

# İSTATİSTİK ARAŞTIRMA DERGİSİ

JOURNAL OF STATISTICAL  
RESEARCH

JSRTR

Cilt  
Volume **14**

Sayı  
Issue **2**

Yıl  
Year **2024**

**İSTATİSTİK ARAŞTIRMA DERGİSİ  
ULUSLARARASI HAKEMLİ DERGİ**

JOURNAL OF STATISTICAL RESEARCH  
INTERNATIONAL PEER-REVIEWED JOURNAL

**ISSN: 2791-7614  
CİLT/VOL. 14 • SAYI / ISSUE 2 • 2024**

**Türkiye İstatistik Kurumu Adına Sahibi / Owner on Behalf of Turkish Statistical Institute  
Erhan ÇETİNKAYA**

**Baş Editör / Editor - in - Chief  
TÜİK Uzmanı Orçun AYDIN**

**Editör Yardımcısı / Assistant Editor  
Prof. Dr. Nurcan METİN**

**Sorumlu Müdür / Responsible Manager  
Stratejik Planlama, İzleme ve Değerlendirme Grup Başkanı Gönül KALENDER**

**Yabancı Dil Editörü / Foreign Language Editor  
TÜİK Uzmanı Nilgün DORSAN**

**İletişim / Communication**  
Türkiye İstatistik Kurumu  
Devlet Mah.  
Necatibey Cad. No: 114  
06420 Çankaya/Ankara/TÜRKİYE

**Tel. / Phone:** +90 312 454 83 72  
**Belgeç / Fax:** +90 312 454 82 96  
**E-posta / E-mail:** journal@tuik.gov.tr  
**İnternet Sayfası / Web Site:**  
<https://dergipark.org.tr/tr/pub/jsstr>

Türkiye İstatistik Kurumu yayını olan İstatistik Araştırma Dergisi, yılda iki kez yayımlanan uluslararası hakemli bir dergidir. Makalelerin bilimsel ve etik kurallara uygunluğu yazarların sorumluluğundadır; Türkiye İstatistik Kurumu sorumlu tutulamaz.

Journal of Statistical Research, the publication of the Turkish Statistical Institute, is an international peer-reviewed journal published twice a year. Compliance of the articles with scientific and ethical rules is the responsibility of the authors; Turkish Statistical Institute cannot be held responsible.

**©Türkiye İstatistik Kurumu Yayını  
©Turkish Statistical Institute Publication**

## **Baş Editör'den**

Değerli okuyucular,

İstatistik Araştırma Dergisi, Türkiye İstatistik Kurumu'nun 2002 yılında yayınlanmaya başladığı uluslararası akademik bir dergisidir.

Yayın politikası yenilenen ve kısa bir aradan sonra Temmuz 2021'de yeniden yayınlanmaya başlayan İstatistik Araştırma Dergisi'nde istatistik yanında ekonometri, yöneylem araştırması ve istatistiğin araç olarak kullanıldığı diğer bilim dallarında teorik çalışmalara yer verdiğimiz gibi, uygulamalı orijinal çalışmalara yer verilmektedir.

İstatistik Araştırma Dergisi'nin hedefi söz ettiğimiz konularda yayınlayacağı makaleler ile literatüre katkı sağlamaktır. Bu doğrultuda, derginin farklı endekslerde taranmasına yönelik çalışmalara da yer verilmektedir.

Bu sayıda makaleleri ile destek veren yazarlarımıza, dergimizin bilimsel çizgisine yön veren hakemlerimize, dergimizin Editörler Kuruluna üye olmayı kabul ederek bizi onurlandıran değerli bilim insanlarına, desteklerinden ötürü TÜİK Başkanı Sayın Erhan ÇETİNKAYA'ya ve emeği geçen herkese teşekkürlerimizi sunarız.

Dergimizin Aralık 2024 (Cilt 14, Sayı 2) sayısının bilim dünyasına katkı sağlamasını diler, bundan sonraki sayılarımıza da ilgi ve desteğinizin devamını diler, saygılarımı sunarım.

**Orçun AYDIN**  
**Baş Editör**

## Editörler / Editors

- Prof. Dr. Ebru ÇAĞLAYAN AKAY, Marmara Üniversitesi, İstanbul  
Fikri AKDENİZ, Çağ Üniversitesi, Mersin  
Cem CANEL, University of North Carolina Wilmington, Amerika Birleşik Devletleri  
Mehmet Ali CENGİZ, Ondokuz Mayıs Üniversitesi, Samsun  
Meral ÇETİN, Hacettepe Üniversitesi, Ankara  
Özlem İLK DAĞ, Orta Doğu Teknik Üniversitesi, Ankara  
Burak GÜRİŞ, İstanbul Üniversitesi, İstanbul  
Jamal HUSEİN, Angelo State University, Amerika Birleşik Devletleri  
Cem KADILAR, Hacettepe Üniversitesi, Ankara  
Safet KOZAREVIC, University of Tuzla, Bosnia and Herzegovina, Bosna-Hersek  
Sakkthivel Annamalai MANICKAM, Skyline University College, Birleşik Arap Emirlikleri  
Ünal Halit ÖZDEN, İstanbul Ticaret Üniversitesi, İstanbul  
Ömer ÖZTÜRK, The Ohio State University, Amerika Birleşik Devletleri  
Muthucattu Thomas PAUL, Papua New Guinea University of Technology, Papua Yeni Gine  
Mustafa SEVÜKTEKİN, Bursa Uludağ Üniversitesi, Bursa  
Ram SHANMUGAM, Texas State University, Amerika Birleşik Devletleri  
Kutluk Kağan SÜMER, İstanbul Üniversitesi, İstanbul  
Semra ERPOLAT TAŞABAT, Mimar Sinan Güzel Sanatlar Üniversitesi, İstanbul  
Mustafa TEKİN, İstanbul Üniversitesi, İstanbul  
Münevver TURANLI, İstanbul Ticaret Üniversitesi, İstanbul
- Prof. Emeritus Barry C. ARNOLD, University of California, Amerika Birleşik Devletleri  
Hanna DUDEK, Warsaw University of Life Sciences, Polonya  
Nikolai KOLEV, University of Sao Paulo, Brezilya  
Debasis KUNDU, Indian Institute of Technology, Hindistan  
W. Y. Wendy LOU, University of Toronto, Kanada
- Doç. Dr. İbrahim DEMİR, Türkiye İstatistik Kurumu Başkanlığı, Ankara  
Dr. Ojonugwa USMAN, Federal College of Education (Technical), Nijerya  
Subhash Kumar YADAV, Babasaheb Bhimrao Ambedkar University, Hindistan

**2024 Yılında ‘İstatistik Araştırma Dergisi’ne Bilimsel Katkı Sağlayan Hakemler**  
**Referees Who Provided Scientific Contributions to The Journal of Statistical Research in 2024**

Prof. Dr.

Burak GÜRİŞ, İstanbul Üniversitesi, İstanbul  
Cemil ÇOLAK, İnönü Üniversitesi, Malatya  
Erol TERZİ, Ondokuz Mayıs Üniversitesi, Samsun  
Ferda Yerdelen TATOĞLU, İstanbul Üniversitesi, İstanbul  
Halil TUNALI, İstanbul Üniversitesi, İstanbul  
Mehmet Kenan TERZİOĞLU, İstanbul Üniversitesi, İstanbul  
Mesut DOĞAN, Bilecik Şeyh Edebali Üniversitesi, Bilecik  
Mustafa SEVÜKTEKİN, Bursa Uludağ Üniversitesi, Bursa  
Münevver TURANLI, İstanbul Ticaret Üniversitesi, İstanbul  
Nazan ŞAK, Marmara Üniversitesi, İstanbul  
Serpil AKTAŞ, Hacettepe Üniversitesi, Ankara  
Sinan ÇALIK, Fırat Üniversitesi, Elazığ

Doç. Dr.

Muhammed Hanifi VAN, Van Yüzüncü Yıl Üniversitesi, Van  
Murat AYDOS, Hacettepe Üniversitesi, Ankara  
Serdar DEMİR, Muğla Sıtkı Koçman Üniversitesi, Muğla  
V. Sinem Arıkan KARGI, Bursa Uludağ Üniversitesi, Bursa

Dr.

Havva GÜLTEKİN, Trakya Üniversitesi, Trakya  
Leyla YÜCEL, İstanbul Üniversitesi, İstanbul  
Şaban KIZILARSLAN, Van Yüzüncü Yıl Üniversitesi, Van  
Uğur BELLİKLİ, Giresun Üniversitesi, Giresun

**Dergimize gönderilen makalelerin tümü intihal tespit aracıyla incelenmektedir.**

*All articles submitted to our journal are analyzed by plagiarism detection tools.*

## İÇİNDEKİLER / CONTENTS

### ARAŞTIRMA MAKALELERİ / ORIGINAL RESEARCH ARTICLES

- 1-13 Beyda DEMİRCİ, Ferda YERDELEN TATOĞLU  
*Türkiye’de Bölgesel İşsizliğin Belirleyicilerinin Mekânsal Panel Veri Modelleriyle Analizi / Spatial Panel Data Analysis of Determinants of Regional Unemployment in Turkey*
- 14-29 Sibel ÖRK ÖZEL, Dilek VEYSİKARANI  
*Sürdürülebilir Kalkınma Kapsamında Dijitalleşme ve Çevre Ekonomisi İlişkisinin Analizi / Analysis of the Relationship between Digitalization and Environmental Economics within the Scope of Sustainable Development*
- 30-45 Turgut ÜN, Özlem TAŞSEVEN, Dilek ALTAŞ KARACA  
*Avrupa Birliği Ülkeleri ve Türkiye’de Doğrudan Yabancı Sermaye Yatırımları ile Gayri Safi Yurtiçi Hasıla Arasındaki Panel Veri Nedensellik Analizi / Panel Data Causality Analysis in the European Union And Turkey Between Foreign Direct Investments And Gross Domestic Product*
- 46-58 Yeliz YALÇIN, Burak BAYSAL  
*Web Kazıma Yöntemlerinin Tüketici Fiyat Endeksi Çalışmalarında Kullanılması Türkiye Örneği / Using Web Scraping Methods in Consumer Price Index Studies, Türkiye Example*
- 59-72 Semih ERGİŞİ, Beyza DOĞANAY, Yasemin YAVUZ  
*Performance Comparison of Least Squares, Ridge, Lasso and Principal Component Regression for Addressing Multicollinearity in Regression Analysis / Regresyon Analizinde Çoklu Doğrusallığın Ele Alınması için En Küçük Kareler, Ridge, Lasso ve Temel Bileşen Regresyonunun Performans Karşılaştırması*



## Türkiye’de Bölgesel İşsizliğin Belirleyicilerinin Mekânsal Panel Veri Modelleriyle Analizi<sup>1</sup>

Beyda DEMİRCİ

İstanbul Üniversitesi, Ekonometri Bölümü / Doktora Öğrencisi

FMV Işık Üniversitesi İşletme Bölümü / Araştırma Görevlisi

beydademirci@gmail.com

Orcid No: 0000-0001-9476-3102

Ferda YERDELEN TATOĞLU

İstanbul Üniversitesi, Ekonometri Bölümü / Prof. Dr.

yerdelen@istanbul.edu.tr

Orcid No: 0000-0002-7365-3649

### Özet

İşsizlik, ekonomik kalkınmanın önünde büyük bir engel olup, bireylerin yaşam kalitesini doğrudan etkileyen önemli bir sorundur. İşgücü piyasasındaki dengesizlikler, yetersiz istihdam olanakları ve bölgesel farklılıklar, işsizlik oranının dalgalanmasına yol açmaktadır. Bu nedenle, bölgesel düzeyde işsizliği etkileyen sosyo-ekonomik faktörlerin belirlenmesi, işgücü piyasasında etkili politikalar geliştirmek için önemlidir. Bu çalışmada, Türkiye’nin İBBS Düzey 2’deki 26 bölgesi için 2014-2022 yıllarına ait verilerle işsizlik oranında mekânsal etkilerin olup olmadığı ve işsizlik oranını açıklayan faktörler mekânsal panel veri modelleri yardımıyla incelenmiştir. Test sonuçlarına göre, mekânsal hata modeli (SEM) uygun model olarak belirlenmiş ve bu modelin tahmin sonuçlarına göre bölgesel büyüme oranı, girişim sayısı ve genç nüfustaki artışın bölgesel işsizliği azalttığı; diğer yandan üniversite mezun sayısı ve yoksulluk oranındaki artışın işsizlik oranını arttırdığı tespit edilmiştir. Bu sonuçların, bölgesel işsizliği azaltmaya yönelik politikaların belirlenmesinde katkı sunacağı düşünülmektedir.

*Anahtar sözcükler: Mekânsal Panel Veri Analizi, Mekânsal Hata Modeli, İşsizlik, Türkiye Düzey2.*

**Sorumlu Yazar / Corresponding Author:** Beyda DEMİRCİ, İstanbul Üniversitesi, Ekonometri Bölümü.

**Atıf / Citation:** DEMİRCİ B., YERDELEN TATOĞLU F. (2024). Türkiye’de Bölgesel İşsizliğin Belirleyicilerinin Mekânsal Panel Veri Modelleriyle Analizi *İstatistik Araştırma Dergisi*, 14 (2), 1-13.

<sup>1</sup> Bu çalışma 7-9.11.2024 tarihlerinde gerçekleşen EMIDWORLD 2nd International Congress on Economics Public Finance Business & Social Sciences’da sözlü bildiri olarak sunulmuştur.

## Spatial Panel Data Analysis of Determinants of Regional Unemployment in Turkey

### Abstract

Unemployment is a significant barrier to economic progress and is an important issue that directly affects individuals' quality of life. Imbalances in the labor market, insufficient employment opportunities, and regional disparities lead to fluctuations in the unemployment rate. Therefore, identifying the socio-economic factors affecting unemployment at the regional level is important for developing effective policies in the labor market. This study used spatial panel data models to examine the spatial effects of the unemployment rate and the factors explaining it in Turkey's 26 regions at the NUTS-2 level for the years 2014–2022. The test results identified the spatial error model (SEM) as the appropriate one. Based on its estimation results, we found that the regional growth rate, the number of enterprises, and the increase in the young population reduce regional unemployment, while the increase in university graduates and the poverty rate increase it. It is believed that these results will contribute to the formulation of policies aimed at reducing regional unemployment.

*Keywords: Spatial Panel Data Analysis, Spatial Error Model, Unemployment, Türkiye NUTS2.*

### 1. Giriş

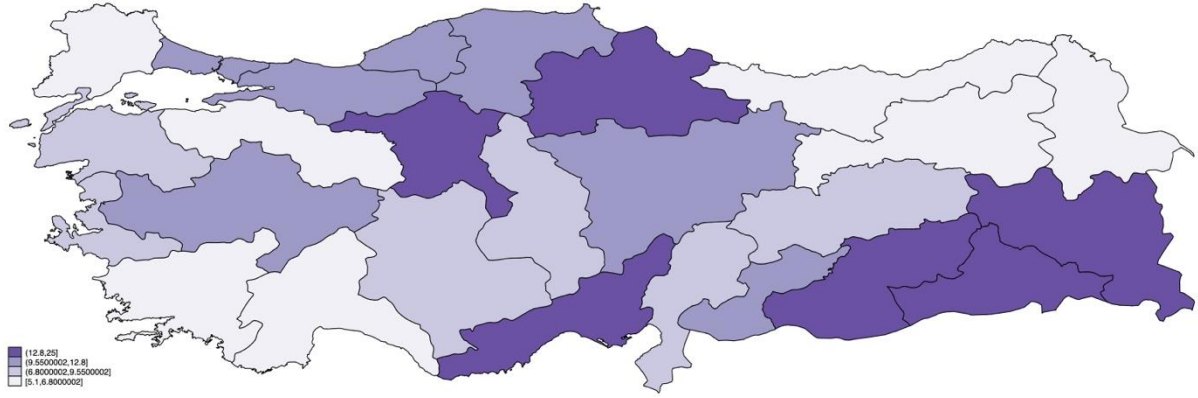
İşsizlik, bireyler ve toplumlar üzerinde geniş kapsamlı ekonomik ve sosyal etkilere neden olan önemli bir olgudur. İşsizlik oranının doğru şekilde ölçülmesi, işgücü piyasasının dinamiklerini anlamak ve etkili istihdam politikaları geliştirmek için kritik öneme sahiptir. Uluslararası Çalışma Örgütü’nün (ILO) 1982’de aldığı karara göre, işsizlik üç temel koşula dayanmaktadır. Bunlar:

- Kişinin işsiz olması, yani herhangi bir ücretli işte ya da serbest meslek faaliyetinde bulunmaması,
- İş bulduğunda hemen çalışabilecek durumda olması, yani iş için fiziksel ve mental açıdan hazır olması,
- Kişinin aktif olarak iş araması, yani yakın dönemde iş bulma amacıyla başvuru yapma, iş bulma kurumlarına kaydolma ya da kendi işini kurmak için girişimlerde bulunması,

gibi somut adımlar atmış olması şeklinde sıralanabilir. Bu koşullar işgücü piyasasının sınırlı olduğu ya da düzenli olmadığı yerlerde iş arama kriterinin esnetilmesine olanak tanımaktadır, böylece işsizlik, bu bölgelerde daha geniş bir anlamda ele alınabilmektedir (ILO, 1982). TÜİK’in tanımına göre ise, çalışma çağına olup, referans dönemi içinde istihdamda bulunmayan, son dört hafta içinde aktif olarak iş arayan ve iki hafta içinde işbaşı yapabilecek durumda olan bireyler, işsiz nüfus olarak tanımlanmaktadır (TÜİK, 2021). TÜİK, işsizlik oranını hesaplamada bireylerin aktif iş arama faaliyetlerini zorunlu bir kriter olarak ele alırken, ILO’nun tanımı işgücü piyasasının sınırlı olduğu bölgelerde daha esnek bir değerlendirmeye olanak tanımaktadır. Bu tanımlar arasındaki farklılık, Türkiye gibi bölgesel ekonomik eşitsizliklerin bulunduğu ülkelerde işsizlik oranının daha doğru bir şekilde değerlendirilmesi açısından önem taşımaktadır.

Diğer yandan Türkiye’de sanayileşmenin bazı bölgelerde yoğunlaşması, hava ve iklim koşulları gibi çeşitli nedenlerle işsizlik oranının dağılımında bölgeler bazında farklılıklar bulunmaktadır. Türkiye’de örnek olarak alınan 2018 yılında işsizlik oranının bölgesel dağılımını gösteren aşağıdaki ısı haritasına göre, en koyu renk en yüksek işsizlik oranının olduğu bölgeleri gösterirken, renk açıldıkça işsizlik oranı da azalmaktadır. Haritadan işsizlik oranının yüksek ve düşük olduğu bölgelerin benzer noktalara kümelenmiş olduğu görülmektedir, dolayısıyla farklı bölgeler arasında mekânsal kümelenmeler belirgindir. Bu ısı haritası, işsizlik oranı değişkeninin mekânsal bir özelliğe sahip olduğunu göstermektedir; Güneydoğu Anadolu, İç Anadolu ve Batı Karadeniz bölgelerinde işsizlik oranı yüksekken, diğer bölgelerde nispeten daha düşüktür.





**Şekil 1.** 2018 Yılı Türkiye NUTS 2 Düzeyinde İşsizlik Oranının Dağılım Haritası (Yazarlar tarafından hazırlanmıştır).

Bu mekânsal özelliklerinden dolayı bu çalışmada, Türkiye’de bölgesel işsizliğin belirlenmesinde, Türkiye’nin NUTS-2 seviyesindeki 26 bölgesinde 2014-2022 yılları arasında işsizlik oranının belirleyicileri mekânsal panel veri analizi yardımıyla incelenmiştir. Dengeli panel veri seti oluşturmak amacıyla çalışmada ele alınan 2014-2022 dönemi, Türkiye’de işsizlik oranını etkileyen önemli ekonomik ve politik gelişmelerin yaşandığı bir zaman dilimidir. 2014 yılı itibariyle mülteci sayısındaki artış, 2018’deki döviz krizi, 2020-2021 pandemi dönemi ve pandemi süresince hükümetin aldığı kararlar, 2021 yılı sonrasında hızla artan enflasyon ve gelir dağılımındaki eşitsizlikler gibi faktörler, işsizlik oranında değişimlere yol açmıştır. Ancak yine de 2022 yılında görülen bu mekânsal kümelenmenin benzeri, bazı farklılıklar bulunsa da genel olarak analize dahil edilen tüm dönemlerde gözlemlenmiştir.

