülkelere göre daha güçlü bir ekonomik yapısı olduğundan diğer faktörlerden fazla etkilenmemekte ve sabit etki parametresi daha küçük çıkmaktadır. Türkiye ise 6 indisine sahip ülkedir ve İspanya ile Yunanistan'dan sonra en yüksek değerli sabit etki parametresine sahip ülkedir.

Çizelge 5. Parametrik sabit etkili panel veri modeli $\hat{Y}_{it}$ tahminleri

i	t	$\hat{Y}_{it}$	i	t	$\hat{Y}_{it}$	i	t	$\hat{Y}_{it}$
1	1	2.898	3	1	-1.064	5	1	4.420
1	2	1.011	3	2	-0.129	5	2	4.347
1	3	0.990	3	3	-0.153	5	3	4.768
1	4	0.634	3	4	2.176	5	4	3.486
1	5	-0.002	3	5	1.335	5	5	1.437
1	6	-2.831	3	6	-2.465	5	6	-6.759
2	1	3.549	4	1	4.465	6	1	5.459
2	2	4.970	4	2	3.669	6	2	5.005
2	3	5.096	4	3	2.308	6	3	5.293
2	4	4.707	4	4	2.331	6	4	4.529
2	5	1.334	4	5	3.660	6	5	4.257
2	6	-8.056	4	6	1.968	6	6	-0.742

Çizelge 5'de tahmin sonuçları verilmiştir. Tahmin sonuçlarına göre özellikle 2009 yılına ait verilerde büyük sapmalar yaşanmıştır. $\hat{Y}_{it}$ tahmin değerleri ile Y_{it} değerleri grafik üzerinde gösterilirse



Model $Y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it}$	**HKO** = 4.783

Şekil 12. Parametrik sabit etkili panel veri model grafiği

Şekil 12' de tahmin değerleri ile gözlem değerlerinin birlikte grafiği gösterilmiştir. Özellikle 2009 yılına ait veriler için tahminlerin daha kötü çıktığı görülebilir. Hata kareler ortalaması parametrik sabit etkili panel veri modeli için 4.783 olarak hesaplanmıştır.

Parametrik modelde panel veri kullanımının bir diğer yolu rasgele etkili model oluşturmaktır. α_i parametresinin rasgele olduğu düşünülerek hata terimine eklenmesi ile oluşturulan parametrik rasgele etkili panel veri modeli

$$Y_{it} = x_{1it}\beta_1 + x_{2it}\beta_2 + x_{3it}\beta_3 + v_{it} \qquad i = 1,...,6 \quad t = 1,...,6$$

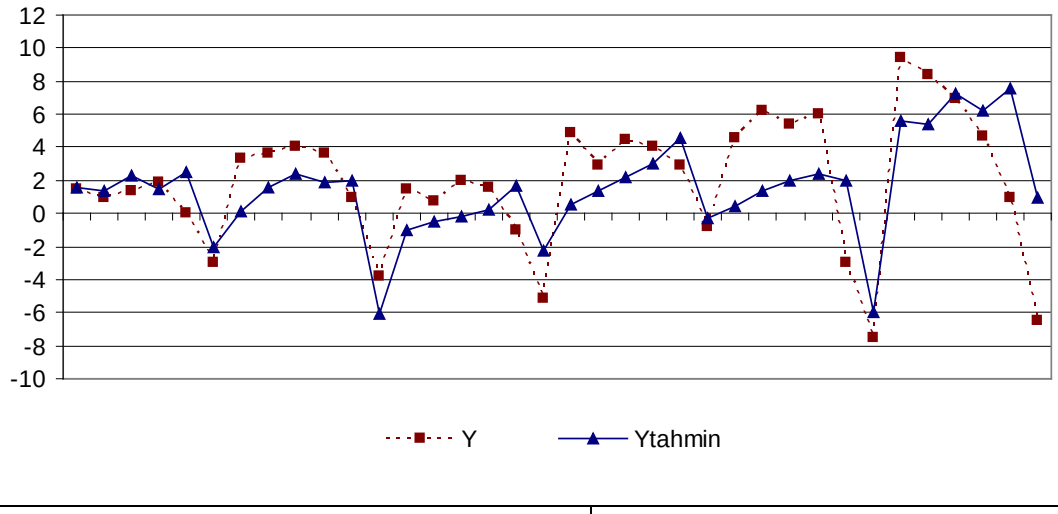
biçimindedir. Modelin tahminini dönüşüm uygulayarak bulmak istersek öncelikle $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ varyanslarının tahmin edilmesi gerekir. Tahminler için gerekli olan $\hat{\beta}_{FE}$ sabit etkili modelin parametre tahminleri $\hat{\beta}_1$ =0.395, $\hat{\beta}_2$ =-1.360, $\hat{\beta}_3$ = 0.481 olarak hesaplanmıştı. $\hat{\beta}_G$ gruplar arası modelden elde edilen tahmin değerleri sırasıyla $\hat{\beta}_{1G}$ =0.5214, $\hat{\beta}_{2G}$ = -0.1056 ve $\hat{\beta}_{3G}$ =-0.146 olarak hesaplanmış ve bunlara bağlı olarak $\hat{\mu}_G$ = 0.1003 olarak bulunmuştur. Bulanan değerler ile $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ tahminleri; $\hat{\sigma}_u^2$ =6.377, $\hat{\sigma}_\alpha^2$ = 0.995 olarak tahmin edilmiştir. $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ değerleri kullanılarak dönüşümü yapmakta kullanacağımız θ parametresi; θ=0.281 olarak hesaplanmıştır. Dönüşüm uygulanmış değişkenler ile yapılan EKK tahmini sonucunda bulunan rasgele etkili modele ait parametre tahminleri sırasıyla, $\hat{\beta}_1$ = 1.018 , $\hat{\beta}_2$ = -0.392 ve $\hat{\beta}_3$ = -0.229 olarak elde edilmiştir.

Parametrelerin tahmin değerleri yardımıyla elde edilen modele bağlı olarak hesaplanan $\hat{Y}_{it}$ değerleri ile dönüştürülmüş değişkenler $Y_{it}^*, x_{1it}^*, x_{2it}^*, x_{3it}^*$ değerleri çizelge 6'da verilmiştir.

Çizelge 6. Parametrik rasgele etkili panel veri modeline ilişkin değerler

i	t	Y_{it}^{*}	x_{1it}^{*}	x_{2it}^{*}	x_{3it}^{*}	$\hat{Y}_{it}$
1	1	1.373	1.933	4.777	-4.857	1.543
1	2	0.773	1.533	5.377	-6.757	1.337
1	3	1.273	2.433	5.477	-7.257	2.329
1	4	1.773	1.833	5.777	-6.657	1.462
1	5	-0.127	2.133	5.377	-9.357	2.544
1	6	-3.127	-1.167	7.277	-7.157	-2.067
2	1	2.756	2.317	7.781	-3.082	0.097
2	2	3.056	2.617	6.081	-5.182	1.551
2	3	3.456	2.817	5.481	-6.782	2.357
2	4	3.056	2.017	5.181	-7.782	1.890
2	5	0.356	3.317	8.181	-7.382	1.945
2	6	-4.344	-1.083	15.081	-3.782	-6.069
3	1	1.514	1.695	6.896	-0.267	-1.021
3	2	0.714	1.595	5.896	-1.067	-0.547
3	3	2.014	1.595	5.596	-1.967	-0.223
3	4	1.614	1.395	3.896	-1.767	0.194
3	5	-0.986	2.895	4.596	-2.767	1.677
3	6	-5.086	0.095	6.896	-1.867	-2.284
4	1	4.037	2.151	7.144	-2.855	0.580
4	2	2.037	2.651	7.344	-4.355	1.355
4	3	3.637	2.451	6.944	-8.155	2.180
4	4	3.137	2.151	5.744	-11.255	3.057
4	5	2.037	3.351	5.044	-11.455	4.599
4	6	-1.663	0.251	6.944	-7.055	-0.313
5	1	4.051	1.756	3.598	0.333	0.401
5	2	5.651	1.656	2.598	-2.567	1.356
5	3	4.851	2.156	2.398	-2.667	1.967
5	4	5.451	2.356	2.798	-4.367	2.404
5	5	-3.549	2.556	4.398	-4.267	1.957
5	6	-8.049	-2.144	10.298	-0.767	-5.948
6	1	8.284	6.171	7.181	-2.401	5.566
6	2	7.284	5.771	7.081	-3.301	5.405
6	3	5.784	7.171	6.781	-4.701	7.269
6	4	3.584	6.371	7.181	-4.501	6.252
6	5	-0.216	7.971	7.881	-4.401	7.584
6	6	-7.616	3.771	11.681	-0.601	0.944

Parametrik rasgele etkili panel veri modelinde en yüksek değerli sapmalar Yunanistan için yapılan tahminler sonucunda oluşmuştur. Tahmin değerleri ile gözlem değerlerini birlikte grafik yardımıyla gösterilirse Şekil 13'de ki gibi görülür.



