

REGRESYON SÜREKSİZLİĞİ YÖNTEMİNDE PARAMETRİK OLMAYAN TAHMİN

NON-PARAMETRIC ESTIMATION IN REGRESSION DISCONTINUITY DESIGN

SERDAR DİNDAŞ

PROF. DR. SEVİL BACANLI

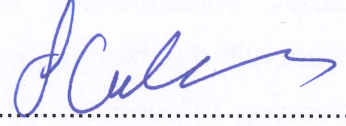
Tez Danışmanı

Hacettepe Üniversitesi
Lisansüstü Eğitim - Öğretim ve Sınav Yönetmeliğinin
İstatistik Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ
olarak hazırlanmıştır.

SERDAR DİNDAŞ'ın hazırladığı “**Regresyon Süreksizliği Yönteminde Parametrik Olmayan Tahmin**” adlı bu çalışma aşağıdaki jüri tarafından **İSTATİSTİK ANABİLİM DALI**'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

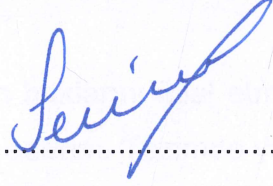
Prof. Dr. Serpil CULA

Başkan



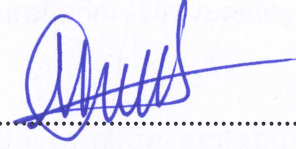
Prof. Dr. Sevil BACANLI

Danışman



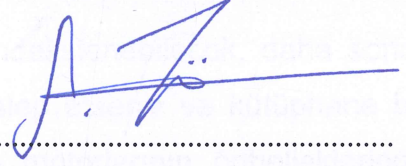
Prof. Dr. Meral ÇETİN

Üye



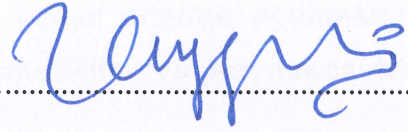
Doç. Dr. Semra TÜRKAN

Üye



Doç. Dr. Duygu İÇEN

Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU

Fen Bilimleri Enstitüsü Müdürü

YAYINLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

- ☐ Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etmeniz ve kütüphane bu talebinizi yerine getirirse bile, tezinin arama motorlarının önbelleklerinde kalmaya devam edebilecektir.)

- ☒ Tezimin/Raporumun 11/06/2020 tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı ve ya tamamının fotokopisi alınabilir)

- ☐ Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum, ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.

- ☐ Serbest Seçenek/Yazarın Seçimi

11 / 06 / 2018

Serdar DİNDAŞ



ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

11.06/2018

Serdar DİNDAŞ



ÖZET

REGRESYON SÜREKSİZLİĞİ YÖNTEMİNDE PARAMETRİK OLMAYAN TAHMİN

Serdar DİNDAŞ

Yüksek Lisans, İstatistik Bölümü

Tez Danışmanı: Prof. Dr. Sevil BACANLI

Haziran 2018, 81 sayfa

Regresyon süreksizlik yöntemi kontrol ve etki gruplarının önceden oluşturulmadığı durumlarda kullanılan yarı deneysel bir etki değerlendirme yöntemidir. Yöntemde kontrol grubu ve etki grubu “eşik değeri” olarak adlandırılan bir kritere göre belirlenmektedir. Eşik değeri çevresinde yer alan bireyler, programın uygulanması dışında bir farkı olmadığı, birbirlerine en yakın özelliğe sahip oldukları varsayılmaktadır. Bireylerin etki ya da kontrol grubuna atanma oranı yöntemin tasarımı belirlemektedir. Eşik değerine göre konumu bireyin atamasında tamamen belirleyici olduğu tasarım Keskin regresyon süreksizliği tasarımıdır. Süreksizliğin ilk adımda görsel olarak tespit edilmesi için grafiksel gösterimden faydalanılmaktadır. Programın etkisi hesaplanırken eşik değeri çevresinde ne kadar gözlemin yöneme dahil edileceğini gösteren bant genişliği hesaplanmaktadır. Bant genişliği içerisinde, eşik değerinin altında ve üstünde yer alan bireylere ait regresyon modelleri tahmin edilerek eşik değerinde bir süreksizlik ya da kırılmanın olup olmadığı incelenmektedir. Kırılma, uygulanan programın etkisi olarak değerlendirilmektedir.

Anahtar Kelimeler: Regresyon Süreksizliği Yöntemi, Yarı Deneysel Etki Değerlendirme Yöntemi

ABSTRACT

NON-PARAMETRIC ESTIMATION IN REGRESSION DISCONTINUITY DESIGN

Serdar DİNDAS

Master of Science, Department of Statistics

Supervisor: Prof. Dr. Sevil BACANLI

June 2018, 81 pages

Regression discontinuity is a quasi-experimental impact evaluation method of assessment which is used in cases where control and effect groups can not be generated before the implementation of the program. In this method, the control group and effect group are determined according to a criterion called the threshold. Individuals, which close to threshold value are assumed to have the closest characteristic to each other, except for the application of the program. The proportion of individuals assigned to an effect or control groups determines the design of the method. In regression discontinuity design, the assignment is completely based on individuals position according to threshold. Graphical representation is used to visualize discontinuity in the first step. As the effect of the program is evaluated, the bandwidth is calculated to inclusion of how many individuals are close to the threshold to analyze. Within this bandwidth, the regression models of the individuals above and below the threshold are estimated in the threshold value, whether there is a break or discontinuity. Discontinuity is considered as the effect of the applied program.

Keywords: Regression Discontinuity Design, Quasi Experiment Impact Evulation Method

TEŞEKKÜR

Tez çalışmam sırasında bilgi, birikim ve tecrübeleri ile bana yol gösterici ve destek olan değerli danışman hocam Sayın Prof. Dr. Sevil BACANLI 'ya sonsuz teşekkür ve saygılarımı sunarım.

Değerli çalışma arkadaşlarım Ercan ORHAN, Hülya BAŞESEN ve Ahmet PEMBEGÜL'e

Çalışmalarım boyunca manevi destekleriyle beni hiçbir zaman yalnız bırakmayan yol arkadaşım değerli eşim Zeliha DİNDAŞ'a,

Çocukluğumdan bugüne dek hayatımın her anında yanımda olan çok kıymetli babam Mehmet DİNDAŞ'a, annem Fehminaz DİNDAŞ'a, teyzem Belgin KAFTAN'a ve kardeşim Nesli DİNDAŞ'a sonsuz teşekkürler ederim.

İÇİNDEKİLER

Sayfa

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR.....	iii
İÇİNDEKİLER.....	iv
ÇİZELGELER.....	vi
ŞEKİLLER	vii
KISALTMALAR.....	viii
1. GİRİŞ	1
2. GENEL BİLGİLER	4
2.1 Etki ve Etki Değerlendirmesi Kavramları	4
2.1.1 Etki Değerlendirmesinde Karşıt Gerçeklik.....	6
2.2. Etki Değerlendirme Yöntemleri.....	7
2.2.1. Deneysel Yöntem.....	8
2.2.2. Yarı Deneysel Yöntemler	9
3. REGRESYON SÜREKSİZLİĞİ YÖNTEMİ	16
3.1 Tanım ve Kavramlar	16
3.1.1 Analizin İçsel Geçerliliği İçin Varsayımlar.....	20
3.1.2 Regresyon Süreksizliği Yöntemi ve Potansiyel Çıktı Çerçevesi	22
3.1.3 Regresyon Süreksizliği Yöntemi ve Deneysel Yöntemin Karşılaştırılması	24
3.2 Regresyon Süreksizliği Yönteminin Uygulanmasında Yapılan Çalışmalar ..	27
3.3 Grafiksel Gösterim	32
3.3.1 Optimal Grup Sayısının Belirlenmesi	35
3.4 Regresyon Süreksizliği Yönteminde Tahmin.....	38
3.4.1 Regresyon Süreksizliği Yönteminde Model ve Tahmin Yöntemi Seçimi	40
4. SAYISAL ÖRNEK	52
4.1 Veri.....	53
4.2 Grafiksel Gösterim	54
4.3 Regresyon Süreksizliği Yönteminde Tahmin Sonuçları.....	56

5. SONUÇ VE TARTIŞMA.....	61
KAYNAKLAR.....	64
EKLER	67
EK 1: R kodları.....	67
ÖZGEÇMİŞ	69



ÇİZELGELER

Sayfa

Çizelge 1: Farkın Farkı Yöntemi.....	12
Çizelge 2 : Çekirdek Fonksiyonlar.....	46
Çizelge 3 : Tanımlayıcı İstatistikler	54
Çizelge 4 : Optimal Grup Genişliği	56
Çizelge 5 : Parametrik Olmayan Regresyon Süreksizliği Tahmin Sonuçları	57



ŞEKİLLER

Sayfa

Şekil 1: Rasgele Dağıtım Yöntemi ile Etki Değerlendirmesi	8
Şekil 2: Program Uygulanmadan Önce Regresyon Süreksizlik Yöntemi	13
Şekil 3: Program Uygulandıktan Sonra Regresyon Süreksizlik Yöntemi.....	13
Şekil 4: Keskin Regresyon Süreksizliği Yöntemi Tasarımı	17
Şekil 5: Fuzzy Regresyon Süreksizliği Yöntemi Tasarımı.....	18
Şekil 6: Mali Yardım Programından Önce Gelir Endeksi – Günlük Gıda Harcamaları Arasındaki İlişki	18
Şekil 7: Mali Yardım Programına Uygun Olan-Olmayan Bireyler	19
Şekil 8: Mali Yardım Programından Sonra Gelir Endeksi – Günlük Gıda Harcamaları Arasındaki İlişki	20
Şekil 9: Regresyon Süreksizliği Yönteminde Ortalama Beklenen Değerler	23
Şekil 10: Rasgele Dağıtım Yöntemi ile Program Etkisi	26
Şekil 11: Regresyon Süreksizliği Yöntemi ile Program Etkisi	27
Şekil 12: Ham Verinin Saçılım Grafiği	33
Şekil 13: Gruplandırılmış Ortalamaya Ait Verinin Saçılım Grafiği	33
Şekil 14: Hatalı Model Tahminin Oluşturduğu Süreksizlik.....	39
Şekil 15: Parametrik Tahmin ve Parametrik Olmayan Tahminin Yaklaşımları	40
Şekil 16: Farklı Bant Genişliğinin Model Tahminine Etkisi	47
Şekil 17: Çapraz Geçerlilik Yöntemi	48
Şekil 18: Ham Verilerin Grafikselsel Gösterimi	55
Şekil 19: Optimal Grup Genişliğinin Grafikselsel Gösterimi	56
Şekil 20: Plug-in Yöntemi İle 1. Dereceden Yerel Polinom Modeline Ait Grafikselsel Gösterim.....	58
Şekil 21: Plug-in Yöntemi İle 2. Dereceden Yerel Polinom Modeline Ait Grafikselsel Gösterim.....	58
Şekil 22: Çapraz Geçerlilik Yöntemi İle 1. Dereceden Yerel Polinom Modeline Ait Grafikselsel Gösterim	59
Şekil 23: Çapraz Geçerlilik Yöntemi İle 2. Dereceden Yerel Polinom Modeline Ait Grafikselsel Gösterim	59

KISALTMALAR

Kısaltmalar

BHKO

Birleřtirilmiř Hata Kareler Ortalaması



1. GİRİŞ

Etki değerlendirme çalışmaları uygulanan bir programın bireyler üzerinde etkisi olup olmadığını araştırmaktadır. Program uygulayıcıları tarafından etkinin ölçülmesi için oluşturulan etki grubu ve kontrol grubunun program sonrası gözlemlenebilir özelliklerinin karşılaştırılmasına dayanmaktadır. Dolayısıyla, etki değerlendirme çalışmalarında gruplar önceden oluşturulamamaktadır.

Regresyon süreksizliği yöntemi etki değerlendirme çalışmalarında kontrol ve etki gruplarının önceden oluşturulamadığı durumlarda kullanılan bir yarı deneysel etki değerlendirme yöntemidir. Yöntemde etki değerlendirmesi yapılırken kontrol grubu ve etki grubunun belirlenmesinde eşik değeri olarak adlandırılan bir kriter kullanılmaktadır. Kriter bir gruba ait gözlemlenebilir bir özelliğin aldığı değerler arasından seçilir. Seçilen kriterin altında kalan gözlemler ve üstünde kalan gözlemler ile etki ve kontrol grubu oluşturulmaktadır. Endeks değişkeni gruba ait gözlemlenebilir bir özelliğin tanımlandığı sürekli bir değişkendir. Sonuç değişkeni, endeks değişkeni ile ilişkili bir değişken olup hedeflenen etki ile değişmesi beklenmektedir.

Regresyon süreksizliği yöntemi, regresyon ve süreksizlik kavramlarını içermektedir. Yöntem uygulanırken, endeks değişkeni bağımsız değişken ve sonuç değişkeni bağımlı değişken olmaktadır. Eşik değerinin altında ve üstünde iki değişken kullanılarak ayrı ayrı tahmin edilecek regresyon modellerinin, eşik değeri üzerinde alacağı değerler tespit edilerek eşik değerinde bir süreksizlik ya da kırılmanın olup olmadığı incelenmektedir. Oluşabilecek kırılmanın uygulanan etkiden ya da etki grubunun kontrol grubundan farklı olarak sahip olduğu özellikten kaynaklandığı düşünülmektedir.

Bireylerin etki grubuna ya da kontrol grubuna hangi oranda atanacağı regresyon süreksizliği yönteminin tasarımını belirler. Keskin regresyon süreksizliği yöntemi tasarımında eşik değerinin altında yer alan bireyler ya da üstünde yer alan bireylerin tamamı sadece bir gruba atanırken, fuzzy regresyon süreksizliği yöntemi tasarımında eşik değerinin hem altında hem de eşik eğerinin üstünde yer alan bireyler oransal olarak etki grubuna ya da kontrol grubuna atanmaktadır.

Regresyon süreksizliği yönteminin çerçevesi ilk olarak Thistlethwaite ve Campbell [1] çalışmasında oluşturulmuştur. Regresyon süreksizliği yöntemi, araştırmacıların kontrol grubu ve etki grubunu rasgele dağılım ile seçemediği durumlarda kullanabileceği bir yöntem olarak önerilmiş ve uygulanan bir burs programının öğrencilerin akademik çıktılarına etkisini incelemek amacıyla kullanılmıştır

Regresyon süreksizliği yönteminin uygulanmasında grafiksel gösterim ilk ve önemli bir adımdır. Bu adımda ham verinin karmaşasından kurtularak verinin daha basitçe tanımlanması sağlanmaktadır. Kolay gözlemlenebilir noktalara indirgenen veri ile eşik değeri üzerinde bir kırılma olup olmadığı tespit edilebilir. Grafiksel gösterimde bu kırılma gözlemlenebiliyorsa analizin ilerleyen bölümlerinde uygulanan istatistiksel yöntemlerde de bu kırılma tespit edilir, fakat bu kırılma grafiksel gösterimde ilk aşamada tespit edilemiyorsa uygulanacak istatistiksel yöntemlerde anlamlı bir kırılmanın tespit edilmesi zor olabilmektedir.

Regresyon süreksizliği yöntemi uygulanırken iki tahmin yönteminden birisi tercih edilmektedir. Bu yöntemler, parametrik tahmin yöntemi ve parametrik olmayan tahmin yöntemidir. Tahmin yöntemleri sadece kullanıldığı veri sayısına göre farklılık göstermektedir. Regresyon süreksizliği yönteminde parametrik tahmin'de bütün veri setinden faydalanılır. Eşik değerine en uzak olan verilerinde yöntemde dahil edilmesi sağlanmaktadır. Böylece tahmin edilecek regresyon modeli bütün veriye ait bilgiyi taşıyacaktır, fakat uygulamalarda bütün verinin kullanılması yüksek dereceden regresyon modellerinin tahmin edilmesine sebep olmakta ve analizin gücünün düştüğü düşünülmektedir.

Parametrik olmayan regreyon süreksizliği yöntemi etki değerlendirme mantığına daha çok uyan bir yöntemdir. Bu yöntemde eşik değerinin hemen altında ve hemen üstünde yer alan veriler karşılaştırılmaktadır. Böylece etki grubu ve kontrol grubu birbirine en yakın özellikteki bireylerden oluşacaktır. Parametrik olmayan regresyon süreksizliği tahmininde 3 önemli bileşen bulunmaktadır. Bu bileşenler, yöntemde eşik değerinin altında ve üstünde olmak üzere ne kadarlık bir pencerede yer alan verilerin dahil olacağını belirleyen bant genişliği, bu bant genişliği içerisinde yer alan gözlemlerin eşik değerine göre konumlarının ağırlıklandırılmasını sağlayan çekirdek fonksiyonu ve bu bant genişliği içerisinde tahmin edilecek yerel polinom regresyon modelinin derecesidir.

Tez çalışması beş ana bölümden oluşmaktadır. Birinci bölümde regresyon süreksizliği yöntemine dair kısa bir giriş yapıldıktan sonra ikinci bölümde, etki değerlendirme kavramları ve etki değerlendirme yöntemleri incelenmiştir. Üçüncü bölümde regresyon süreksizliği yöntemine ait kavramlar ve açıklamaları grafikler üzerinde anlatılmış, analizin geçerlilik koşullarından bahsedilmiştir. Aynı bölüm altında grafiksel gösterim ve gösterimde yer alan uygun grup sayısının belirlenmesine yönelik yöntemlerden bahsedilmiş, parametrik ve parametrik olmayan yöntemler kullanılarak regresyon süreksizliğinde tahmin yöntemleri ve bu yöntemlere ait bileşenlerin hesaplanması için önerilen yöntemler incelenmiştir. Dördüncü bölümde seçim sonuçlarına ait bir veri seti üzerinde [2] bir partinin iktidar olma avantajının sonraki seçimdeki etkisi R programında regresyon süreksizliği yönteminden faydalanılarak incelenmiştir. Son bölümde ise analizin sonuçları özetlenmiş ve yorumlanmıştır.

2. GENEL BİLGİLER

2.1 Etki ve Etki Değerlendirmesi Kavramları

Etki; “Bir kimse veya nesnenin başka bir kişi veya başka bir şey üzerindeki gücü veya tesiri” , “Nedensellik bağlantısı içerisinde, bir nedenin sonucu olarak düşünülen olay” veya “Bir etken veya sebebin sonucu” olarak tanımlanmaktadır. Buradan yola çıkarak etki, bir müdahale sonucunda doğrudan veya dolaylı, amaçlanmış veya amaç dışı olarak üretilen olumlu veya olumsuz sonuçlardır [3].

Etki değerlendirmenin literatürde kullanılan birçok tanımı mevcuttur. Etki değerlendirme temel olarak müdahale edilen durum ile müdahale edilmeyen durum arasındaki net fark olarak tanımlanmaktadır [4]. Gerçekleşen etki sonucunda gözlemlenen sonuçların müdahaleden kaynaklı olup olmadığı, etkinin olmadığı durumda ne olacağının tahmini değerinin karşılaştırıldığı bir sonuç değerlendirmesidir.

Etkin olan programa göre oluşturulacak politikalar; kamu kaynaklarının doğru kullanımı, uygulanan programların daha iyi tasarlanması ve ihtiyacı karşılayacak doğru politikaların uygulanması açısından son derece önemlidir. Bu noktada belirlenen probleme göre oluşturulan politikaların etki değerlendirmesi, politika oluşturma sürecinin önemli bir basamağıdır [5].

Etki değerlendirmesi çalışmaları bir programın hedef kitlede arzulanan etkileri oluşturup oluşturmadığına ve hedef kitlede gözlenen değişikliklerin ne kadarının programa dayandırılabilceğini belirlemeyi amaçlar [6].

Program, bir hedef doğrultusunda, program katılımcısı bireyler tarafından oluşturulan etki grubu olarak adlandırılan gruba uygulanan müdahale ya da etki olarak tanımlanabilir. Müdahale sonucunda bireylerde gözlenen değişime ise çıktı ya da sonuç denilmektedir.

Etki değerlendirme günümüzde politika belirleyicileri, program yöneticileri, akademisyenler tarafından kullanılan kanıta dayalı politika belirmenin önemli bir parçası ve iktisadi değerlendirme aracıdır. Araştırmalarda girdi ve çıktı arasındaki ilişkinin değerlendirilmesi önemli bir noktadır. Özellikle başta kamu hizmeti sağlayıcıları olmak üzere bu girdi-çıkıtı ilişkisinin gözetilmesi, kamusal politikaların nasıl uygulanması gerektiğini de şekillendirmekte ve kısıtlı girdi ile en iyi sonucu

veren politikaya ulaşmak hedeflenmektedir. Etki değerlendirmede sonuçlar sadece bir sonraki hedeflerin belirlenmesi için kullanılmamakta aynı zamanda yöneticiler tarafından açıklanabilirliği artırmak, kaynakların uygun bir şekilde dağılımını sağlamak ve sürdürülebilir politika belirlemek amacıyla da kullanılmaktadır.

İzleme ve değerlendirme kanıta dayalı politika belirlenmesinin temelini oluşturmaktadır. İzleme, uygulanan bir program boyunca neler olduğunun takip edildiği ve ileride değerlendirme yapılmak üzere verilerin toplandığı sürekli bir süreçtir. Burada elde edilen veriler bilgilendirme, yönetim ve karar verme amacıyla kullanılır. Değerlendirme, planlanan, devam eden ya da tamamlanmış program, proje veya politikaların periyodik ve objektif olarak incelenmesidir. İzlemeden farklı olarak değerlendirme belirli zamanlarda yapılır ve teknik uzmanların bilgisinden yararlanır. Değerlendirme tasarımı ve yöntemleri, programa yönelik cevaplanması istenen sorular çerçevesinde değişiklik gösterebilir. Bu soruları, programın ne olduğu; amacının ve önceliklerinin değerlendirildiği tanımlayıcı sorular, programın güncel durumu ile hedeflenen durumunun karşılaştırıldığı normatif sorular ve program sonucunda uygulanan müdahalenin çıktılar üzerindeki etkisinin incelendiği neden-sonuç ilişkisini içeren sorular olmak üzere üç başlık altında toplanabilir. Etki değerlendirme, neden-sonuç ilişkisine yönelik soruların cevaplarının arandığı değerlendirme yöntemidir. Etki değerlendirmede bir soru üzerine yoğunlaşılır: “Programın ilgilenilen çıktılar üzerindeki doğrudan etkisi nedir?”. Programın bireyler üzerindeki etkisinin tahmin edebilmesi için kullanılacak herhangi bir yöntemde, program katılımcısının programa katıldığı duruma ait çıktılara karşılık programa katılmadığı duruma ait çıktılarında biliniyor olması gerekmektedir [7].

Literatürde etki değerlendirme yöntemleri temel olarak nitel (qualitative) ve nicel (quantative) yöntemler olmak üzere ikiye ayrılır. Nitel yöntemler; uzman panelleri, vaka çalışmaları, alan araştırması, ağ analizleri, teknoloji değerlendirmeleri ve kıyaslama gibi yöntemler iken nicel yöntemler; etkisi ölçülecek olan program ya da politikanın niteliğine göre makro ya da mikro ekonometrik ölçüm yöntemlerini içermektedir [8].

Etki değerlendirme amacına göre uygulama zamanı değişkenlik gösterir. Bu kapsamda nicel etki değerlendirme yöntemleri ikiye ayrılır, bunlar: Ön

değerlendirme (ex-ante evaluation), ve son değerlendirme (ex-post evaluation) olarak ifade edilmektedir. Ön değerlendirme, program uygulanmadan önce yapılmaktadır. Güncel duruma ait veriler kullanılarak ileride uygulanması planlanan programın olası çıktıları ölçülmeye çalışılmaktadır. Güncel durum yapısal ekonomik modeller ile modellenir ve programın olası çıktılarının simülasyonu yapılır. Uygulanması planlanan bir mali yardım programının çocukların okula katılım oranına etkisinin program uygulanmadan önce incelenmesi bir ön değerlendirme örneğidir.