Çalışma kapsamında, işsizlik oranının bölgesel dağılımına etki eden faktörler ve bu dağılımdaki mekânsal bağımlılık analiz edilmiştir. Bu doğrultuda, ikinci bölümde daha çok yerli ve yabancı literatürde yer alan işsizlik konusu ile ilgili yapılan ve mekânsal ekonometrik analiz içeren çalışmalar incelenmiştir. Üçüncü bölümde metodolojiye, dördüncü bölümde ise, analiz sonuçlarına yer verilmiştir. Elde edilen bulgular, sonuç bölümünde tartışılmıştır.

## 2. Literatür Özeti

İşsizlik, iktisat literatüründe sıklıkla çalışılan bir konudur. Ancak, mekânsal ekonometri alanında işsizlik üzerine yapılan çalışmalar nispeten daha yenidir ve sınırlıdır. Bu çalışmanın literatür bölümünde, önce yurtdışında yapılan özellikle mekânsal ekonometrik analizler içeren işsizlik ile ilgili çalışmalara ve ardından Türkiye’deki benzer çalışmalara yer verilmiştir.

Niebuhr (2002, 2003) ve Dönmez ve Sugözü (2022) ve Rios (2017) tarafından yapılan çalışmalar Avrupa için mekânsal ekonometrik yöntemler kullanılarak yapılan çalışmalara örnek olarak verilebilir. Niebuhr (2003), bölgesel işsizlik oranının mekânsal etkileşim dikkate alınarak belirleyicilerini araştırmıştır. 1986-2000 dönemi ve 359 Avrupa bölgesi üzerine yaptığı analiz sonuçları, bölgesel işsizlik oranının coğrafi yakınlık ve komşu bölgelerle olan ticaret, göç ve ulaşım gibi faktörlerden etkilendiğini ortaya koymaktadır. Mekânsal bağımlılığın yüksek olduğu bölgelerde, işsizlik oranının komşu bölgelerle benzer seviyelerde kümelendiği gözlemlenmiştir. Çalışmada, mekânsal ekonometrik yöntemler ve mekânsal otokorelasyon ölçütleri kullanılarak, yüksek işsizlik oranına sahip bölgelerin genellikle bir arada bulunduğu ve bunun sonucunda kalıcı işsizlik farklılıklarının oluştuğu vurgulanmıştır. Bu mekânsal bağımlılığın, bölgeler arası ticaret ve göç gibi nedenlerle meydana gelen mekânsal etkileşimden kaynaklandığı sonucuna ulaşılmıştır. 2002 yılında yaptığı çalışmada ise (Niebuhr (2002)), Avrupa Birliği’ndeki bölgesel işsizlik oranının mekânsal bağımlılığını incelemiştir. Çalışmada, 1986-1998 yılları arasında işsizlik oranı mekânsal ekonometrik yöntemler ile analiz edilmiştir. İşsizlik oranını açıklamak için istihdam büyümesi, nüfus yoğunluğu ve sektörel kompozisyon yer almıştır. Sonuçlar, yüksek işsizlik oranına sahip bölgelerin ve düşük işsizlik oranlı bölgelerin mekânsal olarak kümelendiğini göstermiştir. Ayrıca, komşu bölgelerdeki işgücü piyasalarının birbirini önemli ölçüde etkilediği ve bu mekânsal etkileşimin işgücü göçü ve ticaret gibi faktörler sebebiyle gerçekleştiği bulunmuştur. Dönmez ve Sugözü (2022), Avrupa Birliği (AB) ülkelerinde işsizlik ve ekonomik büyüme arasındaki ilişkiyi, Okun Yasası çerçevesinde incelemiştir. Çalışmada mekânsal panel veri analizi kullanılarak, 2003-2016 döneminde 26 AB ülkesi arasındaki komşuluk ilişkilerinin etkisi değerlendirilmiştir. Analizde kullanılan değişkenler arasında işsizlik oranı, tam istihdam düzeyinde işsizlik oranı ve potansiyel çıktı ile fiili çıktı arasındaki fark yer almaktadır. Elde edilen sonuçlar, AB ülkeleri arasında mekânsal bağımlılığın varlığını ve Okun katsayısının komşu ülkelerde benzerlik gösterdiğini ortaya koymaktadır.

Rios (2017), 2000-2011 yılları arasında 23 Avrupa ülkesinin 241 (NUTS 2) bölgesine ait işsizlik farklılıklarını dinamik mekânsal panel veri modeli yardımıyla analiz etmiştir. Dinamik Mekânsal Durbin Modeli (SDM) kullanılmış; işsizlik oranı, ücret büyüme oranı, işgücü büyüme oranı, sektörel istihdam payları, yaş grupları (55-64 ve 16-25), eğitim seviyesi, nüfus yoğunluğu ve göç gibi faktörler modele dahil edilmiştir. Sonuçlar, bölgesel işsizlik oranının önemli ölçüde mekânsal bağımlılık gösterdiğini ve kurumsal faktörlerin işsizlik farklılıkları üzerindeki etkisinin, özellikle 2009-2011 yılları arasındaki kriz döneminde daha belirgin hale geldiğini ortaya koymaktadır.

Literatürde çeşitli ülkelerin bölgesel işsizliğini mekânsal ekonometrik modeller ile inceleyen çalışmalar da bulunmaktadır. Vega ve Elhorst (2016), bölgesel işsizlik oranını incelerken değişkenin dinamik yapısı, mekânsal bağımlılık ve ortak faktörleri aynı anda dikkate alan bir model geliştirerek, 1973-2013 döneminde Hollanda’nın 12 bölgesi için analizlerini gerçekleştirmişlerdir. Araştırmada, bölgesel işsizlik oranı mekânsal dinamik panel veri modelleri ile analiz edilerek hem mekânsal bağımlılığın hem de ortak şokların (ortak faktörlerin) güçlü etkileri olduğu gözlemlenmiştir. Çalışma sonucunda, mekânsal bağımlılık ve ortak şokların bölgesel işsizliği anlamada kritik rol oynadığı vurgulanmaktadır. Hollanda için yapılan bir başka çalışma ise, Elhorst ve Emili (2022)’nin Hollanda’daki 12 bölge için 1974-2018 dönemini kapsayan verilerle, işsizlik oranı değişimi ile çıktı büyümesi arasındaki ilişkiyi iki dinamik mekânsal panel veri modeliyle inceledikleri çalışmadır. Çalışmanın sonuçlarına göre, çıktı büyümesinin işsizlik oranını düşürmedeki etkisi büyüktür ve bölgesel etkileşimler önemli rol oynamaktadır. Pietrzak ve Balcerzak (2016), Polonya’daki girişimcilik ve yatırımların işsizlik oranı üzerindeki etkilerini mekânsal ekonometrik yöntemlerle analiz etmiş; bu amaçla 2015 yılına ait 66 bölgeye (NUTS 3) ait veriler kullanılmıştır. Mekânsal Otoregresif (SAR) model sonuçlarına göre, girişimcilik ve yatırım düzeyindeki artışlar işsizlik oranını düşürmektedir. Özellikle, bir bölgede girişimcilikteki artışın, o bölgedeki işsizlik oranını doğrudan düşürdüğü ve bu etkinin komşu bölgelere de yayıldığı bulunmuştur. Çalışma, bu sonuçların bölgesel politikalarda dikkate alınması gerektiğini vurgulamaktadır. Cracolici vd. (2007), İtalya’da 1998 ve 2003 yıllarına ait işsizlik oranındaki bölgesel farklılıkları mekânsal ekonometrik modeller kullanarak analiz etmişlerdir. Çalışmada, üretim ve hizmet sektörlerindeki istihdam payı, kadın iş gücü katılım oranı gibi denge değişkenleri ile işgücü büyüme oranı gibi dengesizlik faktörleri kullanılarak, işsizlik oranı arasındaki mekânsal bağımlılığın ve kümelenme eğilimlerinin etkileri incelenmiştir. Çalışmanın sonuçları, kısa vadeli istihdam değişikliklerinin işsizlik farklılıklarını büyük ölçüde açıkladığını ve komşu bölgelerdeki işsizlik oranının birbirine benzeme eğiliminde olduğunu göstermiştir. Conley ve Topa (2002), 1980-1990 yılları arasında Chicago’daki işsizlik oranının mekânsal paternlerini incelemişlerdir. Çalışmada kullanılan değişkenler arasında işsizlik oranı, etnik bileşim, eğitim seviyesi ve meslek dağılımı bulunmaktadır. İşsizlik oranının mekânsal olarak yüksek derecede bağımlı olduğunu ve komşu bölgelerde işsizlik paternlerinin büyük benzerlik gösterdiğini ortaya koymuşlardır. Mitchell ve Bill (2004), Avustralya’nın 1991-2001 yılları arasındaki bölgesel işsizlik oranındaki mekânsal bağımlılığı inceledikleri çalışmada, komşu bölgelerdeki işsizlik oranının birbiri üzerinde güçlü bir etkisi olduğu ve bu etkinin mesafe arttıkça azaldığı ortaya konulmuştur. Lopez-Bazo vd. (2002), 50 İspanyol ili (NUTS 3) için 1985 ve 1997 yıllarına ait işsizlik oranının mekânsal analizini yapmışlardır. İşsizlik oranını, istihdam artışı, net göç, birim işgücü maliyeti, tarım ve imalat sektöründeki istihdam oranı, beşerî sermaye (eğitim durumu), genç nüfus oranı ve kadın/erkek işgücüne katılım oranı değişkenleriyle inceledikleri çalışmanın sonuçları, bölgesel işsizlik oranının güçlü bir mekânsal bağımlılık gösterdiğini ve bu bağımlılığın 1985’ten 1997’ye arttığını ortaya koymuştur. Kosfeld vd. (2008), Beveridge eğrisi kullanılarak işsizlik oranı ile açık iş pozisyonları arasındaki ilişkiyi mekânsal ekonometrik modellerle inceledikleri çalışmanın sonuçları; eğrinin dışa doğru kaymasının büyük ölçüde ekonomik döngü dalgalanmalarından kaynaklandığını, bölgesel iş uyumsuzluğunun ise önemli bir açıklama getirmediğini ortaya koymuştur. Patuelli vd. (2011), Almanya’daki bölgesel işsizlik oranını mekânsal bağımlılık ve heterojenliği göz önünde bulundurarak mekânsal filtreleme yöntemiyle analiz etmişlerdir. 1996-2004 yıllarına ait 493 NUTS-3 bölgesini kapsayan bu çalışmada, işsizlik oranını açıklamak için, günlük ortalama ücretler, tam zamanlı çalışan sayısı ve çalışma çağındaki nüfus (15-65 yaş arası) değişkenleri kullanılmıştır. Çalışmanın sonuçları; mekânsal filtrelerin, işsizlik oranındaki mekânsal bağımlılığı büyük ölçüde açıkladığını ve bu bağımlılığın zaman içinde tutarlılık sergilediğini göstermektedir.

Bronars ve Jansen (1987), Luna ve Genton (2004) tarafından yapılan çalışmalar ise ABD için yapılan çalışmalara örnek olarak verilebilmektedir. Luna ve Genton (2004); 1980-2002 yılları arasında aylık olarak ABD’nin dokuz bölgesine ait işsizlik oranı, nüfus yoğunluğu ve bölgesel istihdam oranı gibi sosyo-ekonomik faktörler kullanarak analizlerini gerçekleştirmişlerdir. Çalışmanın bulguları, mekânsal otoregresif modellerin işsizlik oranındaki mekânsal bağımlılık ve zamana bağlı özellikleri açıklamada oldukça başarılı olduğunu göstermektedir. Bronars ve Jansen (1987); ABD’deki işsizlik oranının coğrafi dağılımı mekânsal-zaman serisi analizi ile incelenmiştir. Çalışmada Mekânsal Otoregresif (SAR) model ve Koşullu Otoregresif (CAR) model kullanılmıştır. 1977-1983 yılları arasında ABD’nin 3100’den fazla ilçesinden elde edilen işsizlik oranı verilerine dayanarak bu çalışmada; işsizlik oranı, işgücü büyüklüğü ve demografik bilgiler yer almıştır. Çalışmanın sonuçları, işsizlik oranında meydana gelen geçici bir şokun komşu bölgelerde kısa vadede önemli bir etki yarattığını, ancak bu etkinin altı

çeyrek dönem sonra neredeyse tamamen ortadan kalktığını göstermektedir. Kalıcı değişikliklerin ise 250 mil uzaktaki işgücü piyasalarında bile uzun süreli etkileri olduğu tespit edilmiştir.

Türkiye için zaman serisi ile yapılan çalışmalar yanında iller bazında ya da bölgesel düzeyde yapılan çalışma sayısı oldukça fazla olsa da, komşuluk ilişkilerini dikkate almak için mekânsal ekonometrik analiz ve mekânsal panel veri analizi kullanan oldukça az sayıda çalışma bulunmaktadır. Aral ve Aytaç (2018), 2011 yılı için Türkiye’de 81 il bazındaki işsizlik oranının mekânsal etkileşimi incelemişlerdir. Çalışmada, Mekânsal Gecikme (SAR) modeli kullanılarak işsizlik oranının komşu bölgelerdeki etkileri analiz edilmiştir. Kullanılan değişkenler arasında işsizlik oranı, kadın ve erkek işgücüne katılım oranı, lise veya daha yüksek eğitilmiş nüfus oranı, genç nüfus oranı (15-24 yaş) ve net göç hızı yer almaktadır. Analiz sonuçları, kadınların işgücüne katılımının işsizlik üzerinde negatif bir etkisi olduğunu, genç nüfus oranının ve eğitilmiş nüfus oranının ise işsizlik oranını pozitif yönde etkilediğini göstermiştir. Karaalp Orhan ve Gülel (2016), 2008-2012 yılları özelinde Türkiye’nin 26 bölgesinin (NUTS 2) işsizlik oranını mekânsal panel veri analizi ile incelemişlerdir. İşsizlik oranını açıklamak için erkek ve kadın işgücüne katılım oranı, genç nüfus oranı (15-24 yaş), tarım ve imalat sektöründeki istihdam oranı ve eğitim seviyeleri değişkenleri kullanılmıştır. Sonuçlar, genç nüfus oranının işsizlik oranını artırdığını, tarım sektörünün istihdam oranının ise işsizlik oranını azalttığını göstermiştir. Ayrıca, bölgeler arasında pozitif bir mekânsal bağımlılık tespit edilmiştir.

Türkiye’de ve dünyada işsizlik üzerine yapılan olan çalışmalar, işsizliği etkileyen çok sayıda faktör olduğunu göstermektedir; ekonomik büyüme, enflasyon, döviz kuru, faiz oranı ve kamu harcamaları gibi makroekonomik faktörler literatürde öne çıkmaktadır. Yaş, cinsiyet ve eğitim düzeyi gibi bireysel ve hane özelliklerinin işsizlik üzerindeki etkisi de literatürde tartışılmıştır. İlâveten, genç işsizlik için spesifik çalışmalar bulunmaktadır. Ayrıca literatür, işsizlik oranının bölgesel analizinde mekânsal bağımlılık içermesi sebebiyle, mekânsal ve zamansal bağımlılıkların dikkate alınmasının, politika önerileri geliştirilmesinde daha etkili sonuçlar verdiğini göstermiştir. Yabancı literatürde, çeşitli ülkeler için bölgesel düzeyde mekânsal panel veri analizi kullanan çalışmalar bulunsa da, görüldüğü gibi, Türkiye için çok az sayıdadır. Bu çalışmanın bu kısıtlı literatüre, değişken çeşitliliği ve konuya bakış açısı ile katkı sunacağı düşünülmektedir. Dolayısıyla bu çalışmada, işsizlik oranını açıklamak için literatür takip edilerek bölgesel olarak verilerine ulaşılabilen bölgesel gayrisafı katma değer gelişme hızı, üniversite mezun sayısı, ortalama hanehalkı büyüklüğü, girişim sayısı, yoksulluk oranı ve genç nüfus değişkenleri kullanılmış ve ayrıca mekansallık dikkate alınarak analizler gerçekleştirilmiştir.

### 3. Metodoloji

Jean Paelinck, 1978’de çok bölgeli ekonometrik modelleri tahmin etme ve test etme zorluklarına odaklanan bölgesel bilim araştırmalarını tanımlamak için "mekânsal ekonometri" terimini ortaya atmıştır. Daha sonra mekânsal ekonometri, coğrafi etkileşimleri ve bölgeler arası bağlantıları ekonometrik modellere dahil ederek ekonomik ilişkileri daha kapsamlı bir şekilde ele almayı sağlayan ekonometrinin alt bilim dalı olarak tanımlanmıştır. Anselin (1988) ve LeSage ve Pace (2009) tarafından yapılan çalışmalar, bölgeler arasındaki mekânsal ilişkileri dikkate almak için 0 ve 1 olarak iki değer alan mekânsal ağırlık matrisinin oluşturulması ve uygulanmasını ele alan ilk çalışmalardandır. Ekonometrik modellerdeki mekânsal etkileşimlerin ölçülmesinde temel olan bu yaklaşımlar, mekânsal örüntülerin ve bağımlılıkların analizini kolaylaştırmaktadır.

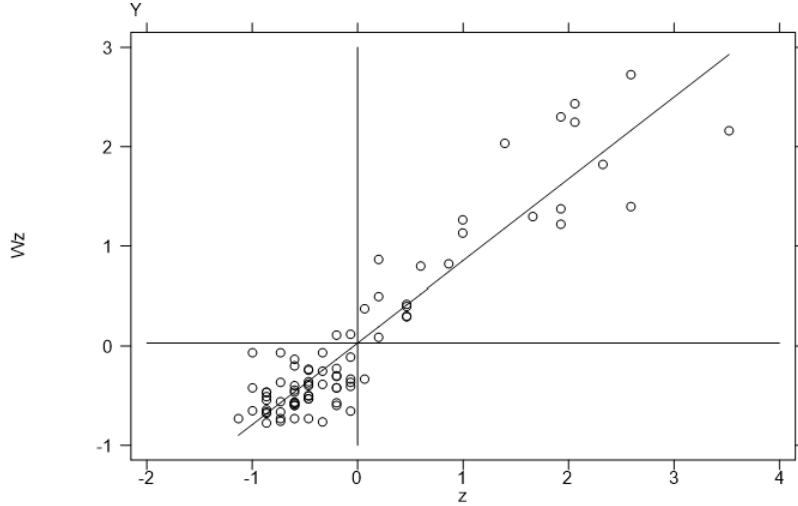
Gelinen noktada bölgesel çalışmalarda, öncelikle mekanlar arasındaki etkileşim incelenmeli ve eğer varsa dikkate alınmalıdır. Cliff ve Ord (1972) ve Hordijk (1974), hatalardaki ve değişkenlerdeki mekânsal otokorelasyonun tespiti için Moran’ın I testini önermişlerdir. Pozitif ve negatif değer alabilen Moran I istatistiği bir z değişkeni için aşağıdaki gibi hesaplanmaktadır:

$$I = \frac{n}{W_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{j=1}^n (z_j - \bar{z})^2} \quad (1)$$

I ile gösterilen Moran's I istatistiğinin tanım aralığı korelasyon katsayısında olduğu gibi -1 ile +1 arasındadır; değer -1'e yaklaştıkça negatif mekânsal otokorelasyonun, +1'e yaklaştıkça ise pozitif mekânsal otokorelasyonun varlığını ifade etmektedir. Ayrıca, Denklem 1'de i ve j farklı konumları,  $w_{ij}$  mekânsal ağırlık matrisini, n gözlem sayısını,  $W_0$  standardizasyon vektörünü ve  $z_i$  ve  $z_j$  sırasıyla değişkenin i ve j konumlarındaki değerini temsil etmektedir (Cliff ve Ord, 1981: 17; Fischer ve Wang, 2011: 23).

Moran I istatistiği yukarıdaki gibi global olarak hesaplanabildiği gibi bölgesel (lokal) bazda da hesaplanabilmektedir. Her bölge için lokal Moran I değerleri hesaplandıktan sonra Moran I serpilme diyagramı yardımıyla görsel olarak da izlenebilmektedir. Şekil 2’de yer alan Moran I serpilme diyagramında Moran I değerlerinin dağılımı görülmektedir. Diyagramın sol üst tarafı kendisinin aldığı değer düşükken komşularında yüksek olan kısmı ifade eden “düşük-yüksek” bölge olmak üzere, sağ üst kısım “yüksek-yüksek”, sol alt kısım “düşük-düşük” ve sağ alt kısım ise “yüksek-düşük” bölge olarak tanımlanmaktadır. Pozitif otokorelasyon,

gözlemler çoğunlukla sağ üst ve sol alt bölgelere düştüğünde görülürken, eğer gözlemlerin çoğunluğu diğer bölgelerde yer alıyorsa negatif otokorelasyon tespit edilmektedir. Bu bölgesel (lokal) bazda incelemeler, lokal istatistikler (LISA) ismini almakta ve tüm bölgeler için Denklem 1’deki gibi genel (global) bir istatistiğin (GISA) oluşturulmasında kullanılmaktadır. Global Moran I değeri pozitifse (pozitif otokorelasyon), serpilme diyagramında pozitif bir eğime işaret etmekte; aksi takdirde negatif bir eğime sahip olmaktadır (Yerdelen Tatoğlu, 2022: 124-125). Aşağıdaki örnek grafikte pozitif mekânsal otokorelasyon görülmektedir.