Model $Y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + v_{it}$	**HKO** = 8.732

Şekil 13. Parametrik rasgele etkili panel veri modeli grafiği

Grafikte özellikle Yunanistan ve Türkiye için tahminlerde hataların yüksek olduğu görülmektedir. Modelin genel olarak hata kareler ortalaması 8.732 olarak hesaplanmıştır. Hata kareler ortalaması sabit etkili modele göre daha yüksek çıkmıştır.

Parametrik Olmayan Panel Verili Regresyon Modeli

Parametrik olmayan regresyonda Panel veri analizi için kullanılan ilk model olan sabit etkili panel veri modeli

$$Y_{it} = \alpha_i + m(x_{1it}, x_{2it}, x_{3it}) + u_{it}, \qquad i = 1,...,6 \quad ; \quad t = 1,...,6$$

biçiminde kurulmuştur. Parametrik olmayan modelin tahmini lokal lineer kernel tahmini ile yapılmıştır. Kernel fonksiyonu olarak standart normal kernel fonksiyonu seçilmiştir. h bant genişliği seçimi için normal referans kuralı kullanılmıştır. Normal referans kuralı seçimi için verilere ait 1. ve 3. çeyreklikler ile standart sapmalar ve hesaplanan h bant genişlikleri tabloda verilmiştir.

Çizelge 7. h bant genişlikleri

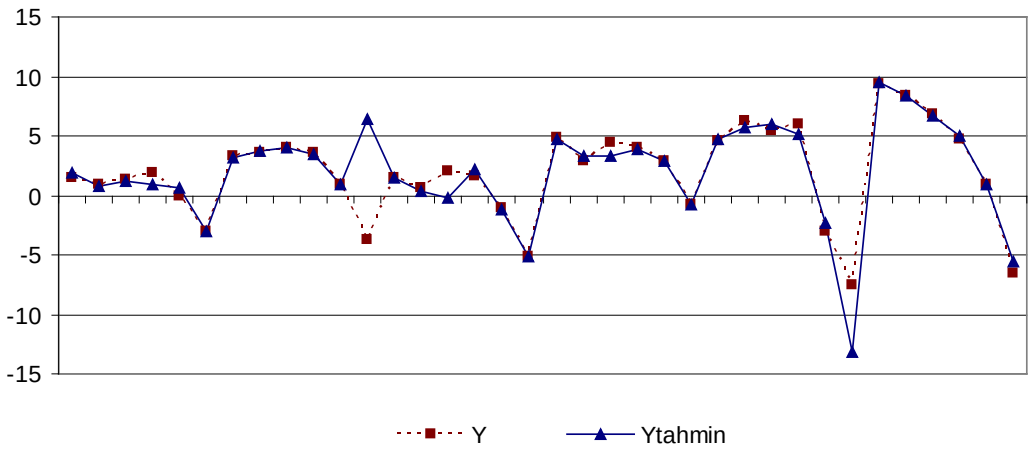
	$Q_{.25}$	$Q_{.75}$	S	$\hat{Q}$	$\hat{Q}/1.34$	h
x_1	2.20	3.55	2.71	1.35	1.01*	0.58
x_2	7.60	10.05	2.75	2.45	1.83*	1.06
x_3	-9.55	-3.45	3.72*	6.10	4.55	2.16

Bulunan h bant genişlikleri ve parametrik olan sabit etkili modelin elde edilmesinde bulunan M_D matrisi yardımıyla parametrik olmayan sabit etkili panel veri modelinin lokal lineer kernel tahmini elde edilmiştir. Hesaplanan $\hat{\alpha}_i$ ve $\hat{m}(x_{1it}, x_{2it}, x_{3it})$ tahminlerinin bağımsız değişkenlerin değerleri yerine koyulduğunda aldıkları değerler ve $\hat{Y}_{it}$ tahminleri Çizelge 8'de verilmiştir.

Sabit etkili parametrik olmayan panel verili model kullanımında 2009 yılına ait gözlemlerin tahmin değerleri kötü sonuçlar vermiş özellikle İspanya'nın 2009 yılı büyüme oranının tahmini diğer ülkelerin tahminlerine göre çok kötü çıkmıştır. Ancak genel olarak bakıldığında parametrik sabit etkili panel veri modelinden daha iyi sonuçlar elde edilmiştir. Elde edilen $\hat{Y}_{it}$ tahminleri ile Y_{it} değerlerinin grafiği ise Şekil 14'de verilmiştir.

Çizelge 8. Parametrik olmayan sabit etkili panel veri modeli hesaplamaları

i	t	$\hat{\alpha}_i$	$\hat{m}(x_{1it}, x_{2it}, x_{3it})$	$\hat{Y}_{it}$
1	1	-1.891	3.829	1.938
1	2	0.385	0.418	0.802
1	3	-0.346	1.563	1.217
1	4	0.219	0.643	0.862
1	5	-0.957	1.569	0.611
1	6	-9.527	6.526	-3.001
2	1	2.847	0.360	3.207
2	2	2.559	1.212	3.772
2	3	2.174	1.771	3.944
2	4	2.438	0.962	3.400
2	5	2.159	-1.221	0.938
2	6	1.974	4.384	6.358
3	1	2.347	-0.863	1.485
3	2	-2.766	3.134	0.367
3	3	-5.080	4.804	-0.276
3	4	-5.785	7.902	2.117
3	5	-10.797	9.614	-1.182
3	6	-0.985	-4.110	-5.095
4	1	3.166	1.560	4.726
4	2	2.850	0.508	3.358
4	3	1.442	1.824	3.266
4	4	0.773	3.036	3.809
4	5	0.889	2.055	2.944
4	6	1.493	-2.265	-0.772
5	1	-4.817	9.467	4.650
5	2	-6.826	12.518	5.692
5	3	-8.828	14.781	5.953
5	4	-12.939	18.088	5.149
5	5	-14.705	12.388	-2.316
5	6	-1.920	-11.215	-13.135
6	1	10.584	-1.080	9.504
6	2	12.297	-3.973	8.324
6	3	11.596	-4.917	6.679
6	4	8.063	-3.088	4.975
6	5	13.337	-12.431	0.906
6	6	8.300	-13.836	-5.537



Model $Y_{it} = \alpha_i + m(x_{1it}, x_{2it}, x_{3it}) + u_{it}$	**HKO** = 4.081

Şekil 14. Parametrik olmayan sabit etkili panel veri modeli grafiği

Parametrik olmayan sabit etkili panel veri modeli ile yapılan tahminler sonucunda hata kareler ortalaması 4.081 olarak bulunmuştur. Hataların büyük kısmının birkaç gözlem üzerinde yoğunlaştığı Şekil 14 de daha net görülmektedir. Bu gözlemler göz ardı edilerek bakıldığında parametrik modele oranla çok daha iyi sonuçlar elde edildiği görülmektedir.