Son değerlendirme program tamamlandıktan belli bir süre sonrasında uygulanır ve programın oluşturduğu gerçek etkiyi belirlemeye olanak sağlar. Son değerlendirme, uygulanan programa katılan ve katılmayan bireylere ait veri toplanmasını gerektirdiğinden ön değerlendirmeye göre daha maliyetlidir. Ön değerlendirmede hesaplanmış ve bulunmuş bir program etkisi her zaman son değerlendirmede bulunmayabilir [9].

2.1.1 Etki Değerlendirmesinde Karşıt Gerçeklik

Etki değerlendirme çalışmalarının temel amacı programın katılımcılar üzerindeki nedensel etkisini ölçmektir. Programın etkisi, program katılımcısının programa katıldığı ve katılmadığı durumlara ait çıktıların karşılaştırılmasına dayanmaktadır. Fakat bu ölçüm sırasında karşılaşılan önemli zorluklardan birisi, bireyin hem programa katıldığı hemde programa katılmadığı durumun aynı anda gözlemlenememesidir. Bu durumda etki değerlendirmesinde programa katılan grubu diğer bir ifadeyle etki grubunu karşılaştırmak amacıyla makul ve ikna edici bir kontrol grubu yani karşıt gerçeği (counterfactual) oluşturulmalıdır [9].

Programın herhangi bir Y sonuç değişkeni üzerinde etkisi, Eşitlik 2.1’de gösterildiği şekilde ifade edebilir:

$$\alpha = (Y|P = 1) - (Y|P = 0) \quad (2.1)$$

Eşitlik 2.1’de P=1 programa katılma durumunu, P=0 ise programa katılmama durumunu temsil etmektedir. (α) ise programın nedensel etkisini göstermektedir [7].

Katılımcının programa katıldığı ve katılmadığı durumun aynı anda gözlemlenememesi “karşıt gerçeklik problemini” oluşturmaktadır. Uygulanan bir

programa katılan bireyin program sonucunda elde ettiđi çıktı kolaylıkla gözlemlenebilir. Dolayısıyla Eşitlik 2.1’de ilk bölümde yer alan ($Y/P=1$) terimi gerçek ve gözlemlenebilir bir veridir. Eşitlik 2.1’de yer alan ($Y/P=0$), katılımcının programa katılmasaydı elde edeceği çıktıyı temsil etmektedir. Bu durum doğrudan gözlemlenemediđi için bu durumun yerine uygun bir kontrol grubu tahmin edilmelidir.

Kontrol grubunun belirlenmesi etki analizinde önemli bir yere sahiptir. Başarılı bir etki değeriendirmesinin önemli bir parçası da uygun kontrol grubunun bulunmasıdır. Araştırmacılar kontrol grubunun tahmin edilmesinde iki yaklaşım üzerinde durmaktadırlar. Bunlardan birincisi program henüz uygulanmadan önce etki alacak grup ile kontrol grubunun birbirinden farklılık göstermeyecek şekilde belirlenmesi, ikincisi ise kontrol grubunun istatistiksel yöntemlerden faydalanılarak bulunmasıdır. İleride değeriilecek etki değeriendirme yöntemlerinin her biri kontrol grubunun belirlenmesinde farklı yöntemler önermektedir [9].

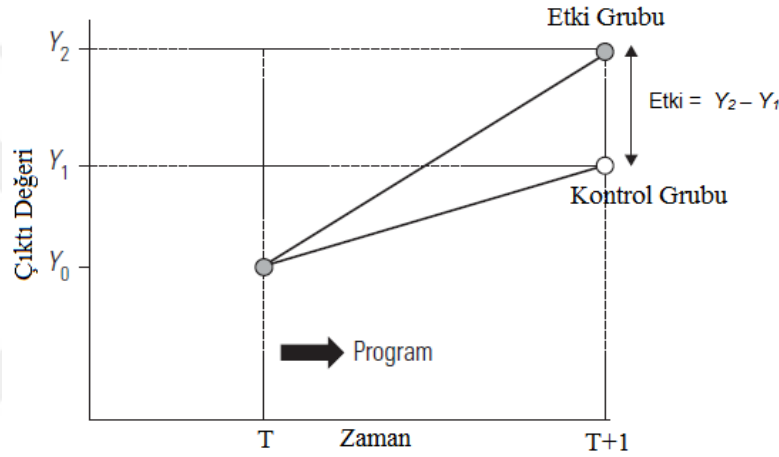
Etki grubunun belirlenmesi her zaman rasgele olmayabilir. Bunun sebebi programın belirli bir amaca yönelik olması ya da program katılımcılarının seçiminin program yöneticileri tarafından yapılmasıdır. Programa katılımı belirleyen etkenlerin programın sonucunda elde edilen çıktılar ile ilişkili olması seçim yanlılığına sebep olmaktadır. Yanlı seçimler ile oluşturulmuş bu durum programın etkisinin çok düşük ya da çok yüksek olmasına sebep olabilmektedir. Bu sorun sebebiyle istatistiksel olarak anlamlı olan etkinin anlamsız ya da tersi durumunun oluşturduğu sonuçlara yol açar. Kontrol grubu ile etki grubunun benzer özellikte olması dışında müdahaleye aynı tepkiyi göstermesi beklenmektedir. Son olarak değeriendirme sırasında müdahale grubuna belirlenen politika etkisi dışında bir etki uygulanmamalıdır [9].

2.2. Etki Değeriendirme Yöntemleri

Etki değeriendirme yapılması için kontrol grubunun belirlenmesine yönelik birçok yöntem bulunmaktadır. Her yöntemin, program etkisini tahmin edebilmek amacıyla oluşturulan kontrol grubunun seçim yanlılığından olabildiğince uzaklaşması için kendine ait farklı uygulamaları vardır. Yöntemler kesim 2.2.1 ve kesim 2.2.2’de detaylı olarak incelenmiştir.

2.2.1. Deneysel Yöntem

Rasgeleleştirme ya da rasgele dağıtım olarak da bilinen deneysel yöntem (Experimental Techniques) genellikle etki değerlendirme yöntemlerinin en gelişmiş olarak düşünülmektedir. Rasgele dağıtım iki aşama üzerine kurulmuştur. İlk aşamada henüz program düzenlenmeden önce, programdan faydalanmaya uygun bireylerin oluşturduğu kitleden, rasgele seçim yöntemiyle, kitleyi temsil edebilecek büyüklükte bir örneklem seçilir. İkinci aşamada ise ilk aşamada belirlenen örneklemde, istatistiksel olarak benzer özellik gösteren etki grubu ve kontrol grubu seçilir. Etki değerlendirmesi bu iki grubun karşılaştırılmasına dayanmaktadır [4], [9].



Şekil 1: Rasgele Dağıtım Yöntemi ile Etki Değerlendirmesi

Rasgele dağıtım yöntemi ile etki değerlendirmesi Şekil 1’de görsel olarak ifade edilmiştir. Rasgele dağıtımın ilk aşamasında kitleden örneklem seçildikten sonra ikinci aşamaya geçilir [9]. İkinci aşama, Şekil 1’de T zamanında, program başlamadan önce, rasgele seçim ile etki grubunun oluşturulması ile başlar. Örneklemde seçilmeyen bireyler ise kontrol grubunu oluşturur. T ile T+1 aralığı boyunca etki grubuna program uygulanmaktadır. T+1 zamanında etki uygulanan grubun çıktı değeri Y₀’dan Y₂ seviyesine gelirken, kontrol grubunun çıktı değeri Y₀’dan Y₁ seviyesine gelmiştir. Şekil 1’de görüldüğü üzere programın tahmin edilen etkisi doğrudan Y₂ – Y₁ arasındaki farka eşittir [9].

Rasgele dağıtımın ilk aşamasında kitleyi temsil edecek örneklem seçilmesi önemlidir. Kitleyi temsil eden uygun büyüklükte örneklem seçilmesi ile rasgele dağıtım yönteminin dış geçerliliği (external validity) sağlanmış olur. Dış geçerliliğin anlamı, tahmin edilen etkinin aynı zamanda kitle içinde geçerli olmasıdır. Rasgele

dağıtımın ikinci aşamasında uygulanan rasgele seçim ile örneklem etki grubu ve kontrol grubu olarak ikiye ayrılmaktadır. Aynı özellikleri taşıyan örneklem bireylerinin etki ve kontrol gruplarına rasgele atanması ile rasgele dağıtım yönteminin iç geçerliliği (internal validity) sağlanmaktadır. Etki ve kontrol gruplarının program uygulanmadan benzer özellik gösterdiği varsayıldığında, kontrol grubunun, etki grubunun tam karşıt gerçekliği olduğu kabul edilmektedir. Böylece kontrol grubuna ait çıktının zamanla değişimi ile etki grubuna ait çıktının zamanla değişimi arasındaki tek fark programın uygulanması olacaktır. Rasgele dağıtım yönteminin bu özelliğinden dolayı etki doğrudan tam olarak hesaplanabilmektedir [7].

Rasgele dağıtım yöntemi dış geçerlilik ve etik açıdan bazı dezavantajlara sahiptir. Dış geçerlilik açısından, kitleden seçilen ve kitleyi en uygun temsil eden örneklemin büyüklüğünün fazla olması etki uygulanacak birey sayısının da artacağı anlamına gelmektedir. Program uygulanacak birey sayısının artması programın maliyetini de arttıracaktır. Genellikle gelişmekte olan ülkelerde yoksulluk vb. konulara yönelik uygulanan bazı programların rasgele dağıtım yöntemi ile uygulanması etik açıdan tartışmalı bulunmaktadır. Aynı yoksulluk ve gelir grubuna ait bireyler arasından rasgele dağıtım yöntemi ile programa seçilen bireyler ile seçilmeyen bireyler arasında bir adaletsizlik olduğu düşünülebilir.

2.2.2. Yarı Deneysel Yöntemler

Yarı deneysel yöntemler, rasgele dağıtım yönteminde olduğu üzere etki ve kontrol grubunun doğrudan oluşturulmadığı durumlarda kullanılmaktadır. Yarı deneysel yöntemler temelde gözlemlenen özellikler bakımından etki grubuna benzeyen kontrol grubu oluşturulmasına dayanır. Yarı deneysel yöntemlerde, etki ve kontrol grubu genellikle program sonrası rasgele olmayan yöntemlerle oluşturulur. Etki ve kontrol grubunun oluşturulması sırasında bu iki grubun olabildiğince benzer olmasını sağlayacak şekilde istatistiksel yöntemler kullanılır [10].

Yarı deneysel yöntemler uygulanabilirliği açısından bazı avantajlara sahiptir. Uygulanan bir program sonrasında elde edilen mevcut veri kullanılarak hızlı ve az maliyetle etki değerlendirme yapılmasına olanak sağlaması en önemli avantajıdır. Diğer bir avantajı, çeşitli istatistiki teknikler kullanılarak birbirine daha benzer müdahaleli ve müdahalesiz grupların seçimine olanak sağlamasıdır. Ayrıca bu

yöntemlerin büyük çaplı örneklemelere uygulanabilmesi ve bunun neticesinde sonuçların karar alıcılar tarafından genelleştirilip kullanılabilmesi de önemli bir avantajdır [4], [10].

İleri istatistiksel yöntemlerin kullanılması, kolay uygulanabilir olmaması yarı deneysel yöntemlerin dezavantajları arasındadır. Dolayısıyla yarı deneysel yöntemlerin uygulanması ve sonuçların yorumlanması önemli ölçüde uzmanlık gerektirmektedir. Yarı deneysel yöntemlerin uygulanmasında karşılaşılan bir diğer problem seçim yanlılığı problemidir. Rasgele dağıtım yönteminde görülmeyen seçim yanlılığı, yarı deneysel yöntemler için önemli bir sorundur [9].

Seçim yanlılığının çözümüne yönelik farklı yaklaşımlar içeren ve literatürde sıkça karşılaşılan yarı deneysel etki değerlendirme yöntemleri:

1. Eğilim Skoru Eşleştirme Yöntemi(Propensity Score Matching-PSM)
 2. Farkın Farkı Yöntemi(Difference in Differences – DD)
 3. Regresyon Süreksizlik Tasarımı (Regression Discontinuity Design – RDD)
- dır [9].

2.2.2.1 Eğilim Skoru Eşleştirme Yöntemi

Eğilim skoru eşleştirme yöntemi teorik olarak her geçen gün geliştirilmekte, uygulamada yarı deneysel yöntemler arasında kullanımı artmaktadır. Ayrıca bu yöntemin klasik regresyon yaklaşımının alternatifi olduğu da belirtilmektedir [4].

Eğilim skoru eşleştirme yönteminde etki değerlendirmesi program uygulandıktan sonra yapılmaktadır. Program etki grubu üzerinde uygulanıp sonuçları elde edildiği için bu sonuçların karşılaştırılacağı ve etki değerlendirmesinin yapılacağı kontrol grubunun oluşturulması gerekmektedir. Yöntem, etki grubunun program uygulanmadan önceki gözlemlenen özelliklerinden faydalanılarak oluşturulan istatistiksel model ile kontrol grubunun oluşturulmasına dayanmaktadır. Kontrol grubunun belirlenmesiyle, program öncesinde benzer özellik gösteren kontrol ve etki grubunun program sonrasında karşılaştırılması doğrudan programın etkisi olarak görülmektedir [9].

Eğilim skoru eşleştirme yöntemine göre, etki değerlendirme, iki aşamada yapılmaktadır. İlk aşamada, programa katılan etki grubunun özelliklerinden faydalanarak, programdan yararlanma olasılığını ortaya koyan bir model

oluşturulur. Model oluşturulduktan sonra, tahmin edilen programdan yararlanma olasılığına eğilim skoru denilmektedir. Bu model kullanılarak etki alan gruba benzer özellik gösteren, birbirine yakın eğilim skoruna sahip bireylerden oluşan bir kontrol grubu oluşturulur. Etki ve kontrol grupları arasındaki benzerlik sonuçların güven düzeyini de arttırmaktadır. İkinci aşamada ise, etki grubunun program sonrasında sahip olduğu çıktı ile kontrol grubunun program uygulanmadan sahip olduğu çıktı değerlerinin farkı alınır. Bu fark müdahalenin etkisi olarak tanımlanır [4].

Müdahale grubuna karşı kontrol grubunun, oluşturulduğu farklı eğilim skoru eşleştirme yöntemleri bulunmaktadır. En yakın skor eşleştirme, etki alanı eşleştirme, tabakalı eşleştirme, çekirdek eşleştirme, koşullu eşleştirme yöntemleridir. Yöntemler regresyon tabanlıdır ve eğilim skoruna dayalı bir ağırlıklandırma kullanılarak etkin sonuçlara ulaşılmaktadır [4], [9].

Eğilim skoru eşleştirme yönteminde, etki grubunun program uygulanmadan önce sahip olduğu gözlemlenebilir özelliklerin kontrol grubu oluşturulmasında kullanılmaktadır. Kontrol grubu ile etki grubu arasında gözlemlenemeyen özellikler bakımından farklılıklar ihmal edilebilir seviyedeysse yöntemi rasgele dağıtım yöntem kadar başarılı bir etki değerlendirmesi sunmaktadır.

2.2.2.2 Farkın Farkı Yöntemi

Farkın farkı yöntemi etki grubu ve kontrol grubunun, program uygulanmadan önce ve program uygulandıktan sonra olmak üzere iki defa karşılaştırılmasına dayanmaktadır. Program uygulanmadan öncesinde kontrol grubu ve etki grubuna gözlemlenebilen özelliklerine ait veri mevcut olmalıdır. Program uygulandıktan sonra kontrol grubu ve etki grubunun bu özellikleri kıyaslanmak üzere iki defa farkı alınarak etki değerlendirmesi yapılır [10].

Etki grubunun program uygulanmadan önce sahip olduğu veriler ile program uygulandıktan sonra elde edilen verilerin farkının alınmasına birinci fark denilmektedir. Birinci fark programın ve program boyunca geçen zamanın etki grubu üzerindeki etkisini göstermektedir. Zamanın etkisi henüz dışlanmadığı için birinci fark programın etkisini doğrudan verememektedir. Zamanın etkisini hesaplayabilmek için kontrol grubuna ait öncesi-sonrası verilerinin farkı alınır ve bu fark ikinci fark olarak adlandırılır.

Birinci farkın içerdiği zamanın etkisinin temizlenmesi için birinci farktan ikinci farkın çıkarılması gerekmektedir. Böylece programın etkisi yanlı olmayan şekilde tahmin edilmektedir [7].

Farkın farkı yöntemine ait alınan fark işlemleri Çizelge 1’de verilmiştir. Çizelge de verilen farklar aşağıda verildiği gibi tanımlanmaktadır. Bu farklar;

Etki Grubundaki Fark (Birinci Fark)	: $Y_1 - Y_0$
Kontrol Grubundaki Fark (İkinci Fark)	: $K_1 - K_0$
Uygulanan Programın Etkisi	: $(Y_1 - Y_0) - (K_1 - K_0)$

Çizelge 1: Farkın Farkı Yöntemi

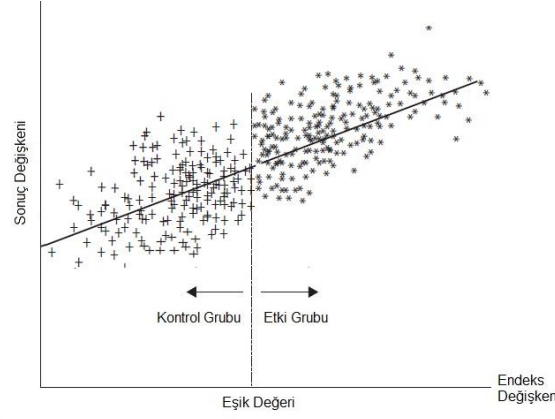
	Program(Etki) Öncesi Değer	Program (Etki) Sonrası Değer	Değişim
Etki Grubu	Y_0	Y_1	$Y_1 - Y_0$
Kontrol Grubu	K_0	K_1	$K_1 - K_0$
Fark		$Y_1 - K_1$	$= (Y_1 - Y_0) - (K_1 - K_0)$

Farkın farkı yöntemi paralel eğilim varsayımına dayanmaktadır. Bu varsayım dâhilinde, etki uygulanmamış olsaydı etki grubundaki zaman içindeki değişim, kontrol grubu ile benzerlik göstermesi beklenmektedir. Bu varsayımın doğru olması bu yöntemin yansız tahminler yapmasını sağlarken, varsayımın sağlanmadığı durumda yöntem geçersiz olacaktır. Bu sebeple farkın farkı yöntemi genellikle tek başına tercih edilmemektedir. Eğilim skoru eşleştirme yöntemi, farkın farkı yöntemi uygulanmadan önce uygulanırsa, seçilen benzer özellikteki kontrol ve etki gruplarının paralel eğilim varsayımına olabildiğince yaklaştığı düşünülmektedir [11].

2.2.2.3. Regresyon Süreksizliği Yöntemi

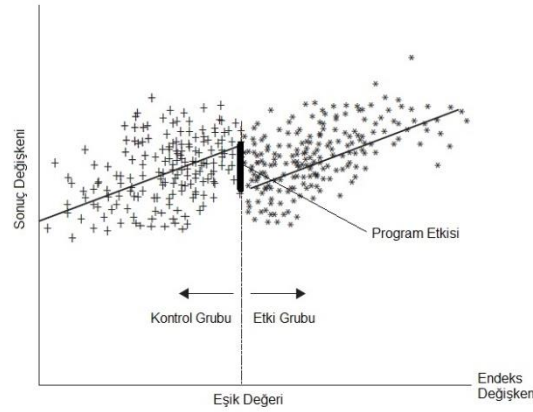
Regresyon süreksizliği yöntemi eşik değeri olarak adlandırılan bir kritere göre etki grubu ve kontrol grubunun belirlenmesine ve karşılaştırılmasına dayanmaktadır. Eşik değeri, endeks değişkeni üzerinde bir noktanın temsil ettiği değerdir. Endeks değeri, programa katılma şansı olan bireylerin belirli bir özelliğinin (yaş, gelir, başarı puanı vb.) tanımlandığı sürekli bir değişkendir [7].

Regresyon süreksizliği yönteminde programın etkisi, eşik değerinin hemen çevresinde yer alan kontrol grubu ve etki grubuna ait sonuç değişkenlerinin karşılaştırılması ile hesaplanmaktadır [7]. Sonuç değişkeni, endeks değişkeni ile ilişkili bir değişken olup programın etkisi ile değişmesi beklenmektedir.



Şekil 2: Program Uygulanmadan Önce Regresyon Süreksizlik Yöntemi

Regresyon süreksizliği yönteminde kavramlar Şekil 2 ve Şekil 3'te verilmiştir [9]. Programın henüz uygulanmadığı, Şekil 2'de eşik değeri, endeks değişkeni üzerinde bir değer olarak gösterilmektedir. Şekilde yer alan her bir nokta bir bireyi temsil etmekte olup eşik değerinin üstünde ya da altında olmasına göre etki grubuna ya da kontrol grubuna atanmışlardır. Endeks değişkeni ve sonuç değişkeni arasında doğrusal ve sürekli bir regresyon doğrusu görülmektedir.



Şekil 3: Program Uygulandıktan Sonra Regresyon Süreksizlik Yöntemi

Programın uygulanmasından sonraki durum Şekil 3'te gösterilmiştir. Etki grubunun sonuç değişkeni aşağı yönlü değişmiş, sürekli regresyon eşik değerinde kırılarak bir süreksizlik oluşturmuştur. Programın uygulanmasından sonra sonuç ve endeks değişkenleri kullanılarak, kontrol grubu ve etki grubu için ayrı ayrı regresyon

tahminleri yapılır. Tahminlerin eşik değeriindeki farkı programın etkisini vermektedir.

Endeks değişkeni ile sonuç değişkeni arasındaki ilişki sebebiyle, bireylerin etki ve kontrol gruplarına yerleştirilmesinde seçim yanlılığı oluşabilmektedir [11]. Bu durum küçük ölçekli işletmelerin uygulanacak bir yatırım teşvik desteği programının firmaların kârları üzerindeki etkisinin incelendiği bir örnek ile açıklanabilir. Firmaların çalışan sayısı(endeks değeri) baz alınarak, 15'den(eşik değeri) az çalışana sahip olan firmaların(etki grubu) destek programından yararlanabilirken, çalışan sayısı 15'den fazla olan firmaların(kontrol grubu) destek programından yararlanmadığı varsayılın. Çalışan sayısı ile firmaların kârları doğru orantılı olduğu düşünüldüğünde, küçük ölçekli firmalar destek programının hedefi olarak seçilmiştir. Destek programı uygulandıktan sonra etki grubunda yer alan tüm firmalar ile kontrol grubunda yer alan tüm firmaların kârlarının karşılaştırılması seçim yanlılığına sebep olmaktadır. Bunun sebebi etki grubu içerisinde yer alan ve çok az sayıda çalışana sahip olan firmalar ile kontrol grubunda yer alan ve fazla sayıda çalışana sahip olan firmalar analize dahil edilerek karşılaştırma yapılmış olmasıdır. Çalışan sayısı çok az olan firmalar ile çalışan sayısı çok fazla olan firmalar, programın uygulanması dışında da birbirlerine benzer özellik göstermezler. Karşılaştırmanın bu firmaları da kapsamaması programın etkisi hesaplanırken yanıltıcı sonuçlar vermesine sebep olur. Eşik değerinin çevresindeki gözlemlerin karşılaştırılmasında seçim yanlılığı sorunu ortadan kalkmaktadır. Seçim yanlılığı eşik değeri çevresinde 0 olduğu varsayılmaktadır [12]. Aynı örnekte, çalışan sayısı 14 olan ve destek programı uygulanan firmalar ile çalışan sayısı 15 olan ve destek programı uygulanmayan firmaların program sonrası kârları karşılaştırıldığında seçim yanlılığı sorunu en aza indirgenmiş olur. Bu iki grup arasındaki 1 çalışanlık farkın firmaların kârlarına etkisinin çok küçük olduğu dolayısıyla bu iki grubunun birbirinin karşıt gerçeği olduğu düşünülebilir. Bu durumda etki grubunda yer alan firmalar ile kontrol grubunda yer alan firmaların kârları karşılaştırıldığında elde edilen artış doğrudan programın etkisi olmaktadır [12].