Şekil 2. Moran I Serpilme Diyagramı.

Bölgeler arası bağlantılar mekânsal ekonometrik modellerde çeşitli şekillerde temsil edilebilmektedir. Manski’nin 1993’te geliştirdiği Genel Yuvalanmış Mekânsal (GNS) model, mekânsal etkileşimlerin tümünü içeren kapsamlı bir modeldir ve aşağıdaki gibi ifade edilmektedir:

$$\text{GNS: } Y = \rho WY + X\beta + WX\theta + u \quad u = \lambda Wu + \varepsilon \quad (2)$$

Bu modelde  $W = I_T \otimes W_N$  iken  $I_T$ : T boyutlu birim matrisi,  $W_N$ :  $N \times N$  boyutlu ağırlık matrisi,  $\lambda$  mekânsal hata katsayısı,  $\rho$  mekânsal gecikme katsayısı ve  $\theta$  ise, mekânsal bağımsız değişkenlerin katsayısını ifade etmektedir. Bu modele farklı kısıtlamalar getirilerek alt grup mekânsal ekonometrik modeller türetilmektedir. Örneğin,  $\lambda=0$  kısıtı uygulandığında, Mekansal Durbin Model (SDM);  $\theta=0$  kısıtı getirilirse, Mekansal Otoregresif model (SAC);  $\rho=0$  kısıtı ile ise Mekansal Durbin Hata Modeli (SDEM) elde edilmektedir. İki katsayı kısıtlandığında ise daha alt modeller elde edilmektedir:  $\lambda=\theta=0$  olduğunda Mekansal Otoregresif (SAR) model;  $\rho=\theta=0$  kısıtıyla Mekansal Hata Modeli (SEM) ve  $\rho=\lambda=0$  iken Bağımsız Değişkeni Mekansal Gecikmeli (SLX) Model oluşturulmaktadır. Bu kısıtlamalar, modellerin farklı mekânsal etkileşim türlerini temsil etmesini sağlamakta ve analizdeki mekânsal bağımlılığın doğasına göre seçim yapılmasını mümkün kılmaktadır (Yerdelen Tatoğlu, 2022: 53-57). Bahsi geçen modeller sırasıyla aşağıda sunulmaktadır:

$$\text{SDM: } Y = \rho WY + X\beta + WX\theta + u \quad (3)$$

$$\text{SAC: } Y = \rho WY + X\beta + u \quad u = \lambda Wu + \varepsilon \quad (4)$$

$$\text{SDEM: } Y = X\beta + WX\theta + u \quad u = \lambda Wu + \varepsilon \quad (5)$$

$$\text{SAR: } Y = \rho WY + X\beta + u \quad (6)$$

$$\text{SEM: } Y = X\beta + u \quad u = \lambda Wu + \varepsilon \quad (7)$$

$$\text{SLX: } Y = X\beta + WX\theta + u \quad (8)$$

Hangi mekânsal etki ya da etkilerin anlamlı olduğunu, bir başka ifade ile en uygun modeli belirlemek için çeşitli testler kullanılabilir ve Lagrange Çarpanı (LM) testi bu testler arasında en çok kullanılanlardan birisidir (Burridge (1980), Anselin (1988)). (9) ve (10) eşitliklerinde mekânsal hatanın varlığını sınamak için kullanılan sırasıyla LM ve robust (dirençli) LM test istatistikleri yer almaktadır:

$$LM_{\lambda} = \frac{\left( \frac{\tilde{u}'(I_T \otimes W)\tilde{u}}{\hat{\sigma}_u^2} \right)^2}{tr(WW + W'W)} \sim X^2(1) \quad (9)$$

$$LM_{\lambda}^* = \frac{\left( \frac{\hat{u}'(I_T \otimes W)\hat{u}}{\hat{\sigma}_u^2} \right) - \left( \frac{\hat{u}'W_y}{\hat{\sigma}_u^2} \right) B^{-1} tr(WW + W'W)}{tr(WW + W'W)(1 - Btr(WW + W'W))} \sim X^2(1) \quad (10)$$

Burada  $\hat{u}$ : Mekânsal regresyon modelinin tahmininden elde edilen kalıntılardır. T, gözlem sayısı olmak üzere;  $I_T$ ,  $T \times T$  boyutlu birim matris, W: Mekânsal ağırlık matrisi,  $\hat{\sigma}_u^2$ : Mekânsal regresyon modelinden elde edilen kalıntı varyansdır.  $tr()$ : Bir matrisin köşegen elemanlarının toplamını temsil eden iz operatörüdür. Test istatistiği “mekânsal hata yoktur”  $H_0$  hipotezi altında 1 serbestlik dereceli ki-kare dağılımına uymaktadır ( $X^2(1)$ ). Eşitlik (10)’de:  $W_y$ : Mekânsal gecikmeli bağımlı değişkeni temsil eden matris ve B: Mekânsal gecikmeli bağımlı değişkenin katsayı matrisidir.

LM testleriyle mekânsal hata tespit edilebildiği gibi, mekânsal gecikmenin varlığı da sınanabilmektedir. Aşağıda mekânsal gecikmenin testi için, LM ve robust LM istatistikleri yer almaktadır: (Anselin vd., 1996: 83–84)

$$LM_{\rho} = \frac{\left( \frac{\hat{u}'WY}{\hat{\sigma}_u^2} \right)^2}{B} \sim X^2(1) \quad (11)$$

$$LM_{\rho}^* = \frac{\left( \left( \frac{\hat{u}'W_y}{\hat{\sigma}_u^2} \right) - \left( \frac{\hat{u}'(I_T \otimes W)\hat{u}}{\hat{\sigma}_u^2} \right) \right)^2}{B - tr(WW + W'W)} \sim X^2(1) \quad (12)$$

Mekânsal hata ve gecikmenin varlığının tespit edilmesinin yanısıra, panel veri ile çalışıldığında birim etkinin varlığı da F, LM ya da LR tipi bir test ile sınanmalıdır (bu çalışmada LM testi tercih edilmiştir).1980’de Breush Pagan’ın geliştirdiği ve “birim etkilerinin varyansı sıfırdır” temel hipotezini sınanan Rao (1938) Score testi olarak da bilinen test istatistiği şöyledir:

$$LM = \frac{NT}{2(T-1)} \left[ 1 - \frac{\sum_{i=1}^N \left( \sum_{t=1}^T \hat{v}_{it} \right)^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2} \right] \quad (13)$$

Burada  $\hat{v}_{it}$ , havuzlanmış en küçük kareler modelinin tahmininden elde edilen kalıntılardır. Birim etkinin varlığı anlaşıldıktan sonra, sabit ve tesadüfi etkiler tahmincileri arasında seçim yapılmaktadır. “Sabit ve tesadüfi etkiler tahmincileri arasında sistematik fark yoktur” temel hipotezini sınanan Hausman test istatistiği aşağıdaki gibidir:

$$H = (\hat{\beta}_{SE} - \hat{\beta}_{TE})' [Avar(\hat{\beta}_{SE}) - Avar(\hat{\beta}_{TE})]^{-1} (\hat{\beta}_{SE} - \hat{\beta}_{TE}) \quad (14)$$

SE ve TE alt indisleri sırasıyla sabit ve tesadüfi etkiler modellerinin tahmininden elde edilen parametre vektörleridir. Burada  $Avar(\hat{\beta}_{SE})$  ve  $Avar(\hat{\beta}_{TE})$ , sabit ve tesadüfi etkiler modeli tahminlerinden elde edilen asimptotik varyans kovaryans matrisleridir. Bu istatistik,  $\hat{\beta}_{SE}$  ve  $\hat{\beta}_{TE}$ ’deki ortak parametre sayısına eşit serbestlik derecesi ile asimptotik  $\chi^2$  dağılmaktadır (Yerdelen Tatoğlu, 2020: 196-197)

#### 4. Veri ve Analiz

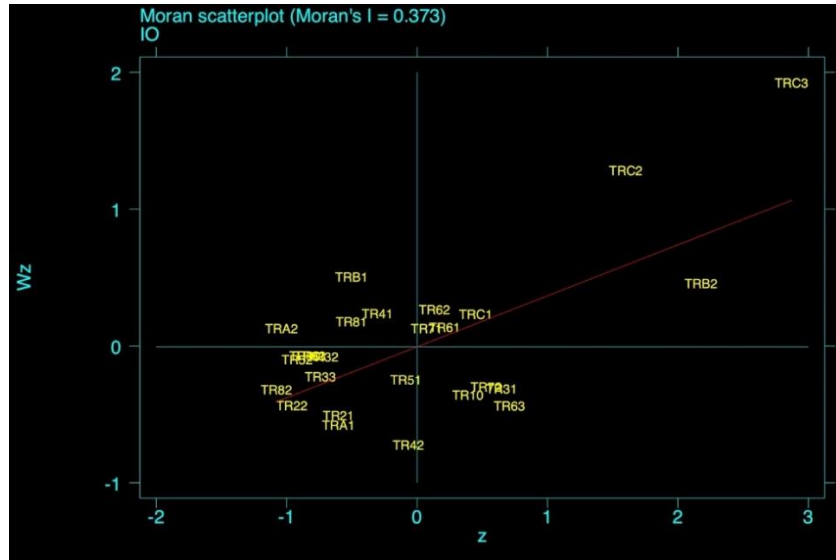
Bu çalışmada, Türkiye’de NUT2 düzeyinde (26 Bölge) 2014-2022 yıl aralığında işsizlik oranını etkileyen faktörleri belirlemek için mekânsal panel veri analizi kullanılmıştır. Mekânsal etkileri modele dahil etmek için NUT2 bölgeleri için sınır komşuluğuna dayanan mekânsal ağırlık matrisi oluşturulmuştur. İşsizliği çok sayıda faktörün etkilediği bilinmesinden dolayı Türkiye’de bölgesel düzeyde verisine ulaşılabilen gayrisafi katma değer gelişme hızı, üniversite mezun sayısı, ortalama hanehalkı büyüklüğü, girişim sayısı, yoksulluk oranı ve genç nüfus değişkenleri bağımsız değişkenler olarak kullanılmıştır. Aşağıdaki tabloda modelde kullanılan değişkenlerin açıklamaları yer almaktadır:

**Tablo 1.** Modelde Kullanılan Değişkenler.

Değişken Adı	Değişken Açıklaması	Veri Kaynağı ve Yıllar
<b>IO</b>	İşsizlik Oranı (%)	
<b>BYM</b>	Gayrisafi Katma Değer Gelişme Hızı (Cari Fiyatlarla, %)	
<b>UMEZ</b>	Üniversite Mezun Sayısı	
<b>HB</b>	Ortalama Hanehalkı Büyüklüğü	
<b>LGS</b>	Girişim Sayısı	
<b>YO</b>	Yoksulluk Oranı	
<b>YO</b>	(Eşdeğer hanehalkı kullanılabilir fert gelirine göre hesaplanmıştır. Gelirlerin referans dönemi bir önceki takvim yılıdır.)	
<b>GNC</b>	Genç Nüfus (15-24 Yaş)	
		TÜİK 2014-2022

Üniversite mezun sayısı, girişim sayısı ve genç nüfus değişkenleri, logaritmaları alınarak modele dahil edilmiştir.

Mekânsal etkinin varlığının testine geçmeden önce Şekil 1’de 2018 yılı için yer alan harita tüm yıllar için incelenmiş ve genel olarak tüm haritalarda bu haritaya benzer şekilde işsizlik oranının yüksek ve düşük olduğu bölgelerin kümelendiği ve genel olarak bir mekânsal etkinin olduğunu görülmüştür. Aşağıda ise, yine örnek olarak 2018 yılında, bağımlı değişken olan işsizlik oranı değişkeni için Moran I serpilme diyagramı yer almaktadır.



**Şekil 3.** İşsizlik Oranı Değişkeni Moran I Dağılım Diyagramı.

Moran I serpilme diyagramı incelendiğinde, gözlemlerin çoğunun “yüksek-yüksek” ve “düşük-düşük” bölgelerde yoğunlaştığı görülmektedir. Bu durum, global Moran I değerinin pozitif olduğunu ve işsizlik oranı değişkeninde çok yüksek olmasa da anlamlı pozitif mekânsal otokorelasyonun varlığını göstermektedir. Aşağıda ise tüm yıllar için Moran I testi sonuçları yer almaktadır.

**Tablo 2.** İşsizlik Oranı Değişkeni Moran I İstatistiği.

Yıl	I	E(I)	Sd(I)	Z	Olasılık
2014	0,202	-0,04	0,122	1,981	<b>0,048</b>
2015	0,17	-0,04	0,121	1,726	<b>0,084</b>
2016	0,226	-0,04	0,116	2,295	<b>0,022</b>
2017	0,189	-0,04	0,117	1,957	<b>0,050</b>
2018	0,378	-0,04	0,125	3,339	<b>0,001</b>
2019	0,382	-0,04	0,122	3,450	<b>0,001</b>
2020	0,368	-0,04	0,114	3,573	<b>0,000</b>
2021	0,286	-0,04	0,113	2,890	<b>0,004</b>
2022	0,231	-0,04	0,125	2,165	<b>0,030</b>

Moran I test istatistikleri 2015 yılı haricinde tüm yıllarda %95 güven düzeyinde, 2015 yılı için ise %90 güven düzeyinde anlamlıdır.  $H_0$  hipotezi olan “mekânsal otokorelasyon yoktur” hipotezi, tüm yıllar için reddedilmektedir; bağımlı değişkendeki mekânsal otokorelasyon anlamlıdır.

Birim etkinin varlığını sınamak için, Breusch Pagan ve Sosa Escudero Yoon LM testleri yapılmıştır. LM testleri olası otokorelasyonun varlığına karşı dirençli değilken, ALM (Adjusted LM) testleri dirençlidir. Tüm testlerin temel hipotezi “birim etki yoktur” şeklinde kurulmuştur. Aşağıdaki tabloda bu testlerin sonuçları yer almaktadır.

**Tablo 3.** Tahmincilerin Belirlenmesi / Birim Etkinin Testi.

	SDM	SAC	SEM	SAR
<b>Breusch Pagan LM</b>	271,514***	321,137***	321,137***	321,137***
<b>Breusch Pagan ALM</b>	136,110***	163,762***	163,762***	163,762***
<b>Sosa Escudero Yoon LM</b>	16,477***	17,920***	17,920***	17,920***
<b>Sosa Escudero Yoon ALM</b>	11,667***	12,796***	12,796***	12,796***

Not: \*\*\* %1 anlamlılık düzeyini ifade etmektedir.

Testler tüm olası modeller üzerinden yapılmıştır. Tüm test sonuçlarına göre,  $H_0$  hipotezi reddedilmiştir, birim etki yani bölge etkisi vardır. Bir sonraki aşamada birim etkinin bağımsız değişkenlerle korelasyonunun sınanması için, SDM ve SAR modelleri üzerinden Hausman testi yapılmıştır ve Tablo 4’te yer verilmiştir.

**Tablo 4.** Tahmincilerin Belirlenmesi / Sabit ya da Tesadüfi Etkinin Testi.

	SDM	SAR
<b>Hausman Testi</b>	77,342***	40,353***

Not: \*\*\* %1 anlamlılık düzeyini ifade etmektedir.

SDM ve SAR modeller üzerinden yapılan Hausman testi sonucuna göre, sabit ve tesadüfi etkiler tahmincileri arasında sistematik bir fark bulunmadığını ileri süren  $H_0$  hipotezi reddedilmiştir. Bu durumda sabit etkiler tahmincisi tutarlı olduğundan kullanılmasının uygun olduğu belirlenmiştir. Sabit etkiler modeli üzerinden uygun mekânsal modelin belirlenebilmesi aşamasında mekânsal hata ve mekânsal gecikmenin varlığı LM ve robust LM testleri yardımıyla sınanmıştır ve aşağıdaki tabloda yer almaktadır.

**Tablo 5.** Uygun Modelin Belirlenmesi İçin Yapılan Testler.

Test	Katsayı	Olasılık
<b>1.Moran I</b>	0,231	0,030**
<b>2.LM<sub>p</sub> (H<sub>0</sub>: <math>\rho=0</math>)</b>	0,599	0,451
<b>3.LM<sub><math>\lambda</math></sub> (H<sub>0</sub>: <math>\lambda=0</math>)</b>	7,793	0,005**
<b>4.Robust LM<sub>p</sub> (H<sub>0</sub>: <math>\rho=0</math>)</b>	0,003	0,954
<b>5.Robust LM<sub><math>\lambda</math></sub> (H<sub>0</sub>: <math>\lambda=0</math>)</b>	7,230	0,007**
<b>6.LM<sub>p,<math>\lambda</math></sub> (H<sub>0</sub>: <math>\rho=\lambda=0</math>)</b>	7,796	0,020**

Not: \*\* %5 anlamlılık düzeyini ifade etmektedir.

Tablo 5’te “mekânsal hata ve mekânsal gecikme yoktur” temel hipotezini sınanan test sonucuna göre (6)  $H_0$  hipotezi reddedilmiştir, mekânsal hata ve/veya mekânsal gecikme vardır. Bu doğrultuda mekânsal gecikme ve mekânsal hatanın varlığı ayrı ayrı sınanmıştır. “Mekânsal gecikme yoktur” hipotezini sınanan LM testi (2) sonucuna göre  $H_0$  hipotezi reddedilemediğinden, mekânsal gecikme yoktur. “Mekânsal hata yoktur” hipotezini test eden LM testi (3) sonucuna göre ise,  $H_0$  hipotezi reddedilmiştir; mekânsal hata vardır. 3 ve 5 numarada yer alan robust LM test sonuçları da, bu sonuçları desteklemektedir. Sonuç olarak, Mekânsal Hata Modeli (SEM) uygundur. Bulunan bu sonucu doğrulamak ve bağımsız değişkenlerde mekânsal gecikmeyi içeren modellerin de uygunluğunu görmek için model seçim kriterleri elde edilmiştir ve aşağıdaki tabloda raporlanmaktadır.

**Tablo 6.** Model Seçim Kriterleri.

Kriter	SAC	SAR	SEM	SDM
$\bar{R}^2$	0,3765	<b>0,3884</b>	0,3871	0,0980
AIC	1242,36	1243,43	1267,47	<b>1235,38</b>
SC	63,2034	46,8952	<b>15,7437</b>	31,5003
HQ	59,4224	44,0897	<b>14,8018</b>	28,0906
RICE	57,1031	42,3689	<b>14,2241</b>	26,1725
Shibata	56,8987	42,2172	<b>14,1732</b>	25,8493

Model seçim kriterlerine göre, en yüksek  $\bar{R}^2$  ve en düşük bilgi kriterine sahip model uygun modeldir. Schwarz (SC), Hannan Quin (HQ), RICE ve Shibata bilgi kriterleri en düşük olan model SEM’dir.  $\bar{R}^2$ ’si en yüksek olan model SAR iken bunu küçük bir farkla SEM modeli takip etmektedir. Akaike bilgi kriteri (AIC) en düşük olan model SDM olmasına rağmen diğer kriterler ve LM test sonuçları SEM modeli desteklediği için SEM uygun model olarak seçilmiştir. Aşağıda sabit etkiler modeli üzerinden yapılan SEM tahmin sonuçları yer almaktadır.

**Tablo 7.** SEM Model Tahmin Sonuçları.

Değişken	Katsayı	Standart Hata	z
HB	-2,051*	1.205	-1,7
BYM	-0,028**	0,008	3,36
LUMEZ	15,348***	3.282	4,68
LGS	-16,489**	6.189	-2,66
YO	0,186**	0,058	3,18
LGNC	-17,053**	4.813	-3,54
$\lambda$	0,333***	0,084	3,96
$\sigma$	2,583***	0,242	10,67
R <sup>2</sup>	0,4221		

Not: \*, \*\* ve \*\*\* sırasıyla %10, %5 ve %1 anlamlılık düzeylerini ifade etmektedir.

Tablo 7’de yer alan SE-SEM tahmin sonuçlarına göre, modelin bağımsız değişkenleri işsizlik oranındaki değişkenliğin %42’sini açıklamaktadır. Modeli açıklamak için kullanılan tüm bağımsız değişkenler istatistiki olarak anlamlıdır. İşsizlik oranını, üniversite mezun sayısı ve yoksulluk oranı pozitif yönde, bölgesel gayrisafi katma değer gelişme hızı, ortalama hane halkı büyüklüğü, girişim sayısı ve genç nüfus negatif yönde etkilemektedir. Bu sonuçların hepsi işsizlik teorisi açısından anlamlı ve literatürle uyumludur. Mekânsal hata katsayısı  $\lambda$  istatistiksel olarak anlamlıdır, değeri komşu bölgelerin hata terimleri arasında %33 pozitif korelasyon olduğu göstermektedir.