Parametrik olmayan regresyon modelinde panel veri kullanımının bir diğer şekli rasgele etkili parametrik olmayan panel veri modeli idi. Rasgele etkili model α_i parametresinin rasgele değişken olduğu düşünülerek u_{it} hatalarına eklenmesi ile

$$Y_{it} = m(x_{1it}, x_{2it}, x_{3it}) + v_{it}, \qquad i = 1,...,6 \quad , \quad t = 1,...,6$$

olarak elde edilmiştir. Modelin tahmini için v_{it} hatalarına ait varyans kovaryans matrisi elde edilmiştir. Bu amaçla $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ ilgili modellerin hata kareler ortalaması

yardımıyla hesaplanır. $\hat{\psi}_k(x)$ değeri parametrik olmayan sabit etkili model için türev fonksiyonunun tahmin değeri olmak üzere

$$\hat{\sigma}_u^2 = \sum_{i=1}^{6}\sum_{t=1}^{6}\left(Y_{it} - \overline{Y}_{i.} - \hat{\psi}_1(x)(x_{1it} - \overline{x}_{1i.}) - \hat{\psi}_2(x)(x_{2it} - \overline{x}_{2i.}) - \hat{\psi}_3(x)(x_{3it} - \overline{x}_{3i.})\right)^2 / 36$$

eşitliği yardımıyla $\hat{\sigma}_u^2 = 9.961$ olarak tahmin edilmiştir. $\hat{\sigma}_\alpha^2$ tahmini ise

$$\hat{\sigma}_\alpha^2 = \sigma_\eta^2 - \hat{\sigma}_u^2/6 = \left(\sum_{i=1}^{6}\left(\overline{Y}_{i.} - \tilde{m}(\overline{x}_{i.})\right)^2/6\right) - \hat{\sigma}_u^2/6$$

tahmin edicisi kullanılarak hesaplanmıştır. $\hat{\sigma}_\eta^2$ tahmini 1.123 olarak bulunmuş, buna bağlı olarak iki varyansın farkı olarak hesaplanan $\hat{\sigma}_\alpha^2$ değeri -0.536 olarak tahmin edilmiştir. Varyansın tahmininin negatif çıkması birimler arasındaki farkın anlamsız olmasından kaynaklanabilmektedir ve bu durumda $\hat{\sigma}_\alpha^2$ değeri 0 alınmaktadır. $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ değerleri yerine koyularak oluşturulan varyans kovaryans matrisi

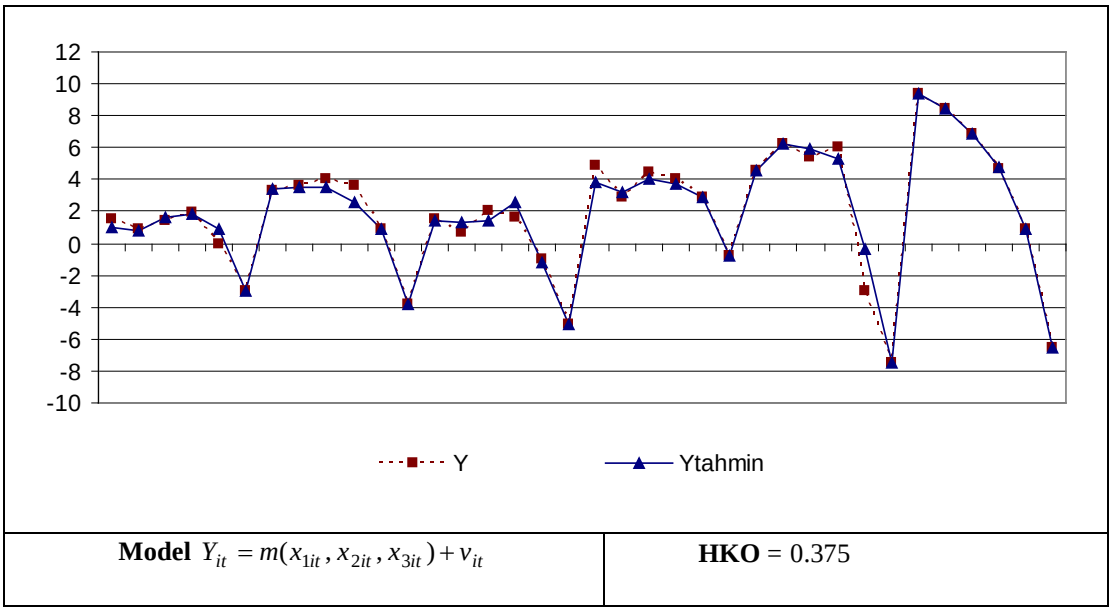
$$M_V = \begin{bmatrix} 9.961 & 0 & \cdots & 0 \\ 0 & 9.961 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 9.961 \end{bmatrix}_{36\times 36}$$

köşegen matris olarak elde edilmiştir. Rasgele etkili modelin tahmininde lokal lineer kernel tahmin yöntemi kullanılmış, Kernel fonksiyonu olarak standart normal kernel alınmış ve *h* bant genişliği normal referans kuralına göre sabit etkili modelde hesaplanan değerler alınmıştır. Tahmin sonucunda elde edilen $\hat{Y}_{it}$ tahminleri her bir *i.* ve *t.* gözlem için tablo olarak verilmiştir.

Çizelge 9. Parametrik olmayan rasgele etkili panel veri modeline ilişkin tahminler

i	t	$\hat{Y}_{it}$	i	t	$\hat{Y}_{it}$	i	t	$\hat{Y}_{it}$
1	1	0.972	3	1	1.397	5	1	4.567
1	2	0.809	3	2	1.294	5	2	6.211
1	3	1.656	3	3	1.401	5	3	5.955
1	4	1.854	3	4	2.542	5	4	5.248
1	5	0.871	3	5	-1.241	5	5	-0.375
1	6	-3.000	3	6	-5.098	5	6	-7.500
2	1	3.385	4	1	3.794	6	1	9.418
2	2	3.485	4	2	3.216	6	2	8.395
2	3	3.477	4	3	4.026	6	3	6.897
2	4	2.586	4	4	3.683	6	4	4.764
2	5	0.902	4	5	2.907	6	5	0.900
2	6	-3.800	4	6	-0.783	6	6	-6.500

Çizelge 9'da görüldüğü gibi tahminler parametrik olmayan sabit etkili modele oranla çok daha iyi çıkmıştır. Rasgele etkili model için $\hat{Y}_{it}$ tahminlerinin Y_{it} değerleri ile birlikte grafiği Şekil 8' deki gibidir.



Şekil 15. Parametrik olmayan rasgele etkili panel veri modeli grafiği

Şekil 15'de tahminlerin önceki modellere göre daha iyi sonuçlar verdiği görülmektedir. Parametrik olmayan rasgele etkili modelde $\hat{\sigma}_{\alpha}^{2}$ tahmininin sıfır olması parametrik olmayan model için α_i değerinin anlamsız olduğunun bir göstergesidir. Bunun sonucunda parametrik olmayan rasgele etkili modelin parametrik olmayan sabit etkili modelden daha iyi sonuç vermesi beklenmektedir. Rasgele etkili modelin hata kareler ortalaması 0.375 olarak hesaplanmıştır. Hata kareler ortalamasının sabit etkili modelden daha iyi olduğu görülmektedir.

Yarıparametrik Panel Verili Regresyon Modeli

Yarıparametrik regresyonda ilk olarak bağımsız değişkenlerden parametrik kısma dahil edilecekler ile parametrik olmayan kısma dahil edileceklerin belirlenmesi gerekmektedir. Bu amaçla bağımlı değişken ile bağımsız değişkenler arasındaki tekli, ikili ve üçlü parametrik regresyon modelleri kurulmuştur. Kurulan modeller en küçük kareler yöntemi ile tahmin edilmiş ve parametrelerin anlamlılıkları incelenmiştir. Bu işlem için sırasıyla;

(1) $Y_{it} = \beta_0 + \beta_1 x_{1it} + \varepsilon_{it}$

(2) $Y_{it} = \beta_0 + \beta_1 x_{2it} + \varepsilon_{it}$

(3) $Y_{it} = \beta_0 + \beta_1 x_{3it} + \varepsilon_{it}$

(4) $Y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \varepsilon_{it}$

(5) $Y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{3it} + \varepsilon_{it}$

(6) $Y_{it} = \beta_0 + \beta_1 x_{2it} + \beta_2 x_{3it} + \varepsilon_{it}$

(7) $Y_{it} = \beta_0 + \beta_1 x_{1it} + \beta_2 x_{2it} + \beta_3 x_{3it} + \varepsilon_{it}$, $i = 1,2,...,6$; $t = 1,2,...,6$

modelleri kurulmuş ve 7 adet model için elde edilen tahminler ve parametrelerin anlamlılık testleri Çizelge 10'da gösterilmiştir.