Regresyon süreksizliği yöntemi diğer yarı deneysel etki değerlendirme yöntemleri arasından karşıt gerçeklik problemine en kuvvetli çözümleri sunmaktadır [11]. Yöntem, uygulanan etkinin, seçim varsayımlarında, model formunda ya da

hataların dağılım varsayımlarında herhangi bir kısıtlama olmadan tespit edilmesine olanak tanımaktadır. Regresyon süreksizliği yöntemi eşik değerinin özelliği sebebiyle etki analizinde rasgeleleştirme yönteminin en iyi alternatifi olarak görülmektedir [12].

Yöntemin kullanımını sınırlayan bazı durumlar da söz konusudur. Bunlara eşik değerinin yeterince açık olarak belirlenmemesi, verinin yeterince büyük olmaması gösterilebilir. Regresyon süreksizliği yönteminin eşik değeri etrafında bölgesel olarak incelenmesi, eşik değerinden uzak verileri göz ardı etmesi ayrı bir sorun olarak görülebilmektedir [11].



3. REGRESYON SÜREKSİZLİĞİ YÖNTEMİ

3.1 Tanım ve Kavramlar

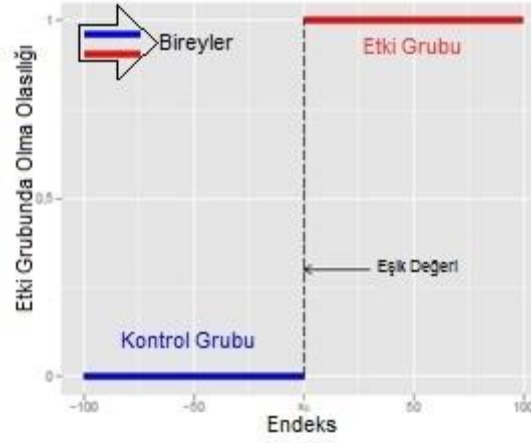
Regresyon süreksizliği yöntemi, bireyler üzerinde uygulanan program etkisini tahmin etmek amacıyla kullanılan bir yarı deneysel etki değerlendirme yöntemidir. Yöntemde, bireylerin etki grubuna ya da kontrol grubuna atanması, bireylerin sahip olduğu endeks değerleri ve bu değerlerin kıyaslandığı bir kriter olan eşik değeri aracılığıyla belirlenir.

Regresyon süreksizliği yöntemi ilk olarak Thistlethwaite ve Campbell [1] tarafından geliştirilmiş ve araştırmacıların kontrol grubu ve etki grubunu rasgele dağılım ile seçemediği durumlarda kullanabileceği bir yöntem olarak önerilmiştir. Çalışmada bir burs programının öğrencilerin akademik çıktılarına etkisini gösteren bir uygulama verilmiştir.

Regresyon süreksizliğinin üç temel elemanı vardır. Endeks değişkeni, eşik değeri ve etkidir. Endeks değişkeni, program uygulanmadan önce, bireylerin herhangi bir özelliğine ait sürekli bir değişkendir. Yaş, gelir, sınav puanı gibi sürekli değişkenler endeks değeri olarak kullanılabilirler. Eşik değeri, endeks değişkeninin sınırları içerisinde program uygulayıcıları tarafından belirlenen bir değerdir. Eşik değeri etki ve kontrol grubunun belirlenmesinde önemlidir. Bireylerin endeks değişkeni üzerinde aldığı değerler eşik değerinin altındaysa (üstünderse) birey etki grubuna atanır ya da üstünderse (altındaysa) birey kontrol grubuna atanır. Kesim noktası, süreksizlik noktası vb. isimlerde kullanılan eşik değeri, programın etkisinin hesaplandığı yer olması sebebiyle ayrıca önem taşımaktadır [13], [14], [15].

Bireylerin endeks değerinin, eşik değerine göre konumu (üstünde ya da altında olması) göz önüne alınarak, bireylerin hangi oranda etki grubuna ya da kontrol grubuna atanacağı regresyon süreksizliği yönteminin tasarımını belirler. Keskin regresyon süreksizliği yöntemi tasarımı, eşik değerinden büyük ya da eşit endeks değerine sahip bireylerin tamamının etki grubuna atandığı, eşik değerinden küçük endeks değerine sahip bireylerin tamamının kontrol grubuna atandığı tasarımıdır. Trochim [16] tarafından önerilen fuzzy regresyon süreksizliği yöntemi tasarımı, eşik değerinin hem altında yer alan bireylerin hem de eşik eğerinin üstünde yer alan bireylerin oransal olarak etki grubuna ya da kontrol grubuna atandığı durumlarda kullanılmaktadır. Eşik değerine göre konumu, bir bireyin program için

seçilme ya da atanma şansı verirken, programa seçilme olasılığını regresyon süreksizliği yönteminin tasarımı belirler [15].

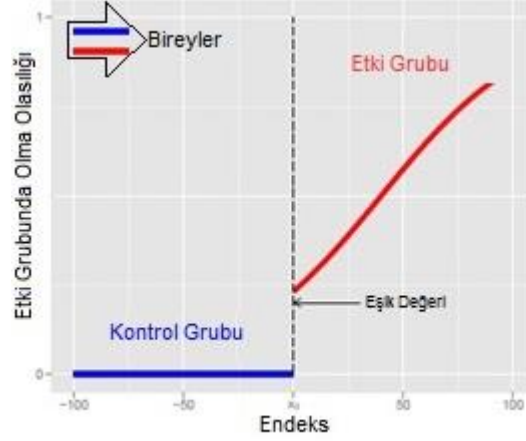


Şekil 4: Keskin Regresyon Süreksizliği Yöntemi Tasarımı

Şekil 4'te Keskin regresyon süreksizliği yöntemi tasarımı ve Şekil 5'te fuzzy regresyon süreksizliği yöntemi tasarımı karşılaştırılmıştır [15].

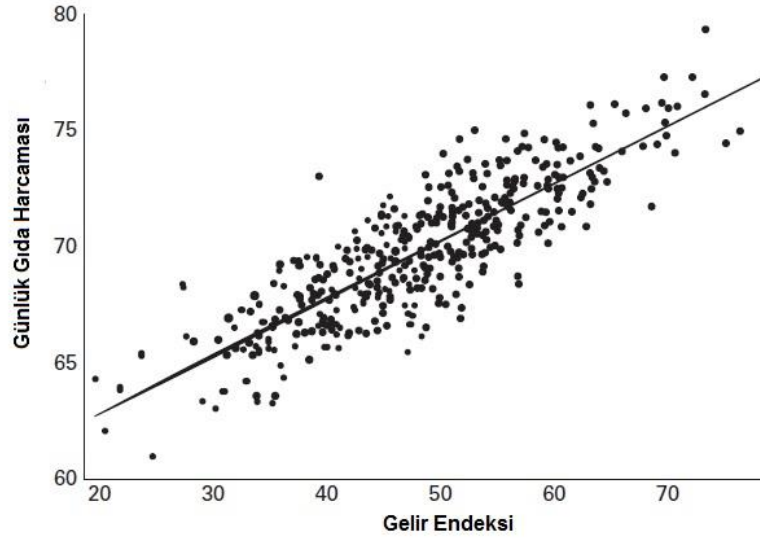
Şekil 4'te ve Şekil 5'te bireylerin aldığı endeks değerinin eşik değerine göre konumunun, etki ya da kontrol grubuna atanması olasılığı gösterilmiştir. Şekil 4'te keskin regresyon süreksizliği düzeninde etki grubunda olma olasılıkları sadece eşik değerinde değişiklik göstererek sıfırdan bire çıkmıştır. Eşik değerinin altında bulunan tüm gözlemlerin etki alma olasılığı sıfır iken üstünde bulunan tüm gözlemlere bu etki uygulanacaktır. Şekil 5'te gösterilen durumda da eşik değerinin altında benzer şekilde etki grubunda olma olasılığı sıfır iken eşik değeri sonrası etki grubuna atanma olasılığı sıfır ile bir arasında değişkenlik göstermiştir. Çalışmanın ilerleyen bölümlerinde verilen regresyon süreksizliği yöntemi, keskin regresyon süreksizliği yöntemi tasarımıdır.

Tasarımın uygulanması, özet grafiksel gösterimin yapılması ile başlamaktadır. Daha sonra etkinin tahmin edilmesi yer almaktadır. Etkinin tahmin edilmesinde, eşik değerinin belirlenmesi ve uygun bant genişliğinin hesaplanması gerekmektedir [14], [17].



Şekil 5: Fuzzy Regresyon Süreksizliği Yöntemi Tasarımı

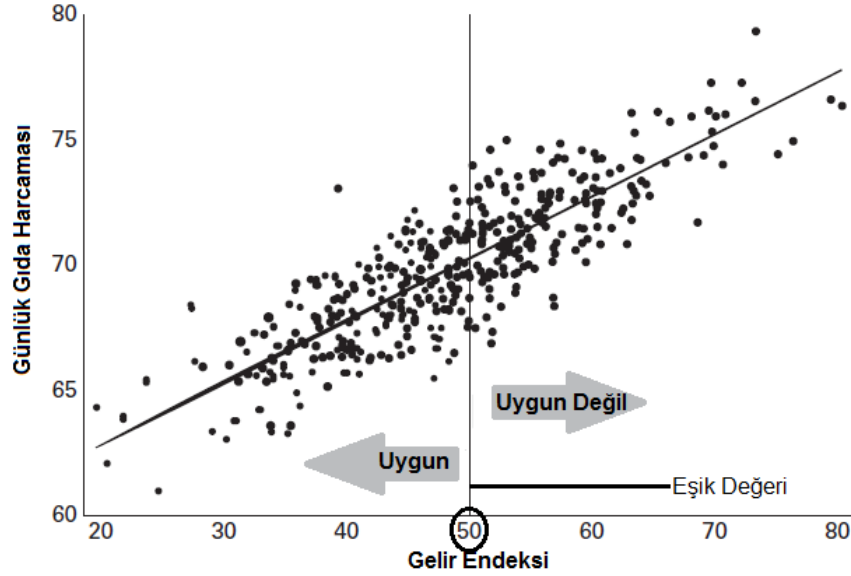
Regresyon süreksizliği yöntemi, Şekil 6'da ve Şekil 7'de gösterilen örnek üzerinde incelenmiştir. Örneğin; ailelerin gıda tüketimlerini arttırmaya yönelik bir mali yardım programı olduğu düşünölsün. Ailelerin gıda tüketimi, gelirleri ile doğru orantılı olduğu varsayıldığında, belli bir seviyenin altındaki ailelere bu program kapsamında mali yardım yapıldığında düşük seviyelerdeki gıda harcamalarını arttırmaları sağlanabilir. Şekil 6'da mali yardım yapılmadan önce gelir endeksi ile günlük gıda harcamaları arasındaki doğrusal ilişki gösterilmektedir [7].



Şekil 6: Mali Yardım Programından Önce Gelir Endeksi – Günlük Gıda Harcamaları Arasındaki İlişki

Program uygulayıcıları tarafından, yapılacak mali yardım programına ait eşik değeri “50” gelir endeksi seviyesine olarak belirlenmiştir. Bu gelir endeksi

seviyesinin altında kalan aileler bu mali yardım programında faydalanabileceklerdir. Mali yardım programı sayesinde sadece gelir endeksi 50 ve altında bulunan ailelerin gıda harcamalarını artırılması sağlanabilir. Bu durum, Şekil 7’de görüldüğü üzere mali yardım programı alacak aileler endeks değeri altında yer alırken, endeksin üstünde yer alan aileler mali yardımdan muaf tutulacaktır [7].

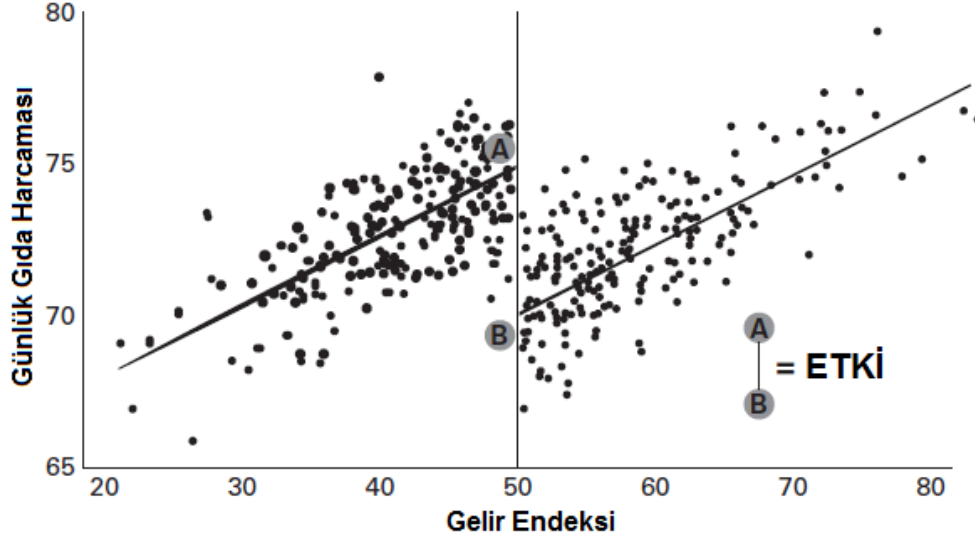


Şekil 7: Mali Yardım Programına Uygun Olan-Olmayan Bireyler

Örnekte gelir endeksi, regresyon süreksizliğinde endeks değerine karşılık gelirken 50 olarak belirlenen değer ise eşik değerine karşılık gelmektedir. Yapılan mali yardım ise etki olarak tanımlanmaktadır. Yapılması planlanan mali yardım programı sonrasında artış beklenen günlük gıda harcaması ise sonuç değişkenidir. Yapılması planlanan yardımın eşik değeri altındaki tüm ailelere yapılıp, eşik değerinin üstünde kalan ailelerin tamamına yapılmamış olması keskin regresyon süreksizliği yöntemi tasarımı olduğunu göstermektedir.

Mali yardım programının uygulanması ile oluşan değişim Şekil 8’de gösterilmiştir. Yapılan mali yardımının günlük gıda harcamalarını da arttıracığı düşünüldüğünden eşik değeri civarında bir sıçrama göze çarpmaktadır. Yapılan mali yardım ile gelir endeks değeri 50 ve altında olan ailelerin günlük gıda harcamaları ortalama 60-70 aralığından, 65-75 aralığına kadar çıkarken, endeks değeri 50 üzerinde olan ailelerde bir değişim gözlenmemiştir. Şekil 8’den

anlaşılacağı üzere etki sonucunda oluşan kırılım mali yardım sayesinde oluşmaktadır [7]. Kırılımın eşik değeri civarında yer alan aileler üzerinde gösterdiği farklılık regresyon süreksizliği ile ölçülmesi planlanan etki miktarıdır.



Şekil 8: Mali Yardım Programından Sonra Gelir Endeksi – Günlük Gıda Harcamaları Arasındaki İlişki

3.1.1 Analizin İçsel Geçerliliği İçin Varsayımlar

Regresyon süreksizliği yöntemi birçok yönden araştırmacılara çekici gelmektedir. Regresyon süreksizliği yöntemi etki değerlendirmesi yöntemi olarak sıklıkla kullanılmaktadır. Regresyon süreksizliği yarı deneysel etki değerlendirme yöntemi olduğu için mevcut veriler ile uygulanabilir. Bu durum yöntemin deneysel yöntemle göre daha az maliyetli olmasını sağlamaktadır. Ne kadar kullanışlı bir yöntem olsa da yarı deneysel bir yöntem olması sebebiyle bazı koşulları sağlaması gerekmektedir [14].

Bu koşullar aşağıdaki gibidir:

- 1- Endeks değişkeni, program uygulanmadan önce belirlenmeli ve program boyunca değişiklik göstermemelidir.
- 2- Eşik değeri endeks değişkeninden bağımsız olarak, program uygulayıcıları tarafından belirlenmeli ve bireyin etki grubunda ya da

kontrol grubunda olması sadece eşik değerine bağlı olarak belirlenmelidir.

- 3- Endeks değişkeni gözlemlenebilir bir değer olduğu için program bu değişkene bağlı olarak oluşturulmaktadır. Uygulanacak program ile sonuç değeri üzerinde bir değişim hedeflenmektedir.
- 4- Endeks değişkeni ve sonuç değişkeni arasındaki ilişkiyi gösteren fonksiyonel form doğru tahmin edilmelidir. Belirlenen fonksiyon eşik değerinde etkinin olduğu durumun tahmin edilmesinde kullanılır.
- 5- Regresyon süreksizliği yöntemi uygulanır iken önemli varsayımlardan birisi de bireylerin endeks değişkeni üzerinde ve dolayısıyla eşik değeri üzerinde etkisi olmamasıdır. Bireylerin endeks değeri üzerinde kontrolü olur ve yapılacak etki grubuna girmelerini sağlar ise bu durum regresyon süreksizliği yöntemini geçersiz kılmaktadır [14], [18].

Örneğin, bir sınav ve sınava bağlı olarak verilecek bir burs olduğu varsayalım. Buna göre tasarlanan programda bursun ileri dönemlerde öğrencilerin başarısına etkisi incelenmek istensin. Programda, eşik değeri 50 olarak belirlensin, buna göre sınavdan 50 ve üzeri alan öğrenciler bursu hak edeceklerdir. Sınava girecek iki tip öğrenci olduğu varsayılmaktadır. Bunlardan A tipi öğrencilerin sınavdan 50 ve üzeri puan almalarının burs için yeterli olacağını bildikleri, B tipi öğrencilerin ise burs ve eşik değerinden haberi olmadıkları varsayılmaktadır. Sınavın standart bir yapıda olduğu ve her öğrencinin benzer dikkatsizlikler ile hatalar yapabileceği, sadece bir kez soruları gözden geçirmenin, hataları gidermede önemli etkisi olduğu bilinmektedir. O halde A tipi öğrencilerin eşik değeri ve bursun bilincinde olduğu için hata yapma olasılığına karşın soruları tekrar gözden geçirmeye hevesli olması beklenirken, B tipi öğrencilerin kontrol için o kadar istekli olmayacakları beklenmektedir. Bunun sonucunda eşik değerini geçebilen öğrenciler A ve B tipi öğrencilerin karışımı olurken, eşik değeri altında kalan öğrencilerin yoğunluğu daha çok B tipi öğrencilerden oluşacaktır. Dolayısıyla eşik değerinin üstünde kalan ve etki grubuna atanan bireyler ile eşik değerinin altında kalan ve kontrol grubuna atanan bireyler arasında farklılık oluşacaktır. Henüz burs programı uygulanmadan kontrol grubu ile etki grubu arasında farklılık olması regresyon süreksizliği yönteminin uygulanabilirliğini ortadan kaldıracaktır.

Bahsedilen örnek ile bireylerin endeks değişkeni ve eşik değeri hakkında bilgisi olması durumunda regresyon süreksizliği yönteminin bozulmasına nasıl yol açtığı gösterilmiştir. Burada bahsedilen bilgi tam bir bilgi olmak durumundadır. Çünkü örnekte görüldüğü gibi bireylerin her zaman endeks değeri üzerinde ufak da olsa bir kontrolü vardır. Buna sınava daha fazla hazırlanmak, özel ders almak vb. durumlar etki edebilse de endeks değeri(sınav puanı) üzerinde tam bir kontrol olmaması gerekmektedir.

3.1.2 Regresyon Süreksizliği Yöntemi ve Potansiyel Çıktı Çerçevesi

Etki analizinde bahsedilen karşıt gerçeklik kavramının bir diğer adı potansiyel çıktı çerçevesidir. Program uygulanan bir bireye etki uygulanmış olur. Etki uygulanan birey, kendisi ile benzer etkiyi alan bireylerle etki grubunu oluşturur. Programın etkisinin ölçülebilmesi için kontrol grubuna ihtiyaç duyulur. Etki değerlendirme bölümünde anlatıldığı üzere bu kontrol grubuna karşıt gerçeklik ya da potansiyel çıktı çerçevesi denilmektedir.

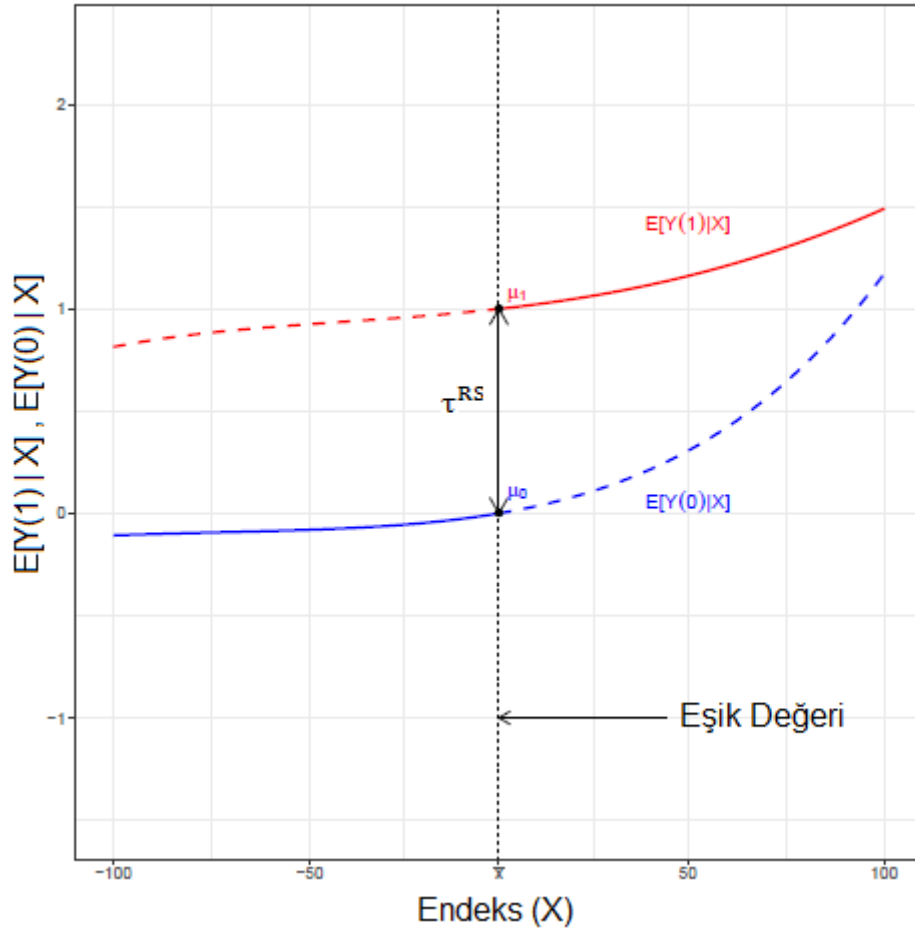
Etki değerlendirmesinde her bir bireyin etki altında olduğu ve olmadığı durum düşünülerek uygulanan program ile bireylerin elde ettiği sonuçlarda ne kadar değişim gösterdiği hesaplanmak istenmektedir. Böylece basit anlamda programın etkisi kolayca ölçülebilmüş olur. Eşitlik 3.1'de x_0 eşik değeri, etki grubunda olma durumu için 1, kontrol grubunda olma durumu için 0 değerleri kullanıldığında; i . bireyin etki grubunda olması halinde sonuç değeri $Y_i(1)$ olurken, kontrol grubunda olduğu durumda alınan sonuç değeri $Y_i(0)$ olacaktır. Sonuç değeri, endeks değeri (X_i) ve etkinin uygulanmasına bağlıdır [15].

$$Y_i = \begin{cases} Y_i(0) & X_i < x_0 \\ Y_i(1) & X_i \geq x_0 \end{cases} \quad (3.1)$$

Karşıt gerçeklik problemi gereğince bireyin etki grubunun üzerinde uygulandığı ve uygulanmadığı durumun aynı anda görmenin mümkün değildir. Bu sebeple birbirine olabildiğince benzer bireylerden oluşan bir grubu, etki grubu ve kontrol grubu olarak ikiye ayırarak programın etkisi sonucunu ölçülebilir. Fakat bu durum her zaman mümkün olmayabilir. Bahsedilen gruplara ayırma işlemi, program uygulanmadan önce hazırlanmış olması gerekmektedir, Ayrıca gruplara ayırma işlemi sırasında rasgele dağıtım yönteminin kullanılması, bireyler tarafından adil bulunmayabilir, programı uygulamak zorlaşabilmektedir.