## 5. Sonuç

Bu çalışmada Türkiye’nin 2014-2022 yılları arasında 26 NUTS-2 bölgesinde işsizlik oranının belirleyicileri analiz edilmiştir. İşsizlik oranını açıklamak amacıyla, mekânsal panel veri analizinden yararlanılmıştır. Bu kapsamda, bölgesel gayrisafi katma değer büyüme hızı, üniversite mezun sayısı, ortalama hanehalkı büyüklüğü, girişim sayısı, yoksulluk oranı ve genç nüfus gibi faktörlerin işsizlik oranı üzerindeki etkileri değerlendirilmiştir. Analiz sonuçları, işsizlik oranında mekânsal bir bağımlılığın olduğunu ve bölgeler arasında karşılıklı etkileşimlerin varlığını göstermektedir. Moran I testleri, pozitif mekânsal otokorelasyonu kanıtlamıştır. Yapılan testler sabit etkili mekânsal hata modelinin (SE-SEM) en uygun model olduğunu ortaya koymuştur ve bu model tahmin edilmiştir.

Tahmin sonuçlarına göre, üniversite mezun sayısı ve yoksulluk oranı, işsizlik oranını artırıcı bir etkiye sahipken; bölgesel gayrisafi katma değer büyüme hızı, ortalama hanehalkı büyüklüğü, girişim sayısı ve genç nüfus, işsizlik oranını azaltıcı yönde etkileyen faktörler olarak öne çıkmaktadır. Çalışmada kullanılan bağımsız değişkenlerin işsizlik üzerindeki etkisi, literatürdeki araştırmacıların bulgularıyla benzerlik göstermektedir. Bu bulgular, politika yapıcılar için önemli ipuçları sunmaktadır. Üniversite mezun sayısının işsizlik oranını artırıcı etkisi, eğitim politikalarının işgücü piyasası talepleriyle uyumlaştırılmasının önemini ortaya koymaktadır. Bölgesel gayrisafi katma değer büyüme hızının işsizlik oranını azaltıcı etkisi ise, işsizliğin yüksek olduğu bölgelerde ekonomik teşviklerin, altyapı yatırımlarının ve özel sektörü destekleyici politikaların istihdam yaratmadaki önemini



vurgulamaktadır. Girişim sayısının işsizlik oranını düşürmesi, girişimcilik faaliyetlerinin artırılması için KOBİ’lere yönelik desteklerin ve finansal erişim imkanlarının güçlendirilmesi gerektiğini göstermektedir. Yoksulluk oranının işsizlikle pozitif ilişkisi, işsizlikle mücadelede gelir dağılımını iyileştiren sosyal destek programlarının önemini vurgulamaktadır. Ortalama hanehalkı büyüklüğünün işsizlik üzerindeki negatif etkisi, demografik yapıların işgücü piyasası üzerindeki etkisini dengelemek adına sosyal politikaların işgücüne katılımı artırmada rol oynayabileceğini işaret etmektedir.

Çalışmada, Türkiye’de işsizlik oranı mekânsal bağımlılık bağlamında ele alınarak, bölgeler arası etkileşimler mekânsal ekonometrik modellerle kapsamlı bir şekilde analiz edilmiştir. Literatürdeki sınırlı sayıda çalışmaya ek olarak, Türkiye’de işsizlik analizi kapsamında mekânsal etkinin varlığını ortaya koyulmuştur. Literatürde Türkiye’nin işsizlik oranı üzerine yapılan çalışmalarda mekânsal ekonometrik analizin sınırlı kullanımını dikkate aldığında, bu çalışma mekânsal otokorelasyonu doğrudan ele alması ve güncel bir veri seti kullanmasıyla önemli bir katkı sunmaktadır. Analiz sonuçları, işsizlik oranında mekânsal bir bağımlılığın ve bölgeler arasında karşılıklı etkileşimlerin olduğunu gösterdiğinden, işsizliğin analizinde mekânsal etkilerin varlığını göz ardı etmemek gerekmektedir; işsizlik oranının bölge bazında değil, komşu bölgeleri de kapsayacak şekilde bölgesel olarak azaltılması için çaba gösterilmelidir. Bu özellikleriyle bu çalışmanın, Türkiye’de işsizlikle mücadelede bölgesel farklılıkları dikkate alan politika önerilerinin geliştirilmesini desteklediği söylenebilmektedir. Elde edilen sonuçlar doğrultusunda, bölgesel düzeyde farklı ihtiyaçlara uygun stratejiler geliştirilmesi ve işgücü piyasasında uyumlu bir yapının sağlanması önerilmektedir.

Bundan sonraki çalışmalarda, Türkiye’nin bölgesel işsizlik oranını daha ayrıntılı incelemek için bazı yöntemsel ve içeriksel genişlemeler yapılabilir. Özellikle, sektörel işsizlik verileri ile yaş ve cinsiyet gibi demografik alt gruplar üzerinden analizler gerçekleştirilerek işsizlik oranındaki farklılıkların daha detaylı olarak ele alınması ve buna ek olarak, dinamik mekânsal panel veri modelleri kullanılarak işsizlik oranının zaman içindeki değişimlerinin dikkate alınması öneriler arasında sayılabilir. Ayrıca, bölgesel ya da il bazında veri mevcudiyeti sağlandığı takdirde bağımsız değişkenlerin kapsamı genişletilerek teknoloji, altyapı kaynakları, enerji kullanımı, bölgesel teşvikler, fabrika sayıları gibi değişkenlerin modele dahil edilebilmesi ile, işsizlik oranını etkileyen daha geniş bir faktör yelpazesi analiz edilerek daha kapsamlı sonuçlar elde edilebilir. Bu tür genişlemeler, işsizlikle mücadelede yerel yönetimlere ve politika yapıcılara daha etkili ve hedeflenmiş çözümler geliştirme konusunda yol gösterici olabilir.

**Yazarların çalışmadaki katkı oranları eşittir.**

**Çalışma herhangi bir destek almamıştır. Teşekkür edilecek bir kurum veya kişi bulunmamaktadır.**

**Çalışma kapsamında herhangi bir kurum veya kişi ile çıkar çatışması bulunmamaktadır.**

## Kaynaklar

- Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20(1), 1–17. <https://doi.org/10.1111/j.1538-4632.1988.tb00159.x>
- Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77–104. [https://doi.org/10.1016/0166-0462\(95\)02111-6](https://doi.org/10.1016/0166-0462(95)02111-6)
- Aral, N., & Aytaç, M. (2018). Türkiye’de işsizliğin mekânsal analizi. *Öneri Dergisi*, 13(49), 1-20.
- Arbia, G. (2006). *Spatial Econometrics: Statistical foundations and applications to regional convergence*. Springer Science & Business Media.
- Breusch, T. S., & Pagan, A. R. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies*, 47(1), 239-253.
- Bronars, S. G., & Jansen, D. W. (1987). The geographic distribution of unemployment rates in the US: A spatial-time series analysis. *Journal of Econometrics*, 36, 251–279.
- Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(1), 107–108. <https://doi.org/10.1111/j.2517-6161.1980.tb01108.x>
- Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: models & applications*. Pion.
- Cliff, A., & Ord, K. (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4(3), 267–284. <https://doi.org/10.1111/j.1538-4632.1972.tb00475.x>
- Conley, T. G., & Topa, G. (2002). Socio-economic distance and spatial patterns in unemployment. *Journal of Applied Econometrics*, 17(4), 303–327. <https://doi.org/10.1002/jae.670>
- Cracolici, M. F., Cuffaro, M., & Nijkamp, P. (2007). Geographical distribution of unemployment: An analysis of orovincial differences in Italy. *Growth and Change*, 38(4), 649–670. <https://doi.org/10.1111/j.1468-2257.2007.00391.x>
- Dönmez, B., & Sugözü, H. İ. (2022). The relationship between unemployment and economic growth under Okun’s Law: A spatial econometric analysis on EU countries. *Reforma*, 1(93), 35-44.
- Elhorst, J. P., & Emili, S. (2022). A Spatial econometric multivariate model of Okun’s Law. *Regional Science and Urban Economics*, 93, 103756. <https://doi.org/10.1016/j.regsciurbeco.2021.103756>
- Fischer, M. M., & Wang, J. (2011). *Spatial data analysis: models, methods and techniques*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-21720-3>
- Hordijk, L. (1974). Spatial correlation in the disturbances of a linear interregional model. *Regional and Urban Economics*, 4(2), 117–140. [https://doi.org/10.1016/0034-3331\(74\)90025-6](https://doi.org/10.1016/0034-3331(74)90025-6)
- Karaalp Orhan, H. S., & Gülel, F. E. (2016). Türkiye’de bölgesel işsizlik: Mekansal panel veri analizi. *Sosyal Güvenlik Dergisi*, 6(2), 47–67.
- Kosfeld, R., Dreger, C., & Eckey, H.-F. (2008). On the stability of the German Beveridge Curve: A spatial econometric perspective. *The Annals of Regional Science*, 42(4), 967–986. <https://doi.org/10.1007/s00168-007-0190-y>
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman & Hall/CRC.
- Lopez-Bazo, E., Del Barrio, T., & Artis, M. (2002). *The regional distribution of spanish unemployment: A spatial analysis*. Papers in Regional Science, 81, 365–389.
- Luna, X., & Genton, M. G. (2004). Spatio-temporal autoregressive models for US unemployment rate. In *Spatial and Spatiotemporal Econometrics* (279-294). Emerald Group Publishing Limited.

- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3), 531-542.
- Mitchell, W., & Bill, A. (2004). *Spatial dependence in regional unemployment in Australia*. Working Paper No. 04(11), Australia: Centre of Full Employment and Equity The University of Newcastle.
- Niebuhr, A. (2002). *Spatial dependence of regional unemployment in the European Union* (No. 186). HWWA Discussion Paper.
- Niebuhr, A. (2003). Spatial interaction and regional unemployment in Europe. *European Journal of Spatial Development*, 1(5), 1–26.
- International Labour Organization(ILO). (1982). *Resolution concerning statistics of the economically active population, employment, unemployment and underemployment*. Thirteenth International Conference of Labour Statisticians, October 1982.
- Paelinck, J. (1978). Spatial econometrics. *Economics Letters*, 1(1), 59–63.
- Patuelli, R., Griffith, D. A., Tiefelsdorf, M., & Nijkamp, P. (2011). Spatial filtering and eigenvector stability: space-time models for German unemployment data. *International Regional Science Review*, 34(2), 253–280. <https://doi.org/10.1177/0160017610386482>
- Pietrzak, M. B., & Balcerzak, A. P. (2016). *A spatial SAR model in evaluating influence of entrepreneurship and investments on unemployment in Poland*, Research Working Papers No:2, Institute of Economic Research (IER), Toruń (Poland)
- Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *In Mathematical Proceedings of the Cambridge Philosophical Society*, 44(1), 50-57). Cambridge University Press.
- Rios, V. (2017). What drives unemployment disparities in European regions? A dynamic spatial panel approach. *Regional Studies*, 51(11), 1599–1611. <https://doi.org/10.1080/00343404.2016.1216094>
- Türkiye İstatistik Kurumu (TÜİK). (2021). *Hanehalkı işgücü araştırması genel açıklama*. Erişim adresi: <https://data.tuik.gov.tr/>
- Vega, S. H., & Elhorst, J. P. (2016). A regional unemployment model simultaneously accounting for serial dynamics, spatial dependence and common factors. *Regional Science and Urban Economics*, 60, 85–95. <https://doi.org/10.1016/j.regsciurbeco.2016.07.002>
- Yerdelen Tatoğlu, F. (2020). *Panel Veri Ekonometrisi Stata Uygulamalı*. Beta Basım Yayım Dağıtım AŞ, İstanbul.
- Yerdelen Tatoğlu, F. (2022). *Mekansal Ekonometri Stata Uygulamalı*. Beta Basım Yayım Dağıtım AŞ, İstanbul.



## Sürdürülebilir Kalkınma Kapsamında Dijitalleşme ve Çevre Ekonomisi İlişkisinin Analizi

Sibel ÖRK ÖZEL

Çukurova Üniversitesi, Ekonometri Bölümü / Doç. Dr.

sork@cu.edu.tr

Orcid No: 0000-0002-7030-3512

Dilek VEYSİKARANI

Munzur Üniversitesi, Uluslararası Ticaret ve İşletmecilik Bölümü / Arş. Gör. Dr.

dilekveysikarani@munzur.edu.tr

Orcid No: 0000-0001-8071-0720

### Özet

Küresel çevresel sorunlar ve sürdürülebilir kalkınma hedefleri günümüzde gittikçe daha fazla önem kazanırken, ülkelerin bu doğrultuda stratejiler geliştirmesi ve uygulaması gerekmektedir. Bu çalışmada, sürdürülebilir kalkınma perspektifinden çevre ekonomisini etkileyen faktörlerin analiz edilmesi amaçlanmıştır. Bu amaçla dijitalleşmeyi temsil eden çeşitli faktörlerin çevre ekonomisinde önemli bir değişken olan karbon emisyonları üzerindeki etkileri lojistik regresyon analizi ile incelenmiştir. Çalışmada 131 ülkenin Karbondioksit Emisyonu, Elektriğe Erişim Oranı, Düzeltilmiş Tasarruflar: Enerji Tükenmesi, GSYİH, İnternet Kullanımı ve Yenilenebilir Enerji Tüketimi değişkenlerine ait 2020 yılı verileri kullanılmıştır. Bu doğrultuda ele alınan değişkenlerin karbon emisyonları üzerine etkisini belirlemek maksadı ile yapılan lojistik regresyon analizinde, GSYİH'nın ve yenilenebilir enerji tüketiminin karbon emisyonlarını azaltıcı etkileri bulunurken, düzeltilmiş tasarruflar: enerji tükenmesinin karbon emisyonlarını artırdığı saptanmıştır. Diğer yandan, elektriğe erişim ve internet kullanımının karbon emisyonları üzerinde anlamlı bir etkisi bulunmamıştır. Modelin doğruluk oranı %79 olarak elde edilmiş ve 131 ülkenin yaklaşık 104'ünün karbon emisyonları doğru olarak öngörülmüştür. Bu bulgular, ekonomik büyüme ve yenilenebilir enerji yatırımlarının çevresel sürdürülebilirliği destekleyebileceğini ortaya koymaktadır.

*Anahtar sözcükler: Sürdürülebilir Kalkınma, Karbon Emisyonları, Lojistik Regresyon Analizi.*

**Sorumlu Yazar / Corresponding Author:** Sibel ÖRK ÖZEL, Çukurova Üniversitesi, Ekonometri Bölümü.

**Atıf / Citation:** ÖRK ÖZEL, S., VEYSİKARANI, D. (2024). Sürdürülebilir Kalkınma Kapsamında Dijitalleşme ve Çevre Ekonomisi İlişkisinin Analizi, *İstatistik Araştırma Dergisi*, 14 (2), 14-29.

## **Analysis of the Relationship between Digitalization and Environmental Economics within the Scope of Sustainable Development**

### **Abstract**

As global environmental issues and sustainable development goals are becoming increasingly important, it is essential for countries to develop and implement strategies in this direction. This study aims to analyze the factors affecting environmental economics from a sustainable development perspective. For this purpose, the impacts of various factors representing digitalization on carbon emissions, a significant variable in environmental economics, were examined using logistic regression analysis. The study used 2020 data from 131 countries, including variables such as Carbon Dioxide Emissions, Access to Electricity, Adjusted Savings: Energy Depletion, GDP, Internet Usage, and Renewable Energy Consumption. In the logistic regression analysis conducted to determine the effects of these variables on carbon emissions, GDP and renewable energy consumption were found to have a reducing effect on carbon emissions, while adjusted savings: energy depletion was found to increase emissions. On the other hand, access to electricity and internet usage did not have a significant effect on carbon emissions. The model achieved an accuracy rate of 79%, correctly predicting carbon emissions for approximately 104 of the 131 countries. These findings suggest that economic growth and investments in renewable energy can support environmental sustainability.

*Keywords: Sustainable Development, Carbon Emissions, Logistic Regression Analysis.*

### **1. Giriş**

Kalkınma, basit bir tabirle kişi başı gelirin artırılması olarak tanımlanabilmektedir. Bu tarz bir kalkınma klasik kapitalist modelin belirlediği bir kalkınma şeklidir. Bu modelde üretim sınırsız olup amaç, kaynakları tüketerek yüksek kar elde etmektir. Bu şekilde kalkınan bir ülkede gelecek yok sayılmakta, kaynaklar hunharca harcanmaktadır. Bu kalkınma şekli ile birlikte fosil yakıt kaynaklı kaynaklara olan bağımlılık, doğal kaynakları düşüncesiz bir şekilde tüketme, bu tüketim sonucunda ortaya çıkan atıklar, çevrenin zarar görmesi, sağlıksız bir biçimde kentleşme, küresel ısınma, iklimlerin bozulması, suların kirletilmesi, hayvanların soylarının tükenmesi ve doğal bitki örtüsünün özelliklerini kaybetmesi gibi sonuçlar olağan bir şekilde beraberinde gelmektedir (Ozmehmet, 2008: 1854).

Sürdürülebilir kalkınma kavramı ise yukarıda bahsedilen bu çevresel bozulmaların gün geçtikçe daha fazla artması nedeni ile son yıllarda en çok tartışılan konulardan biri haline gelmiştir. Sürdürülebilirlik; çevre, ekonomi ve enerji kavramlarının ortasında bulunmakta, bundan dolayı da hükümetlerin, politika yapıcılarının ve küresel kurumların incelediği konu olarak merkezlerinde yer almaktadır. Bundan hareketle sürdürülebilirliğin ölçülmesi için yaklaşımlarda bulunulmakta ve çeşitli amaçlar oluşturulmaktadır (Yeni, 2014: 183). Günümüzde hemen hemen herkes tarafından bilinen, tüm ülkelerin bu amaçları gerçekleştirmek için çabaladığı “Sürdürülebilir Kalkınma Amaçları” Birleşmiş Milletler tarafından tanımlanan amaçlar dizisidir. Tüm ülkeler bu amaçları 2030 yılına kadar gerçekleştirmeyi hedeflemektedir. 17 hedeften oluşan Sürdürülebilir Kalkınma Amaçları Şekil 1’de gösterilmiştir.



Şekil 1. Sürdürülebilir Kalkınma Amaçları, <https://turkiye.un.org/tr/sdgs>, Erişim Tarihi: 19.07.2024.

Şekil 1’den de anlaşılacağı üzere sürdürülebilir kalkınmanın başarılı olabilmesi için üç temel boyut üzerinde durulması gerekmektedir. Bu boyutlar çevre boyutu, ekonomik boyut ve sosyal boyuttur. Çevresel boyut, yukarıda ifade edilen kaynakların korunarak tüketilmesini gerektirir. Özellikle de gelişmekte olan ülkeler için kalkınmak beraberinde doğal kaynakların korunması zorunluluğunu da getirmektedir. 1972 Stockholm Çevre Konferansı, 1992 Rio Konferansı ve 1997 Kyoto Protokolü çevre ve sürdürülebilir kalkınma ilişkisinin ele alındığı çalışmalardır. 1997 Kyoto Protokolü, sera gazı salınımının azaltılmasını, enerji kaynaklarının kullanımında yenilenebilir kaynakların tercih edilmesini, çevre dostu doğal enerji kaynaklarının kullanılmasını, yakıt tüketimini fazla yapan ve fazlaca karbon üretimi yapanlardan daha çok vergi alınmasını ifade eden uluslararası bir anlaşmadır (Tıraş, 2012). 1970’lere gelinmesiyle beraber sürdürülebilir kalkınma anlayışı çevreye daha fazla duyarlı olma anlayışını ve bununla beraber ekonomik büyüme anlayışını da getirmiştir. Ekonomik kalkınmada sürdürülebilir kalkınmanın hem araç olarak kullanılabileceği hem de çevre dostu üretim politikaları yapmaya teşvik edeceği görülmeye başlanmıştır (Toprak, 2006: 147).

Geçilen bu yıllar beraberinde teknolojik gelişmeleri de getiren dönemlerdir. Dijitalleşme dönemi olarak da ifade edebileceğimiz bu dönem itibarı ile dijitalleşme, yaşanan yeni dönemde hayatın tüm alanlarında kendini ve etkisini gösteren, giderek yayılan bir süreç olarak ifade edilebilmektedir (Atatanır, 2022: 78). Şu anda içinde bulunduğumuz beşinci sanayi devrimi yani Toplum 5.0 içinde yaşadığımız bu dijitalleşme sürecinin adıdır. Bahsi geçen bu süreç hızlı bir biçimde teknolojik değişimle günlük yaşamımızı etkileyen her alanda değişimlere sebep olmaktadır. Hem yaşam tarzını hem devlet politikalarını hem de işletme yönetimlerini değiştiren dijitalleşme sürdürülebilir kalkınma için de temel bir değer oluşturmaktadır. Bu süreç istihdam kavramını, emeğin verimliliği, fiziki ve beşerî sermaye kalitesini ve sürdürülebilir kalkınma kavramlarını etkileyerek politika yapımcılar için büyük bir önem arz etmektedir (Demirkıran vd., 2022: 31).