Çizelge 10. Model belirlemeye ilişkin değerler

Model	$\hat{\beta}_k$		R^2	**Modelin Anlamlılığı**		**Bireysel Parametrelerin Anlamlılığı**		
				F	*p değeri*	$\hat{\beta}_k$	***t***	*p değeri*
(1)	$\hat{\beta}_0$	-0.63	0.261	12.02	0.001	$\hat{\beta}_0$	-0.68	0.497
	$\hat{\beta}_1$	0.73				$\hat{\beta}_1$	3.46	0.001
(2)	$\hat{\beta}_0$	6.83	0.155	6.25	0.017	$\hat{\beta}_0$	3.30	0.002
	$\hat{\beta}_1$	-0.55				$\hat{\beta}_1$	-2.50	0.017
(3)	$\hat{\beta}_0$	1.11	0.013	0.45	0.503	$\hat{\beta}_0$	-0.84	0.403
	$\hat{\beta}_1$	-0.12				$\hat{\beta}_1$	-0.67	0.503
(4)	$\hat{\beta}_0$	4.91	0.48	14.81	0.0001	$\hat{\beta}_0$	2.86	0.007
	$\hat{\beta}_1$	0.81				$\hat{\beta}_1$	4.46	0.0001
	$\hat{\beta}_2$	-0.65				$\hat{\beta}_2$	-3.64	0.001
(5)	$\hat{\beta}_0$	-1.49	0.473	6.33	0.005	$\hat{\beta}_0$	-1.09	0.283
	$\hat{\beta}_1$	0.74				$\hat{\beta}_1$	3.47	0.001
	$\hat{\beta}_2$	-0.13				$\hat{\beta}_2$	-0.85	0.398
(6)	$\hat{\beta}_0$	6.05	0.277	3.34	0.047	$\hat{\beta}_0$	2.59	0.014
	$\hat{\beta}_1$	-0.56				$\hat{\beta}_1$	-2.48	0.018
	$\hat{\beta}_2$	-0.12				$\hat{\beta}_2$	-0.72	0.472
(7)	$\hat{\beta}_0$	4.04	0.169	10.22	0.0001	$\hat{\beta}_0$	2.11	0.43
	$\hat{\beta}_1$	0.82				$\hat{\beta}_1$	4.48	0.0001
	$\hat{\beta}_2$	-0.65				$\hat{\beta}_2$	-3.64	0.001
	$\hat{\beta}_3$	-0.13				$\hat{\beta}_3$	-1.01	0.318

Çizelge 10 incelendiği zaman değişkenlerin tek tek alındığı modelde x_3 değişkeni yardımıyla kurulan modelde F ve t testlerine göre parametrelerin anlamsız çıktığı görülmektedir. Ayrıca her üç değişkeninde modele alındığı 7 numaralı modelde x_1 ve x_2 değişkenlerine ait parametreler bireysel parametrelerin testi sonucunda anlamlı çıkmasına rağmen x_3 değişkenine ait parametre anlamsız çıkmıştır. Bu durumda x_3 değişkenine ait parametrelerin modellerde anlamsız çıktığı görülmektedir. x_1 ve x_2 bağımsız değişkenlerini yarıparametrik modelin parametrik kısmına, x_3 değişkenini parametrik olmayan kısmına dahil edersek yarıparametrik panel verili regresyon modeli

$$Y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + m(x_{3it}) + \varepsilon_{it}$$

biçiminde elde edilir.

Yarıparametrik modelde, bireyler arasındaki farklılıklardan kaynaklanan değişim sabit olarak alınırsa yarıparametrik sabit etkili panel veri modeli

$$Y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + m(x_{3it}) + u_{it}, \qquad i = 1,...,6 \;\; ; \;\; t = 1,...,6$$

olarak yazılabilir. Sabit etkili modelin parametrik kısmında yer alan β_1 ve β_2 parametrelerinin tahmini için öncelikle parametrik olmayan kısımdaki değişken x_{3it} ye göre koşullu beklenen değer alınır ve dönüşüm yapılırsa parametrik model

$$Y_{it} - E(y_{it}|x_{3it}) = (x_{1it} - E(x_{1it}|x_{3it}))\beta_1 + (x_{2it} - E(x_{2it}|x_{3it}))\beta_2 + u_{it}$$

olarak elde edilir. Koşullu beklenen değerler $E(Y_{it}|x_{3it})$, $E(x_{1it}|x_{3it})$ ve $E(x_{2it}|x_{3it})$ lerin yerine parametrik olmayan

$$Y_{it} = m(x_{3it}) + u_{it}$$

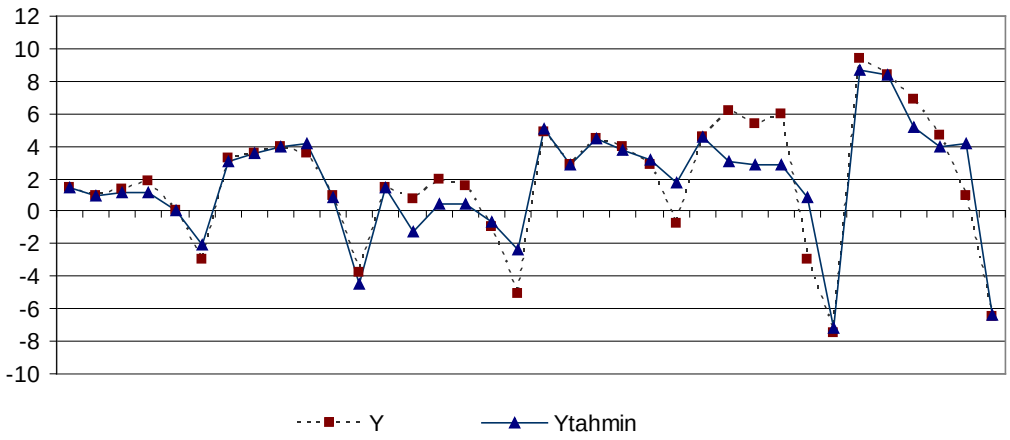
$$x_{1it} = m(x_{3it}) + u_{it}$$
$$x_{2it} = m(x_{3it}) + u_{it}$$

Modellerinden elde edilen Nadayara-Watson kernel tahmin değerleri kullanılmıştır. Nadaraya-Watson kernel tahmininde $K(.)$ fonksiyonu standart normal kernel fonksiyonu, h bant genişliği de 0.10 olarak alınmıştır. Dönüştürülmüş değişkenler yardımıyla elde edilen modele klasik en küçük kareler yöntemi uygulanmış $\hat{\beta}_1$ ve $\hat{\beta}_2$ tahminleri sırasıyla $\hat{\beta}_1 = 0.528$, $\hat{\beta}_2 = -0.696$ olarak elde edilmiştir. Parametrik kısım tahmin edildikten sonra parametrik olmayan kısmın tahminine geçilir. Bunun için öncelikle $Y_{it}^{*} = Y_{it} - 0.528x_{1it} - (0.696)x_{2it}$ eşitliği yardımıyla bağımlı değişkenin yeni değerleri hesaplanır. Elde edilen $Y_{it}^{*} = \alpha_i + m(x_{3it}) + u_{it}$ modelinden $\hat{\alpha}_i$ ve $\hat{m}(x_{3it})$ tahminleri lokal lineer kernel tahmin yöntemi ile hesaplanır. Çizelge 11'de uygulamada kullanılan $E(Y_{it}|x_{3it})$, $E(x_{1it}|x_{3it})$, $E(x_{2it}|x_{3it})$, Y_{it}^{*} değerleri ile bağımsız değişkenlerin aldığı değerler yerine koyulduğunda bulunan $\hat{\alpha}_i$, $\hat{m}(x_{3it})$ ve $\hat{Y}_{it}$ tahminleri verilmiştir.

Çizelge 11'de verilen tahminler incelendiğinde gerek parametrik gerekse parametrik olmayan sabit etkili panel veri modellerinden daha iyi tahminler elde edildiği görülmektedir. Herhangi bir gözlem değeri için aşırı bir sapma bulunmamaktadır ve hatalar bütün gözlemler üzerinde dengeli dağılmışlardır. Tahminlerin uyumunu inceleyebilmek amacı ile $\hat{Y}_{it}$ tahminleri ile $\hat{Y}_{it}$ değerleri grafik üzerinde Şekil 16' da gösterilmiştir.