Regresyon süreksizliği yöntemi, oluşabilecek bu tür zorluklara çözüm olarak kullanılabilir. Eşik değeri yardımıyla belirli bir amaç doğrultusunda belirlenen etki grubu ile benzer özellik gösteren kontrol grubu karşılıklı olarak eşik değerinin her iki yanında yer almaktadır. Dolayısıyla program etkisi bu iki grubun eşik değeri üzerinde karşılaştırılması ile hesaplanabilmektedir.

Belirlenen eşik değerinin hemen altındaki bireyler ile hemen üstündeki bireyler benzer özellikler göstermektedir. Bu sebeple eşik değerine göre hemen üstünde olan bir bireyin etki grubunda olması ve eşik değerinin hemen altında yer alan bir bireyin kontrol grubunda yer alması, programın etkisinin hesaplanmasında kullanılmaktadır.



Şekil 9: Regresyon Süreksizliği Yönteminde Ortalama Beklenen Değerler

Şekil 9'da ve Eşitlik 3.2'de ortalama beklenen değerler $E[Y_i(1)|X_i = x]$ ve $E[Y_i(0)|X_i = x]$ görülmektedir [15]. Düz kırmızı çizgilerle gösterilen $E[Y_i(1)|X_i = x]$ fonksiyonu eşik üzerinde bulunan etki grubunda yer alan bireylerin gözlenen

değerleri ile oluşturulmuştur. Eşik değerinin altında yer alan ve kesikli kırmızı çizgi ile gösterilen $E[Y_i(1)|X_i = x]$ fonksiyonda, gözlemlenebilir değerin olmadığı sadece fonksiyonun tahmininin yer aldığı görülmektedir. Aynı durum düz mavi ile eşik değeri altında yer alan bireylerin oluşturduğu kontrol grubuna ait $E[Y_i(0)|X_i = x]$ fonksiyonu içinde geçerlidir. Kontrol grubunda bireylerin gözlemlenebilir değerleri düz mavi ile gösterilirken gözlemlenemeyen değerlere ait fonksiyon yer almaktadır [15].

$$E[Y_i|X_i] = \begin{cases} E[Y_i(0)|X_i] & X_i < x_0 \\ E[Y_i(1)|X_i] & X_i \geq x_0 \end{cases} \quad (3.2)$$

Şekil 9'da görüldüğü gibi herhangi bir endeks değeri düzeyindeki ortalama etki $E[Y_i(1)|X_i = x]$ ile $E[Y_i(0)|X_i = x]$ arasındaki farka karşılık gelmektedir. Bu fark eşik değerinin her iki tarafı için aynı anda gözlemlenen değer bulunmadığından doğrudan hesaplanamamaktadır. Fakat eşik değeri üzerinde iki eğrinin de neredeyse gözlemlendiği özel bir durum oluşmaktadır. Şekil 9'da program, eşik değerinin altına uygulanmadığı için eşik değerinin altında, eşik değerine çok küçük bir ε değeri kadar yakın gözlemlerin eşik değeri üzerindekiymiş gibi değer aldığı düşünülebilir.

Eşik değeri üzerindeki bu özel durum $E[Y_i(1)|X_i = x]$ ve $E[Y_i(0)|X_i = x]$ fonksiyonlarının x_0 üzerinde sürekli olduğu varsayımı altında Eşitlik 1'de ifade edilmiştir [13].

$$\lim_{x \downarrow x_0} E[Y_i | X_i = x] - \lim_{x \uparrow x_0} E[Y_i | X_i = x] = E[Y_i(1) - Y_i(0) | X_i = x_0] \quad (3.3)$$

Eşitlik 3.3'de etki grubu ve kontrol grubu değerlerinin eşik değeri yakınlarında alınan farkın, eşik değerinde, ortalama program etkisine eşit olduğu gösterilmektedir. Eşitlik 3.4, Eşitlik 3.2'nin sadeleştirilmiş halidir [15]:

$$\tau^{RS} \equiv E[Y_i(1) - Y_i(0) | X_i = x_0] \quad (3.4)$$

3.1.3 Regresyon Süreksizliği Yöntemi ve Deneysel Yöntemin Karşılaştırılması

Eşitlik 3.3'de görüldüğü üzere program etkisi, eşik değerinin hemen çevresinde bulunan veriler kullanılarak hesaplanmıştır. Deneysel yöntem ya da diğer adıyla rasgele dağıtım yönteminde bir bireyin etki grubunda olması ya da kontrol grubunda olması rastgele dağıtım ile belirlenmektedir. Regresyon süreksizliği

yönteminde, eşik değeri çevresinde etki grubuna ve kontrol grubuna ait bireyler endeks değişkeni açısından büyük farklılıklar göstermediği için regresyon süreksizliği yöntemi eşik değeri çevresinde rasgele dağıtım yöntemine benzetilebilir [15].

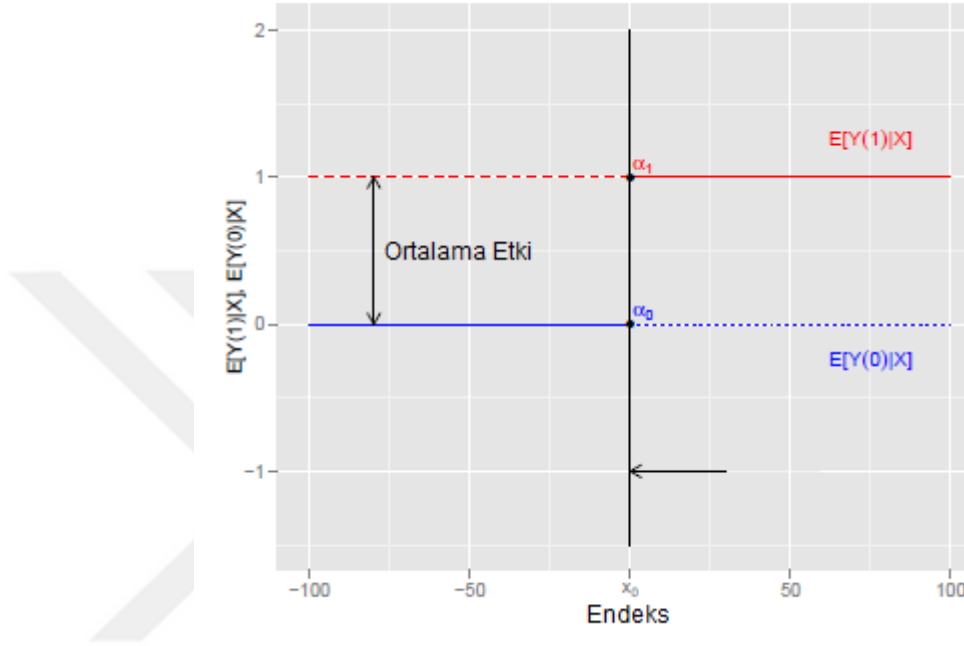
Etki analizinin “altın standartları” olarak görülen bu durum yarı deneysel etki değerlendirme yöntemleri arasında sadece regresyon süreksizliği yönteminde sağlanmaktadır. Bu özelliğinden dolayı rasgele dağıtım yöntemine en çok yaklaşan yarı deneysel etki değerlendirme yöntemi regresyon süreksizliği yöntemidir. Rasgele olmayan seçimde rasgele deney düzeni incelemesi yapmıştır. Regresyon süreksizliğinin diğer yarı-deneysel etki analizi yöntemlerine göre daha zayıf koşullar altında analiz yaptığını ve eşik değeri üzerinde yapılacak süreksizlik incelemesinin rasgele deney özelliği incelenmiştir. Regresyon süreksizliğinin temel bir özelliği olan bölgesellik kavramını, etki uygulanmadan önceki endeks değerlerinin gözlemlerin karakteristik özelliklerini gösteren ve bilinen sürekli bir dağılım olduğu, bunun sonucunda etkinin uygulanması ile eşik değerinde oluşacak sıçramanın çevresindeki gözlemlerin rasgele deney düzenine benzer şekilde yorumlanabileceği belirtilmiştir [13], [18].

Regresyon süreksizliği yöntemi ve rasgele dağıtım yöntemi arasındaki benzerlikler ve farklılıkların iyi yorumlanması yapılacak analiz için önemlidir. İki yöntem arasındaki farklardan en önemlisi regresyon süreksizliği yönteminin regresyon fonksiyonuna bağlı olması iken rasgele dağıtım yönteminde bu bilgiye ihtiyaç duyulmamaktadır [15].

Rasgele dağıtım yöntemi, bireylerin etki grubu ve kontrol grubu olarak rasgele ayrılmasına dayanmaktadır. Bireylerin etki grubuna ya da kontrol grubuna atanma şansı birbirine eşittir. Örnek olarak aynı sektör ve ölçekte olan 100 farklı firmanın yarısına teşvik kredisi verilmesine yönelik program uygulansın. Firmalar rasgele 1’den 100’e kadar numaralandırılarak sıra numarası 50 ve üzerinde olan firmalara teşvik uygulansın. Firmalar başta rasgele sıralandığı için sıra numarası 50 ve üzerinde olan firmaların 50’nin altında olan firmalardan herhangi bir farkı yoktur [15].

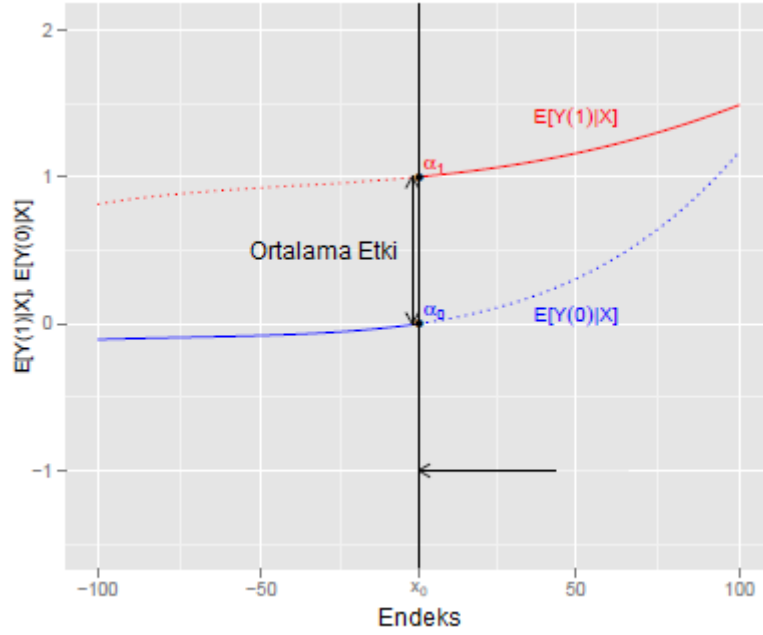
Regresyon süreksizliği yöntemi ile rasgele dağıtım yöntemi arasında önemli bir farkı ise rasgele dağıtım yönteminde beklenen ortalama sonuç değerinin endeks

değerinden tamamen bağımsız olmasıdır. Bu sebeple aşağıda Şekil 10'da $E[Y_i(1) | X_i]$ ve $E[Y_i(0) | X_i]$ endeks değeri boyunca aynı değeri aldığı görülmektedir. Rasgele dağıtım yönteminde ifade edilen etki ve kontrol değerleri arasındaki farkın herhangi bir nokta alınmasıyla etki her zaman aynı büyüklükte hesaplanabilmektedir [15].



Şekil 10: Rasgele Dağıtım Yöntemi ile Program Etkisi

Rasgele dağıtım yöntemi ile regresyon süreksizliği yönteminin, eşik değerinde programın etkisini aynı olduğu Şekil 10 ve Şekil 11'de gösterilmektedir [15]. Regresyon süreksizliği yönteminde, endeks değeri ile sonuç değeri arasında ilişki olduğu bilinmektedir. Endeks değişkeni ile sonuç değişkeni arasındaki bu ilişki regresyon süreksizliği yönteminin karakteristik bir özelliği olarak oluşmaktadır. Endeks değişkeni olarak belirlenen değişkenler genellikle bireylerin yeteneklerine kaynakların performansına göre değişkenlik gösterir. Bu sistematik değişkenlik sebebiyle yüksek endeks değerine ait bireylerin uygulanan program sonucunda elde ettiği sonuç değeri ile düşük endeks değerine sahip bireylerin uygulanan program sonucunda elde ettiği sonuç değeri arasında belirgin bir farklılık vardır.



Şekil 11: Regresyon Süreksizliği Yöntemi ile Program Etkisi

Şekil 10 ve Şekil 11’de gösterilen iki durumda görülen önemli bir farklılık ise fonksiyona ait bilgidir. Şekil 11’de ve Eşitlik 3.5’de ortalama etki, eşik değerinin hemen çevresinde yer alan gözlemlerin sonuç değerleri kullanılarak hesaplanır.

$$\tau^{RS} = \lim_{x \downarrow x_0} E[Y_i | X_i = x] - \lim_{x \uparrow x_0} E[Y_i | X_i = x] \quad (3.5)$$

Eşik değeri üzerinde etkinin belirlenmesi için eşik değeri üzerindeki noktadaki değeri tahmin etmek gerekmektedir. Bunun içinde eşik değerine yakın gözlemlerden faydalanılarak regresyon fonksiyonunun belirlenmesi önemlidir. Yanlış bir fonksiyon belirlenirse eşik değeri üzerindeki tahmini sonuç değeri yanlış hesaplanabilir bu durum etkinin yanlış ölçülmesine sebep olur.

3.2 Regresyon Süreksizliği Yönteminin Uygulanmasında Yapılan Çalışmalar

Thistlethwaite ve Campbell [1] regresyon süreksizlik yöntemini öğrencilere uygulanan bir sınavda, alınan puanların endeks değişkeni olarak belirlendiği ve belirli bir puanın altında kalan öğrenciler ile üstünde kalan öğrencilerin burs kazanma oranları, kariyer planları, yükseköğretime başlamaları vb. konularda karşılaştırılması için incelemiştir. Çalışmada eşik değeri olarak belirli bir puanın üstünde kalan öğrencilerin, altında kalan öğrencilere göre burs kazanma oranlarının anlamlı olarak fazla olduğu gözlemlenmiş fakat öğrencilerin ileri dönem kariyer planlarını ve yükseköğretime katılma oranlarını etkilememiştir.

Black [19] çalışmasında regresyon süreksizliği çalışmalarına farklı bir bakış açısı sunmuştur. Regresyon süreksizliğinde endeks değeri olarak uzaklık birimi kullanılarak yöntemin coğrafi alanlar üzerinde uygulanabileceğini gösteren bir uygulama yapmıştır. Çalışmasında bir okulun kalitesini ölçmek için öğrencilerin puanları yanında bölgedeki ev fiyatlarının ilişkisi kullanılmıştır. Öğrenci okul eşleştirmesi sistemine göre öğrenciler kendi bölgelerindeki okullara gitmek zorundadır. Bunun sonucunda birbirleri ile sınır olan bölgelerde sınırın hemen her iki tarafında bulunan ev fiyatları ve okul başarılarının farklılık gösterip göstermediği analiz edilmiştir. Analize başlamadan önce farklı bölgelerde bulunan sınır komşu evler özellik bakımında karşılaştırılmış ve farklılık olmadığı tespit edilmiştir. Çalışmada temel sonuç olarak, sınıra 0,15 mil uzaklıkta farklı bölgelerde bulunan aynı özellikteki evlerin fiyatları ve öğrenci puanları karşılaştırıldığında, öğrenci puanlarının %5 artışına karşılık ev fiyatlarında %2,1'lik bir artış tespit edilmiştir. Çalışma ders bazına indirildiğinde sınav puanlarında 1 puanlık artış için ev fiyatlarında %1,5'lik bir artış görülmüştür.

Carpenter [20] çalışmasında yasal olarak belirlenen en düşük alkol kullanım yaşından faydalanarak alkol tüketiminin ölüm oranları üzerindeki etkisini regresyon süreksizliği düzenini kullanarak incelemiştir. Otoritelerce 21 olarak belirlenen yasal alkol kullanım yaşının bu yaşın öncesi ve sonrası için gençlerin alkole ulaşımında önemli bir fark yaratması, diğer etkenlerin sabit artışı göz önüne alındığında 21 yaşında görülen alkol kullanımı ve ölüm oranlarındaki artışın birbirleriyle ilişkili olduğu düşünülmüştür.

Çalışmada 21 yaş, eşik değeri, yaş endeks değeri olarak kullanılmıştır. Yaşa bağlı olarak; bir yıl içerisinde alkol kullanılan günlerin oranı, bir yıl içerisinde yüksek miktarda alkol kullanan gençlerin yaş grubuna oranı ve bir yıl içerisinde 12'den fazla sayıda alkol kullanan gençlerin yaş grubuna oranı sonuç değer olarak alınmıştır. Sonuçlar genel görüşü destekler nitelikte, alkol tüketimi artışının ölüm oranlarındaki artışa neden olduğu yönündedir. 21 yaşına girmeyi takiben gençlerin yıl içerisindeki alkol kullandığı günün oranı 21 yaşın hemen öncesine göre %21 oranında artmaktadır. Bu oran yüksek miktarda alkol alımı için de aynı düzeydedir. Çalışmada elde edilen önemli bir sonuçta yasal sınır olan 21 yaşındaki alkol kullanımındaki artışın ölüm oranlarında %9'luk bir artışa neden olmasıdır.

Lee [18] çalışmasında A.B.D.'de yapılan Temsilciler Meclisi seçim sonuçlarını regresyon süreksizlik yöntemi ile analiz etmiştir. Aynı bölge için bir önceki dönemde yapılmış seçimde, küçük bir farkla göreve gelen parti ile seçimi küçük bir farkla kaybeden rakip parti birlikte yarışacakları sonraki seçime “görevde olma” avantajı dışında bir farklılık olmadan girmektedirler. Çalışmada önceki seçimi kazanan partinin bir sonraki seçime “görevdeki parti” olma avantajıyla girmesinin sonraki seçim sonuçlarına etkisi incelemiştir. Seçime giren ve oyların büyük çoğunluğunu alan iki partiden biri olan Demokratik Partinin bir önceki seçimde en yakın rakibine göre oy oranlarının farkı, çalışmada endeks değeri olarak kullanılırken, bu farkın sıfır olduğu değeri eşik değeri olarak kullanılmıştır. Önceki seçimde oy farkının Demokratik Parti lehine pozitif olduğu bölgelerde Demokratik Parti “görevdeki parti” avantajına sahip olarak sonraki seçime girmiştir. Sonraki seçimdeki oy oranının sonuç değeri olarak belirlendiği çalışmada Demokrat Parti geçmiş seçimde az bir farkla kazandığı bölgelerde görevde olma avantajının etkisiyle az bir farkla seçimi kaybettiği bölgelere göre bir sonraki seçimde %7-8 arasında daha fazla oy oranına sahip olduğu sonucuna ulaşılmıştır.

Bronzini [21] çalışmasında Kuzey İtalya bölgesinde uygulanan Ar-Ge teşvik programının etkinliği ölçülmek istenmiştir. Yaklaşık 1200 firmanın başvuruda bulunduğu programda projelerden belirli bir puan üzerinde olanlar programa kabul edilmiş ve parasal destek sağlanmıştır. Çalışmada uygulanan teşvik programının firmaların yatırım harcamalarında nasıl bir değişikliğe sebep olduğu araştırılmıştır.

Çalışmada firmaların aldığı puan endeks değeri olarak belirlenmiştir. Proje değerlendirmesinden 75 ve üzerinde alan firmaların projesinin kabul edilmesi ile bu değeri eşik değeri olarak belirlenmiştir. Firmaların yatırımlarının satışlara oranı ise sonuç değeri olarak belirlenmiştir. Değerlendirme puanı 75'in hemen altında ve hemen üstünde olan firmaların karşılaştırıldığı, firma ölçeği ayırmadan tüm firmaların incelendiği analizde yatırım oranında anlamlı bir artış olmadığı gözlemlenmiştir. Fakat aynı analiz küçük ölçekli firmalar için yapıldığında eşik puanını geçerek destek alan firmaların destek almayan firmalara göre yatırım oranlarında iki kat artış görülmüştür.

Caliendo [22] çalışmasında işsizlik sigortasının farklı uzunluklarda uygulanmasının işsizlik süresi ve yerleşilen iş kalitesine gösterdiği farklılıkları incelenmiştir. İşsizlik

sigortasının süresi uzadıkça kişinin iş aramaya yönelik davranışlarında azalma görüleceği yönündeki genel düşünceye karşı bu sürenin daha uygun bir iş bulmaya yönelik kullanılabileceği görüşü ortaya atılmıştır.

Çalışmada Almanya'da uygulanan işsizlik sigortası süreleri kullanılmıştır. Belirlenen bu süre 45 yaşının öncesi için 12, sonrası için 18 aydır. Böylece yaşın endeks değeri olarak alındığı çalışmada 45 yaşı ise eşik değeri olarak kullanılabilmektedir. Eşik değerinin hemen altındaki yaşlar ve hemen üstündeki yaşlar sadece uygulanan işsizlik sigortası süresi bakımından farklılık göstermektedirler. Analiz sonucunda erkek bireyler genelinde daha uzun süre işsizlik sigortasından faydalananların daha kısa süre faydalananlara göre yeni işlerinden ayrılma süreleri 22-24 ay uzun sürmektedir. Ayrıca çalışmada daha uzun süre işsizlik sigortası hakkına sahip olan bireylerin daha yüksek maaş alacakları işlere yerleştiği görülmüştür.