Bundan hareketle çalışmada sürdürülebilir kalkınma kapsamında hem çevre ekonomisini etkileyebileceği düşünülen hem de dijitalleşme göstergeleri olduğu düşünülen elektriğe erişim oranı, düzeltilmiş tasarruflar: enerji tükenmesi, gayrisafi yurtiçi hasıla (GSYİH), internet kullanımı ve yenilenebilir enerji tüketimi değişkenleri bağımsız değişkenler olarak seçilerek bu değişkenlerin çevre ekonomisini temsilen seçilen karbondioksit emisyonu bağımlı değişkeni üzerine etkileri incelenmiştir. Bu amaçla lojistik regresyon analizi kullanılmıştır. Kullanılan değişkenler için ortak yıl olması sebebi ile Dünya Bankası tarafından yayımlanan 2020 yılı verileri kullanılmış ve ülkemizin de içinde bulunduğu 131 ülke için istatistiksel analiz yapılmıştır.

Çalışmanın ikinci bölümünde konuya ilişkin son yıllarda literatürde öne çıkan çalışmalardan bahsedilmiştir. Ardından çalışmada kullanılan yöntemden ve veri setinden bahsedilmiştir. Sonuç kısmında ise elde edilen bulgular yorumlanarak birtakım önerilerde bulunulmuştur.

## 2. Literatür Taraması

Sürdürülebilir kalkınma kapsamında hem çevre ekonomisi hem de dijitalleşme kavramlarını ele alan çalışmalar üç farklı durumla karşımıza çıkmaktadır. İlk olarak sürdürülebilir kalkınma ve çevre ilişkisini inceleyen kavramlar sıklıkla literatürde görülmektedir. Bir diğer türlü bakış olarak sürdürülebilir kalkınma ile dijitalleşme arasındaki ilişki incelenmiştir. Ayrıca literatürde dijitalleşme ve çevre ekonomisi arasındaki ilişkiyi inceleyen çalışmalar diğer durumlara göre daha az da olsa görülmektedir.

Bu çalışmanın farkı sürdürülebilir kalkınma kapsamında dijitalleşme göstergeleri olduğu düşünülen değişkenlerin çevre ekonomisi üzerine etkisini incelemektir. Bu açıdan çalışmanın literatüre katkı sağlayacağı düşünülmektedir. Bu bölümde 2000’li yıllardan itibaren özellikle de son dönemlerde yapılan, yukarıda ifade edilen kapsamlardaki yerli ve yabancı çalışmalara yer verilmeye çalışılmıştır.

Mutlu (2007), çalışmasında ilk olarak çevre muhasebesini incelemiş ardından hem çevre muhasebesini hem de çevre ekonomisini sürdürülebilir kalkınma çerçevesinde araştırmıştır.

Omer (2008) makalesinde daha temiz bir enerji teknolojisine yönelik kritik ihtiyaca yanıt olarak sabit ve taşınabilir enerji piyasasında bu tür entegre sistemlerin potansiyelini tartışmıştır. Gelecekteki enerji kullanımına ilişkin öngörülen modeller ve bunun sonucunda ortaya çıkan çevresel etkiler bu makalede kapsamlı bir şekilde tartışmıştır. Tema boyunca yenilenebilir enerjiler, çevre ve sürdürülebilir kalkınma ile ilgili çeşitli konular hem mevcut hem de gelecek perspektiflerinden incelenmiştir.

Yalçın (2010) çalışmasında, sürdürülebilir kalkınma ile iklim değişikliği mücadelesi için düşük karbon emisyonlu bir ekonomi modelinin olması gerektiğini ve düşük karbon emisyonlu ekonomi modelinin nasıl bir çerçevede olması gerektiğini ifade etmeyi amaçlamıştır. Çalışmada kuramsal manada bir değerlendirme araştırması yapılmıştır. Çalışmanın sonucunda düşük karbon emisyonlu ekonomi modelinin, iklim değişikliği ile mücadele ve sürdürülebilir kalkınma kapsamında en makul serbest piyasa çözümü olduğu ifade edilmiştir.

Marangoz, Önce ve Aydın (2015) yaptıkları çalışmada dünyadaki e-atık uygulamalarını değerlendirilmekte, geri dönüşüm oranları ve ekonomik etkilerini incelenmekte, e-atık yönetiminin finansal tetikleyici olarak önemini vurgulanmakta, e-atık yönetiminin sadece kısa vadeli değil uzun vadeli etkileri de irdelenmiştir. Sürdürülebilirlik açısından öneriler verilen çalışmada uygulayıcılar ve yasa koyucular için geri dönüşüm süreçlerinin iyileştirilmesi amacıyla ele alınan konu tartışılmıştır.

Yılmaz (2018) ise sürdürülebilir kalkınma sürecinden yeşil büyüme sürecine kadar yaşanan gelişmelerden sürdürülebilir kalkınmanın kavramının nasıl ölçüleceğinden, sürdürülebilir kalkınma ve yeşil büyüme göstergelerini ifade ederek bu iki kavram arasındaki ilişkinin açıklanmasını amaçlamıştır. Sürdürülebilir kalkınmanın yeşil büyüme yani çevre ekonomisinin temelinde olduğu ifade edilen çalışmada yeşil büyümenin üç temel zorunluluğu oluşumdan bahsedilmektedir.

Al (2019) yeşil ekonominin sürdürülebilir kalkınma için temel stratejisi olarak kabul edilmesi ile yeşil ekonomide çevresel endişelerle ekonomik hedefler arasındaki ikilemin ortadan kalkacağı düşünülüğünü, bu nedenle yeşil ekonominin ölçülüp değerlendirilmesinin uygulanabilecek politikalar açısından büyük önem taşıdığını çalışmasında belirtmiştir. Çalışmada, ülkemiz için yeşil ekonomi performansının ölçülmesi ve değerlendirilmesi amaçlanmıştır. Elde edilen bulgular neticesinde bahsedilen dönemde ülkemizde yeşil ekonomideki performansının arttığı ifade edilmiştir.

Polasky vd. (2019), sürdürülebilir kalkınmayı başarmanın zorluğunun büyük ve acil olduğunu, dünya sistemlerinde devam eden büyük değişikliklerin insan refahı açısından potansiyel olarak büyük olumsuz sonuçlara neden olabileceğini ve bu çevresel değişimler karşısında yoksulluğun nasıl azaltılacağı ve artan eşitsizliğin nasıl ele alınacağı, çevresel boyutların yanı sıra sürdürülebilir kalkınmanın önemli sosyal ve ekonomik yönlerini de gündeme getirdiğinden bahsetmiştir. Bu çalışmada konu ile ilgili makaleler incelenerek sürdürülebilir kalkınma zorluklarının üstesinden gelmek için gereken çalışma türüne ilişkin olumlu örnekler sunulmuştur.

Mete (2020) çalışmasında lojistik sektördeki faaliyetlerin gelişiminin sera gazı ve karbondioksit emisyonu üzerindeki etkisini araştırmayı amaçlamıştır. Bu kapsamda, 2010 ila 2018 arası döneme ait Avrupa Birliği’ne üye ülkeler ve Türkiye’de gerçekleşen lojistik faaliyetleri ele alarak bu faaliyetlerin karbondioksit emisyonları ve sera gazı üzerindeki etkisini panel regresyon modelini kullanarak analiz etmiştir. Analizde elde edilen sonuçlara göre, incelenen ülkelerin lojistik faaliyetlerinde yaşanan gelişim ile karbondioksit emisyonları ve sera gazı arasında hem negatif hem de anlamlı bir ilişki elde edilmiştir. Ayrıca çalışmada yenilenebilir enerji kaynakları kullanım oranları ile de karbondioksit emisyonları ve sera gazı arasında hem negatif hem de anlamlı bir ilişki bulunmuştur.

Özaslan (2023), nüfusun artmasının çevre üzerine etkisinin de olduğunu ve bunun olumsuz bir şekilde devam ettiğini çalışmasında ifade etmiştir. Çevre üzerine olumsuz etkilerden arınabilmek için ülkelerin yeşil ekonomi ve yeşil büyüme yaklaşımlarına yöneldiği yine çalışmada ifade edilmiştir. Yeşil ekonomi ile karbondioksit salınım oranı azaltılacak, iklim değişikliğinin yarattığı olumsuz etkiler azaltılabilecektir. Bu çalışmada toplumsal refahı sağlayabilmek adına uygulanacak olan bu yaklaşımların sürdürülebilir kalkınma çerçevesinde açıklanması amaçlanmıştır. Ayrıca şehirlerde sürdürülebilir çevrenin nasıl oluşturulabileceğinin bir değerlendirilmesi de yapılmıştır.

Jovanović, Dlačić ve Okanović (2018) çalışmalarında dijitalleşme ile sürdürülebilir kalkınma arasındaki ilişkiyi incelemekte ve ulusların dijital rekabet gücünü ölçmek için kullanılan bileşik endeksi - Dijital Ekonomi ve Toplum Endeksi’ni (DESI) sunmaktadır. Bu makale DESI metodolojisini incelemiş ve dijital performansın nasıl olduğunu gözlemlenmiştir. Ayrıca çalışmada, DESI ve sürdürülebilirlik bileşenleri olan ekonomik, sosyal ve çevresel etkenleri ölçen diğer bileşik endekslerin korelasyonlarını araştırılmıştır. Bununla birlikte araştırmada Hofstede’nin

kültürel boyutları ile dijital performans arasındaki ilişkiler de incelenmiştir. Makale, toplumun sürdürülebilir kalkınmasının bir diğer önemli bileşeni olarak dijitalleşmenin önemini vurgulamıştır.

Konu (2020) çalışmasında teknolojik gelişmelerin takibinin günümüzde çok önemli olduğunu ifade etmiştir. Dijitalleşmenin hem ekonomik hem de sosyal açıdan bir güç unsuru haline geldiği ifade edilen çalışmada dijital ekonominin önemi sürdürülebilir kalkınma bağlamında ele alınmıştır. Bu amaçla 2018 yılına ait veriler AB ülkeleri için kullanılarak yatay kesit analiz metodu ile dijital ekonominin sürdürülebilir kalkınma üzerine etkisi incelenmiştir. Elde edilen bulgulara göre dijital ekonomiyi temsilen kullanılan dijital ekonomi ve toplum endeksinin, sürdürülebilir kalkınmayı temsilen ele alınan sürdürülebilir kalkınma endeksi üzerinde hem olumlu hem de istatistiksel olarak anlamlı bir etkiye sahip olduğu ifade edilmiştir.

Çayrağası ve Sakıcı (2021), Amerika Birleşik Devletleri'nde yaşanan politika ile ilgili süreçlerden ve 2015 yılında yapılan Paris Anlaşması'na kadar yaşanan süreçlerden, bunlarla birlikte Avrupa Komisyonu ve İngiltere sürecine dayanan Green Deal (Yeşil Mutabakat) ve 2030 yılına kadarki hedef olan Birleşmiş Milletlerin sera gazı emisyonu azaltma hedeflerinden bahsettikleri çalışmalarında bunların tamamı ile birlikte Covid-19 sürecinin de sürdürülebilirliği artırma hedeflerinin tetiklendiğinden söz etmişlerdir. Bahse konu süreçler dijital teknolojileri ve yeşil kavramını öne çıkaran süreçler olmakta olup DIGITALEUROPE adlı kuruluşun bu aşamalarda kilit rol olduğunu ifade etmişlerdir. Bundan dolayı çalışmada Web of Science veri tabanında konu ile ilgili literatür tarama çalışması yapılarak sürdürülebilir dijital pazarlama stratejileriyle alakalı gene bir kavramsal çerçeve oluşturmak amaçlanmıştır.

Mondejar vd. (2021) dijitalleşmenin toplum ve çevre için potansiyel faydalar sağlayan, kullanılmamış büyük verilerden oluşan entegre bir ağa erişim sağladığını ifade ettikleri çalışmalarında nesnelerin internetine bağlı akıllı sistemlerin geliştirilmesi, adil, çevresel olarak sürdürülebilir ve sağlıklı bir toplum sağlamak amacıyla Birleşmiş Milletler Sürdürülebilir Kalkınma Hedefleri (SDG'ler) ile bağlantılı zorlukların stratejik olarak ele alınması için benzersiz fırsatlar yarattığını detaylı olarak ele almıştır. Bu açıdan dijitalleşmenin geleceğin sürdürülebilir toplumunu inşa etmeye yönelik sağlayabileceği fırsatları yani sürdürülebilir gıda üretimi, temiz ve güvenli içme suyuna erişim ve yeşil enerji üretimi ile kullanımını çalışmada tartışmışlardır. Son olarak nesli tükenmekte olan gezegen biyoçeşitliliği ve iklim değişikliğinden kaynaklanan ciddi zorlukların ele alınmasına nasıl katkıda bulunabileceğine dair bütünsel bir bakış sunarak dijitalleşmenin faydalarını açıklamışlardır.

Demirkıran, Beyoğlu, Terzioğlu ve Yaşar (2022), çalışmalarında dijitalleşmeyi ve dijitalleşmenin verimliliğe olan etkisini hem teorik olarak hem de pratikte incelemeyi amaçlamışlardır. Yapay sinir ağ kullanılarak bu süreçte etkili olan değişkenler ve hatta değişkenlerin önem düzeylerine göre sıralamaları bulunmak istenmiştir. Bu çalışma ülkemiz için 2005 ila 2020 arası dönemde yıllık veriler kullanılarak yapılmıştır. Çalışmanın bulgularına göre dijitalleşme sürecinde verimlilik artışının teşviki ile sürdürülebilir olarak kalkınmaya da katkı sağlanacağı sonucuna ulaşılmıştır.

Akyol, Akar ve Akar (2023), çalışmalarında 27 adet Avrupa Birliği ülkesinde 2000-2020 arası dönem için dijitalleşmenin ekonomik büyümeye etkisini panel veri analizi metodunu kullanarak incelemeyi amaçlamışlardır, Driscoll-Kraay (1998) standart hatalar tahmin edicisi kullanılarak değişkenler arasındaki ekonometrik ilişki tahmin edilmiştir. Çalışmada dijitalleşme bağımsız değişken olarak ele alınmış ve dijitalleşme göstergeleri olarak bilgi ve iletişim teknolojisi ürünleri ihracatı cep telefonu abone sayısı ve internet kullanan birey sayısı seçilmiştir. Ekonomik büyüme ise bağımlı değişken olarak düşünülmüş ve onu temsilen ise kişi başı reel GSYH kullanılmıştır. Çalışmada elde edilen bulgulara göre dijitalleşmenin ekonomik büyümeye hem anlamlı hem de pozitif bir etkisi olduğu belirlenmiştir.

Kuvvetli Yavaş (2024), günümüzde insanlığın çeşitli çevresel, sosyal ve ekonomik sorunlarla karşı karşıya olduğundan bahsettiği çalışmasında 2010 yıllarına gelindiğinde bu sorunlara istinaden çözüm arayışlarına gidildiğinden, böylece dijital dönüşüme ve yeşil dönüşüme politika önerilerinin ortaya çıktığından söz etmiştir. Çalışmada; 2015 yılında Birleşmiş Milletler tarafından ortaya konulan Sürdürülebilir Kalkınma Hedefleri (Amaçları), 2011 yılında Almanya'da Endüstri 4.0 kavramı, 2016 yılında Japonya'da ortaya çıkan Toplum 5.0, 2021 yılında Avrupa Birliği'nin ortaya çıkardığı Endüstri 5.0 açılımı bu politikalara örnek olarak gösterilmiştir. Tüm bu ifadelerin merkezinde olan dijitalleşmeden dolayı ilgili çalışmada dünyadaki dijitalleşme sürecinin seyri, bahsi geçen kavramlar ve bunların çevreye ve sosyal hayata yansımaları değerlendirilmiştir.

Sazanova (2021) çalışmasında modern ekonomide iş organizasyonundaki oynadıkları rol, ekonomik varlıkların faaliyetlerini ve ekonominin sürdürülebilir kalkınmasını nasıl etkiledikleri açısından iş ekosistemi gibi yeni iş organizasyonu biçimlerini araştırmıştır. Makalenin amacı, ekosistemi ekonomik bir olgu, ülkenin sosyo-ekonomik sisteminin bir bileşeni olarak, ekonominin dijitalleşmesi bağlamında sürdürülebilir ekonomik kalkınmanın olası bir itici gücü olarak incelemektir. Çalışma, analiz, sentez yöntemi, bilimin rasyonel yeniden inşası yöntemi, karşılaştırmalı analiz yöntemi, genelleme yöntemi ve diğerleri gibi genel bilimsel ve özel araştırma yöntemlerinin kullanımına dayanmaktadır. Karşılaştırmalı analiz yöntemi, yerli ve yabancı bilim adamlarının ekonomik ekosistemlerin özüne ilişkin görüş farklılıklarını ortaya çıkarmıştır. Sentez yöntemi, modern ekonominin bilimde ekonomik ekosistemlerin incelenmesine yönelik yaklaşımları belirlememize olanak sağlamıştır. Bilimsel bilginin rasyonel olarak yeniden inşası yöntemi, doğu ve batı ekonomik düşüncesine ilişkin ekonomik çalışmalarda



ekonomik kalkınmanın bir faktörü olarak bilgiyle ilgili farklılıkların tespit edilmesini mümkün kılmıştır. Genelleme yöntemi ise araştırma sonuçlarının formülasyonunun temeli haline gelmiştir.

Lee, He ve Yuan (2023), dijital gelişimin yeşil toplam faktör verimliliği üzerindeki etkisini araştırmak için Çin'in Yangtze Nehri Deltası'nın 2002'den 2020'ye kadar olan panel verilerini kullanarak bir çalışma gerçekleştirmiştir. İlk olarak, bu çalışmada, yeşil toplam faktör verimliliğini kapsamlı ve sistematik bir şekilde hesaplamak için beklenmedik çıktılara sahip süper SBM yöntemini sunmuştur. İkinci olarak, mekansal-zamansal heterojenliği ayırt edebilen, dijital gelişimin GTFP'yi nasıl iyileştirdiğini analiz edebilen ve bu ilişkide insan sermayesi ve endüstriyel yapı iyileştirmesinin etkisini değerlendirebilen kısmi doğrusal fonksiyon katsayısı modeli kullanılmıştır. Ayrıca bu delta bölgesi ile Çin'deki diğer bölgeler arasında karşılaştırmalı bir çalışma yürütmüştür. Bu makalenin sonuçları şu şekildedir: (1) Dijitalleşmenin tüm örnekler için GTFP üzerinde anlamlı ve olumlu bir etkisi vardır. (2) HC ve IS, Yangtze Nehri Deltası'ndaki şehirlerin GTFP'sini teşvik etmede dijital kalkınmanın rolünü etkiliyor, ancak bunların etkisi kuzeydoğudaki üç ilde önemli değildir. (3) Stimülasyon etkisi, HC ve IS'nin etkisi altında önemli zamansal ve mekansal heterojenlik gösterir. Eyalet başkentleri ve ekonomik açıdan gelişmiş şehirler en belirgin teşvik edici etkiye sahiptir. Son olarak, dijital kalkınma ile GTFP arasındaki parametrik olmayan ilişkiye dayanarak, bu makale GTFP'yi iyileştirmek ve yeşil ekonomik kalkınmayı teşvik etmek için yapıcı ve pratik politika önerileri sunmaktadır.

Yeşilyaprak (2023), dünyada dijitalleşme ile birlikte e-ihracat uygulamalarının arttığını ve bununla birlikte lojistik faaliyetlerinin de artış gösterdiğini, bu durumun ise çevre kirliliğinin artmasına yol açtığını ifade ettiği çalışmada yeşil lojistik alanında birçok firmanın yapmış olduğu uygulamalara yer verilmiştir. Ayrıca hem yurt dışı hem de da Türkiye'nin bazı e-ihracat ve e-ticaret verilerini vererek karşılaştırmalarda bulunmuştur.

### 3. Yöntem ve Veri

Çalışmanın bu bölümünde, araştırmada kullanılacak olan çok değişkenli istatistiksel analiz yöntemi olan lojistik regresyon analizi ile bağımlı ve bağımsız değişken setleri hakkında bilgi verilecektir.