Çizelge 11. Yarıparametrik sabit etkili panel veri modeline ilişkin hesaplamalar

i	t	$E(y_{it}\|x_{3it})$	$E(x_{1it}\|x_{3it})$	$E(x_{2it}\|x_{3it})$	Y_{it}^{*}	$\hat{\alpha}_i$	$\hat{m}(x_{3it})$	$\hat{Y}_{it}$
1	1	1.761	2.616	7.288	5.052	-1.086	6.137	1.498
1	2	1.173	2.730	8.724	5.081	2.105	3.035	0.959
1	3	0.660	1.813	8.668	5.175	-1.764	6.679	1.139
1	4	1.475	2.428	8.117	6.201	3.815	1.605	1.119
1	5	0.008	2.700	7.600	3.864	0.114	3.750	0.000
1	6	-0.156	1.283	8.864	3.929	-1.620	6.516	-2.033
2	1	2.870	3.024	7.328	9.250	2.442	6.587	3.080
2	2	3.193	3.365	9.273	8.208	2.021	6.180	3.594
2	3	3.999	3.600	8.600	8.085	2.072	6.013	4.000
2	4	0.660	1.813	8.668	7.898	1.761	6.679	4.141
2	5	0.953	3.246	9.753	6.600	3.349	3.227	0.876
2	6	1.932	4.826	13.555	9.026	2.095	6.269	-4.462
3	1	1.534	2.300	9.058	6.619	-8.979	15.538	1.440
3	2	-3.596	0.674	10.351	5.176	5.940	-2.725	-1.262
3	3	-0.503	1.662	8.121	6.267	2.958	1.726	0.417
3	4	-0.702	1.563	7.277	4.790	-20.346	23.946	0.410
3	5	2.051	3.021	5.749	1.885	-9.246	11.486	-0.645
3	6	-1.318	1.467	7.922	0.864	-12.355	15.993	-2.326
4	1	3.672	6.731	10.592	10.067	3.732	6.539	5.104
4	2	3.153	3.456	9.618	7.942	1.755	6.188	2.901
4	3	4.500	3.300	9.500	9.370	-8.589	17.958	4.500
4	4	3.869	3.143	8.217	8.193	1.435	6.480	3.722
4	5	3.031	4.057	7.683	5.972	3.206	3.009	3.144
4	6	0.660	1.813	8.668	5.231	1.059	6.679	1.707
5	1	4.566	2.300	5.342	7.074	-3.281	10.363	4.608
5	2	4.318	3.034	5.240	8.031	-5.542	10.415	3.042
5	3	6.271	4.141	5.909	6.828	-10.013	14.328	2.887
5	4	2.870	3.024	7.328	7.601	-2.102	6.587	2.885
5	5	1.193	3.045	6.977	-0.391	-3.896	7.312	0.807
5	6	-3.596	0.674	10.351	1.697	4.772	-2.725	-7.150
6	1	7.703	6.032	7.669	12.028	-7.199	18.525	8.699
6	2	8.400	8.200	10.200	11.170	-427.368	438.536	8.399
6	3	1.932	4.826	13.555	8.722	0.717	6.269	5.164
6	4	3.672	6.731	10.592	7.222	-0.044	6.539	3.972
6	5	3.023	7.899	10.487	3.065	-0.348	6.709	4.196
6	6	-5.840	4.943	13.788	0.527	5.111	-4.489	-6.405



Model $Y_{it} = \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + m(x_{3it}) + u_{it}$	**HKO** = 2.221

Şekil 16. Yarıparametrik sabit etkili panel veri modeli grafiği

grafiği elde edilir. Yarıparametrik sabit etkili panel veri modelinin hata kareler ortalaması 2.221 olarak hesaplanmıştır.

Bireyler arasındaki farklardan kaynaklanan değişim rasgele olarak alınırsa yarıparametrik rasgele etkili panel veri modeli

$$Y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + m(x_{3it}) + w_{it}, \qquad i = 1,...,6 \;\; ; \;\; t = 1,...,6$$

olarak elde edilir. Öncelikle rasgele etkili modelin parametrik kısmında yer alan β_1 ve β_2 parametrelerinin tahmini için gerekli dönüşümler yapılarak $R^{*yx_3}_{it}$, $R^{*x_1x_3}_{kit}$ ve $R^{*x_1x_3}_{kit}$ değerleri elde edilir. bu değerleri elde etmek için M_V matrisinin hesaplanması gerekmektedir. M_V matrisinin elemanları olan $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ ilgili modellerin hata kareler ortalaması yardımıyla hesaplanır. $\hat{\beta}_{1FE} = 0.528$, $\hat{\beta}_{2FE} = -0.696$ şeklinde hesaplanan sabit etkili modelin parametre tahminleri $\overline{R}_{i.}$ değerleri $i.$ bireye ait R_{it} değerlerinin ortalaması olmak üzere

$$\hat{\sigma}_u^2 = \sum_{i=1}^{6}\sum_{t=1}^{6}\left(R_{it}^{yx_3} - \overline{R}_{i.}^{yx_3} - \hat{\beta}_{1FE}\left(R_{it}^{x_1x_3} - \overline{R}_{i.}^{x_1x_3}\right) - \hat{\beta}_{2FE}\left(R_{it}^{x_2x_3} - \overline{R}_{i.}^{x_2x_3}\right)\right)^2 / 36$$

yardımıyla hesaplanmış ve $\hat{\sigma}_u^2 = 2.130878$ olarak bulunmuştur. $\hat{\sigma}_\alpha^2$ değeri ise $\hat{\beta}_{1G}$ ve $\hat{\beta}_{2G}$ tahmini $\overline{R}_{i.}^{yx_3} = \alpha_i + \overline{R}_{i.}^{x_1x_3}\beta_{1G} + \overline{R}_{i.}^{x_2x_3}\beta_{2G} + u_{it}$ modelinin en küçük kareler tahmini olmak üzere

$$\hat{\sigma}_\alpha^2 = \sigma_\eta^2 - \hat{\sigma}_u^2/6 = \left(\sum_{i=1}^{6}\left(\overline{R}_{i.}^{yx_3} - \hat{\beta}_{1G}R_{i.}^{x_1x_3} - \hat{\beta}_{2G}R_{i.}^{x2x_3}\right)^2/6\right) - \hat{\sigma}_u^2/6$$

yardımıyla hesaplanmıştır. $\hat{\sigma}_\eta^2$ tahmini 0.0495 olarak bulunmuş, buna bağlı olarak iki varyansın farkı olarak hesaplanan $\hat{\sigma}_\alpha^2$ değeri -0.3056 olarak tahmin edilmiştir. Varyansın tahmininin negatif çıkması birimler arasındaki farkın anlamsız olmasından kaynaklanabilmektedir ve bu durumda $\hat{\sigma}_\alpha^2$ değeri 0 alınmaktadır. $\hat{\sigma}_u^2$ ve $\hat{\sigma}_\alpha^2$ değerleri yerine koyularak oluşturulan M_V matrisi

$$M_V = \begin{bmatrix} 2.1308 & 0 & \cdots & 0 \\ 0 & 2.1308 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 2.1308 \end{bmatrix}_{36\times 36}$$

köşegen matris olarak elde edilmiştir. M_V matrisinin hesaplanması ile parametrik kısma ait β_1 ve β_2 parametrelerinin tahminleri $\hat{\beta}_1 = 0.528$, $\hat{\beta}_2 = -0.696$ olarak elde edilmiştir. M_V matrisinde $\hat{\sigma}_\alpha^2$ değeri sıfır olduğundan köşegen matris ile yapılan dönüşüm parametrik kısmın tahmininde bir değişiklik yapmamış ve sonuçlar sabit etkili model ile aynı bulunmuştur. Parametrik olan kısma ait $\hat{\beta}$ tahminleri bulunduktan sonra $m(x_{3it})$ fonksiyonunun tahmin edilmesi işlemine geçilir. İlk olarak $Y_{it}^* = Y_{it} - 0.528x_{1it} - (0.696)x_{2it}$ eşitliği yardımıyla bağımlı değişkenin yeni değerleri

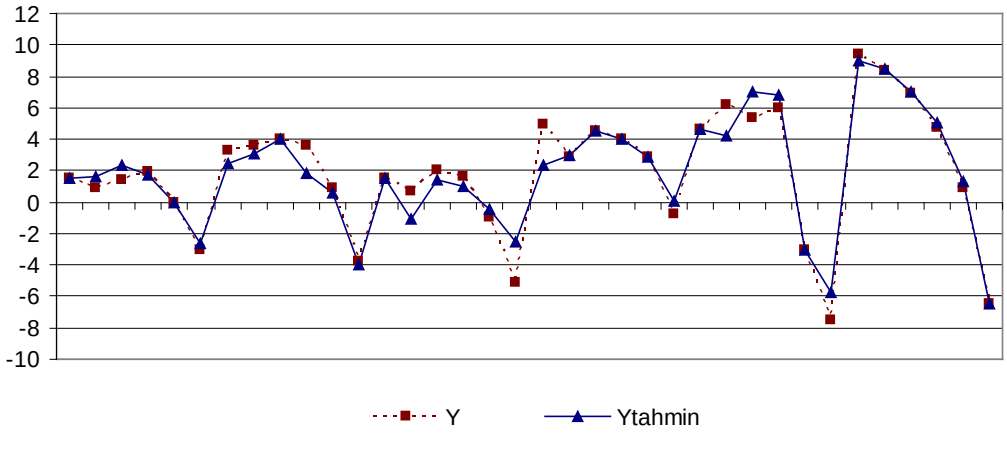
hesaplanır. $Y_{it}^{*} = m(x_{3it}) + v_{it}$ modeli için $\hat{m}(x_{3it})$ tahmini lokal lineer kernel tahmin yöntemi ile hesaplanır. Parametrik kısmın tahmininde kullanılan $R^{*yx_3}_{it}$, $R^{*x_1x_3}_{kit}$ ve $R^{*x_2x_3}_{kit}$ değerleri, lokal lineer kernel tahmininde kullanılan Y_{it}^{*} değerleri ve bağımsız değişkenlerin aldığı değerler yerine koyularak rasgele etkili model için hesaplanan $\hat{m}(x_{3it})$ ve $\hat{Y}_{it}$ tahminleri Çizelge 12'deki gibidir.