Anderson [23] çalışmasında internet ortamında kişilerin yaptıkları görüş ve puanlamaların aynı hizmetten faydalananak diğer kişilerin seçimlerini nasıl etkilediğini analiz etmiştir. Restoran, eğlence merkezi vb hizmetlerin kullanıcılar tarafından bir ile beş puan arasında değerlendirildiği geniş çevrelerce kullanılan bir uygulamada yer alan puanlamalar analizde kullanılmıştır. Uygulamada kullanıcıların verdiği puanların ortalamasından elde edilen restoran puanı, uygulamanın gösterge panelinde buçuk değer ve katları halinde gösterilmekte olup, ortalama puanı 3,24 olan bir restoran gösterge puanı 3'e yuvarlanırken, ortalama puanı 3,26 olan restoran gösterge puanı 3,5'e yuvarlanmaktadır. Bu kadar yakın ortalamalara sahip olan firmaların gösterge panelinde çok farklı kalitede gösterilmesi, hizmeti kullanmak isteyen kullanıcıların bu göstergeleri kullanarak tercihlerini nasıl oluşturdukları araştırılmak istenmiştir.

Çalışmada ortalama kullanıcı puanları endeks değeri olarak kullanılırken yuvarlamanın yapıldığı 0,25'lik ve 0,75'lik değerler eşik değeri olmuştur. Sonuç değeri olarak restoran boş olma oranına karşılık gelen ortalama uygunluk oranı alınmıştır. Analiz sonucunda yüksek gösterge puanına sahip restoranların en çok talep olduğu akşam 7 saatinde %19 oranında daha fazla doluluk oranına sahip olduğu görülmüştür.

Zhao [24] çalışmasında Amerika Birleşik Devleti Michigan eyaletinde otoritelerce yürütülen Ar-Ge destek programı fonlamasının firmalar üzerindeki etkisini incelemiştir. Başlangıç düzeyinde olan firmaların desteklendiği programda, Ar-Ge desteği alan firmaların maddi sıkıntılarının azalması ve bunun neticesinde daha uzun süre faaliyet göstermeleri çalışmanın temel motivasyonunu oluşturmuştur.

Çalışmada fonlama eşik değeri olarak alınmış olup sonuç değerleri olarak firmanın teşvik sonrası 3 yıl boyunca faaliyet göstermesi olasılığı ve üretkenliği gösteren patent çalışmaları alınmıştır. Firmalar arasında Ar-Ge desteği almaya hak kazanan firmaların 3 yıl daha fazla faaliyet gösterme olasılığı %15 - %25 aralığında artış gösterirken, patent sayılarında anlamlı bir fark gözlemlenmemiştir.

Ozier [25] çalışmasında regresyon süreksizliği yaklaşımı yardımıyla Kenya'da liseye devam etmenin öğrenciler üzerindeki etkisini incelemiştir. Bu incelemenin önemli girdilerin biri de ülke genelinde 8. sınıf düzeyindeki öğrencilere uygulanan sınav sonuçlarıdır. Sınav sonuçlarını ortalamaya göre standartlaştırdığında eşik değeri (standartlaştırılmış ortalama olan 0) üstünde kalan öğrencilerin liseyi başarı ile tamamlama oranının ortalamadan hemen altında kalan öğrencilere göre %15 oranında daha fazla olduğu görülmüştür. Ayrıca lise öğretimi alan bireylerin lise öğrenimi almayan öğrencilere göre %23 ile %43 arasında daha yüksek oranla bir işte çalıştığı gözlemlenmiştir. Çalışmada liseyi bitiren kadınlarda erken doğum oranlarının %9 oranında düşüş gösterdiği belirtilmiştir.

Aysoy [26] Merkez Bankası Ekonomi Notları kapsamında hazırlanan çalışmada taze meyve-sebze tedarik zincirindeki engelleri azaltıcı tedbirlerin fiyatlar üzerindeki etkisini regresyon süreksizliği ve diğer bir yarı deneysel yöntem olan farkların farkı yöntemlerini kullanarak incelemiştir. Uygulanan tedbirin hemen öncesi ve hemen sonrası karşılaştırılmak istendiği için regresyon süreksizliği yönteminin farkların farkı yöntemine göre daha avantajlı olduğuna değinilmiştir.

Çalışmada endeks değişkeni görev değişkeni olarak adlandırılmış olup zaman olarak tanımlanmıştır. Eşik değeri olarak tedbir politikasının uygulandığı tarih alınmıştır. Bant genişliğinin 2 ay olarak tanımlandığı aralıkta politikanın uygulanmasından hemen sonra ürün fiyatlarında %17 ile %24 arasında keskin bir düşüş gözlemlenmiştir. Bu düşüşün etkiye bağlı olduğunu göstermek adına bir

önceki yılın aynı dönemindeki fiyat hareketleri incelenmiş olup herhangi bir keskin değişikliklere rastlanmamıştır.

Arslan [27] çalışmasında regresyon süreksizliği yöntemini kullanarak Türkiye’de il bazında terörizm ve iktisadi kalkınma arasındaki ilişkiyi incelemiştir. Çalışmada elde edilen sonuçlara göre terörizm ve iktisadi kalkınma arasında ters yönlü bir ilişki olduğu saptanmıştır. Ekonomik büyümenin %1 artış göstermesinin terör endeksi üzerinde bölgenin sahip olduğu değerin %2,42’lik bir düşüşüne neden olduğu görülmektedir.

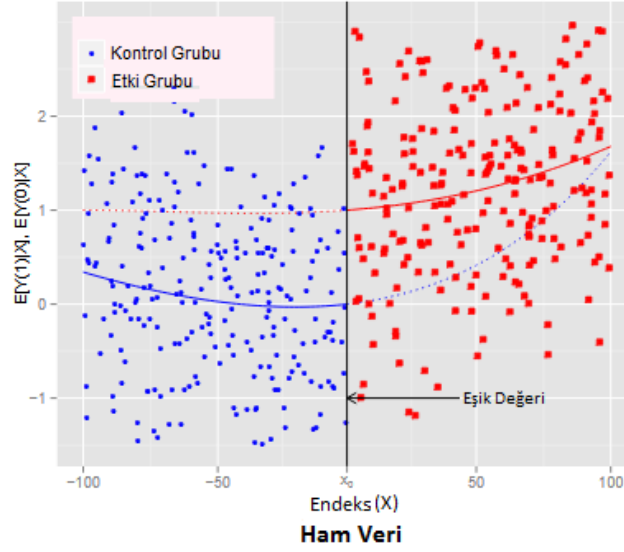
3.3 Grafikselsel Gösterim

Grafikselsel gösterim regresyon süreksizliği yönteminin ilk adımı olarak görülmektedir. Verinin grafik üzerindeki dağılımı herhangi bir noktada süreksizlik olup olmadığına ait fikir vermesi açısından önemlidir.

Grafikselsel gösterim regresyon süreksizliği yönteminin ayrılmaz bir parçasıdır [17]. Grafikselsel gösterimin regresyon süreksizliğindeki önemi ve uygulamasını detaylı olarak incelemiştir. Grafikselsel gösterimin regresyon süreksizliğini daha cazip bir hale getirdiğine değinilmektedir. Grafikselsel gösterim ile veri daha kolay ve detaylı incelenebileceği için çalışmalarda kullanılması tavsiye edilmektedir [15], [28].

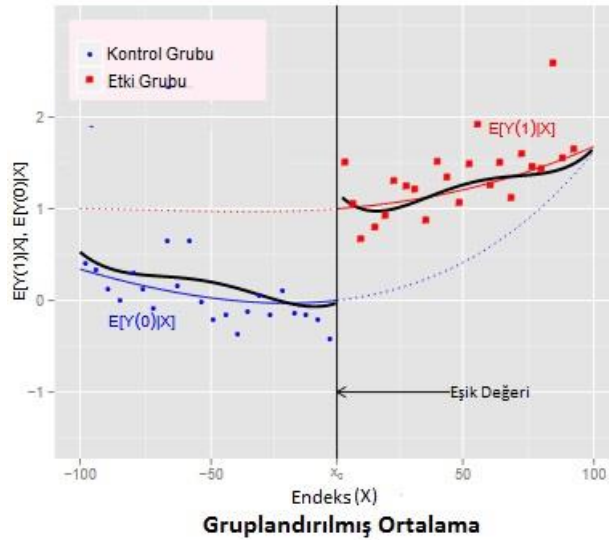
Grafikselsel gösterimin regresyon süreksizliği yönteminde uygulamasının temelinde 3 faydası vardır:

1) Fazla sayıda bulunan verinin grafik üzerinde gösterilmesi, eşik değeri üzerinde olası bir süreksizlik gözden kaçmasını engellemektir. Süreksizliğin daha kolay tespit edilebilmesi için endeks değişkeni boyunca eşik değerinin altında ve üstünde kalan alan ayrı ayrı belirli ölçülerde bölünerek verilerin bu bölünmüş gruplarda tek bir nokta halinde temsil edilmesi önerilmektedir.



Şekil 12: Ham Verinin Saçılım Grafiği

Şekil 12’de her bir bireye ait ham verilerin saçılım grafiği gösterilmektedir [15]. Eşik değerinin üstünde gösterilen etki grubuna ait bireyler ile eşik değerin altında yer alan kontrol grubuna ait bireylerin özellikle eşik değeri yakınlarında birbirine yakınlık göstermesi eşik civarındaki bir sıçramanın göz ile tespit edilmesini zorlaştırmaktadır.



Şekil 13: Gruplandırılmış Ortalamaya Ait Verinin Saçılım Grafiği

Şekil 13’te, etki ve kontrol gruplarına ait gruplandırılmış ortalamalar ile oluşturulan saçılım grafiğinde ise eşik değeri çevresinde süreksizlik net bir şekilde tespit edilebilmektedir [15]. Şekil 12’de eşik değerinin alt alanı ve eşik değerinin üst alanı

için ayrı ayrı hesaplanan regresyon modellerine göre Şekil 13'de aynı alanlar için hesaplanan regresyon değerinin daha tutarlı olduğu görülmektedir. Sonraki bölümlerde verilecek regresyon süreksizliği yönteminde tahmin bölümünde kullanılacak yöntemler, grafiksel gösterimin sadece biraz daha karmaşık hali olarak değerlendirilmiştir.

2) Regresyon süreksizliği yönteminde tahmin amacıyla kullanılan regresyon fonksiyonun derecesine ait bir fikir vermesidir. Gruplandırılmış ortalamalar aracılığıyla oluşturulan yeni saçılım grafiği ham veriye göre daha tutarlı ve kullanışlı bir regresyon fonksiyonunun hesaplanmasında kullanılabilir.

3) Belirlenen eşik değeri dışındaki bir ya da birden fazla noktada süreksizliğin olup olmadığını tespit etmektir. Eşik değeri dışında olup, eşik değeri ile karşılaştırılabilir süreksizliklerin olması halinde süreksizliğin başka sebeplerden dolayı oluşmuş olduğu düşünülmektedir [17], [29].

Şekil 12'de ham verinin saçılım grafiği gösterilmiş, Şekil 13'de ise veri endeks değeri üzerinde belirli ölçülerde bölünerek gruplandırılmıştır.

Gruplandırma sonucunda elde edilecek grafik belirlenen grup genişliğine bağlı olarak farklılıklar gösterebilmektedir. Grup genişliği endeks değişkeni üzerinde alınan grupların büyüklüğünü ifade eder. Farklı grup genişliklerinde farklı sonuçlar alınabileceği için bilgi kaybı ya da bilgi karmaşasının uygun düzeylerde tutulması gerekmektedir. Gruplandırılmış ortalamaların hesaplanma aşamaları aşağıda verilmiştir.

- 1- Endeks değişkeni, önceden belirlenen grup genişliğine göre eşit aralıklarla bölünür. Bölme işlemini başlangıç noktası eşik değeri olup, eşik değerinin altı için alt sınıra kadar, eşik değerinin üstü için üst sınıra kadar gider. Bölünme sonrası oluşan bu aralıklara küme ya da grup adı verilmektedir. Bir veri belirlenen gruba ait aralık dışında başka bir grupta yer almamalı yalnızca bir gruba ait olmalıdır. Özellikle eşik değerinin altında ve üstünde yer alan ilk gruplar arasında verilerin dağıtılmasına dikkat edilmelidir.

- 2- Her bir grup için, gruplarında yer alan verilerin sonuç değerlerinin ortalaması alınır. Ayrıca grubun endeks değeri üzerindeki aralığında orta noktası belirlenir.
- 3- Sonuç değeri ortalaması ile Y eksenini üzerinde, endeks değeri ortalaması ile X eksenini üzerinde bir nokta oluşturulur. Bu nokta o grubun temsil edildiği yeni noktadır. Ayrıca grupta kaç veri olduğunu gösteren ağırlık değeri de hesaplanır. Bu işlemler her bir grup için yapılır [14].

Yukarıda sıralanan işlemlerin yapılmasıyla verinin karmaşık yapısı giderilerek daha belirgin ve kolay çözümlenebilir bir hale dönüştürülmüştür.

3.3.1 Optimal Grup Sayısının Belirlenmesi

Grup genişliği belirlenirken görsel olarak grupların veride karmaşa yaratmayacak kadar dar, eşik değeri üzerinde karşılaştırılabilir olacak kadar da geniş olması gerekir. Karmaşık verinin düzlemde rahatlıkla anlaşılabilir seviyeye indirgenmesi ve aynı zamanda taşıdığı bilgide kayıp yaşamaması istenir.

Optimal grup sayısı, karmaşık ham veriyi düzlem üzerinde gruplar halinde temsil edilebilecek en uygun grup sayısı anlamına gelmektedir. Böylece verinin görsel gücü en üst seviyeye çıkarılır.

Optimal grup sayısının belirlenmesi sırasında birbirinden farklı grup genişlikleri hem görsel hem de istatistiksel olarak test edilir [29].

Optimal grup sayısının belirlenmesi için F-testi ya da birleştirilmiş hata kareler ortalaması kullanılmaktadır.

3.3.1.1 F-Testi

Uygun grup büyüklüğünün tespit edilmesi için F-testinin yapılması önerilmektedir. Bu test ile belirlenen grup genişliğinin veriyi düzleştirme seviyesine göre yeterli olup olmadığının tespiti yapılır. Eğer grup genişliği fazla ise ham veri fazlasıyla düzleşmiş olur. Bu sebeple daha küçük aralıklar içeren grup genişliklerinin denenmesi gerekmektedir. F testi bu aşamada hangi grup genişliğinin yeterli düzeyde düzleştirme yaptığının belirlenmesinde kullanılmaktadır [29].

F testi adımları aşağıdaki gibidir.

- 1- Test edilmek istenilen grup büyüklüğü h için, eşik değerinin iki tarafında gruplandırmalar yapılarak K adet grup için değişkenler oluşturulur. Eşik değerinin solunda oluşan grupların toplam sayısı K_0 , sağında oluşan grupların sayısı K_1 olduğu varsayılır ise;

$$K = K_0 + K_1$$

kadar grup ve dolayısıyla değişken oluşturulur.

- 2- Endeks değerlerinin her biri ait olduğu gruba göre K değişken için tanımlanır ve K değişken için sonuç değerinin tahmin edildiği regresyon analizi yapılır (Regresyon 1)
 - 3- Her bir grup büyüklüğü yarıya indirilir ve grup sayısı önceki değerinin 2 katına çıkarılmış olur. Oluşturulan yeni grup önceki K sayıdaki grubu kapsamaktadır.
 - 4- Yeni oluşturulan $2K$ grup için 2. adımda uygulanan regresyon analizi uygulanır. (Regresyon 2)
- Yapılan iki regresyon için edilen R^2 değerleri sırasıyla R_r (Sınırlandırılmış R^2) ve R_u (Sınırlandırılmamış R^2) olarak hesaplanır.

Aşağıda belirtilen F testi yapılır:

$$F_{Hesap} = \frac{(R_u^2 - R_r^2)/K}{(1 - R_u^2)/(n - K - 1)} \sim F_{Tablo(K;n-K-1)} \quad (3.6)$$

Eşitlik 3.6'da n , veri setinde yer alan toplam gözlem sayısını, K birinci regresyon için kullanılan grup sayısını göstermektedir.

- 5- F -testi ile kurulan 1. ve 2. regresyon modelleri arasında farkın olup olmadığı test edilir. Yapılan testin red edilmesi verinin yeterli derece düzleştirilmediği yani seçilen h grup genişliğinin olması gerekenden büyük olduğunu gösterir.
- 6- Uygun grup genişliği F -Testi ile bulunana kadar adımlar tekrarlanır.

3.3.1.2. Birleştirilmiş Hata Kareler Ortalaması (BHKO)

Birleştirilmiş Hata Kareler Ortalaması (BHKO) (Integrated Mean Square Error – IMSE) yöntemi ham verinin belirli grup genişliği ile gruplandırılmasıyla oluşan grupların, ham verinin temsil ettiği varsayılan uygun dağılıma göre hata kareler ortalamasının minimum olduğu durumu temsil etmektedir. Bu yöntem ile eşik değerinin sağında ve solunda yer alan veriler ayrı ayrı düşünülerek gruplandırılır.

Eşik değerin altında kalan bölüm için farklı, üstünde kalan bölüm için farklı grup genişlikleri hesaplanabilir [28].

Grup sayısının büyük bir değer olarak belirlenmesi, verilerin bulunduğu aralıkların küçük gruplara bölünmesi anlamına gelmektedir. Bu küçük gruplar için alınacak ortalamalar ile veriyi temsil edebilecek uygun değerler elde edilebilir. Böylece regresyon denkleminin tahmini için düşük yanlılıkta grup sayısına ulaşılmış olur. Fakat bu durumun bir dezavantajı vardır. Grup genişliklerinin küçük olmasıyla gruplara düşen veri sayısı azalır, grup içi değişkenlik yükselir [28].

Calonico [2015] çalışmasında, eşik değerin altı ve üstü için gerekli optimal sabit grup genişliğini, yanlılık ve varyans arasındaki ilişkiyi göz önüne alarak Eşitlik 3.7’de tanımlamıştır:

$$\delta^{BHKO} = \frac{K}{n} Varyans + \frac{1}{K} Yanlılık^2 \quad (3.7)$$

Verilerin gruplandırılması için eşit aralıklı (Evenly spaced (ES)) ve eşit verili (Quantile Spaced (QS)) olmak üzere iki yöntem önerilir. 3.3.1.1 bölümünde Lee [29] çalışmasında önerilen gruplandırma yöntemi ile benzerliğinden dolayı, çalışmada önerilen eşit aralıklı (ES) gruplandırma yöntemi incelenmiştir.

Bu yöntemde, eşik değerin alt bölümü (K_0) için Varyans ve Yanlılık hesaplamaları Eşitlik 3.8 ve Eşitlik 3.9’da gösterilmiştir:

$$Varyans_- = \frac{1}{\bar{x} - x_l} \int_{x_l}^{\bar{x}} \frac{\sigma^2(x)}{f(x)} d(x) \quad (3.8)$$

$$Yanlılık_- = \frac{(\bar{x} - x_l)^2}{12} \int_{x_l}^{\bar{x}} \left(\mu^{(1)}(x) \right)^2 d(x) \quad (3.9)$$

Eşitliklerde yer alan parametreler aşağıda verilmektedir.;

x_l : x endeks değerin taban değeri,

$\bar{x}$: eşik değeri

$\mu^{(1)}(x)$:1. dereceden türevi alınabilen $E[Y_i(0)|X_i = x]$

σ^2 : $\text{Var}[Y_i(0)|X_i = x]$

Eşitlik 3.7’de görüleceği gibi grup sayısı K ’nın artması Varyans teriminin artmasına ve yanlılık teriminin azalmasına sebep olmaktadır. Optimal BHKO değerinin bulunması için karşılaştırılmak istenilen K_i grup sayısı değerleri sırasıyla uygulanarak elde edilen BHKO değerleri karşılaştırılır ve aralarından BHKO değerinin en küçük yapan grup sayısı belirlenir.

$$\text{Eşik değerinin solu için uygun grup sayısı ;} \quad K_0^{BHKO} = \left\lceil \delta_-^{BHKO} n^{\frac{1}{3}} \right\rceil$$

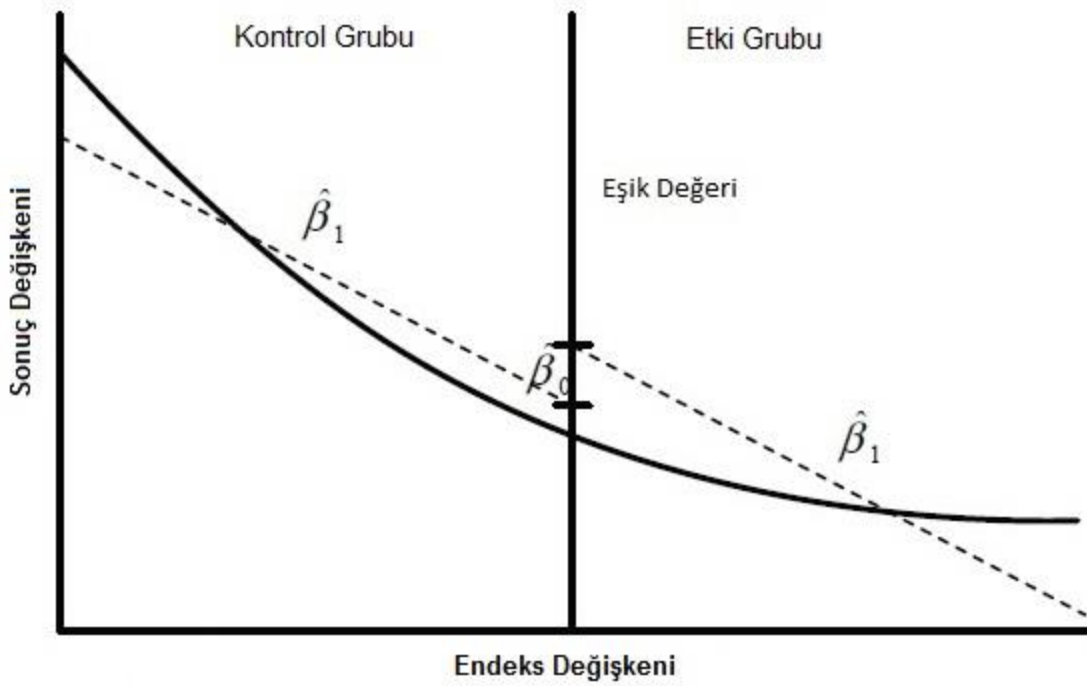
$$\text{Eşik değerinin sağı için uygun grup sayısı ;} \quad K_1^{BHKO} = \left\lceil \delta_+^{BHKO} n^{\frac{1}{3}} \right\rceil$$

3.4 Regresyon Süreksizliği Yönteminde Tahmin

Regresyon süreksizliği yönteminin temel amacı etkinin tahmin edilmesidir. Etki değerlendirmesi yapılırken, yarı deneysel etki değerlendirme yöntemlerinin tamamında olduğu gibi regresyon süreksizliği yönteminde de kontrol grubunun oluşturulması sırasında görülen seçim yanlılığı problemi ile karşılaşmaktadır. Eğer kontrol grubu bireylerinin seçim süreci tamamen bilinir ve seçimden kaynaklı etki ölçülebilirse, doğrudan programın yansız etkisi tahmin edilebilmektedir [14].

Regresyon süreksizliği yönteminde endeks değişkenine ait bilgi varken, endeks değişkeni ile sonuç değişkeni arasındaki ilişkiyi tanımlayan doğru model fonksiyonu hakkında bilgi bulunmamaktadır. Doğru regresyon model fonksiyonun bilinmesi kontrol grubuna ait bireylerin seçiminin nasıl yapıldığının bilinmesini sağlayacak ve yansız tahmin yapılmasına olanak tanıyacaktır [14].

Modelin doğru formda tahmin edilmesi ve eşik değeri üzerindeki durumunun incelenmesi gerekmektedir. Doğrusal olmayan bir modeli yanlış tahmin ederek doğrusal bir model gibi tahmin edersek, oluşturulan modelin büyük olasılıkla eşik değeri üzerinde yaratacağı süreksizlik yanılmaya sebep olacaktır.