#### 3.1. Yöntem: Lojistik Regresyon Analizi

Regresyon analizi, bağımlı ve bağımsız değişkenler arasında var olduğu öngörülen ilişkinin matematiksel modellerle açıklanmasıdır. Bir bağımlı değişkenin değişimi, iki veya ikiden daha çok bağımsız değişkenle açıklandığında çoklu regresyon kullanılır. Modelde kullanılan değişkenler arasındaki ilişkiler doğrusal olarak belirlendiğinde "Doğrusal Regresyon Analizi" şeklinde ifade edilmektedir, aksi durumda ise "Doğrusal Olmayan Regresyon Analizi" şeklinde adlandırılır. Doğrusal regresyonda bağımsız değişkenler sürekli veya kesikli nicel, nitel veri türünde olabilirken, bağımlı değişken sürekli niceldir. Uygulamada kesikli nicel bağımlı değişkenlere de rastlanmaktadır. Bağımlı değişkenin iki veya daha fazla düzeyde nitel veri özelliği gösterdiği durumlarda ise doğrusal regresyon analizi kullanılamaz. Söz konusu böyle durumlarda lojistik regresyon analizi kullanılmaktadır (Alpar, 2011: 409; Büyüköztürk, 2018: 91). Lojistik regresyon analizi süreklilik, normallik, çok değişkenli normallik ve eşvaryanslık gibi varsayımlara gerek duymadan regresyon modelinin kurulmasını sağlayan bir istatistiksel analiz tekniğidir. Bu analiz, sürekli veya kesikli yordayıcı değişkenlerin bir arada kullanılabilirliği ve bağımlı değişkenin kesikli olduğu durumlarda uygulanmaktadır. Ayrıca, sürekli bir bağımlı değişken gerektiğinde kesikli bir değişkene dönüştürülerek lojistik regresyon analizi yapılabilmektedir (Tabachnick ve Fidell, 2020: 521).

Lojistik regresyon analizinde, logit dönüşümünün uygulandığı bağımlı (yordayıcı) değişkenin yapısı, analizinin sınıflandırılmasında oldukça önemli bir rol oynar. Bu analiz türü, bağımlı değişkenin sahip olduğu yapıya göre üç farklı modele ayrılır. Bu modellerin ilki olan "İkili Lojistik Regresyon Modeli", bağımlı değişkenin iki kategoriye sahip olduğu durumlarda (örneğin, sigara içiyor-içmiyor) kullanılmaktadır. İkincisi ise "Multinomial Lojistik Regresyon Modeli", bağımlı değişkenin birden fazla kategoriye sahip olduğu durumlarda (örneğin, eğitim durumu; ilkökul-ortaokul-lise-üniversite) uygulanmaktadır. Bağımlı değişkenin birden fazla kategorik ve sıralı yapıya sahip olduğu durumlarda (örneğin, sağlık durumu; kötü-orta-iyi) ise "Sıralı Lojistik Regresyon Modeli" tercih edilir (Alpar, 2011: 615; Şenel ve Alatl, 2014).

İkili lojistik regresyon analizi, modelde kullanılan bağımlı değişkenin iki kategorili olduğu durumlar (var/yok; evet/hayır, büyük/küçük vb.) için özel olarak kullanılan bir lojistik regresyon türüdür. Lojistik regresyon analizi, temelde regresyon mantığına dayanan yapısı ve diğer analizlere kıyasla daha kullanışlı olması nedeniyle kategorik veri analizinde ön plana çıkar (Çokluk, 2010: 1360). Lojistik regresyon analizi, bağımlı değişkenin iki veya ikiden fazla kategorilerde gözlemlendiği durumlarda, bağımsız değişkenlerle olan ilişkisini belirlemek için kullanılan parametrik olmayan bir yöntemdir. Bu analiz, bağımsız değişkenlere göre bağımlı değişkenin beklenen değerini olasılık olarak hesaplamaktadır ve ayrıca olasılık kurallarına göre sınıflandırma yapmaktadır (Tabakan ve Avcı, 2021: 547).

Bağımlı değişkenin kategorik yapıya (iki, üç veya daha fazla) sahip olduğu durumlarda, lojistik regresyon analizi bağımlı değişkenin bağımsız değişkenlerle arasındaki neden-sonuç ilişkisini saptamak ve bağımsız değişkenlerin bağımlı değişken üzerinde var olduğu öngörülen etkilerini odds oranı ile açıklamak için kullanılan bir tekniktir.

Odds oranı ise bir olayın gerçekleşme olasılığının, ilgili olayın gerçekleşmeme olasılığına bölünmesiyle elde edilir ve bu durum  $\text{Exp}(\beta)$  ile temsil edilir. Lojistik regresyon analizinde önemli görülen bir diğer kavram logit'tir ve odds oranının doğal logaritması alınarak hesaplanan bir değerdir (Yerdelen Kaygın, Tazegül ve Yazarkan, 2016: 150; Erkilic ve Aksoy, 2020).

Lojistik regresyon modelinde, bağımlı değişkenin olasılığı  $p$ :

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}} \quad (1)$$

Denklem 1'de yer alan bağımlı değişkenin olasılığını ifade eden  $p$ , bağımlı değişkenin belirli bir kategoriye ait olma olasılığını (örneğin, başarılı olma olasılığı) ifade etmektedir.  $\beta_0$  modelin sabit terim katsayısını,  $\beta_1, \beta_2, \dots, \beta_k$  katsayıları ise modelde yer alan  $k$  adet bağımsız değişkenlerin katsayılarıdır.  $X_1, X_2, \dots, X_k$  değerleri ise  $k$  adet bağımsız değişkenleri göstermektedir. Gerekli düzeltmelerin ardından lojistik regresyon formülü şu şekilde hesaplanır:

$$\ln \frac{p_k}{1-p_k} = \beta_0 + \sum_k \beta_k X_k \quad (2)$$

Maksimum Likelihood (ML) tahmin yöntemi kullanılarak yapılan lojistik regresyon analizlerinde, hataların normal dağıldığı ve sabit varyanslı varsayımı gibi koşullar aranmamaktadır.  $\text{Exp}(\beta)$  parametreleri,  $X=1$  olan bireylerin  $X=0$  olanlara kıyasla bağımlı değişkenin ( $y$ ) ne kadar daha sık görüldüğünü gösterir ve bu bilgi üzerinden değerlendirme yapılır. Eğer  $\text{Exp}(\beta) > 1$  ise, bu, parametrenin referans gruba kıyasla bağımlı değişkenin ne kadar çok fazla olduğunu ifade eder;  $\text{Exp}(\beta) < 1$  ise,  $1/\text{Exp}(\beta)$  hesaplanarak referans gruba göre bağımlı değişkenin ne kadar daha az görüldüğü anlaşılır. Pozitif bir katsayı, riskin diğer bir duruma göre daha büyük olduğunu gösterir. Lojistik modelin parametre tahminleri, bağımlı değişken üzerinde bağımsız değişkenlerin gösterdikleri etkilerinin yönünün anlaşılmasına yardımcı olur (Mok, Gan ve Sanyal, 2007; Evcim, Güneş ve Karaalp-Orhan, 2020).

Lojistik regresyon analizi, bağımsız değişkenlerin dağılımlarına yönelik herhangi bir varsayım şartı aramamaktadır, ancak bazı varsayımlar ve gereklilikler vardır (Dean, 1992; Tabachnick ve Fidell, 2020):

1. **Kayıp ve Uç Değerler:** Modelde kayıp ve uç değerler var ise incelenerek gerekli düzeltmeler yapılmalıdır. Tüm kategorik değişken çiftlerinde beklenen frekans 1'den yüksek olmalı ve 5'ten küçük olan gözeneğin sayısı %20'yi aşmamalıdır.
2. **Çoklu Doğrusal Bağlantı:** Oluşturulan lojistik regresyon modelinde yer alan ilgili bağımsız değişkenler arasında yüksek korelasyon ( $r > .90$ ) olmamalıdır. Bu sorunu çözmek için değişkenler faktör analizi ile gruplanabilir veya çoklu bağlantıya neden olan değişkenler modelden çıkarılabilir.
3. **Doğrusal İlişki Varsayımı:** Lojistik regresyon analizi, yordayıcı değişkenler ile bağımlı değişkenin logit değeri arasında lineer bir ilişki olduğunu öngörmektedir. Bu, yordayıcı ve log dönüşümü arasındaki ilişkinin istatistiksel olarak anlamlılığına bakılarak değerlendirilmektedir.
4. **Hataların Bağımsızlığı:** Hataların bağımsız olması gerekir. Aşırı yayılım kavramı, gözlenen varyansın beklenen varyansın daha yüksek olmasıdır ve bu, standart hataların küçülmesine ve regresyon modelindeki yordayıcıların güven aralıklarının daralmasına yol açabilir. Aşırı yayılım, Sapma ki kare istatistiğinin veya uyum iyiliği değerlerindeki Pearson istatistik değerinin serbestlik derecesine oranının hesaplanmasıyla tespit edilebilir. Bu oran 1'den büyükse aşırı yayılım vardır ve 2 veya daha yüksek değerler problem oluşturabilir.

### 3.2. Veri Seti

Sürdürülebilir kalkınma çerçevesinde dijitalleşme göstergelerinin çevre ekonomisi üzerinde etkisinin var olup olmadığının araştırıldığı ilgili çalışmada 2020 yılı verileri baz alınarak hesaplama yapılmıştır. Çevre ekonomisini temsil eden ve bağımlı değişken olan karbondioksit emisyonunun değerleri Dünya Bankası ([www.worldbankdata.com](http://www.worldbankdata.com)) veri tabanından alınmıştır. Ayrıca, araştırmanın evrenini oluşturan 131 ülke için basit ortalaması alınarak söz konusu bağımlı değişken lojistik regresyon analizi için kategorik hale getirilmiştir. Bu bağlamda ortalama değerden düşük olanlar "1" olarak, ortalamadan yüksek olanlar ise "2" olarak kodlanmıştır. Diğer yandan dijitalleşme göstergeleri olarak ise literatürde sıklıkla kullanılan Elektrik Erişim Oranı, Düzeltilmiş Tasarruflar: Enerji Tükenmesi (DT: Enerji Tükenmesi), GSYİH, İnternet Kullanımı ve Yenilenebilir Enerji Tüketimi değişkenleri seçilmiştir. Lojistik regresyon modelinde GSYİH'nın doğal logaritmasının ( $\ln(\text{GSYİH})$ ) kullanılmıştır. Böylelikle GSYİH değişkeninin dağılımını normalleştirir ve aşırı uç değerlerin etkisini azaltmaktadır. Bu durum, bağımlı değişken üzerindeki etkisinin daha anlamlı hale gelmesine yardımcı olmaktadır. Bahse konu değişkenler ve veri kaynakları Tablo 1'de özetlenmiştir.

**Tablo 1.** Araştırmada Kullanılan Değişken ve Kaynakları.

Değişken Adı	Birimi	Veri Tipi	Veri Tabanı	Yıl
Karbondioksit Emisyonu	Kiloton (kt)	Kategorik	Dünya Bankası	2020
Elektriğe Erişim	Nüfusun Yüzdesi	Sürekli	Dünya Bankası	2020
DT: Enerji Tükenmesi	GSMH'nın Yüzdesi	Sürekli	Dünya Bankası	2020
Gayrisafi Yurtiçi Hasıla (GSYİH)	Kişi Başına Düşen	Sürekli	Dünya Bankası	2020
İnternet Kullanımı	Nüfusun Yüzdesi	Sürekli	Dünya Bankası	2020
Yenilenebilir Enerji Tüketimi	Toplam Nihai Enerji Tüketiminin Yüzdesi	Sürekli	Dünya Bankası	2020

Tablo 1’de yer alan çalışmanın analizinde kullanılan bağımlı ve bağımsız değişkenlere ait değerler lojistik regresyon analizi ile IBM SPSS 22 paket programında hesaplanacaktır. Çalışmanın örneklemini oluşturan ülkeler ise Tablo 2’de yer almaktadır.

**Tablo 2.** Çalışma Evrenini Oluşturan Ülkeler.

Çalışma Grubuna Dahil Olan Ülkeler				
Afganistan	Danimarka	İsviçre	Lüksemburg	Polonya
Almanya	Ekvador	İtalya	Macaristan	Portekiz
Amerika Birleşik Devletleri	Endonezya	İzlanda	Madagaskar	Romanya
Arjantin	Ermenistan	Jamaika	Malavi	Rusya Federasyonu
Arnavutluk	Estonya	Japonya	Maldivler	Senegal
Avustralya	Etiyopya	Kamboçya	Malezya	Sırbistan
Avusturya	Fas	Kamerun	Mali	Singapur
Azerbaycan	Fildişi Sahili	Kanada	Malta	Slovak Cumhuriyeti
Barbados	Filipinler	Karadağ	Mauritius	Slovenya
Belarus	Finlandiya	Katar	Meksika	Sri Lanka
Belçika	Fransa	Kazakistan	Mısır, Arap Cum.	Sudan
Belize	Gabon	Kenya	Moğolistan	Suudi Arabistan
Benin	Gana	Kıbrıs	Moldova	Şili
Birleşik Arap Emirlikleri	Guatemala	Kırgız Cumhuriyeti	Moritanya	Tanzanya

**Tablo 3.** Çalışma Evrenini Oluşturan Ülkeler (devam).

Çalışma Grubuna Dahil Olan Ülkeler				
Birleşik Krallık	Güney Afrika	Kiribati	Mozambik	Tayland
Bolivya	Gürcistan	Kolombiya	Myanmar	Togo
Bosna Hersek	Haiti	Kongo, Dem. Temsilci	Namibya	Tunus
Botsvana	Hırvatistan	Kore Cum.	Nepal	Türkiye
Brezilya	Hindistan	Kosta Rika	Nijer	Uganda
Bulgaristan	Hollanda	Kuzey Makedonya	Nijerya	Ukrayna
Burkina Faso	Honduras	Lao PDR	Nikaragua	Umman
Burundi	İran, İslam Cum.	Lesotho	Norveç	Ürdün
Cabo Verde	İrlanda	Letonya	Özbekistan	Vietnam
Cezayir	İspanya	Liberya	Pakistan	Yeni Zelanda
Çekya	İsrail	Litvanya	Paraguay	Yunanistan
Çin	İsveç	Lübnan	Peru	Zambiya
				Zimbabve

Tablo 2’de yer alan ülkeler kullanılan değişkenler için ortak yıl 2020 olması sebebi ile seçilmiştir. Ülkemizin de içerisinde yer aldığı toplamda 131 ülke söz konusudur. Lojistik regresyon analizi için örneklem sayısı, bağımsız değişken sayısını  $k$  olarak kabul edersek, önerilen minimum örneklem sayısı  $N \geq 10 * k$  şeklinde hesaplanabilir (Peduzzi vd., 1996). Bu bağlamda, ilgili Çalışma için örneklem sayısının oldukça yeterli olduğu ve modelin tahmin performansı için kabul edilebilir olduğu söylenebilir.

#### 4. Uygulama

Lojistik regresyon modelinin kurulması safhasında SPSS paket program içeriğinde bulunan İkili (Binary) Lojistik Regresyon aracı kullanılmıştır. Bağımlı değişken olarak karbondioksit emisyonu ve modele yer alabileceği ön görülen bağımsız değişkenler kümesi “Enter” metodu seçilerek analiz gerçekleştirilmiştir. Lojistik regresyon modelinin anlamlılığına ilişkin elde edilen Bulgular Tablo 3’te yer almaktadır.

**Tablo 4.** Anlamlılık İstatistikleri- Omnibus testi.

	Ki-kare	Serbestlik derecesi	p-değeri
Adım	56,541	5	0,000
Block	56,541	5	0,000
Model	56,541	5	0,000

Omnibus testi, modelin istatistiksel olarak anlamlılığının sınanması için lojistik regresyon analizinde kullanılan bir testtir. Söz konusu teste ait hipotezler şöyle verilmektedir:

$$H_0: \beta_0 = \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad \text{Model istatistiksel olarak anlamlı değildir.}$$

$$H_1: \beta_0 \neq \beta_1 \neq \beta_2 \neq \dots \neq \beta_k \neq 0 \quad \text{Model istatistiksel olarak anlamlıdır.}$$

Tablo 3 incelendiğinde modele ilişkin anlamlılık değerleri %5’ten küçük olarak elde edilmiştir. Bu bağlamda, ulaşılan sonuca göre kurulan model istatistiksel olarak % 95 güven seviyesinde anlamlı bulunmuştur.

Tablo 4’te modelde kullanılan değerlere göre dijital göstergeler olan bağımsız değişkenlerin çevre ekonomisini temsil eden ve bağımlı değişken olan karbondioksit emisyonunu açıklama yüzdeleri yer almaktadır.

**Tablo 5.** Lojistik Regresyon Analizi ile Oluşturulan Modelin Özeti.

Adım	-2 Log likelihood	Cox & Snell $R^2$	Nagelkerke $R^2$
1	115,598	0,351	0,479

Tablo 4’e göre dijital göstergelerin, bağımlı değişken olan karbondioksit emisyonunu açıklama yüzdesi Cox & Snell  $R^2$  ‘ye göre %35 iken Nagelkerke  $R$  ‘ye göre ise %48’dir. Bu iki değer için literatürde yer alan ve sosyal bilimler alanı için kabul edilebilir bir model performansı gösteren eşik değerlerinin minimum 0,2-0,4 arasında olmasıdır (Cox & Snell, 1989; Nagelkerke, 1991). Bu doğrultuda, elde edilen sonuçların kabul edilebilir olduğu gözlemlenmiştir.

Araştırmada kullanılan verilerin uyumunu gösteren Hosmer ve Lemeshow Uyum İyiliği Testi Tablo 5’te yer almaktadır.

**Tablo 6.** Hosmer ve Lemeshow Uyum İyiliği Testi.

Adım	Ki-Kare	Serbestlik Derecesi	p-değeri
1	7,961	8	0,479

Fosmer ve Lemeshow Uyum İyiliği Testi için yokluk hipotezi  $H_0$ : *Verilerin uyumluluğu sağlanmıştır* şeklinde iken alternatif hipotez ise  $H_1$ : *Verilen uyumluluğu sağlanmamıştır* şeklinde ifade edilebilir. Bu bağlamda Tablo 5 değerlendirildiğinde, olasılık değerinin 0,479 olup 0,05 değerinden büyük olduğundan istatistiksel olarak anlamsız olduğu gözlemlenmiştir. Bu durumda, verilerin uyumsuz olduğunu söyleyen alternatif hipotez reddedilmiştir. Diğer bir ifade ile araştırmada kullanılan veri setinin uyum iyiliği gerçekleştirilmiştir.

Kurulan lojistik regresyon modelinin sınıflama gücüne ilişkin elde edilen bulgular Tablo 6’da yer almaktadır.

**Tablo 6.** Lojistik Regresyon Modelinin Doğru Sınıflandırma Yüzdeleri.

		Tahmin			
		Karbondioksit Emisyonu			
	Gözlem	Düşük	Yüksek	<b>Toplam</b>	Doğru Sınıflama Yüzdesi
Karbondioksit Emisyonu	Düşük	71	12	<b>83</b>	85,5
	Yüksek	16	32	<b>48</b>	66,7
	<b>Toplam</b>	<b>87</b>	<b>24</b>		<b>78,6</b>

Tablo 6’da lojistik regresyon modelinin doğru sınıflama oranları görülmektedir. Bu bağlamda model yorumlanmak istendiğinde, karbondioksit emisyonu düşük olan ülkelerin % 85,5’i, yüksek olan ülkelerin % 66,7’si doğru tahmin edilmiştir. Genel olarak modelde kullanılan örnekleme oluşturan ülke grubu için % 79 doğru tahmin yapılmıştır. Yani ikili (binary) lojistik regresyon modeli, gözlenen 131 ülkenin % 79’unu (yaklaşık 104’ünün) düşük ya da yüksek karbondioksit emisyonuna sahip olup olmadığını doğru tahmin etmiştir.

Bağımlı değişken olan karbondioksit emisyonu üzerinde etkisi olan değişkenlere yönelik lojistik regresyon analizi bulgularına Tablo7’de yer verilmiştir. Tablo 7’de, modele dahil edilen bağımsız değişkenlerin (elektriğe erişim oranı, enerji tüketimi, gayrisafı yurtiçi hasıla (GSYİH), internet kullanımı ve yenilenebilir enerji tüketimi) katsayı öngörülere, öngörülerin standart hataları (SE( $\beta$ )), Wald istatistiği, Wald istatistiğinin serbestlik derecesi (s.d.), %95 güven aralığında olasılık oranlarının aralık tahminleri ve olasılık oranı (Exp( $\beta$ )) sunulmaktadır. Lojistik regresyon analizinde, Wald test istatistiği yaygın olarak bağımsız değişkenlerin regresyon katsayılarının anlamlılığını değerlendirmek için kullanılır. Bu test, her bir logit katsayısının sifıra eşit olup olmadığını belirlemek için sifir hipotezini test eder (Menard, 2002). Wald test istatistiği, lojistik regresyonda  $\beta$  katsayısının anlamlılığını test etmek amacıyla kullanılır ve şu şekilde ifade edilir (Çokluk, 2010: 1375):

$$W = \left( \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \text{ burada } i = 1, 2, \dots, k \quad (3)$$

Tabloya dahil edilen  $\text{Exp}(\beta)$  ise üstel lojistik regresyon katsayılarını ifade etmektedir.