Yarıparametrik rasgele etkili panel veri modeli ile yarıparametrik sabit etkili panel veri modelinden daha iyi tahmin değerleri elde edilmiştir. Yarıparametrik rasgele etkili modelde $\hat{\sigma}_{\alpha}^{2}$ tahmininin negatif çıkması sabit etki parametresinin yarıparametrik model için anlamsız olduğu şeklinde yorumlanıyordu. Bunun sonucunda yarıparametrik model için rasgele etkili panel veri modelinin, sabit etkili panel veri modelinden daha iyi sonuç vermesi beklenir. Rasgele etkili modelde sabit bir α_i parametresinin bulunması tahminlerin daha iyi çıkmasına imkan vermiştir. Rasgele etkili model için yapılan tahmin sonucunda elde edilen $\hat{Y}_{it}$ tahminleri ile Y_{it} değerleri grafik üzerinde Şekil 17'de gösterilmiştir.

Yarıparametrik rasgele etkili model için elde edilen hata kareler ortalaması 0.955 olarak bulunmuştur ve tahminlerin gözlem değerleri ile uyumu Şekil 17'de daha açık biçimde görülmektedir.

Çizelge 12. Yarıparametrik rasgele etkili panel veri modeline ilişkin hesaplamalar

i	t	$R^{*yx_3}_{it}$	$R^{*x_1x_3}_{kit}$	$R^{*x_2x_3}_{kit}$	Y^{*}_{it}	$\hat{m}(x_{3it})$	$\hat{Y}_{it}$
1	1	-0.179	-0.079	-0.197	5.052	5.458	1.509
1	2	-0.187	-0.431	-0.770	5.0808	5.805	1.623
1	3	0.507	0.814	-0.663	5.1752	5.762	2.323
1	4	0.291	-0.019	-0.080	6.2008	5.771	1.725
1	5	-0.005	0.000	0.000	3.864	3.864	0.001
1	6	-1.948	-1.290	0.436	3.9288	5.288	-2.658
2	1	0.295	0.052	2.447	9.2496	6.133	2.441
2	2	0.279	0.024	-0.050	8.208	7.901	3.119
2	3	0.000	0.000	0.000	8.0848	8.084	4.000
2	4	2.014	0.677	-0.252	7.8984	5.762	1.800
2	5	-0.036	0.585	1.060	6.6	6.095	0.581
2	6	-3.927	-3.512	3.182	9.0256	8.185	-3.938
3	1	-0.023	0.000	0.029	6.6192	6.621	1.500
3	2	2.943	1.045	-1.542	5.176	3.263	-1.040
3	3	1.715	0.368	-0.220	6.2672	4.210	1.377
3	4	1.577	0.299	-0.806	4.7896	3.571	0.977
3	5	-2.090	0.328	0.720	1.8848	4.441	-0.480
3	6	-2.591	-0.526	0.807	0.864	3.410	-2.543
4	1	0.841	-2.556	-0.611	10.0672	7.173	2.369
4	2	-0.173	0.030	0.193	7.9424	7.752	2.968
4	3	0.000	0.000	0.000	9.3696	9.370	4.509
4	4	0.090	-0.098	0.057	8.1928	7.903	4.000
4	5	-0.090	0.098	-0.057	5.972	6.244	2.900
4	6	-1.000	-0.488	0.570	5.2312	5.762	0.067
5	1	0.023	0.000	-0.029	7.0744	7.072	4.600
5	2	1.289	-0.571	-0.644	8.0312	6.310	4.229
5	3	-0.597	-0.987	-1.239	6.828	8.102	7.045
5	4	2.145	-0.085	-1.937	7.6008	6.133	6.790
5	5	-2.873	0.038	-0.601	-0.3912	4.358	-2.999
5	6	-2.674	-1.558	1.130	1.6968	3.263	-5.760
6	1	1.163	1.759	1.802	12.028	9.647	8.968
6	2	0.000	0.000	0.000	11.1696	11.170	8.470
6	3	3.403	3.270	-2.504	8.7216	8.185	7.066
6	4	0.704	1.417	-0.200	7.2224	7.173	5.014
6	5	-1.455	1.713	0.352	3.0648	6.257	1.355
6	6	-0.452	0.861	0.693	0.5272	1.175	-6.500



Model $Y_{it} = \beta_1 x_{1it} + \beta_2 x_{2it} + m(x_{3it}) + v_{it}$	**HKO** = 0.955

Şekil 17. Yarıparametrik rasgele etkili panel veri modeli grafiği

6. SONUÇ VE ÖNERİLER

Çalışmada son yıllarda özellikle ekonometri ve sağlık alanlarında yoğun biçimde kullanılan panel veri teknikleri ele alınmıştır. Panel verilerin analizi için literatürde sıklıkla parametrik regresyon modelleri kullanılmaktadır. Ancak istatistik bilimindeki hızlı gelişme parametrik tekniklerin yetersiz kaldığı noktalarda parametrik olmayan tekniklerin devreye girmesini sağlamıştır. Son yıllarda ise regresyon analizinde hem parametrik hem de parametrik olmayan modellerin birleştirilmesi ile yarıparametrik modeller kullanılmaya başlanmıştır. Yarıparametrik modeller bağımlı değişken ile bağımsız değişken veya değişkenler arasındaki ilişkiyi hem parametrik hem de parametrik olmayan model yardımıyla açıkladığından büyük avantaj sağlamaktadır.

Parametrik regresyon modelleri gerek tahmin kolaylığı gerekse yorumlama kolaylığı açısından tercih edilmektedir. Ancak parametrik regresyonun katı varsayımları bazı durumlarda uygulamada sorunlar çıkarmakta ve hatalı sonuçlar elde edilmesine yol açmaktadır. Parametrik olmayan regresyon ise özellikle veri setinin yapısından etkilenmeden uygun tahmin sonuçları vermekte fakat yorulmada herhangi bir parametre değeri elde edilemediğinden zorluk çıkarmaktadır. Yarıparametrik regresyonda parametrik kısmın varlığı sonucunda elde edilen parametre tahminleri sonucunda yorumlama kolaylığı sağlamakta ve bağımlı değişken ile bağımsız değişken arasındaki ilişkinin türü ortaya koyulabilmektedir. Aynı zamanda yarıparametrik regresyon, parametrik regresyonun eksine, bağımlı değişken ile herhangi bir bağımsız değişken arasında parametrik bir fonksiyon yardımıyla ilişki açıklanamıyorsa onu modelden çıkarmak yerine parametrik olmayan kısma dahil ederek hem bilgi kaybını azaltmakta hem de parametrik modele göre daha iyi sonuçlar vermektedir.