Şekil 14: Hatalı Model Tahminin Oluşturduğu Süreksizlik

Şekil 14'de gösterilen kalın eğri çizgi endeks-sonuç değişkenleri arasında gerçek ilişkiyi göstermektedir [14]. Çizgi doğrusal olmayan ve azalarak ilerleyen bir özelliktedir. Eşik değeri üzerinde bu çizgiye ait herhangi bir süreksizlik görülmemektedir. Dolayısıyla bir etkiden söz edilememektedir. Fakat bu ilişkinin doğrusal olarak değerlendirilmesi ve eşik değeri sınırları arasında kontrol ve etki grupları için ayrı ayrı incelenmesi hataya yol açmaktadır. Kontrol ve etki grupları için şekilde ayrı ayrı modellenen veriler eşik değeri üzerinde süreksizliğe neden olur. Süreksizlik etki olmadığı halde bir etkinin varmış gibi görünmesine sebep olmaktadır [14].

Etkinin tahmininin yapılması için regresyon model fonksiyonunun doğru tahmin edilmesi gerekmektedir. Bu sebeple araştırmacılar genel olarak tüm model fonksiyon çeşitlerini (doğrusal, karesel, kübik, kuadratik vb.) veri seti üzerinde uygulayarak en doğru model fonksiyonunun tahmin edilmesini önermektedirler. Model fonksiyonlarının tahmin edilmesinde iki temel yaklaşım vardır. Birincisi tüm veri seti kullanılarak tahminin yapıldığı parametrik tahmin, ikincisi eşik değeri çevresindeki kısıtlı bir alanda yer alan verileri kullanarak yapılan parametrik olmayan tahmindir [14].

3.4.1 Regresyon Süreksizliği Yönteminde Model ve Tahmin Yöntemi Seçimi

Parametrik tahminde, veri setindeki tüm verileri analize dahil etmeyi amaçlayan kapsayıcı bir özelliği vardır. Böylelikle eşik değerinden uzak verilerinde eşik değeri üzerinden etkinin hesaplanmasına katkı sağlamaktadır. Farklı model formları ile en düşük yanlılık değerine sahip olan uygun model seçilebilmesine de imkan sağlar.

Parametrik olmayan tahmin, bölgesel olarak incelenmektedir. Regresyon süreksizliği yöntemi, etkiden kaynaklı olarak eşik değeri üzerindeki süreksizliğe odaklanması sebebiyle, parametrik olmayan tahmin eşik değeri çevresindeki verilerin kullanıldığı tahminleri içermektedir. Analiz, sadece eşik değerine belli bir uzaklık çerçevesinde olan verileri kapsar ve eşik değerinden uzak veriler göz ardı edilmektedir. Eşik değerinin üstünde ve altında yer alan etki ve kontrol gruplarının birbiri ile benzer özellikler gösterdiği varsayımı altında karşılaştırma olanağı sağlanmaktadır [14].

Parametrik tahminde tüm veri seti model tahmininde kullanıldığı için hassas tahminler yapılmaktadır. Fakat tüm veriler üzerinden tahmin edilen modelin, çoğu zaman 3. ya da 4. dereceden model fonksiyonu olduğu görülmektedir. Bu durum, karmaşık bir model ile yanlı tahminler yapılmasına sebep olmaktadır.

Parametrik olmayan tahminde eşik değeri çevresinde yer alan kısıtlı sayıda veri ile model tahmin edildiği için hassas tahminler yapılamamaktadır. Fakat kısıtlı bir alanda yer alan verilerin kullanılması karmaşık olmayan modellerin tahminine ve bunun sonucunda da yansız tahminler yapılmasına sebep olmaktadır.



Şekil 15: Parametrik Tahmin ve Parametrik Olmayan Tahminin Yaklaşımları

Şekil 15’de Parametrik tahmin ve parametrik olmayan tahmin yaklaşımları modellenmiştir. Parametrik yaklaşım tüm veri setini kullanarak uygun modelin tahmin edilmesine dayanırken, parametrik olmayan tahmin belirlenen basit bir modeli tahmin etmek için eşik değeri çevresindeki verileri içeren genişliğin bulunmasına dayanmaktadır.

3.4.1.1 Regresyon Süreksizliği Yönteminde Parametrik Tahmin

Parametrik tahmin, veri setinin tamamı ile tahmin edilecek bir modeli temel almaktadır. Model, Eşitlik 3.10’da

$$Y_i = \alpha + \beta_0 T_i + f(r_i) + \varepsilon_i \quad (3.10)$$

biçiminde tanımlanır. Eşitlikte,

α = ortalama etki değeri

Y_i = i. gözlem için elde edilen sonuç değeri

T_i = etki grubunda ise 1, kontrol grubunda ise 0,

r_i = i. birey için endeks değeri,

ε_i = i. birey için rasgele hata terimi,

β_0 = eşik değerindeki marjinal program etkisi

dir. Parametrik tahminde model birden çok olası durum için tanımlanabilir. Model; doğrusal, karesel, kübik formunda olabilir ve ayrıca etkileşim terimleri içerebilir. Modeller aşağıda verilmiştir:

1. Doğrusal Form

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \varepsilon_i \quad (3.11)$$

2. Doğrusal Form ve Etkileşim

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i \cdot T_i + \varepsilon_i \quad (3.12)$$

3. Karesel Form

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \varepsilon_i \quad (3.13)$$

3. Karesel Form ve Etkileşim

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \beta_3 \cdot r_i \cdot T_i + \beta_4 \cdot r_i^2 \cdot T_i + \varepsilon_i \quad (3.14)$$

4. Kübik Form

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \beta_3 \cdot r_i^3 + \varepsilon_i \quad (3.15)$$

5. Kübik Form ve Etkileşim

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot r_i + \beta_2 \cdot r_i^2 + \beta_3 \cdot r_i^3 + \beta_4 \cdot r_i \cdot T_i + \beta_5 \cdot r_i^2 \cdot T_i + \beta_6 \cdot r_i^3 \cdot T_i + \varepsilon_i \quad (3.16)$$

Parametrik tahmin yapılırken dikkat edilmesi gereken bir noktada modelin karmaşıklığı ne kadar artar ise analizin gücünün o kadar düşeceği. Bu yüzden olabildiğince basit modellerin seçilmesine dikkat edilmesi gerekmektedir [14].

3.4.1.1.1 Parametrik Tahminde Uygun Modelin Seçilmesi İçin Kullanılan Yöntemler

i) F Testi Yöntemi

Parametrik tahminde uygun modeli belirlemek amacıyla F- testinin kullanılabileceği önerilmiştir. F-testi için belirlenen adımlar sırasıyla aşağıda yer almaktadır [29].

- 1- K-2 grup sayısı için gösterge seti hazırlanır. Kullanılan K değeri grafiksel gösterimde yer alan K ile aynıdır. Doğrusallık probleminde kaçınmak için K-2 kadar grup alınır.
- 2- Yukarıda bahsedilen olası her bir model (6 model) için regresyon analizi yapılır (regresyon 1).
- 3- İlk adımda oluşturulan göstergeler de eklenerek yeniden regresyon analizi yapılır (regresyon 2).

- 4- Sırasıyla regresyon 1 için R_r^2 ve regresyon 2 için R_u^2 değerleri elde edilir.
- 5- F istatistik değeri Eşitlik 3.17'de belirtilen formül ile hesaplanır.

$$F_{Hesap} = \frac{(R_u^2 - R_r^2)/K}{(1 - R_u^2)/(n - K - 1)} \sim F_{Tablo(K;n-K-1)} \quad (3.17)$$

- 6- Eğer sonuç istatistiksel olarak anlamlı değilse, 3. adımda eklenen verilerin modele ek bir bilgi katmadığı ortaya çıkmaktadır bu da modelin veriyi doğru açıklayamadığı anlamına gelmektedir.

Çoğu zaman ilk olarak en basit model olan doğrusal model ile analize başlanır. Örneğin etkileşimli doğrusal modele karşı doğrusal model karşılaştırıldığında ve F-testi sonucu istatistiksel olarak anlamlı çıkmadığında doğrusal modelin veri seti için en uygun model olduğu söylenebilir. Tersine olması durumunda etkileşimli doğrusal model ile bir üst dereceden model karşılaştırılır. Bu işlem yüksek dereceli fonksiyonlarda istatistiksel olarak anlamlı bir sonuç elde edilene kadar devam eder [14].

ii) Akaike Bilgi Kriteri

Parametrik tahminde uygun modelin tespiti için bir diğer yöntem Akaike Bilgi Kriteri (Akaike Information Criteria (AIC)) kullanılabilir. AIC en iyi olabilirlik yöntemini kullanarak tutarlılık-yanlılık arasında en uygun değişimi sağlayarak modeli tahmin eder. AIC'ne ait formül Eşitlik 3.18'de yer almaktadır.

$$AIC = N \ln(\widehat{\sigma_b^2}) + 2p \quad (3.18)$$

$\widehat{\sigma_b^2}$, p parametrelili bir model için tahmin edilen hata varyansıdır. Formülden de anlaşıldığı gibi tahmin edilen varyansın artması ya da parametre sayısının artması AIC değerini arttırmaktadır. AIC değeri tutarlılık-yanlılık arasında en uygun değişimi sağladığı için örneğin yüksek dereceden bir modelin parametre sayısı artacağından bu modellerde tahmini hata varyansı azalacaktır.

AIC değeri karşılaştırılan modeller arasında veriye en uygun olan modelin belirlenmesinde kullanılır. Fakat bir grup arasından seçilen uygun modelin veriye ne kadar uyduğunu belirleyemez. Bütün olası modellerin veriyi çok düşük

açıkladığı bir model kümesi arasında AIC ile seçilen en uygun model bu açıdan çok kullanışlı olmayacaktır. Bu sebeple ilk aşamada F-testi'nin yapılması ardından AIC değeri ile uygun modelin seçilmesi önerilmektedir [14].

3.4.1.2 Regresyon Süreksizliğinde Parametrik Olmayan Tahmin

Parametrik olmayan yöntemler ile model, eşik değerini de kapsayacak şekilde, eşik değerinin altında ve üstünde belirlenen bir çerçeve içerisinde yer alan verileri kullanarak tahmin edilir. Bu çerçeveye bant genişliği ya da süreksizlik örneği adı verilmektedir [14].

Parametrik olmayan yaklaşımda etki hesaplanırken, örneğin eşik değerinin altında yer alan verilerin kontrol grubu ve eşik değerinin üstünde yer alan verilerin etki grubu olduğu varsayıldığında, ilk adımda bu iki gruba ait model denklemleri yukarıda bahsedilen bant genişliği içerisinde yer alan veriler kullanılarak tahmin edilmektedir. İkinci adımda tahmin edilen bu modellerin kontrol grubu için $E[Y_i(0) | X_i = c]$, etki grubu için $E[Y_i(1) | X_i = c]$ olacak şekilde eşik değerinde olacakları beklenen değer bulunmaktadır. Son adımda bu her iki gruba ait bulunan beklenen değerlerin farkı alınarak etki hesaplanmaktadır [15].

Burada önemli olan etki grubu ve kontrol grubuna ait modellerin doğru tahmin edilmesidir. Modellerin tahmin edilmesinde yerel polinom yaklaşımı kullanılmaktadır. Yerel polinom yaklaşımında sadece eşik değerinin hemen yakınlarında yer alan bireylere ait veriler kullanılmaktadır. Yaklaşımda tahmin edilen modeller genellikle doğrusal ya da karesel modellerdir.

Yerel polinom regresyonu, çekirdek regresyonuna benzer ancak ağırlıklandırılmış ortalama olan kestirim değerleri yerine ağırlıklandırılmış en küçük kareler kestirim değerlerini kullanır [30].

Yerel polinom yaklaşımında genelde bant genişliği içerisinde yer alan bireylere ait endeks değerleri, eşik değerine yakınlıklarına göre ağırlıklandırılmaktadır. Ağırlıklandırma, kullanılan çekirdek fonksiyonuna göre farklılık göstermektedir [15].

Yerel polinom yaklaşımıyla etkinin hesaplanması için adımlar aşağıda belirtilmiştir:

- 1- Bant genişliği (h) belirlenir.

- 2- Çekirdek fonksiyonu ($K(\cdot)$) seçilerek bant genişliği içerisinde yer alan her bir bireye ait endeks değeri ağırlıklandırılır.
- 3- Modele ait fonksiyonun derecesi (p) belirlenir.
- 4- Eşik değerinin üstünde yer alan bireylere ait endeks değerleri ($X_i \geq c$) çekirdek fonksiyonu ile ağırlıklandırılır. Endeks değeri ile sonuç değişkeni arasındaki ilişkiyi gösteren ağırlıklandırılmış en küçük kareler regresyon modeli tahmin edilir. Tahmin edilen model kullanılarak, eşik değeri üzerinde beklenen değeri, ($\alpha_1 = E[Y_i(1) | X_i = c]$) Eşitlik 3.19'da gösterilmiştir.

$$\begin{aligned} \hat{Y}_i = & \hat{\alpha}_1 + \beta_1 \cdot (X - c) + \beta_2 \cdot (X - c)^2 + \dots \\ & + \beta_p \cdot (X - c)^p + \varepsilon \quad c \leq X < c + h \end{aligned} \quad (3.19)$$

- 5- Eşik değerinin altında yer alan bireylere ait endeks değerleri ($X_i < c$) çekirdek fonksiyonu ile ağırlıklandırılır. Endeks değeri ile sonuç değişkeni arasındaki ilişkiyi gösteren ağırlıklandırılmış en küçük kareler regresyon modeli tahmin edilir. Tahmin edilen model kullanılarak, eşik değeri altında beklenen değer, ($\alpha_0 = E[Y_i(0) | X_i = c]$) Eşitlik 3.20'de gösterilmiştir.

$$\begin{aligned} \hat{Y}_i = & \hat{\alpha}_0 + \beta_1 \cdot (X - c) + \beta_2 \cdot (X - c)^2 + \dots \\ & + \beta_p \cdot (X - c)^p + \varepsilon \quad c - h \leq X < c \end{aligned} \quad (3.20)$$

- 6- Eşik değeri üzerinde etki $\hat{\tau} = \hat{\alpha}_1 - \hat{\alpha}_0$ şeklinde hesaplanır [15].

Dolayısıyla yerel polinom yaklaşımı uygulanırken, üç ana bileşenin seçiminin yapılması gerekmektedir. Bunlar; çekirdek fonksiyonu $K(\cdot)$, yerel polinom modelin derecesi (p) ve bant genişliğidir(h) [15].

3.4.1.2.1. Çekirdek Fonksiyonu ve Yerel Polinom Modelin Derecesinin Belirlenmesi

Çekirdek fonksiyon, her bir bireye ait endeks değerinin, negatif olmayan değerler ile ağırlıklandırılmasını sağlamaktadır. Ağırlıklandırma bireyin endeks değerinin eşik değerine göre yapılmakta olup, bant genişliği içerisinde yer alan bireylere uygulanmaktadır. Bant genişliği dışarısında yer alan bireylerin ağırlıklandırılmasında çekirdek fonksiyonu 0 değerini almaktadır. Uygulamalarda,

tahmin sonuçlarına seçilen çekirdek fonksiyon çeşidinin etkisinin olmadığı görülmektedir [15].

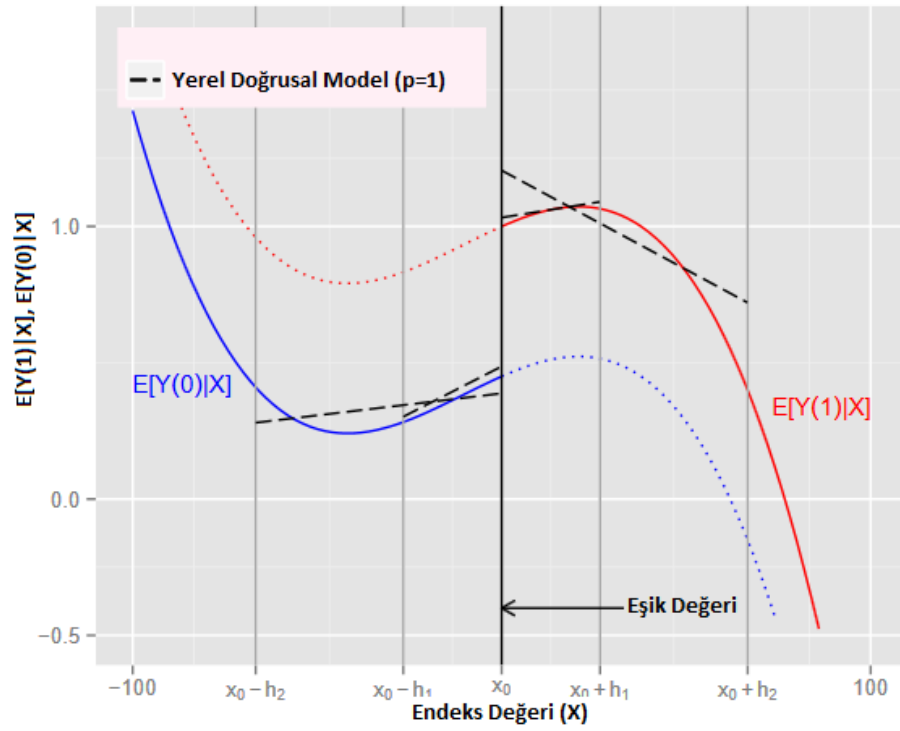
Çizelge 2 : Çekirdek Fonksiyonlar

Uniform Çekirdek Fonksiyon	$K(u) = 1, \quad u \leq 1$
Dikdörtgen Çekirdek Fonksiyon	$K(u) = \frac{1}{2}, \quad u \leq 1$
Üçgen Çekirdek Fonksiyon	$K(u) = (1 - u), \quad u \leq 1$
Epanechnikov Çekirdek Fonksiyon	$K(u) = (1 - u^2), \quad u \leq 1$

Yerel polinom modelinin derecesinin belirlenmesi parametrik olmayan tahminde ayrıca önem taşımaktadır. Verilen bant genişliği içerisinde belirlenen modelilin derecesi arttıkça tahminin doğruluğu artarken, aşırı uyum(overfitting) ve yanlılık problemleri de ortaya çıkmaktadır. Literatürde düşük dereceden modellerin kullanılması tavsiye edilmektedir. Çoğunlukla derecenin 1'e eşit olduğu yerel doğrusal polinom modeli kullanılmakta iken, derecenin 2'ye eşit olduğu yerel karesel polinom modeli de bazı uygulamalarda kullanılmaktadır.

3.4.1.2.2. Bant Genişliğinin Belirlenmesi

Parametrik olmayan yaklaşım ile etkinin hesaplanması sırasında kullanılan bant genişliği kavramı analizin en önemli parçasını oluşturmaktadır. Bant genişliğine göre analize dahil olan veri sayısı değişebilir. Bunun sonucunda bant genişledikçe ya da küçüldükçe verinin özelliği ve temsil edecek model değişebilir.



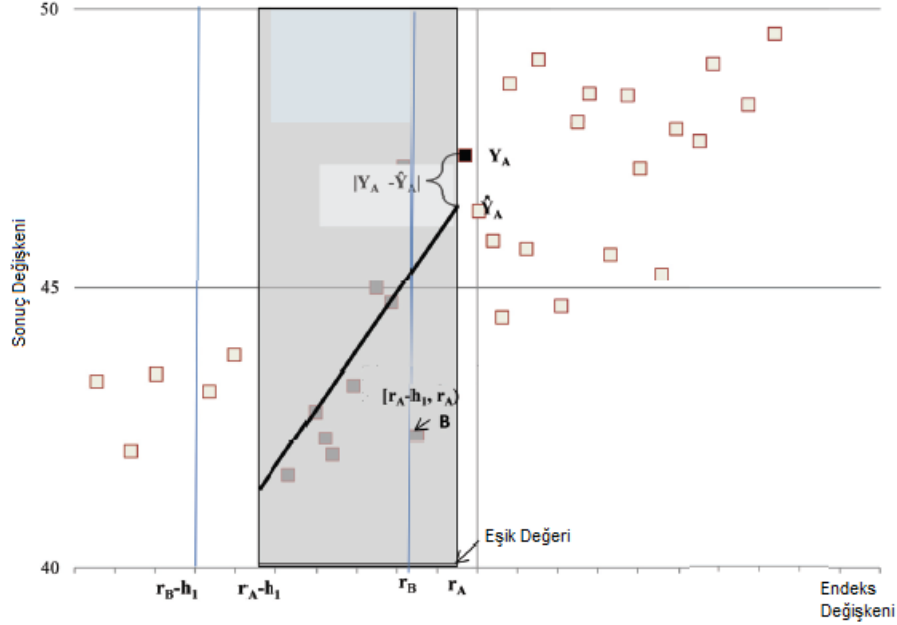
Şekil 16: Farklı Bant Genişliğinin Model Tahminine Etkisi

Şekil 16'da h_1 ve h_2 olmak üzere 2 farklı bant genişliği tanımlanmıştır [15]. İki bant genişliği de içerdiği veriler nedeniyle farklı modeller ile gösterilmektedir [15]. Daha dar bir çerçevede yer alan h_1 bant genişliği ile eşik her iki tarafında artan bir fonksiyon görünümünde iken, daha geniş bir alan içeren h_2 bant genişliğine ait model eşik her iki tarafında farklı özellikler göstermektedir. Ayrıca kullanılacak bant genişliğine göre sadece model de değişim olmaz. Model kullanılarak eşik değeri üzerinde tahmin edilen değerlerde farklı olacaklardır. Şekil 16'da görüldüğü üzere geniş bant genişliği için daha geniş bir etki görülmektedir [15].

Uygun bant genişliğini belirlenmesi için tutarlılık-yanlılık arasında bir denge olması gerekmektedir. Daha geniş bir bant genişliği daha fazla veri içereceği için yüksek bir tahmin gücüne sebep olurken verilerin doğrusal bir modelde açıklanmasını zorlaştırır.

i)Çapraz Geçerlilik Uygulanması

Bant genişliğinin belirlenmesi için uygulanan yöntemlerden biri çapraz geçerlilik yöntemidir.



Şekil 17: Çapraz Geçerlilik Yöntemi

Şekil-17 Çapraz geçerlilik yöntemi özetlemektedir. Yöntemin adımları sırasıyla aşağıda belirtilmiştir [14].

- 1) Başlangıç için h_1 bant genişliği belirlenir.
- 2) Eşik değerinin hemen solunda yer alan bir A noktası alınır. Bu A noktasının endeks değeri üzerinde r_A ve çıktı değeri üzerinde Y_A olarak gösterilir.
- 3) A noktası dışarıda kalacak şekilde endeks değeri üzerinde r_A , r_A-h_1 aralığında yer alan veriler kullanılarak regresyon modeli tahmin edilir.
- 4) A noktası dışarıda bırakılarak hesaplanan regresyon modeli kullanılarak r_A değeri için modelden $\hat{Y}_A$ değeri tahmin edilir.
- 5) Daha sonra bant genişliği h_1 'de sabit kalmak koşulu ile bant eşik değerinin solunda yer alan ikinci veriye kaydırılır. B olarak adlandıracağımız bu veri için tekrar tanımlanan 3. ve 4 basamakta yapılan işlemler gerçekleştirilir.
- 6) Bu adımlar en az iki gözlem bant içerisinde kalana kadar tekrarlanır.