**Tablo 7.** İki Kategorili Lojistik Regresyon Sonuçları.

Değişkenler	$\beta$	SE( $\beta$ )	Wald	s.d.	p-değeri	Exp( $\beta$ )	95% C.I.for EXP(B)	
							Alt	Üst
Elektriğe Erişim	0,016	0,017	0,939	1	0,333	1,016	0,984	1,050
DT: Enerji Tükenmesi	0,537	0,208	6,677	1	<b>0,010*</b>	1,712	1,139	2,573
İnternet Kullanımı	0,010	0,027	0,148	1	0,700	1,010	0,959	1,065
Yenilenebilir Enerji Tüketimi	-0,056	0,016	12,195	1	<b>0,000*</b>	0,946	0,916	0,976
lnGSYİH	-1,458	0,414	12,367	1	<b>0,000*</b>	0,233	0,103	0,525
Sabit	11,384	2,907	15,337	1	<b>0,000*</b>	87893,124		

\*%5 seviyesinde istatistiksel olarak anlamlıdır.

Tablo 7 değerlendirildiğinde DT: Enerji Tükenmesi, Yenilenebilir Enerji Tüketimi ve lnGSYİH değişkenlerinin çevre ekonomisini temsil eden karbondioksit emisyonu bağımlı değişkeni üzerinde istatistiksel olarak anlamlı bir etkisi olduğu ( $p < 0,05$ ) gözlemlenmiştir. Diğer yandan Elektriğe Erişim ve İnternet Kullanımı değişkenlerinin Karbondioksit Emisyonu üzerinde etkisinin anlamsız olduğu saptanmıştır ( $p > 0,05$ ). Bu bağlamda, elde edilen lojistik regresyon modelinin son hali şu şekildedir:

*Karbondioksit Emisyonu*

$$= 11,384 - 1,458(\ln\text{GSYİH}) - 0,056(\text{Yenilenebilir Enerji Tüketimi}) + 0,537(\text{DT: Enerji Tükenmesi})$$

Tablo 7’de,  $\text{Exp}(\beta)$  sütunu her bir bağımsız değişken için hesaplanan olasılık oranlarını ifade etmektedir. Olasılık oranlarının yorumlanması,  $\beta$  katsayılarının yorumlanmasından çoğunlukla daha basit ve kolay olduğu belirtilmektedir.  $\text{Exp}(\beta)$  değerlerinin yorumlanması şu şekildedir: Eğer olasılık oranı 1’den büyükse, bu ilgili sonucun gerçekleşme olasılığının arttığını; 1’den küçükse, olasılığın azaldığını gösterir. Bu doğrultuda, Tablo 7’de anlamlı olarak bulunan değişkenler için aşağıdaki yorumlar yapılabilir (Hosmer ve Lemeshow, 2000).

Tablo 7’ye göre lnGSYİH değişkeninin  $\text{Exp}(\beta)$  değeri 0,233 olarak elde edilmiştir. Bu değer, GSYİH’nın doğal logaritmasındaki her bir birimlik artışın, yüksek karbon emisyonu (yani 2 kodlu) olasılığını önemli ölçüde azaltacağını göstermektedir. Olasılık oranı 1’den küçük olduğu için, ln(GSYİH) yükseldikçe yüksek karbon emisyonu olasılığı yaklaşık %76,7 oranında azalmaktadır ( $1 - 0,233 = 0,767$ ). Bu, ekonomik büyümenin yüksek karbon emisyonları üzerinde önemli bir azaltıcı etkisi olduğunu göstermektedir. Elde edilen negatif  $\beta$  (-1,458) değeri, GSYİH’nın doğal logaritmasındaki her bir birimlik artışın, yüksek karbon emisyonu olasılığında azaltıcı etkiye yol açtığını ortaya koymaktadır. Matematiksel olarak ifade etmek gerekirse, bu değer yüksek karbon emisyonu olasılığının azaldığını ifade eder.

Anlamlı çıkan ikinci bağımsız değişken Yenilenebilir Enerji Tüketimi değişkeninin  $\text{Exp}(\beta)$  değeri = 0,946’dır. Elde edilen bu değer, ikinci anlamlı çıkan bağımsız değişken olan yenilenebilir enerji tüketiminde meydana gelen her bir birimlik artışın, yüksek karbon emisyonu (yani 2 kodlu) olasılığını yaklaşık %5,4 oranında ( $1 - 0,946 = 0,054$ ) kadar azalttığını göstermektedir. Bu, örneklemdaki 131 ülke için yenilenebilir enerji tüketiminin yüksek karbon emisyonları üzerindeki etkisinin negatif olduğunu ifade etmektedir.  $\beta$  katsayısının aynı lnGSYİH bağımsız değişkeni gibi negatif olması (-0,056), yenilenebilir enerji tüketimindeki her bir birimlik artışın yüksek karbon emisyonu olasılığını azalttığını ifade eder. Ancak, elde edilen bu etkinin büyüklüğü sınırlı olup yani  $\text{Exp}(\beta)$  değeri 1’e yakın olduğu için yenilenebilir enerji tüketimindeki artışın yüksek karbon emisyonu olasılığını önemli derecede azaltmadığı gözlemlenmiştir.

Tablo 7’de anlamlı olarak elde edilen üçüncü bağımsız değişken olan DT: Enerji Tükenmesi  $\text{Exp}(\beta)$  değeri 1,016 olarak elde edilmiştir. DT: Enerji Tükenmesi değişkenindeki her bir birimlik artış, yüksek karbon emisyonu (yani 2 kodlu) olasılığını yaklaşık %1,6 ( $1,016 - 1 = 0,016$ , yani %1,6 artış) oranında artırmaktadır.  $\beta$  katsayısının 0,017 olması, bu artışın her bir birimlik artış için küçük bir etki yarattığını gösterir. Bu sonuç, analiz edilen 131 ülke arasında, enerji kaynaklarının tükenmesiyle ilgili tasarrufların artmasının yüksek karbon emisyonları olasılığında küçük ama ölçülebilir bir artışa neden olduğunu göstermektedir.

## 5. Sonuç ve Tartışma

Küresel ölçekte çevresel sorunlar ve sürdürülebilir kalkınma hedefleri giderek daha fazla önem kazanmaktadır. Bu bağlamda, çeşitli ekonomik ve çevresel göstergelerin etkilerini anlamak, ülkelerin sürdürülebilir kalkınma stratejilerini şekillendirmede kritik bir rol oynamaktadır. Ekonomik büyüme, enerji tüketimi ve dijitalleşme gibi faktörler, çevresel sonuçlarla etkileşim içinde bulunarak hem ülke ekonomileri hem de çevresel sürdürülebilirlik açısından önemli sonuçlar doğurmaktadır. Bu çalışma, bu faktörlerin çevresel etkilerinin ne şekilde şekillendiğini ve hangi alanlarda iyileştirme yapılması gerektiğini anlamaya yönelik değerli bilgiler sunmaktadır. Bu bilgilerin, politika yapıcılar ve strateji geliştiriciler için yol gösterici olması beklenmektedir.

Bu çalışmanın temel amacı, sürdürülebilir kalkınma perspektifinden çevre ekonomisini etkileyen faktörleri analiz etmek ve dijitalleşmenin bu faktörler üzerindeki rolünü değerlendirmektir. Dijital göstergelerin çevre ekonomisi üzerinde etkisi olup olmadığının araştırıldığı bu çalışmada, çok değişkenli istatistiksel analizlerden olan ikili lojistik regresyon analizi kullanılmıştır. Çalışmanın çevre ekonomisini temsil eden göstergesi ve bağımlı değişkeni Karbondioksit Emisyonu seçilmiştir. Lojistik regresyon analizinde bağımlı değişkenin kategorik tipte veri yapısına sahip olması gerektiği için çalışmanın örneklemini oluşturan 131 ülkenin Karbondioksit Emisyon değerlerinin basit ortalaması alınarak, ortalamanın aşağısında olan değerler Düşük Karbon Emisyonuna sahip, ortalamanın üstünde olan değerler ise Yüksek Karbon Emisyonuna sahip olduğu belirtilerek analize alınmıştır. Dijital göstergeler olarak Elektriğe Erişim Oranı, DT: Enerji Tükenmesi, GSYİH, İnternet Kullanımı ve Yenilenebilir Enerji Tüketimi değişkenleri seçilmiştir. Lojistik regresyon modelinde GSYİH'nın doğal logaritmasının ( $\ln(\text{GSYİH})$ ) kullanılmıştır. Bağımlı ve bağımsız değişkenlere yönelik tüm değerler 2020 yılına ait olup ilgili değişkenlere ait veriler Dünya Bankası veri tabanından ulaşılmıştır.

Çalışmanın sonucunda, lojistik regresyon modeli kullanılarak çeşitli dijital göstergelerin ve ekonomik değişkenlerin karbondioksit emisyonları üzerindeki etkileri incelenmiştir. Modelin genel anlamlılığı, Omnibus testinden elde edilen p-değerinin 0,000 olarak bulunmasıyla %95 güven seviyesinde istatistiksel olarak anlamlı bulunmuştur. Ayrıca, modelin açıklayıcılığı Cox & Snell  $R^2$  değeri %35 ve Nagelkerke  $R^2$  değeri %48 olarak hesaplanmıştır, bu da modelin kabul edilebilir bir performans sergilediğini göstermektedir. Hosmer ve Lemeshow Uyum İyiliği Testi sonuçlarına göre, verilerin modelle uyumlu olduğu gözlemlenmiştir ( $p > 0,05$ ). Modelin sınıflama gücü ise %79 olarak belirlenmiş olup, bu oran karbondioksit emisyonu düşük ya da yüksek olan ülkelerin doğru tahmin edildiğini ortaya koymaktadır.

Bu çalışmada elde edilen bulgular, bağımsız değişkenlerin karbondioksit emisyonları üzerindeki etkilerini detaylı bir şekilde ortaya koymaktadır. Özellikle, GSYİH'nın doğal logaritması ( $\ln(\text{GSYİH})$ ), yenilenebilir enerji tüketimi ve enerji tıkanması değişkenlerinin karbondioksit emisyonları üzerinde istatistiksel olarak anlamlı etkiler yarattığı görülmüştür. GSYİH'daki artışın yüksek karbon emisyonlarıyla negatif bir ilişki içinde olduğu ve bu artışın yüksek karbon emisyonu olasılığını %76,7 oranında azalttığı bulunmuştur. Bu, ekonomik büyümenin karbon emisyonlarını azaltıcı etkisini vurgular ve yüksek GSYİH seviyelerinin çevresel sürdürülebilirliği destekleyebileceğini ima eder. Yani, ekonomik büyümenin karbon emisyonları üzerinde belirgin bir azaltıcı etkisi vardır. GSYİH'nın karbon emisyonlarını azaltıcı etkisi, ekonomik büyümenin çevresel sürdürülebilirlikle nasıl etkileşime girdiğini gösterir. Genellikle ekonomik büyüme enerji tüketiminde artış ve dolayısıyla karbon emisyonlarında yükselmeye yol açar (Stern, 2004; Dietz ve Rosa, 1997). Stern (2004), ekonomik büyümenin enerji tüketimini artırarak karbon emisyonlarını yükselttiğini belirtirken, Dietz ve Rosa (1997) ekonomik gelişmenin çevresel etkilerinin artış gösterdiğini vurgulamaktadır. Ancak, bu çalışmada GSYİH artışının yüksek karbon emisyonlarıyla negatif ilişki içinde olduğu ve emisyon olasılığını azalttığı bulunmuştur. Bu durum, ekonomik büyümenin belirli bir seviyeden sonra çevresel sürdürülebilirliği destekleyebileceğini gösterir. Yüksek teknoloji ve enerji verimli sektörlerin büyümesi, bu durumu açıklayabilir; bu sektörler, ekonomik büyüme sağlarken daha az karbon emisyonu üretme eğilimindedir. Bu bulgu, Panayotou'nun (1997) çevresel Kuznets eğrisi teorisini destekler. Panayotou, ekonomik büyümenin belirli bir noktadan sonra çevresel iyileşmeler sağladığını öne sürer. Stern (2004), ekonomik büyümenin çevresel sürdürülebilirlik üzerindeki etkilerini incelediği çalışmasında, büyümenin karbon emisyonları üzerindeki etkilerinin zamanla azaldığını ve yüksek büyüme oranlarının çevresel iyileşmelere yol açabileceğini belirtmiştir. Miçooğulları (2022), iki farklı ülke grubunu ele alarak yapmış olduğu çalışmada ekonomik büyümenin karbon emisyonunu azalttığı yönünde bulgulara ulaşmıştır. Bu literatür bulguları, yüksek teknoloji ve enerji verimli sektörlerin ekonomik büyüme ile karbon emisyonlarını azaltmada önemli bir rol oynadığını vurgular.

Yenilenebilir enerji tüketiminin  $\text{Exp}(\beta)$  değeri 0,946 olarak hesaplanmış olup, bu değişkenin yüksek karbon emisyonu olasılığını %5,4 oranında azalttığı gözlemlenmiştir. Bu bulgu, yenilenebilir enerji tüketiminin karbon emisyonlarını azaltma yönündeki etkisini doğrular, ancak etkinin daha büyük bir etki yaratması için ek stratejilerin gerekli olabileceğini de ima eder. Literatürde elde edilen sonucu destekleyen çalışmalar mevcuttur (Mete, 2020). Bucak (2021) çalışmasında, üretim safhasında kullanılan yenilenebilir enerji kaynaklarının karbon emisyonunu azaltarak çevrenin korunmasında rol aldığını ayrıca bu şekilde dış pazarlara yönelik bağımlılığı da azalttığı sonucuna ulaşmıştır.

DT: Enerji Tükenmesi değişkeninin  $Exp(\beta)$  değeri ise 1,016'dır, bu da yüksek karbon emisyonu olasılığında %1,6 oranında küçük bir artışa neden olduğunu göstermektedir. Bu durum, enerji tasarruflarındaki artışın karbon emisyonları üzerindeki etkisinin nispeten küçük olduğunu, fakat ölçülebilir olduğunu ortaya koyar. Özellikle, bu etki ülkeler arasında görece küçük bir artış olarak gözlemlenmiştir ve bu, enerji tasarruflarının yüksek karbon emisyonları üzerinde belirgin bir etki yaratmadığını veya etkisinin sınırlı olduğunu gösterebilir. Literatürde Hussain, Khan ve Zhou'nun (2020) çalışmasında enerji tüketiminin karbon emisyonlarını artırdığına dair bulgular ile Asongu ve Odhiambo'nun (2021) çalışmalarında enerji tüketimi ve doğal kaynak tükenmesinin çevresel bozulmaya ve dolayısıyla karbon emisyonlarının artmasına neden olduğunu gösteren bulgular, elde edilen sonuçları desteklemektedir. Bu çalışmalar, enerji tüketimi ve doğal kaynak kullanımının çevresel etkilerini ortaya koyarak, çevre politikalarının ve sürdürülebilir enerji yönetimi stratejilerinin önemini vurgulamaktadır.

Sonuç olarak, ekonomik büyüme ve yenilenebilir enerji tüketiminin yüksek karbon emisyonları üzerinde azaltıcı etkileri olduğu görülmüştür. Buna karşın, DT: Enerji Tükenmesi gibi faktörlerin etkisi daha sınırlı kalmıştır. Bu bulgular, sürdürülebilir kalkınma politikalarının belirlenmesinde ve çevresel stratejilerin oluşturulmasında dikkate alınması gereken önemli göstergelerdir. Enerji tasarruflarını artıran politikaların, karbon emisyonları üzerindeki etkilerini daha iyi anlamak için ek veriler ve analizler gerekebilir. Ekonomik büyüme ayrıca enerji verimliliği projeleri ve yenilenebilir enerji yatırımlarına daha fazla kaynak ayrılmasını teşvik edebilir.



## Kaynaklar

- Akyol, H., Akar, T., Akar, G. (2023). Dijitalleşme ve Ekonomik Büyüme: AB Ülkelerinden Yeni Kanıtlar. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26(49), 99-114.
- Al, İ. (2019). Sürdürülebilir Kalkınma ve Yeşil Ekonomi: Türkiye için Bir Endeks Önerisi. *Hitit Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 12(1), 112-124.
- Alpar, R. (2011). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*. 3. Baskı, Ankara: Detay Yayıncılık.
- Asongu, S. A., Odhiambo, N. M. (2021). Inequality, Finance and Renewable Energy Consumption in Sub-Saharan Africa. *Renewable Energy*, 165, 678-688.
- Atatanır, H. (2022). Dijitalleşme: İş yaşamında ve Sosyal Güvenlikte. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 12(Dijitalleşme Özel Sayısı), 77-92.
- Birleşmiş Milletler Türkiye (2024). Sürdürülebilir Kalkınma Amaçları, <https://turkiye.un.org/tr/sdgs>, Erişim Tarihi: 19.07.2024.
- Bucak, Ç. (2021). AB15 ülkelerinde ve Türkiye’de Ekonomik Karmaşıklık Endeksi, İnsani Gelişme Endeksi ve Karbon Emisyonu: Panel Veri Analizi. *Ege Stratejik Araştırmalar Dergisi*, 12(1), 71-88.
- Büyüköztürk, Ş. (2018). *Sosyal Bilimler İçin Veri Analizi El Kitabı*, 24. Baskı, Ankara: Pegem Akademi Yayıncılık.
- Cox, D. R., Snell, E. J. (1989). *The Analysis of Binary Data*. Chapman and Hall/CRC.
- Çayrağası, F., Sakıcı, Ş. (2021). Avrupa Yeşil Mutabakatı (Green Deal) ve Birleşmiş Milletler Sürdürülebilir Kalkınma Hedefleri Perspektifinde Sürdürülebilir Dijital Pazarlama Stratejileri. *Gaziantep University Journal of Social Sciences*, 20(4), 1916-1937.
- Çokluk, Ö. (2010). Lojistik Regresyon Analizi: Kavram ve Uygulama. *Kuram ve Uygulamada Eğitim Bilimleri*, 10(3), 1357-1407.
- Dean, C. B. (1992). Testing for Overdispersion in Poisson and Binomial Regression Models. *Journal of American Statistical Association*, 87(418), 451-457.
- Demirkıran, S., Beyoğlu, A., Terzioğlu, M. K., Yaşar, A. (2022). Sürdürülebilir Kalkınma Odaklı Dijitalleşme Belirleyicilerinin Verimlilik Üzerindeki Etkilerinin Yapay Sinir Ağları ile Sınıflandırılması. *Verimlilik Dergisi*, Dijitale Dönüşüm ve Verimlilik Özel Sayısı, 30-47.
- Dietz, T., Rosa, E. A. (1997). Effects of Population and Affluence on Atmospheric Carbon Dioxide Emissions. *Proceedings of the National Academy of Sciences*, 94(1), 175-179.
- Driscoll, J. C., Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80(4), 549-560.
- Erkılıç, C. E., Aksoy, A. (2020). Hastanelerde Finansal Başarısızlık Tahmini: Lojistik Regresyon Modeli ile Kamu Hastaneleri Üzerine Bir Uygulama. *İşletme Araştırmaları Dergisi*, 12(2), 1415-1433.
- Evcim, N., Güneş, S., Karaalp-orhan, H. S. (2020). Türkiye’de Hanehalkı Göreli Yoksulluğunu Etkileyen Faktörler: Lojistik Regresyon Analizi. *Sosyoekonomi*, 28(43), 11-32.
- Hosmer, D. W., Lemeshow, S. (2000). *Applied Logistic Regression*. Wiley.
- Hussain, J., Khan, A., Zhou, K. (2020). The Impact of Natural Resource Depletion on Energy Use and CO2 Emission in Belt & Road Initiative Countries: A Cross-Country Analysis. *Energy*, 199, 117409.
- Jovanović, M., Dlačić, J., Okanović, M. (2018). Digitalization and Society’s Sustainable Development—Measures and Implications. *Zbornik radova Ekonomskog fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu*, 36(2), 905-928.