Uygulamanın birinci kısmında parametrik olmayan tahmin tekniklerinden medyana dayalı tahmin yöntemleri ele alınmıştır. Literatürde yer alan medyana dayalı

tahmin tekniklerinden Brown-Mood yöntemi ile Theil yöntemi çalışmanın ikinci bölümünde ayrıntılı olarak ele alınmıştır. Bu yöntemler, veriler arasındaki ilişkinin yapısının daha çok doğrusal olması durumunda uygun sonuçlar vermektedir. Ancak bağımlı değişken ile bağımsız değişken arasındaki ilişkinin fonksiyonel yapısı doğrusal değil de 2. dereceden bir polinom şeklinde ise bu yöntemler uygun sonuç vermeyecektir. Bu durumlarda kullanılabilecek medyana dayalı bir parametrik olmayan tahmin yöntemi bu çalışmada önerilmiştir. Geliştirilen algoritmanın veri seti üzerinde uygulanabilirliği ve literatürde yer alan yöntemler ile kıyaslanması için 6 farklı durumda veri seti oluşturulmuş ve tahminler karşılaştırılmıştır. Doğrusala yakın ilişkiye sahip veri çiftleri için literatürde yer alan Brown-Mood yöntemi ile bu çalışmada önerilen medyan yöntemi yakın sonuçlar vermiştir. Ancak bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin yapısı doğrusallıktan uzaklaştıkça bu çalışmada önerilen medyan yöntemi Brown-Mood yöntemine göre daha iyi sonuçlar vermiştir. Tahminler ile gözlemler arasındaki hata kareler ortalamasını minimum yapacak en küçük kareler yöntemi ile elde edilen eğriler ile karşılaştırma yapılmış ve EKK tahminlerine yakın sonuçlar elde edilmiştir. Ancak avantaj olarak medyan yöntemi herhangi bir varsayım gerektirmemektedir ve aykırı gözlemlerden EKK yöntemi gibi etkilenmemektedir. Çalışmanın ikinci bölümünde önerilen medyana dayalı tahmin yönteminin algoritması verilmiş ve 3. ve daha yüksek dereceden polinomlar tarafından açıklanabilecek ilişki yapılarında kullanılması açısından genelleştirilmiştir.

Yapılan karşılaştırmalar sonucunda önerilen medyana dayalı tahmin yönteminin iyi sonuçlar verdiği görülmüştür. Bu çalışmada önerilen yöntem konu ile ilgili çalışmalara kaynak olabilecek ve parametrik olmayan tahmin yöntemleri arasında kullanılabilecek bir yöntemdir.

Uygulama kısmının ikinci bölümünde Avrupa kıtasını etkisi altına alan ve ilk olarak Yunanistan da patlak veren ve ardından İspanya ve Portekiz gibi ekonomilerde etkisini göstermeye başlayan ekonomik krizin modellenmesi amaçlanmıştır. Türkiye'nin de içinde bulunduğu kritik 6 ülke için 2004-2009 yılları arasında

ekonomik göstergeler alınmış ve elde edilen panel veri seti için parametrik, parametrik olmayan ve yarıparametrik modeller analiz edilmiştir.

Panel veri analizi için en çok kullanılan parametrik panel verili regresyon modeli çalışmada iki farklı biçimde kurulmuş ve tahminler elde edilmiştir. İlk olarak parametrik sabit etkili panel veri modelinden tahminler elde edilmiş ve değerlendirilmiştir. Yunanistan ve İspanya için sabit etki parametreleri diğer ülkelere oranla yüksek çıkmıştır. Bu ülkeler için, enflasyon, işsizlik ve cari işlemler dengesinin GSYİH ya oranından başka büyüme için diğer faktörlerinde etkili olduğu düşünülebilir. Temel göstergeler dışında büyümeye etki eden yan faktörlerin önemli olması ülke ekonomisinin sağlam yapıda olmadığının göstergesidir. Bu iki ülkeden sonra en büyük sabit parametre değerine sahip ülke Türkiye'dir. Türkiye'nin de yan faktörlerden oldukça fazla oranda etkilendiği görülmektedir. Sabit etkili panel veri modelinde, büyüme ile büyümeye etki eden göstergeler arasında kurulan modelin $\hat{\beta}_k$ tahminlerine bakılarak yorumlama yapılabilir. Büyüme ile işsizlik arasında negatif yönde bir ilişki olduğu görülmektedir. İktisat teorisini destekleyen bu sonuca göre büyüme üzerinde en yüksek etki işsizlik göstergesi ile açıklanabilir.

Parametrik rasgele etkili panel veri modelinde sabit etkiler hata terimine dahil edildiğinden ülkelere ait bir parametre tahmin edilmemiştir. Büyüme oranı için sabit etkili modelden daha kötü tahminler bulunmuştur. Parametrik rasgele etkili panel veri modelinden elde edilen $\hat{\beta}_k$ tahminleri incelendiğinde işsizlik ve cari işlemler dengesinin GSYİH' ya oranı ile büyüme arasında ters yönlü bir ilişki olduğu söylenebilir. Parametrik sabit etkili panel veri modelinde hata kareler ortalaması 4.783, parametrik rasgele etkili panel veri modelde hata kareler ortalaması 8.732 olarak hesaplanmıştır. Parametrik rasgele etkili panel veri modelinde ülkelere ait bireysel parametre mevcut olmadığından ülkeler arası farklılıklar için yorumlama yapılamamaktadır. Parametrik panel veri modelleri için sabit etkili model tercih edilmelidir. Ülkelerin farklılıklarından kaynaklanan değişimin modele sabit olarak eklenmesi hem yorumlama için kolaylık getirmiş hem de daha iyi sonuçlar elde edilmesini sağlamamıştır.

Parametrik olmayan sabit etkili ve rasgele etkili panel veri modeli için yapılan tahminlerde parametrik modele göre çok daha iyi tahminler elde edilmiştir. Parametrik olmayan sabit etkili panel veri modelinin hata kareler ortalaması 4.081, parametrik olmayan rasgele etkili panel veri modelinin hata kareler ortalaması 0.375 olarak bulunmuştur. Parametrik olmayan rasgele panel veri modeli ile büyüme oranlarının tahmini yapılırken $\hat{\sigma}_{\alpha}^{2}$ tahmini anlamsız çıkmıştır. Buna bağlı olarak parametrik olmayan panel veri modeli için ülkeler arasındaki farklılıklardan kaynaklanan değişimin anlamlı olmadığı söylenebilir. Bu durumda parametrik olmayan rasgele etkili panel veri modelinde α_i katsayılarının olmaması tahminlerin sabit etkili panel veri modeline göre daha iyi çıkmasına olanak vermiştir. Ancak parametrik olmayan panel verili modeller ile yapılan tahminlerde yorum zorluğu yaşanmaktadır. Modelde parametre tahmin edilmediğinden büyüme ile büyümeyi etkileyen göstergelerin arasındaki ilişki ile ilgili herhangi bir yorum yapılamamaktadır.

Yarıparametrik panel verili modeller kullanımı ile bu yorumlama problemi kısmen ortadan kaldırılmıştır. Yarıparametrik modelin kurulması aşamasında gerekli testler yapılmış ve parametrik kısma enflasyon ile işsizlik göstergeleri parametrik olmayan kısma ise cari işlemler dengesinin GSYİH' ya oranı dahil edilmiştir. Yarıparametrik sabit etkili panel veri modeli ile yapılan tahminler sonucunda hata kareler ortalaması 2.221 olarak hesaplanmıştır. Yarıparametrik rasgele etkili panel veri modeli ile tahmini yapılırken $\hat{\sigma}_{\alpha}^{2}$ tahmini, parametrik olmayan modelde olduğu gibi anlamsız çıkmıştır. Yarıparametrik panel veri modeli için ülkeler arasındaki farklılıklardan kaynaklanan değişimin anlamlı olmadığı görülmektedir. Buna bağlı olarak yarıparametrik rasgele etkili panel veri modelinden elde edilen tahminler sonucunda hata kareler ortalaması 0.955 olarak hesaplanmıştır. Rasgele etkili modelin sabit etkili modelden daha iyi sonuç verdiği görülmektedir. Parametrik kısımdan elde edilen $\hat{\beta}_k$ tahminlerine bakıldığında yarıparametrik modelde işsizlik ile büyüme arasında negatif yönlü ilişki olduğu görülmektedir.

İktisat teorisi için önemli bir nokta büyüme ile enflasyon arasındaki ilişkinin yorumlanmasıdır. İktisat teorisine göre büyüme ile enflasyon arasındaki ilişki net olarak ortaya koyulamamıştır. Enflasyonun büyümeyi pozitif veya negatif yönde etkileyeceğine dair farklı görüşler mevcuttur. Bu çalışmada kullanılan veri seti için, parametrik ve yarıparametrik panel verili modeller ile elde edilen tahmin sonuçlarına göre büyüme ile enflasyon arasında pozitif yönlü bir ilişkinin olduğu söylenebilir.

Bütün modellerde, krizin etkisini gösterdiği 2009 yılına ait tahminlerin hataları diğer yıllara göre daha yüksek çıkmıştır. Bunun sebebi 2009 yılında Amerika ve Avrupa'nın diğer güçlü ekonomilerinin de oluşturduğu kriz havası ile nispeten daha güçsüz ekonomilerin daha büyük darbeler alması ve dış faktörlerden çok etkilenmeleridir.