7) Çapraz geçerlilik değeri $\zeta G(h_1)$,

$$\zeta G(h_1) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (3.21)$$

formülü ile hesaplanır. Eşitlikte N, veri setinde yer alan toplam gözlem sayısını, h_1 bant genişliğini göstermektedir. Dolayısıyla h_1 bant genişliği için çapraz geçerlilik değeri hesaplanmaktadır.

7- Bu adımlar sırasıyla $h_2, h_3, \dots$ değerleri için tekrarlanır.

8- En küçük CV değerine sahip bant genişliği uygun bant genişliği olarak belirlenir [29], [31].

ii) Plug-In Uygulanması

Imbens [32] çalışmasında veriye dayalı olarak optimal bant genişliğinin bulunması için yöntem önermektedir. Önerilen yöntemde 6 farklı değişken yer almaktadır. Değişkenlerin hesaplanmasında bir başlangıç bant genişliği kullanılır. Giriş bant genişliği kullanılarak oluşturulan değişkenler kullanılan formül sonucunda uygun bant genişliği bulunmaktadır. Bu sürece “Plug-In” ya da IK bant genişliği adı verilmektedir.

Kullanılan verilerin varyans, ortalama, medyan ve sıklığına göre bant genişliği de farklılık göstermektedir.

IK bant genişliği için uygun formül Eşitlik 3.22’de tanımlamıştır.

$$\hat{h}_{opt} = C_K \cdot \left(\frac{\frac{2 \cdot \hat{\sigma}^2(c)}{\hat{f}(c)}}{\left(\hat{m}_+^{(2)}(c) - \hat{m}_-^{(2)}(c) \right)^2 + (\hat{r}_+ + \hat{r}_-)} \right)^{\frac{1}{5}} \cdot N^{-\frac{1}{5}} \quad (3.22)$$

Formül kullanılmaya başlamadan önce yukarıda da bahsedilen giriş bant genişliğinin h_1 tanımlanması gerekmektedir. Imbens [32] çalışmasında bu giriş değeri Eşitlik 3.23’te gösterildiği gibidir:

$$h_1 = 1,84 \cdot S_X \cdot N^{-\frac{1}{5}} \quad (3.23)$$

olarak önerilmiştir. S_x , bütün örneklemin endeks değeri için hesaplanan standart sapma değeridir.

Formülü oluşturan değişkenler ve hesaplama adımları şu şekildedir:

- 1- $\hat{f}(c)$: Endeks değerinin eşik değeri çevresindeki tahmini olasılık yoğunluk fonksiyonu Eşitlik 3.24'te verilmiştir:

$$\hat{f}(c) = \frac{N_{h_1,-} + N_{h_1,+}}{2 \cdot N \cdot h_1} \quad (3.24)$$

Eşik değerinin her iki tarafında giriş eşik değeri kullanılarak bir aralık oluşturulur. Bu aralık içerisinde yer alan endeks değerleri kullanılarak $\hat{f}(c)$ fonksiyonuna ulaşılır. Eşik değerinin solunda yer alan aralıktaki endeks değeri sayısı $N_{h_1,-}$, sağında yer alan endeks değeri sayısı ise $N_{h_1,+}$ şeklindedir.

$$N_{h_1,-} = \sum_{i=1}^N 1_{c-h_1 \leq X_i \leq c} \quad N_{h_1,+} = \sum_{i=1}^N 1_{c \leq X_i \leq c+h_1}$$

- 2- $\hat{\sigma}^2(c)$: Eşik değer üzerindeki koşullu varyans fonksiyonu Eşitlik 3.25'da verilmiştir:

$$\hat{\sigma}^2 = \frac{1}{N_{h_1,-} + N_{h_1,+}} \left(\sum_{i:c-h_1 \leq X_i \leq c} (Y_i - Y_{h_1,-})^2 + \sum_{i:c \leq X_i \leq c+h_1} (Y_i - Y_{h_1,+})^2 \right) \quad (3.25)$$

- 3- $\hat{m}^{(2)}(c)$: Endeks değeri ve çıktı değeri arasındaki ilişkiyi gösteren regresyon denkleminin ikinci dereceden türevidir

$\hat{m}^{(2)}(c)$ değerleri eşik değerinin sağ ve solu için ayrı ayrı hesaplanır. Eşik değerinin sağ ($X_i \geq c$) için alınan medyan değeri X_+ , solu ($X_i < c$) için alınan medyan değeri X_- hesaplanır. Medyan değerlerini içeren $[X_+, X_-]$ aralığında yer alan veriler ile $X_i \geq 0$ için bir değişken eklenilerek 3. dereceden Eşitlik 3.26'de yer alan model denklemi oluşturulur:

$$Y_i = \gamma_0 + \gamma_1 \cdot 1_{X_i \geq c} + \gamma_2(X_i - c) + \gamma_3(X_i - c)^2 + \gamma_4(X_i - c)^3 + \varepsilon_i \quad (3.26)$$

Modelde yer alan $\hat{\gamma}_4$ en küçük kareler tahmin edicisi $\hat{m}^{(3)}(0) = 6 \cdot \gamma_4$ eşitliğinde kullanılarak h_2 denklemi için girdi oluşturulur. Eşik değerinin sağ ve solu için h_2 değerleri ayrı ayrı hesaplanır. Hesaplamada kullanılan formül Eşitlik 3.27’de verilmiştir:

$$h_{2,\mp} = 3,56 \left(\frac{\sigma^2(0)}{\hat{f}(0) \cdot \max\left((\hat{m}^{(3)}(0))^2, 0,01\right)} \right)^{1/7} \cdot N_{\mp}^{-1/7} \quad (3.27)$$

Elde edilen yeni $h_{2,\mp}$ aralıkları kullanılarak her iki aralık için 2. dereceden Eşitlik 3.28’da yer alan model oluşturulur. Bu aralıklardaki gözlem sayıları $N_{2,\mp}$ ’dir.

$$Y_i = \gamma_0 + \gamma_1 \cdot 1 + \gamma_2(X_i - c) + \gamma_3(X_i - c)^2 \quad (3.28)$$

Son basamakta ise $m_{\mp}^{(2)} = 2 \cdot \gamma_{3,\mp}$ eşitliği kullanılır.

- 4- $\hat{r}_+$ ve $\hat{r}_-$ düzeltme katsayılarının hesaplanması için Eşitlik 3.29’da yer alan formülden faydalanılır.

$$\hat{r}_+ = \frac{720 \cdot \hat{\sigma}^2(c)}{N_{2,+} \cdot h_{2,+}^4} \quad \hat{r}_- = \frac{720 \cdot \hat{\sigma}^2(c)}{N_{2,-} \cdot h_{2,-}^4} \quad (3.29)$$

Eşitlik 3.24’de diğer eşitliklerden elde edilendeğişkenler yerleştirilerek uygun bant genişliği bulunur. Eşitlik 3.22’de yer alan C_K sabiti için çalışmasında 3,4375 değerinin kullanılmasını önerilmiştir [32].

4. SAYISAL ÖRNEK

Bu bölümde, regresyon süreksizliği yönteminin uygulamada kullanımını göstermek amacıyla Calonica [2] çalışmasında verilen veri setinden yararlanılmıştır.

Siyasal bilimciler birçok çalışmada daha önce yönetimde olan partinin, iktidar partisi olarak girdiği seçimlerde iktidar partisi olma özelliğinden dolayı avantaja sahip olup olmadıklarını araştırmaktadırlar.

Literatürde iktidar olma avantajına yönelik çok sayıda çalışma bulunmakta olup çalışmaların büyük çoğunluğu Amerika Birleşik Devletleri temsilciler meclisi seçimleri üzerine yoğunlaşmaktadır [33], [34], [35].

Amerika Birleşik Devletleri temsilciler meclisi ABD kongresinin iki yasama organından biridir. Temsilciler meclisi üyeleri iki yılda bir yapılan seçimler ile belirlenmektedir [36]. ABD seçimlerinde oyların büyük çoğunluğu Demokratik parti ve Cumhuriyetçi parti arasında dağılmaktadır. İki partili bir sistemin oluşmasına sebep olmaktadır. Sistemde bir parti seçimde %50'nin üzerinde oy alarak iktidar olmaya hak kazanırken, diğer parti muhalefette yer almaktadır.

Lee [29] çalışmasında, ABD temsilciler meclisi üyeleri seçimlerinde, geçmiş seçimi kazanarak iktidar olma avantajına sahip olan parti ile seçimi kaybederek avantaja sahip olmayan partinin, sonraki seçimde oy oranlarında görülen farklılığa, iktidarda olma avantajının etkisinin olup olmadığı regresyon süreksizliği yöntemi kullanılarak incelemiştir.

İki partili seçim sistemlerinde, bir parti %50 üzerinde oy alarak o dönem yönetimde olma yani iktidar olmaktadır. Rakibi olan diğer parti %50'nin altında kalacağı için yönetimde olma avantajına sahip değildir. Dolayısıyla bir parti seçimde %50 oy oranını aşarsa bir sonraki seçime iktidar olma avantajı ile girmektedir. Burada iktidar olma avantajı, regresyon süreksizliği yönteminde kullanıldığı gibi programdan yararlanmaya benzetilebilir. Partinin iktidar olma avantajına sahip olabilmesi için seçimde %50 oy oranını geçmesi gerektiğinden, %50 oy oranı regresyon süreksizliği yönteminde eşik değeri olarak kabul edilebilir. Endeks değişkeni olarak partinin t döneminde yapılan seçimde oy oranı, sonuç değişkeni olarak partinin $t+1$ döneminde yapılan seçimdeki oy oranı kullanılmaktadır. Regresyon süreksizliği yöntemi kullanılarak partinin t döneminde küçük bir farkla

seçimi kazandığı bölgeler ile küçük bir farkla seçimi kaybettiği bölgelerde yapılacak $t+1$ dönemindeki seçimlerde görülecek oy oranı farkı karşılaştırılmak istenmektedir. Burada görülecek farkın sebebi olarak partinin iktidar olma avantajı düşünülmektedir [37].

4.1 Veri

Uygulamada, R programında paketi olan “Rdrobust” paketinde yer alan, 1914-2010 yılları arasında ABD’de yapılan senato üyeleri seçimlerine ait veri seti kullanılmaktadır [2]. ABD seçimlerinde iki partili sistemin ağırlıklı olması sebebiyle veri seti partilerden biri olan Demokratik partiye göre hazırlanmıştır. Veri seti iki değişkenden oluşmaktadır. Birinci değişken fark değişkeni olup t döneminde yapılan seçimlerde Demokratik partinin aldığı oy oranının Cumhuriyetçi partinin oy oranına olan farkını göstermektedir. Örneğin aynı dönemde belirli bir bölgede yapılan seçimde Demokratik parti %55 oy aldığı bölgede Cumhuriyetçi parti %45 oy almıştır. Burada fark değişkenine ait değer oy oranları farkı olan %10 olmaktadır. Fark değişkeni regresyon süreksizliği yöntemi uygulamasında endeks değişkeni olmaktadır. İkinci değişken oy oranı değişkeni olup $t+1$ döneminde yapılan seçimlerde Demokratik partinin oy oranını göstermektedir. Oy oranı değişkeni regresyon süreksizliği yöntemi uygulamasında sonuç değişkeni olmaktadır. Seçimlerde Demokratik partinin iktidar partisi ya da muhalefet partisi olacağının belirlenmesi fark değişkeni üzerinde 0’dan büyük ya da küçük değer almasına bağlıdır. Eğer fark değişkeni 0’dan büyük olursa Demokratik parti t döneminde yapılan seçimi kazanmış ve bir sonraki seçime iktidar olma avantajı ile girmektedir. Eğer fark değişkeni 0’dan küçük olursa Demokratik parti t döneminde yapılan seçimi kaybetmiş ve bir sonraki seçime iktidar olma avantajı olmadan girmektedir. Bu sebeple partinin aldığı oy farkının 0’a göre konumu partinin iktidar olup olmadığını belirlediğinden 0 değeri çalışmada eşik değeri olarak alınmaktadır. Veri setinde yer alan bir gözleme ait fark değişkeni için -7,69, oy oranı değişkeni 36,10 olarak görülmektedir. Bu değerler sırasıyla; t döneminde yapılan seçimde Demokratik partinin %7,69 farkla seçimi kaybettiğini, dolayısıyla iktidar partisi olma avantajından yoksun olarak girdiği $t+1$ dönemindeki seçimden %36,10’luk bir oy oranına sahip olduğunu göstermektedir [2].

Uygulamada eşik değerinin altında kalan tüm değerlerde Demokratik parti iktidar partisi olma avantajına sahip değilken, üstünde kalan tüm değerlerde ise

Demokratik parti iktidar olma avantajına sahip olmaktadır. Bu durumda uygulanacak regresyon süreksizliği yöntemi bir keskin regresyon süreksizliği yöntemi tasarımıdır.

Çalışmada R programında yer alan “Rdrobust” kütüphanesinden faydalanılmıştır. Kütüphane, regresyon süreksizliği yönteminin uygulanmasına yönelik fonksiyonları içermektedir. Fonksiyonlar grafiksel gösterimde kullanılan “rdplot()” fonksiyonu, parametrik olmayan tahmin yönteminde kullanılan “rdbwselect_2014()” ve etkinin tahmin edildiği “rdrobust()” fonksiyonlarını içermektedir. Fonksiyonlara ilişkin R kodları EK 1’de verilmiştir [2].

Veri setinde fark değişkeni ve oy oranı değişkenine ait 1.297 gözlem bulunmaktadır. Gözlemlere ait tanımlayıcı istatistikler Çizelge 3’de verilmektedir.

Çizelge 3 : Tanımlayıcı İstatistikler

Değişkenler	Oy Oranı	Fark
Ortalama	52,67	7,11
Medyan	50,55	2,16
Minimum	0	-100
Maksimum	100	100

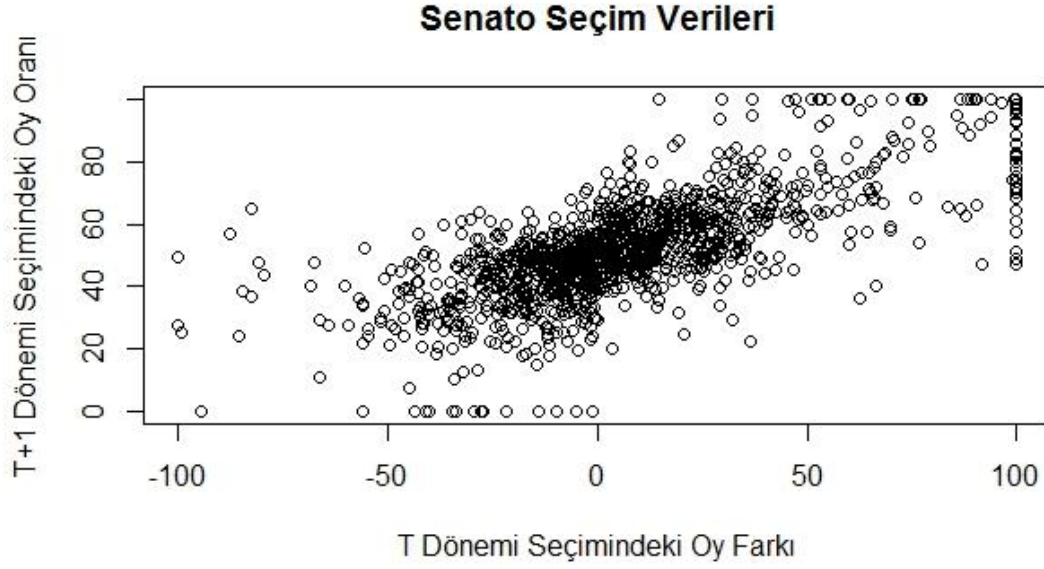
Fark değişkeni ± 100 arasında değer almaktadır. Demokratik partinin t seçim döneminde, seçim bölgelerinde oyların tamamını aldığını ya da oyların tamamını kaybettiğini göstermektedir. Fark değişkeninin pozitif olduğu bölgelerde Demokratik parti seçimi kazanırken, negatif olduğu bölgelerde rakibi olan Cumhuriyetçi parti seçimi kazanmaktadır.

Oy oranı değişkeni 0 ile 100 arasında değişmektedir. Demokrat partinin t+1 seçim döneminde aldığı oyu göstermektedir.

4.2 Grafiksel Gösterim

Regresyon süreksizliği yöntemi uygulamaya başlamadan önce eşik değerinde kırılım olup olmadığı görülmek istenmektedir. Grafiksel gösterim kırılımın olup olmadığı konusunda fikir verdiği için yöntem uygulanmadan önce ilk adımda verinin saçılım grafiğine bakılmaktadır.

Sekil 18’de Senato seçimlerine ait ham verilerin grafiksel olarak gösterimi görülmektedir. Grafik incelendiğinde, eşik değeri üzerinde bir kırılmanın olmadığı görülmektedir.



Şekil 18: Ham Verilerin Grafiksel Gösterimi

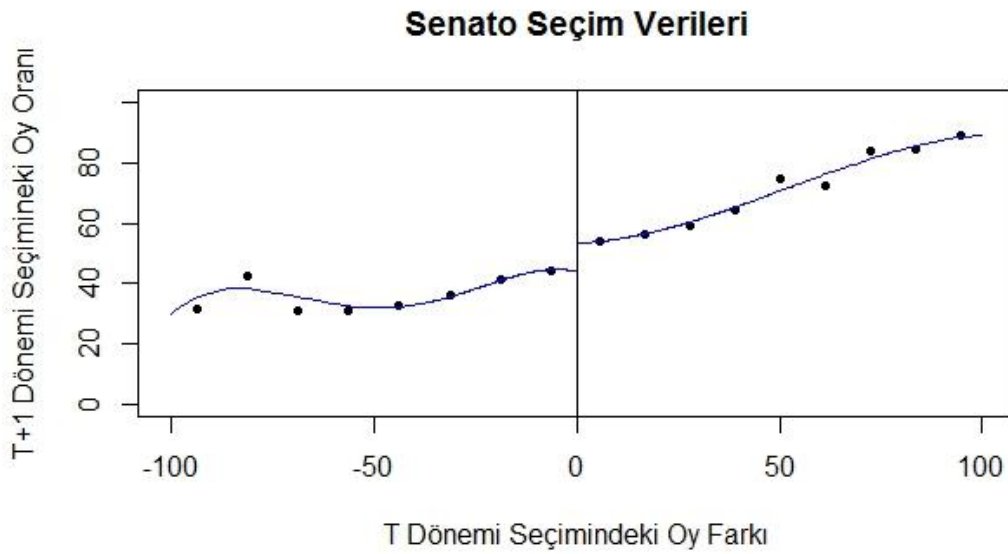
Grafiksel gösterim uygulanırken ham veri karmaşık olarak görüldüğü için, veriler endeks değeri üzerinde gruplandırılması gerekmektedir. Grupların veride karmaşa yaratmayacak kadar dar, eşik değeri üzerinde karşılaştırılabilir olacak kadar da geniş olması gerekir. Optimal grup sayısının belirlenmesi için birleştirilmiş hata kareler ortalaması kullanılmıştır.

BHKO yöntemine göre optimal grup genişliği sayısı Çizelge 4’de verilmektedir. Eşik değeri altında 595 gözlem bulunurken eşik değeri üstünde 702 gözlem bulunmaktadır. Eşik değerinin altında yer alan gözlemler, 12,49’luk grup genişliği olacak şekilde 8 grup altında gösterilirken, eşik değerinin üstünde yer alan gözlemler, 11,107’luk grup genişliği olacak şekilde 9 grup altında gösterilmektedir.

Çizelge 4 : Optimal Grup Genişliği

	Eşik Değeri Altı	Eşik Değeri Üstü
Gözlem Sayısı	595	702
Polinom Derecesi	4	4
Optimal Grup Sayısı	8	9
Optimal Grup Genişliği	12,490	11,107

Optimal grup genişliğine göre hazırlanan grafiksel gösterim Şekil 19'da verilmiştir.



Şekil 19: Optimal Grup Genişliğinin Grafiksel Gösterimi

Şekil 19 incelendiğinde eşik değerinde kırılma görülmektedir. Dolayısıyla bu durum iktidar olma avantajının bir sonraki seçimde anlamlı bir etkisinin olabileceği yönünde yorumlanabilir.

4.3 Regresyon Süreksizliği Yönteminde Tahmin Sonuçları

Uygulamada regresyon süreksizliğinde parametrik olmayan tahmin yöntemi uygulanmıştır. Parametrik olmayan tahminin 3 önemli bileşeni olan, bant genişliği(h), çekirdek fonksiyon(K) türünün ve yerel polinom modeline ait fonksiyonun derecesinin(p) belirlenmesi gerekmektedir.

Çekirdek fonksiyonunun türü tahmin sonuçlarına etkisinin olmadığı düşünülmektedir [15]. Bu nedenle çalışmada üçgen fonksiyon çekirdek fonksiyon olarak seçilmiştir.

Bant genişliğinin belirlenmesi için çapraz geçerlilik yöntemi ve plug-in yöntemleri kullanılmıştır. Yöntemlerde elde edilen farklı bant genişlikleri kullanılarak 1. dereceden ve 2. dereceden yerel polinom modelleri tahmin edilmiş ve programın etkisi hesaplanmıştır. Hesaplanan değerler Çizelge 5'te verilmiştir.

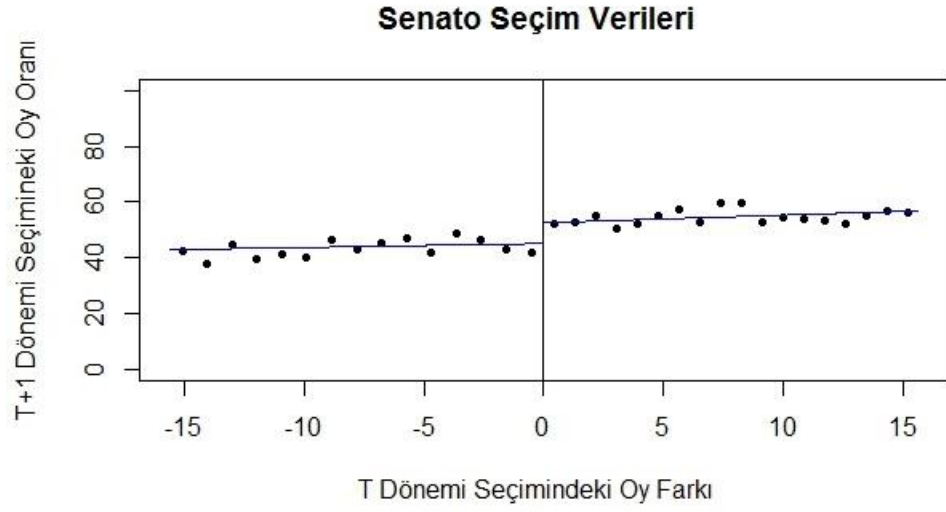
Çizelge 5 : Parametrik Olmayan Regresyon Süreksizliği Tahmin Sonuçları

	Plug-In Yöntemi İle Bant Genişliği		Çapraz Geçerlilik Yöntemi İle Bant Genişliği	
	1. Dereceden Yerel Polinom	2. Dereceden Yerel Polinom	1. Dereceden Yerel Polinom	2. Dereceden Yerel Polinom
Yerel Polinom Modelinin Derecesi(p)	1	2	1	2
Bant Genişliği (h)	15,6676	20,4367	35,4211	45,3442
Çekirdek Fonksiyonu(K)	Üçgen Fonksiyon	Üçgen Fonksiyon	Üçgen Fonksiyon	Üçgen Fonksiyon
Tahmin Edilen Etki ($P > z $ değeri)	7,43 (0,00)	8,13 (0,00)	7,16 (0,00)	7,89 (0,00)

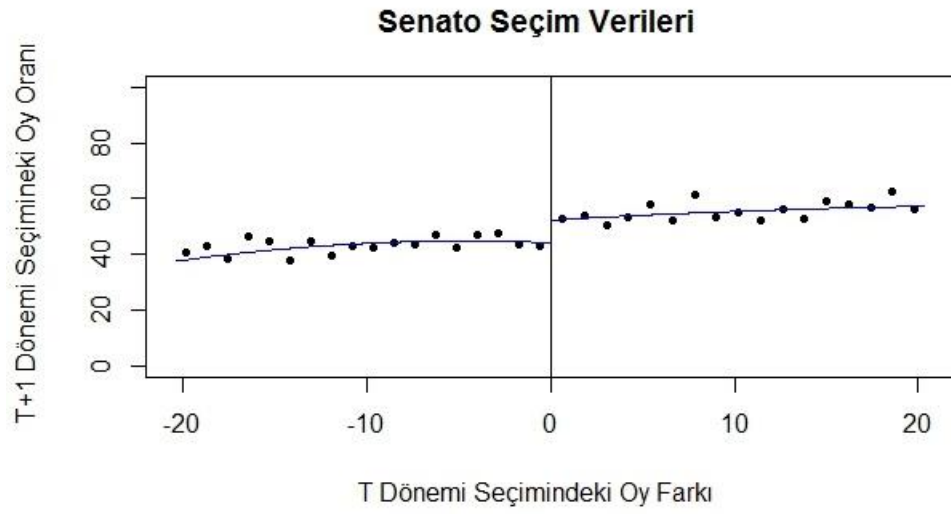
Çizelge 5'de incelendiğinde, plug-in yöntemi ile bulunan bant genişliği, çapraz geçerlilik yöntemi ile bulunan bant genişliğinden daha küçüktür. Her iki yöntemde de bant genişlikleri 1.dereceden yerel polinom modelde, 2. dereceden yerel polinom modele göre daha küçük olduğu görülebilir.