Çalışmada panel veri analizleri için parametrik, parametrik olmayan ve yarıparametrik modeller için bağımsız değişken sayısının birden fazla olduğu durumlar göz önüne alınmıştır. Daha önceleri konu ile ilgili yapılan çalışmalarda Parametrik panel verili modellerde bağımsız değişken sayısının birden fazla olduğu durumların incelendiği görülmektedir. Ancak parametrik olmayan ve yarıparametrik modeller için bağımsız değişken sayısının birden fazla olduğu durumlarda fazla çalışma yapılmamıştır. Bağımsız değişken sayısı birden fazla olduğunda özellikle parametrik olmayan panel verili regresyonda tahminlerin hesaplanması zorlaşmaktadır. Bu çalışmada 3 adet bağımsız değişken için modeller kurulmuş ve tahminler yapılmıştır. Yapılan tahminler sonucunda panel veri analizi için yarıparametrik modellerin kullanımı ile sağlanacak avantajlar gösterilmiştir. Ayrıca parametrik olmayan medyana dayalı yöntemlere alternatif bir yöntem geliştirilmiş ve bağımlı değişken ile bağımsız değişken arasındaki ilişkinin doğrusal bir yapıda olmaması durumunda öne sürülen yöntemin diğer yöntemlerden daha iyi sonuçlar verdiği simülasyon çalışmaları ile gösterilmiştir. Bu çalışmada önerilen medyana dayalı parametrik olmayan tahmin yöntemi üzerinde yapılacak çalışmalar ile geliştirilebilecek ve daha iyi sonuçlar elde edilebilecektir.

KAYNAKLAR

Artan, S., 2008, "Türkiye'de Enflasyon, Enflasyon Belirsizliği ve Büyüme", Uluslararası İktisadi ve İdari İncelemeler Dergisi, Sayı:1.

Balestra, P., Sevestre, P., 1996, "The Econometrics of Panel Data: Handbook of the Theory with Applications, Kluwer Academic Publishers.

Balshaw, R.F. and Dean, C.B., 2002, "A semiparametric model for the analysis of recurrent-event panel data", Biometrics, Vol:58,2 Pg:324.

Bellemare, C., Melenberg, B. and Soest, A.V., 2002, "Semi-parametric Models For Satisfaction With Income", The Institue for Fiscal Studies Department of Economics, Cemmap Working Paper.

Chamberlain, G., 1992, "Efficiency Bounds for Semiparametric Regresion", Econometrica, Vol:60 No:3.

Chambers, D.L., 2004, "The Semiparametric Analysis of Inequality and Growth Using Panel Data", University of California, Ph.D. Thesis, Riverside.

Chen, Y.Q., Wang, M. and Huang, Y., 2002, "Semiparametric Regression Analysis on Longitudinal Pattern of Recurrent Gap Times", U.C., Berkeley Division of Biostatistics Working Paper Series, paper 112, Bepress.

Cheng, S.C., Weii L.J., 2000, "Inferences for a semiparametric model with panel data", Biometrika, Vol:87,1, pg:89, Britain.

Çağlayan, E., 2002, "Yarıparametrik Regresyon Modelleri İle Yaşam Boyu Sürekli Gelir Hipotezinin Türkiye Uygulaması", Marmara Üniversitesi, Doktora Tezi, Sosyal Bilimler Enstitüsü, İstanbul.

Daniel, W.W., 1990, "Applied Nonparametric Statistics" Georgia State University, Boston, 2nd edition, p: 18-20, 426-443.

Darnell, A.C., 1994, "A Dictionary of Econometrics", Edward Elgar Publishing. U.K.

Fox, J., 1997, "Applied Regression Analysis, Linear Models and Related Models", Sage Publications, California.

Fox, J., 2002, "Nonparametric Regression", Thousand Oaks CA:Sage.

Frees, E.W., 2004, "Longitudinal and Panel Data, Analysis and applications in the Social Sciences", Cambridge University Pres, Cambridge, U.K.

Göktaş Yılmaz, Y., 2005, "Türkiye Ekonomisinde Büyüme ile İşsizlik oranı Arasındaki Nedensellik İlişkisi", Ekonometri ve İstatistik Dergisi Sayı:2, İ.Ü. İstanbul Türkiye.

Gujarati, D.N., 2003, "Basic Econometrics", McGraw-Hill Companies Inc. United States.

Hardle, W., 1994, "Applied Nonparametric Regression", Humboldt Univ., Berlin Germany.

Hardle, W., Muller, M., 1997, "Multivariate and Semiparametric Kernel Regression", Universitat zu Berlin, Germany.

Huh, I., 2004, "Essays on Semiparametric Estimation and Testing", University of Rochester, Ph.D. Thesis, New York.

Hussain, S.S., Sprent, P., 1983, "Non-parametric Regression", J.R. Statist. Soc. A, pp, 182-191.

Kaki, B., Yeşilova, A. ve Şen, C., 2004, "Yarıparametrik regresyon yönteminin hayvancılıkta kullanılması", 4. Ulusal Zeoteknik Bilim Kongresi bildiri kitabı, Süleyman Demirel Üniversitesi. Isparta.

Kıldea, D.G., Melbourne, C.S.I.R.O., 1979, "Brown-Mood type Median Estimators for Sımple Regression Models", Annals of Statistics, vol.9, pp 438-442.

Kumar, S., 2000, "Semiparametric Dynamic Panel Data Model: An Application to Cross Country Income Dynamics". University of California, Ph.D. Thesis, Riverside.

Lee, K., 2001, "Essays on Semiparametric Estimation", Louisiana State University, Ph.D. Thesis, Louisiana.

Liu, Z., 1998, "Nonparametric and Semiparametric Estimation and Testing of Econometric Models", The University of Guelph, Ph.D. Thesis.

Manski, C.F., 1987, " Semiparametric Analysis of Random Effects Linear Models From Binary Panel Data", Econometrica, Vol:55, No:2, Pg:357.

Montgomery, D.C. and Peck, E.A., 1992, "Introduction to Linear Regression Analysis", 2 nd. John Wiley and Sons, New York.

Mukherjee, D., 2002, "Nonparametric and Semiparametric Generalized Panel Data Analysis of Convergence and Growth", University of California, Ph.D. Thesis, Riverside.

Müller, M., 2000, “Semiparametric Extensions to Generalized Linear Models”, Humboldt Universitat, Berlin.

Okur, S., 2009, “Parametrik Ve Parametrik Olmayan Basit Doğrusal Regresyon Analiz Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi” Yüksek Lisans Tezi, Çukurova Ünv. Fen Bilimleri Enstitüsü.

Pagan, A., Ullah, A., 1999, “Nonparametric Econometrics”, Cambridge Unıversity Press, Cambridge, U.K.

Racine, J.S., Qui, L., 2007. "A Partially Linear Kernel Estimator for Categorical Data". Unpublished Manuscript, Mcmaster University.

Robinson, P.M., 1988, “Root-N Consistent Semiparametric Regression”, Econometrica, Vol56, No:4 931-954.

Roy, N., 1997, “Nonparametric and Semiparametrc Analysis of Panel Data Models: An Application to Calorie-Income Relation for Rural South India”, University of California, Ph.D. Thesis, Riverside.

Ruppert, D., Wand,M.P. and Carrol, R.J., 2003, “Semiparametric Regression”, Cambridge University Press, Cambridge.

Sayyan, H., 2000, “Dinamik Panel Veri Modelleri Ve OECD Ülkeleri Para Talebi Uygulaması”, Marmara Üniversitesi, Doktora Tezi, Sosyal Bilimler Enstitüsü, İstanbul.

Silverman, B. W., 1990, “Density Estimation for Statistics and Data Analysis”, Chapman and Hall, London.

Simonoff, J, S., 1996, “Smoothing methods in Statistics”, Springer-Verlag, New York.

Sverchkov, M., Pfeffermann, D., 1999, “Parametric and Semi-parametric Estimation of Regression Models Fitted to Survey Data”, The Indian Journal of Statistics, Vol:61, pg:166.

Thompson, J.,R., Tapıa, R.A., 1990, “Nonparametric Function Estimation, modelling and Simulation”, SIAM, Philadelphia.

Wasserman, L., 2006, “All of Nonparametric Statistics”, Springer Science+Business Media Inc, USA.

Yapıcı Pehlivan, N., 2005, “Parametrik Olmayan Regresyonda Alternatif Tahmin Ediciler”, Selçuk Üniversitesi, Doktora Tezi, Konya.

Yatchew, A., 2003, “Semiparametrik Regression for the Applied Econometrician”, Cambridge University Pres, Cambridge.