Elde edilen sonuçlara göre t döneminde yapılan seçimi kazanarak iktidar olma avantajını kazanan Demokratik partinin t+1 döneminde yapılan seçimde elde ettiği oy oranına pozitif olarak yansıdığı görülmektedir.

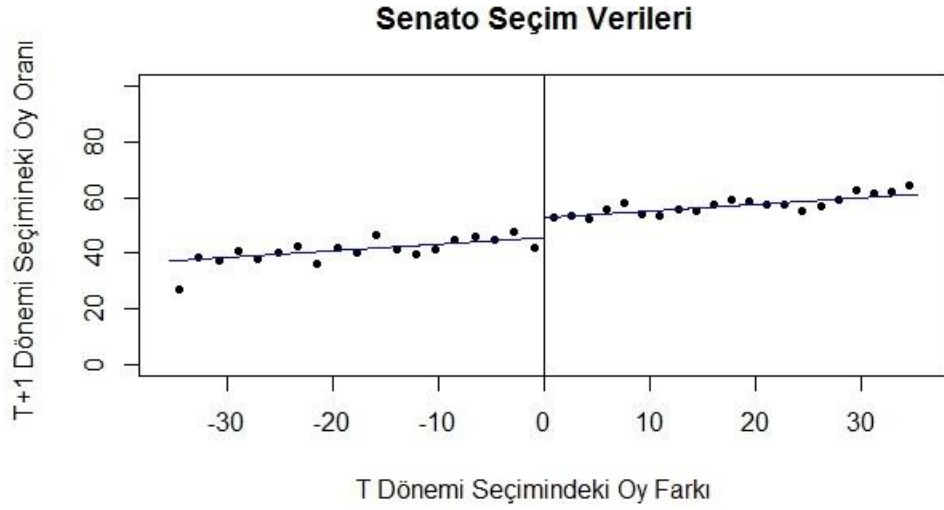
Plug-in yöntemi ve çapraz geçerlilik yöntemleri kullanılarak 1. ve 2. dereceden yerel polinom modellerinin grafiksel gösterimleri Şekil 20-23' verilmiştir.



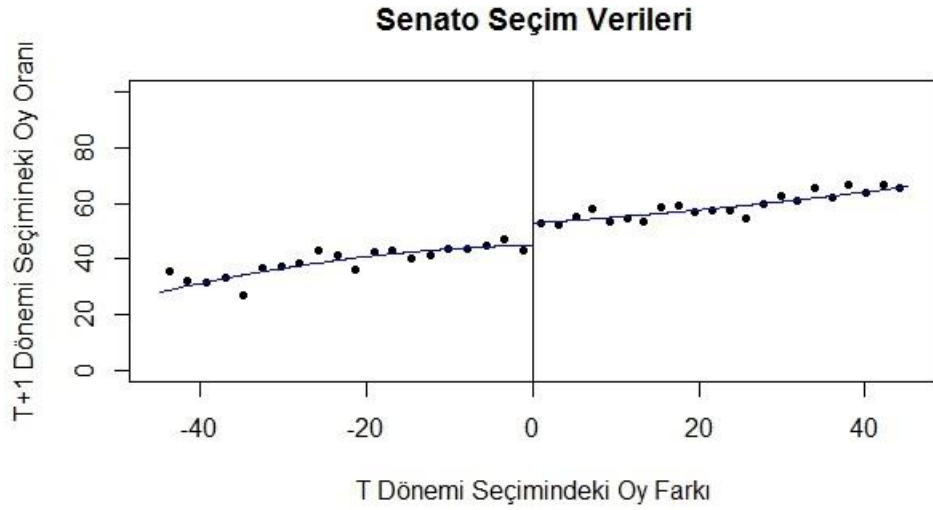
Şekil 20: Plug-in Yöntemi İle 1. Dereceden Yerel Polinom Modeline Ait Grafikselleştirme



Şekil 21: Plug-in Yöntemi İle 2. Dereceden Yerel Polinom Modeline Ait Grafikselleştirme



Şekil 22: Çapraz Geçerlilik Yöntemi İle 1. Dereceden Yerel Polinom Modeline Ait Grafikselsel Gösterim



Şekil 23: Çapraz Geçerlilik Yöntemi İle 2. Dereceden Yerel Polinom Modeline Ait Grafikselsel Gösterim

Plug-in yöntemi ile 1.dereceden yerel polinom kullanılarak elde edilen bant genişliği 15,66 bulunmaktadır. Değer Demokratik partinin t seçiminde aldığı oy oranlarının farkının $[-15,66;15,66]$ lık bir pencerede yer alan gözlemleri içerecek şekilde karşılaştırılacağı anlamına gelmektedir. Bant genişliği içerisinde, eşik değerinin altı ve üstü için ayrı ayrı birinci dereceden yerel polinom modeli tahmin edilir. Tahmin edilen bu model kullanılarak, eşik değerinde alacağı değerler tahmin

edilmektedir. 15,66 deęerinde bir bant geniřlięi iin tahmin edilen program etkisi %5 anlamlılık dzeyinde %7,43'tr. Bu oran (%0 , %15,66] arasında farkla oy alarak kazandıęı seimlere gre [%-15,66 , %0) arasında farkla oy alarak kaybettięi seimler karřılařtırıldıęında, iktidar olma avantajı ile bir sonraki seimde %7,43 daha fazla oy aldıęı anlamına gelmektedir.



5. SONUÇ VE TARTIŞMA

Regresyon süreksizliği yöntemi yarı deneysel bir etki değerlendirme yöntemi olarak kullanılmaktadır. Yöntemde, etki grubu ve kontrol grubuna ait gözlemlerin eşik değerine göre konumu belirlenmektedir.

Bölüm 4’de verilen sayısal örnekte, regresyon süreksizliği yöntemi kullanılarak bir seçimde kazanan partinin iktidar partisi olarak bir sonraki seçimde oy avantajına sahip olup olmadığı incelenmiştir. Literatürde iktidar partisi olma avantajının etkisi birçok çalışmada incelenmiş olup, regresyon süreksizliği yönteminde bu çalışmada kullanılabileceği değerlendirilmiştir.

Regresyon süreksizliği yöntemi uygulanırken ilk aşamada uygulanan grafiksel gösterim, eşik değeri üzerinde bir kırılmanın tespit edilmesi varsa yöntemin tahmin aşamasına geçilmesinde kullanılmaktadır. Birleştirilmiş Hata Kareler Ortalaması kullanılarak elde edilen grafik, ham verinin sahip olduğu grafiğe göre daha anlaşılabilir. Grafiksel gösterimde, uygun grup sayısına göre belirlenen eşik değerinin altında yer alan veriler 12,49 değerinde bir grup genişliği içerisinde yer alan 8 gruptan oluşurken, eşik değerinin üstünde yer alan veriler 11,10 değerinde bir grup genişliği içerisinde yer alan 9 gruptan oluşmaktadır. Uygun grup sayısına göre hazırlanan grafik incelendiğinde, eşik değeri üzerinde görülen kırılım, Demokratik partinin iktidar olma avantajı ile girdiği sonraki seçimden bu avantaj dolayısıyla daha fazla oy aldığını görsel olarak göstermektedir.

Sayısal örnekte, bant genişliğinin belirlenmesinde plug-in yöntemi ve çapraz geçerlilik yöntemleri, 1. ve 2. dereceden yerel polinom modelleri ile birlikte kullanılarak iktidar olmanın sonraki seçimdeki etkisi incelenmiştir.

Plug-in yöntemi kullanılarak elde edilen bant genişliği içerisinde 1. dereceden yerel polinom modeli kullanıldığında, 15,66 değerinde bir bant genişliği için tahmin edilen program etkisi %5 anlamlılık düzeyinde %7,43’tür. Demokratik partinin (%0, %15,66] arasında farkla oy alarak kazandığı seçimlere göre [%-15,66 - %0) arasında farkla oy alarak kaybettiği seçimler karşılaştırıldığında, iktidar olma avantajı ile bir sonraki seçimde %7,43 daha fazla oy aldığı anlamına gelmektedir.

Plug-in yöntemi kullanılarak elde edilen bant genişliği içerisinde 2. dereceden yerel polinom modeli kullanıldığında, 20,43 değerinde bir bant genişliği için tahmin

edilen program etkisi %5 anlamlılık düzeyinde %8,13'tür. Demokratik partinin (%0, %20,43] arasında farkla oy alarak kazandığı seçimlere göre [%-20,43 - %0) arasında farkla oy alarak kaybettiği seçimler karşılaştırıldığında, iktidar olma avantajı ile bir sonraki seçimde %8,13 daha fazla oy aldığı anlamına gelmektedir.

Çapraz geçerlilik yöntemi kullanılarak elde edilen bant genişliği içerisinde 1. dereceden yerel polinom modeli kullanıldığında, 35,42 değerinde bir bant genişliği için tahmin edilen program etkisi %5 anlamlılık düzeyinde %7,16'dır. Demokratik partinin (%0 , %35,42] arasında farkla oy alarak kazandığı seçimlere göre [%-35,42 - %0) arasında farkla oy alarak kaybettiği seçimler karşılaştırıldığında, iktidar olma avantajı ile bir sonraki seçimde %7,16 daha fazla oy aldığı anlamına gelmektedir.

Çapraz geçerlilik yöntemi kullanılarak elde edilen bant genişliği içerisinde 2. dereceden yerel polinom modelinin kullanıldığında, 45,34 değerinde bir bant genişliği için tahmin edilen program etkisi %5 anlamlılık düzeyinde %7,89'dur. Demokratik partinin (%0 , %45,34] arasında farkla oy alarak kazandığı seçimlere göre [%-45,34 - %0) arasında farkla oy alarak kaybettiği seçimler karşılaştırıldığında, iktidar olma avantajı ile bir sonraki seçimde %7,89 daha fazla oy aldığı anlamına gelmektedir.

Analiz sonuçları karşılaştırıldığında plug-in yöntemi ile bulunan bant genişliklerinin 1. ve 2. dereceden yerel polinom modellerinden bağımsız olarak çapraz geçerlilik yöntemi ile bulunan bant genişliklerinden daha küçük olduğu görülmektedir. Bant genişliğinin belirlenmesi için kullanılan aynı yöntem içerisinde, 1. dereceden yerel polinom modeli kullanıldığında elde edilen bant genişlikleri 2. dereceden yerel polinom modeli kullanıldığında elde edilen bant genişliklerinden küçüktür.

Bant genişliğinin belirlenmesi için uygulanan her iki yöntemde, 1. ve 2. dereceden yerel polinom modelleri kullanıldığında elde edilen programın etkisi birbirlerine yakın olduğu görülmektedir. Programın etkisi en yüksek %8,13 olarak plug-in yöntemini kullanılarak elde edilen bant genişliği içerisinde 2.dereceden yerel polinom modeli kullanıldığında bulunmuştur. Programın etkisi en düşük %7,16 olarak çapraz geçerlilik yöntemini kullanılarak elde edilen bant genişliği içerisinde 1.dereceden yerel polinom modelinin kullanıldığında bulunmuştur. Demokratik partinin iktidar olma avantajı ile girdiği seçimlerden, iktidar olma avantajına sahip

olmadan girdiđi seimlere gre %7,16 - %8,13 aralıđında bir oy artışına sahip olduđunu gstermektedir.

Uygulamada 1. dereceden yerel polinom modeline ait tahmin edilen etkinin, 2. dereceden yerel polinom modeline ait tahmin edilen etkiden kk olduđu grlmřtr. Benzer řekilde Arslan [27] alıřmasında model derecesi arttıka tahmin edilen etki deđerinin de arttıđı grlmektedir.



KAYNAKLAR

- [1] Thistlethwaite, D., Campbell D., Regression Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment, *Journal of Educational Psychology*, 51, 309-317, **1960**
- [2] Calonico S., Cattaneo M. D., Titiunik R., rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs, *R Journal*, 7(1), 38-51, **2015**.
- [3] Türk Dil Kurumu, Resmi İnternet Sitesi, http://www.tdk.gov.tr/index.php?option=com_bts&arama=kelime&guid=TDK.GTS.5aeb5dec7c4113.58997469, (Mayıs, **2018**)
- [4] Meydan, M. C., *Kalkınma Ajansları Desteklerinin Değerlendirilmesi : Karşıt Durum Etki Değerlendirme Örneği*, Uzmanlık Tezi, T.C. Kalkınma Bakanlığı, Ankara, **2014**.
- [5] Adıgüzel R. S., Kalaycı K. Ö., Yılmaz P., Etki Değerlendirmeye Genel Bir Bakış, *Anahtar Dergisi*, 27, 319, 4-7, **2015**.
- [6] Yıldız S. B., *Küçük ve Orta Ölçekli İşletmelerin Desteklenmesinde KOSGEB'in Rolü: Bir Etki Değerlendirmesi Araştırması*, Doktora Tezi, Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü, Isparta, **2013**.
- [7] Gertler P. J., Martinez S., Premand, P., Rawlings, L. B., Vermeersch C. M. J., *Impact Evaluation in Practice*, The World Bank, **2010**.
- [8] Belgin, Ö., Karapınar, B., Arge Desteklerinin Etki Analizi, *Anahtar Dergisi*, 27, 319, 8-10, **2015**.
- [9] Khandker S. R., Koolwal G. B., Samad, H. A., *Handbook on Impact Evaluation Quantitative Methods and Practices*, The World Bank, **2010**.
- [10] Baker, J. L., Evaluating the impact of development projects on poverty : a handbook for practitioners, *Directions in development*, World Bank, Washington, **2000**.
- [11] White, H., Sabarwal, S., *Quasi-Experimental Design and Methods*, Unicef, Methodological Briefs Impact Evaluation No 8, Italy, **2014**.
- [12] European Commission, Evalsed Sourcebook: Method and Techniques, http://ec.europa.eu/regional_policy/en/information/publications/evaluations-guidance-documents/2013/evalsed-the-resource-for-the-evaluation-of-socio-economic-development-sourcebook-method-and-techniques (Mayıs, **2018**)
- [13] Hahn, J., Todd P., Klaauw, W.V.D., Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design, *Econometrica*, 69, 201-209, **2001**.
- [14] Jacob, R., Zhu, P., Somers, M. A. and Bloom, H., A Practical Guide to Regression Discontinuity, MDRC, **2012**.
- [15] Skovron, C., Titiunik, R., A Practical Guide to Regression Discontinuity Designs in Political Science, *American Journal of Political Science*, **2015**.
- [16] Trochim, W., M., K., *Research Design for Program Evaluation: The Regression-Discontinuity Approach*, Sage Publications, Beverly Hills, **1984**.

- [17] Imbens, G., Lemieux, T., Regression Discontinuity Designs: A Guide to Practice, *Journal of Econometrics*, 142 (2): 615–635, **2008**.
- [18] Lee, D. S., Randomized experiments from non-random selection in U.S. House elections, *Journal of Econometrics*, 142, 675-697, **2008**.
- [19] Black, S. E., Do Better Schools Matter? Parental Valuation of Elementary Education, *The Quarterly Journal of Economics*, 114, 2, 577-599, **1999**.
- [20] Carpenter C., Dobkin, C., The Effect of Alcohol Consumption On Mortality: Regression Discontinuity Evidence from The Minimum Drinking Age, *The National Bureau of Economic Research*, **2007**.
- [21] Bronzini, R., Iachini, E., Are Incentives for R&D Effects? Evidence From a Regression Discontinuity Approach, *American Economic Journal: Economic Policy*, 6(4):100-134, **2014**.
- [22] Caliendo, M., Tatsiramos, K., Uhlendorff, A., Benefit Duration, Unemployment Duration and Job Match Quality: A Regression-Discontinuity Approach, *Applied Microeconometrics*, 28,4, 604-627, **2013**.
- [23] Anderson, M., Magruder, J., Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database, *The Economic Journal*, 122, 563, 957-989, **2011**.
- [24] Zhao, B., Rosemarie, Z., State Governments as Financiers of Technology Startups: Implications for Firm Performance, <https://ssrn.com/abstract=2060739> (Mayıs, **2018**)
- [25] Ozier, O., The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis, World Bank Group, <http://documents.worldbank.org/curated/en/700151467997577920/The-impact-of-secondary-schooling-in-Kenya-a-regression-discontinuity-analysis> (Mayıs, **2018**)
- [26] Aysoy, C., Halim Kırılı, D., Tümen, S., *Taze Meyve-Sebze Tedarik Zincirindeki Engelleri Azaltıcı Tedbirlerin Fiyatlar Üzerindeki Etkisi*, T.C. Merkez Bankası, Ekonomi Notları, **2015**.
- [27] Arslan N., *Regression Discontinuity Design: An Application in Economics*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2017**.
- [28] Calonico, S., Cattaneo M. D., TITIUNIK R., Optimal Data-Driven Regression Discontinuity Plots, *Journal of the American Statistical Association*, 110, 512, **2015**.
- [29] Lee, D. S., Lemieux, T., Regression Discontinuity Designs in Economics, *Journal of Economic Literature*, 48, 281-355, **2010**.
- [30] Türkan S., *Yarı Parametrik Regresyon Modelinde Etkili Gözlem Analizi*, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2012**.
- [31] Ludwig J., Miller D. L, *Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design*, Working Paper 11702, National Bureau of Economic Research **2005**.

- [32] Imbens, G. W., Kalyanaraman. K., Optimal Bandwidth Choice for the Regression Discontinuity Estimator, *Review of Economic Studies*, 79 (3): 933–959, **2012**.
- [33] Erikson R. S., The Advantage of Incumbency in Congressional Elections, *Polity*, 3, 395-405, **1971**.
- [34] Gelman A., King G., Estimating Incumbency Advantage Without Bias, *American Journal of Political Science*, 34, 1142-1164, **1990**
- [35] Ansolabehere S., Snyder J. M., The Incumbency Advantage in U.S. Elections: An Analysis of State and Federal Offices 1942-2000, *Election Law Journal*, 2002;1 (3), 315-338, **2002**.
- [36] United States House of Representatives Resmi İnternet Sitesi, <https://www.house.gov/the-house-explained> (Mayıs, **2018**)
- [37] Cattaneo M. D., Frandsen B. R., Titiunik R., Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the US Senate, *Journal of Causal Inference*, 3(1), 1-24, **2015**,

EKLER

EK 1: R kodları

Verinin Yüklenmesi ve Verinin Özetlenmesi

```
library(readxl)
veri <- read_excel("R/veri.xlsx", col_types = c("numeric", "numeric"))

> summary(veri$Oy_Orani)

> summary(veri$Fark)
```

RDPlot Ham Veri

```
> plot(veri, main="Senato Seçim Verileri",
xlab="T Dönemi Seçimindeki Oy Farkı", ylab="T+1 Dönemi Seçimindeki Oy Oranı")
```

Grafiksel Gösterim

```
> (rdplot( y = veri$Oy_Orani, x = veri$Fark, binselect="es", title= "Senato Seçim
Verileri", y.label = "T+1 Dönemi Seçimineki Oy Oranı", x.label = "T Dönemi
Seçimindeki Oy Farkı"))
```

Bant Genişliğinin Hesaplanması

Plug-In Yöntemi için:

```
> rdbwselect_2014(y=veri$Oy_Orani, x=veri$Fark, bwselect="IK")
```

Çapraz Geçerlilik Yöntemi için:

```
> rdbwselect_2014(y=veri$Oy_Orani, x=veri$Fark, bwselect="CV")
```

Etkinin Hesaplanması

1.Derecen Yerel Polinom Modeli için:

```
>rdrobust(y=veri$Oy_Orani, x=veri$Fark, p=1, h=*)
```

2..Derecen Yerel Polinom Modeli için:

```
>rdrobust(y=veri$Oy_Orani, x=veri$Fark, p=2, q=3, h=*)
(* ; bulunan bant genişliği girilir)
```

Etkinin Bant Genişliklerinde Grafiksel Gösterimi


```
> rdplot(veri$Oy_Orani[abs(veri$Fark) <= *], veri$Fark[abs(veri$Fark) <= *], p=1,  
kernel="triangular")  
(* ; bulunan bant genişliği girilir)
```



ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Serdar DİNDAS
Doğum Yeri : İzmir
Medeni hali : Evli
E-posta : serdardindas@yahoo.com
Adresi : Tunahan mah. 254 cad. Sarıgül sitesi Etimesgut / ANKARA

Eğitim

Lise : Suphi Koyuncuoğlu Lisesi
Lisans : Hacettepe Üniversitesi, İstatistik Bölümü
Yükseklisans : Hacettepe Üniversitesi, İstatistik Bölümü

Yabancı Dil ve Düzeyi

İngilizce (IELTS 2018: 6.5/9)
Almanca (A2)

İş Deneyimi

2012 - : Türkiye Bilimsel ve Teknolojik Araştırma Kurumu
Teknoloji ve Yenilik Destek Programları Başkanlığı
Bilimsel Programlar Uzman Yardımcısı

Deneyim Alanları

Tezden Üretilmiş Projeler ve Bütçesi

Tezden Üretilmiş Yayınlar

Tezden Üretilmiş Tebliğ ve/veya Poster Sunumu ile Katıldığı Toplantılar



HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
İSTATİSTİK ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 11/06/2018

Tez Başlığı / Konusu: REGRESYON SÜREKSİZLİĞİ YÖNTEMİNDE PARAMETRİK OLMAYAN TAHMİN

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 68 sayfalık kısmına ilişkin, 11/06/2018 tarihinde ~~şahsen~~/tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 3. 'tür.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar ~~hariç~~/dâhil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

11/06/2018
Tarih ve İmza

Adı Soyadı: SERDAR DINDAŞ

Öğrenci No: N14320867

Anabilim Dalı: İSTATİSTİK

Programı: TEZLİ YÜKSEK LİSANS

Statüsü: ☒ Y.Lisans ☐ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

PROF.DR.SEVİL BACANLI