- Atf / Citation:** ÖRK ÖZEL, S., VEYSİKARANI, D. (2024). Sürdürülebilir Kalkınma Kapsamında Dijitalleşme ve Çevre Ekonomisi İlişkisinin Analizi, *İstatistik Araştırma Dergisi*, 14 (2), 14-29.
- Konu, A. (2020). Sürdürülebilir Kalkınma ve Dijital Ekonomi İlişkisi: AB Ülkeleri İçin Bir Araştırma. *Akademik İncelemeler Dergisi*, 15(2), 655-678.
- Kuvvetli Yavaş, H. (2024). Sürdürülebilir Kalkınma ve Dijitalleşmeyle Birlikte Ortaya Çıkan Meydan Okumalar. *AJIT-e: Academic Journal of Information Technology*, 15(2), 138-173.
- Lee, C. C., He, Z. W., Yuan, Z. (2023). A Pathway to Sustainable Development: Digitization and Green Productivity. *Energy Economics*, 124, 106772.
- Marangoz, M., Önce, A. G., Aydın, A. E. (2015). Çevre Ekonomisi ve Sürdürülebilir Kalkınma Açısından E-Atık Yönetiminin Önemi. *International Conference on Eurasian Economies*, 654-662.
- Menard, S. (2002). *Applied Logistic Regression Analysis*. Sage Publications.
- Metek, E. (2020). Sürdürülebilir Kalkınma Kapsamında Yeşil Lojistik: Avrupa Birliği ve Türkiye Örneği. *Karadeniz Sosyal Bilimler Dergisi*, 12(23), 383-396.
- Miçooğulları, S. A. (2022). Yüksek Kurumsal Kalite ve Düşük Karbon Emisyonu ile Yüksek Büyüme Mümkün Mü?. *İzmir İktisat Dergisi*, 37(4), 849-869.
- Mok, T.Y., Gan, C., Sanyal, A. (2007). The Determinants of Urban Household Poverty in Malaysia. *Journal of Social Sciences*, 3(4), 190-196.
- Mondejar, M. E., Avtar, R., Diaz, H. L. B., Dubey, R. K., Esteban, J., Gómez-Morales, A., Garcia-Segura, S. (2021). Digitalization to Achieve Sustainable Development Goals: Steps Towards a Smart Green Planet. *Science of The Total Environment*, 794, 148539.
- Mutlu, A. (2007). Sürdürülebilir Kalkınma ve Çevre Muhasebesi (II). *Muhasebe ve Finansman Dergisi*, (34), 162-173.
- Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determination. *Biometrika*, 78(3), 691-692.
- Omer, A. M. (2008). Energy, Environment and Sustainable Development. *Renewable and sustainable energy reviews*, 12(9), 2265-2300.
- Ozmehmet, E. (2008). Dünyada ve Türkiye Sürdürülebilir Kalkınma Yaklaşımları. *Yaşar Üniversitesi E-Dergisi*, 3(12), 1853-1876.
- Özaslan, A. (2023). Sürdürülebilir Kalkınma Yolunda Yeşil Büyüme Stratejilerinin Kentlere Yansıması. *Sağlık ve Sosyal Refah Araştırmaları Dergisi*, 5(1), 98-113.
- Panayotou, T. (1997). Demystifying the Environmental Kuznets Curve: Turning A Black Box Into a Policy Tool. *Environment and Development Economics*, 2(4), 465-484.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., Feinstein, A. R. (1996). A Simulation Study of The Number of Events Per Variable in Logistic Regression Analysis. *Journal of Clinical Epidemiology*, 49(12), 1373-1379.
- Polasky, S., Kling, C. L., Levin, S. A., Carpenter, S. R., Daily, G. C., Ehrlich, P. R., Heal, G. M., Lubchenco, J. (2019). Role of Economics in Analyzing The Environment and Sustainable Development. *Proceedings of the National Academy of Sciences*, 116(12), 5233-5238.
- Sazanova, S. L. (2021). Socio-economic Ecosystems, Sustainable Economic Development and Digitalization of The Economy. In *Socio-economic Systems: Paradigms for the Future* (pp. 799-808). Cham: Springer International Publishing.
- Stern, D. I. (2004). The Rise and Fall of The Environmental Kuznets Curve. *World Development*, 32(8), 1419-1439.
- Şenel, S., Alatlı, B. (2014). Lojistik Regresyon Analizinin Kullanıldığı Makaleler Üzerine Bir İnceleme. *Journal of Measurement and Evaluation in Education and Psychology*, 5(1), 35-52.

**Atıf / Citation:** ÖRK ÖZEL, S., VEYSİKARANI, D. (2024). Sürdürülebilir Kalkınma Kapsamında Dijitalleşme ve Çevre Ekonomisi İlişkisinin Analizi, *İstatistik Araştırma Dergisi*, 14 (2), 14-29.

Tabachnick, B. G., Fidell, L. S. (2020). *Using Multivariate Statistics*. New York, USA: Harper Collins College Publishers.

Tabakan, G., Avcı, O. (2021). Vergiye Gönüllü Uyumu Etkileyen Faktörlerin Lojistik Regresyon Analizi ile Belirlenmesi. *Sosyoekonomi*, 29(48), 541-561. <https://doi.org/10.17233/sosyoekonomi.2021.02.25>

Tıraş, H. H. (2012). Sürdürülebilir Kalkınma ve Çevre: Teorik Bir İnceleme. *Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 2(2), 57-73.

Toprak, D. (2006). Sürdürülebilir Kalkınma Çevresinde Çevre Politikaları ve Mali Araçlar. *Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2(4), 146-169.

Yalçın, A. Z. (2010). Sürdürülebilir Kalkınma için Düşük Karbon Ekonomisinin Önemi ve Türkiye için Bir Değerlendirme. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 13(24), 186-203.

Yeni, O. (2014). Sürdürülebilirlik ve Sürdürülebilir Kalkınma: Bir Yazın Taraması. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 16(3), 181-208.

Yerdelen Kaygın, C., Tazegül, A., Yazarkan H. (2016). İşletmelerin Finansal Başarılı ve Başarısız Olma Durumlarının Veri Madenciliği ve Lojistik Regresyon Analizi ile Tahmin Edilebilirliği. *Ege Akademik Bakış*, 16(1), 147-159.

Yeşilyaprak, M. (2023). Yeşil Lojistiğin E-İhracattaki Yeri ve Önemi. *Beykoz Akademi Dergisi*, 11(1), 248-265.

Yılmaz, V. (2018). Sürdürülebilir Kalkınma ve Yeşil Büyüme Arasındaki İlişki. *Journal of International Management Educational and Economics Perspectives*, 6(2), 79-89.

## Avrupa Birliği Ülkeleri ve Türkiye’de Doğrudan Yabancı Sermaye Yatırımları ile Gayri Safi Yurtiçi Hasıla Arasındaki Panel Veri Nedensellik Analizi

Turgut ÜN

Marmara Üniversitesi / Dr.Öğr. Üy.  
Ekonometri Bölümü, [turgutun@marmara.edu.tr](mailto:turgutun@marmara.edu.tr)  
Orcid: 0000-0002-7660-1510

Özlem TAŞSEVEN

Marmara Üniversitesi / Prof. Dr.  
Ekonometri Bölümü, [ozlem.tasseven@marmara.edu.tr](mailto:ozlem.tasseven@marmara.edu.tr)  
Orcid: 0000-0002-3759-5074

Dilek ALTAŞ KARACA

Marmara Üniversitesi / Prof. Dr.  
Ekonometri Bölümü, [dilekaltas@marmara.edu.tr](mailto:dilekaltas@marmara.edu.tr)  
Orcid:0000-0001-5103-9018

### Özet

Bu çalışma, Doğrudan Yabancı Sermaye Yatırımlar (DYY) ile Gayri Safi Yurtiçi Hasıla (GSYİH) arasındaki nedensellik ilişkisini, Avrupa Birliği (AB) ülkeleri ve Türkiye’yi kapsayan 2010-2022 dönemine ait panel veriler üzerinde analiz etmiştir. Çalışmanın temel amacı, DYY ve GSYİH arasındaki ilişkiyi hem bölgesel hem de ekonomik farklılıklar bağlamında derinlemesine incelemek ve bu iki değişken arasındaki nedensellik ilişkilerinin dinamiklerini ortaya koymaktır. Çalışmada, ekonomik yapılarındaki benzerliklere göre ülkeler kümelerine ayrılmış ve analizler bu kümeler çerçevesinde gerçekleştirilmiştir. Kümeleme anlamlı ve spesifik sonuçlar elde etmeyi hedeflemiştir. Kümeler içindeki eğimlerin benzerliği analizi sonucu kümelerin homojen eğimlere sahip oldukları sonucuna ulaşılmıştır. Çalışmanın küme bazında gerçekleştirilmesi sonuçların sağlamlığını artıran bir faktör olmuştur. Nedensellik ilişkilerinin belirlenmesi amacıyla gerçekleştirilen analizler, iki grup arasında çarpıcı farklılıklar olduğunu ortaya koymuştur. Bu durum, birinci küme ülkelerinin ekonomik büyüklüğü artırma odaklı politikalarının, yabancı yatırımcıları çekme konusunda daha etkili olduğunu düşündürmektedir. Öte yandan, ikinci kümede yapılan analizlerde doğrudan yabancı yatırımların ekonomik büyüklüğü artırıcı yönünde etkilediği teşvik ettiği gibi ekonomik büyüklüğün de yabancı yatırımları artıran bir faktör olduğunu göstermektedir. Çalışmanın bulguları, politika yapıcılar için önemli çıkarımlar sunmaktadır. Yabancı yatırımları teşvik etmeye yönelik politikaların, her ülkenin kendine özgü ekonomik özellikleri ve dinamikleri dikkate alınarak tasarlanması gerektiği vurgulanmaktadır. Bu çalışma, değişkenler arasındaki ilişkilerin bölgesel farklılıklarına ışık tutarak, gelecekte yapılacak çalışmalara metodolojik bir rehberlik sağlamayı hedeflemektedir.

*Anahtar sözcükler: Doğrudan Yabancı Sermaye Yatırımları, Gayri Safi Yurt İçerisi, Panel Veri Analizi, Panel Nedensellik.*

**Sorumlu Yazar / Corresponding Author:** Dilek Altaş Karaca, Marmara Üniversitesi.

**Atf / Citation:** ÜN T., TAŞSEVEN Ö., KARACA D. A. (2024). Avrupa Birliği Ülkeleri ve Türkiye’de Doğrudan Yabancı Sermaye Yatırımları ile Gayri Safi Yurtiçi Hasıla Arasındaki Panel Veri Nedensellik Analizi. *İstatistik Araştırma Dergisi*, 14 (2), 30-45.

## Panel Data Causality Analysis in the European Union And Turkey Between Foreign Direct Investments And Gross Domestic Product

### Abstract

This study analyzed the causal relationship between Foreign Direct Investments (FDI) and Gross Domestic Product (GDP) on panel data for the period 2010-2022 covering the European Union (EU) countries and Turkey. The main purpose of the study is to examine the relationship between FDI and GDP in depth in the context of both regional and economic differences and to reveal the dynamics of the causal relationships between these two variables. In the study, countries were divided into clusters according to the similarities in their economic structures and the analyses were carried out within the framework of these clusters. Clustering aimed to obtain meaningful and specific results. As a result of the similarity analysis of the slopes within the clusters, it was concluded that the clusters had homogeneous slopes. Conducting the study on a cluster basis was a factor that increased the robustness of the results. The analyses carried out to determine the causal relationships revealed striking differences between the two groups. This suggests that the growth-oriented policies of the first cluster countries are more effective in attracting foreign investors. On the other hand, the analyses conducted in the second cluster show that foreign direct investments encourage economic growth and growth is also a factor that increases foreign investments. The findings of the study offer important implications for policy makers. It is emphasized that policies to encourage foreign investments should be designed by taking into account the specific economic characteristics and dynamics of each country. This study aims to provide methodological guidance for future studies by shedding light on regional differences in the relationships between variables.

*Keywords: Foreign Direct Investment, Gross Domestic Product, Panel Data Analysis, Panel Causality.*

### 1. Giriş

Doğrudan yabancı sermaye yatırımları (DYY), ekonomik büyüme literatüründe önemli bir çalışma alanı olmuştur. Gelişmekte olan ülkeler sürdürülebilir ekonomik büyümeyi gerçekleştirmek için dış sermayeye ihtiyaç duyarlar. Bu sebeple doğrudan yabancı sermaye oluşumu ekonomik büyüme, istihdam, teknoloji transferi, serbest piyasa rekabeti ve beşeri sermaye stokunu arttırabilmek amacıyla önemli bir kaynak olarak görülmektedir. Gelişmekte olan ekonomiler özellikle teknoloji göçü, iş fırsatları ve uluslararası pazara giriş yolu sağladıkları için doğrudan yabancı sermaye yatırımlarına büyük ölçüde bağımlıdır. Özellikle 2003 yılından itibaren doğrudan yabancı sermaye yatırımları Türkiye’yi önemli ölçüde desteklemiştir. Hızlı büyüyen ekonomisi ve çeşitli sektörleri sayesinde Türkiye, özellikle 1980’lerde ekonomisini liberalleştirdikten ve 2000’lerin başında politika reformları yaptıktan sonra önemli miktarda doğrudan yabancı yatırım çekmiştir.

Neoklasik teoriye göre DYY, ekonomik büyümeyi yatırım miktarında meydana getirdiği artış yoluyla sağlarken, içsel büyüme teorisi kapsamında teknolojik gelişim ve beşeri sermaye stokunu arttırmaktadır. Yapılan çalışmaların bir kısmı DYY’nin ekonomik büyüme üzerindeki etkisini incelerken, diğer kısmı DYY’ni etkileyen faktörleri ve daha fazla yabancı sermaye yatırımı çekmenin yollarını araştırmaktadır. Ülkelerin ekonomik, sosyal ve finansal kriterleri sağlaması yabancı sermaye yatırımları açısından gereklidir. Bu kriterler arasında ülkedeki eğitim düzeyi, teknolojik gelişmişlik, finansal, ekonomik ve politik istikrar, özel ve fikri mülkiyet hakları, vergi sistemi, ticaret ve yatırım politikaları, bürokratik yapı, yerel piyasanın özellikleri yer almaktadır.

Türkiye’nin çekmiş olduğu doğrudan yabancı sermaye yatırımları da enflasyon oranı (Demirel, 2006; Aydemir & Genç, 2015) faiz oranı, döviz kuru vs. (Bozdağlıoğlu & Özpınar, 2011) makroekonomik aktörlere ve hukuki düzenlemelere (Çetin, 2024) bağlıdır. Ayrıca bir diğer önemli husus da iş dünyasını düzenleyen kurumlardır. Makroekonomik istikrarın yanı sıra ekonomik büyüklük (Demirel, 2006; Aydemir & Genç, 2015); işgücü maliyetleri (Bilgili, Tülüce & Doğan, 2012) ve bölgesel ekonomilere erişim (Deichmann, Karidis & Sayek 2003), yabancı yatırımcıların tercihlerini belirleyen bazı önde gelen ekonomik unsurları oluşturmaktadır. Ayrıca, Türkiye’nin düzenleyici kurumları, siyasi istikrarı (Dumludağ, 2009) ve Avrupa Birliği ile daha fazla bütünleşme arzusu (Çetin, 2024) da doğrudan yabancı sermaye yatırımlarının çerçevesinin belirlenmesinde etkili olmuştur.

Türkiye'ye yönelik doğrudan yabancı yatırım miktarında son dönemde meydana gelen değişiklikler, yabancı yatırımcıları Türkiye'ye çeken veya Türkiye'den caydıran unsurların anlaşılması gerekliliğinin altını çizmektedir. Bu durum, bu belirleyicilerin hem elverişli bir yatırım ortamı yaratmaya çalışan politika yapımcılar hem de ülkeye girmeyi veya ülkedeki varlıklarını artırmayı amaçlayan doğrudan yabancı yatırımcılar için hayati önem taşıdığı anlamına gelmektedir.

Avrupa Birliği ülkelerine gelen doğrudan yabancı sermaye yatırımları da ekonomik büyüklük, düzenleyici kurumlar, jeopolitik gelişmeler ve yatırımcı tutumlarından etkilenmektedir.

Bu araştırma, 2010-2022 yılları arasında Avrupa Birliği ve Türkiye'nin çekmiş olduğu doğrudan yabancı sermaye yatırımları (DYY) ile Gayri Safi Yurtiçi Hasıla (GSYİH) arasındaki nedensellik ilişkisini panel veri nedensellik yöntemini kullanarak analiz etmektedir. Çalışmanın ikinci bölümünde literatür taraması yer almaktadır. Üçüncü bölümde veri seti ve modelleme, son bölümde ise sonuçlar sunulmuştur.

## 2. Literatür

Doğrudan yabancı sermaye yatırımları (DYY) bir ülkedeki ekonomik büyüme, endüstriyel gelişme ve asgari ücret gibi faktörlerden etkilenen önemli bir ekonomik olgu olarak günümüz küresel ekonomisinde öne çıkmaktadır. Türkiye, coğrafi konumu, genç nüfusu, dinamik iş ortamı ve stratejik sektörlerdeki potansiyeli ile dikkat çeken bir yatırım destinasyonu olmuştur. Türkiye'nin çekmiş olduğu doğrudan yabancı sermaye yatırımlarının belirleyicileri konusunda literatürde birçok araştırma bulunmaktadır.

Türkiye'nin doğrudan yabancı yatırımları çekme kapasitesi çeşitli faktörlerle şekillenmektedir. İlk olarak, makroekonomik istikrarın sağlanması ve sürdürülmesi yabancı yatırımcılar için önemli bir belirleyici olarak öne çıkmaktadır. Literatür, düşük enflasyon, istikrarlı döviz kuru ve sağlam bir mali politika gibi makroekonomik göstergelerin, yatırımcı güvenini artırdığını vurgulamaktadır. Bunun yanı sıra Türkiye'nin hukuki çerçevesi ve düzenleyici ortamı da doğrudan yabancı sermaye yatırımlarını etkileyen önemli unsurlardan biridir. Yatırımcıların adil rekabet koşullarına ve şeffaf bir hukuki sürece güvenmeleri Türkiye'nin yatırım cazibesini artırabilir. Bu bağlamda, literatür, hukuki reformların ve düzenleyici iyileştirmelerin Türkiye'nin yabancı yatırımlarını çekme kapasitesini artırabileceğine işaret etmektedir.

Türkiye'nin sektörel ve bölgesel özellikleri de doğrudan yabancı sermaye yatırımlarının belirleyicileri üzerinde etkilidir. Özellikle enerji, altyapı ve teknoloji gibi stratejik sektörlerde yapılan yatırımların teşvik edilmesi, ülkenin uzun vadeli sürdürülebilir büyüme hedeflerine ulaşmasına katkı sağlayabilir. Literatürdeki, Türkiye'de doğrudan yabancı yatırımların belirleyicileri ve nedensellik ile ilgili yapılan çalışmalar aşağıda özetlenmiştir. Bu alandaki araştırmalar, Türkiye'nin yatırım çekme stratejilerini daha iyi anlamak ve daha etkili politika önlemleri geliştirmek adına önemli bir kaynak oluşturmaktadır.

Yapraklı (2006), eşbütünlüşme ve hata düzeltme modeliyle yaptığı analizde 1970-2006 yılları arasında doğrudan yabancı sermaye yatırımları (DYY) ile gayrisafi yurtiçi hasıla ve ekonominin dışı açıklığı arasında pozitif, asgari ücret, reel döviz kuru ve dış ticaret açığı ile arasında negatif anlamlı bir ilişki tespit etmiştir. Ayrıca, DYY, GSYİH ve reel döviz kuru arasında ise çift taraflı nedensellik de analizde ulaşılan bir başka sonuçtur.

Batmaz ve Tunca (2007) DYY'nin bölgesel belirleyicilerini Türkiye özelinde 1992-2003 yılları arası için belirlemek için VAR ve eşbütünlüşme analizi kullanmışlardır. Gayrisafi yurtiçi hasıla, altyapı yatırımları ve ticari açıklık oranı ile DYY arasında pozitif ilişki tespit edilirken; FDI ile faiz oranları, ücretler ve döviz kuru arasında negatif ilişki bulunmuştur.

Eşiyok (2011) 19 OECD ülkesinin 1982-2007 arasında Türkiye'ye yaptığı yatırımları panel veri analiziyle ele almıştır. Çalışmada birleşik ulusal gelirler, kişi başına gelir farkı, yatırımların liberalizasyonu ve Türkiye'ye olan ihracat maliyetlerinin DYY'nı pozitif anlamlı etkilediği çıkarımları yapılmıştır.

Bilgili vd. (2012) Markov rejim değişim modeliyle DYY'nin belirleyicilerini incelemişlerdir. Türkiye'ye ait DYY'nin GSYİH, ihracat ve güven indeksinden pozitif ve anlamlı şekilde etkilendiği; emek maliyetleri, ithalat, reeskont faizi ve ABD ve AB'nin güven endeksinden ise negatif etkilendiği ulaşılan önemli sonuçlardır.

Aydemir ve Genç (2015) çalışmasında 1991-2014 yılları için belirleyicileri için dinamik OLS ve çoklu yapısal kırılmalı kointegrasyon testleri uygulanmıştır. Analiz sonuçlarına göre GSYİH, birim emek maliyeti ticari açıklık DYY'na pozitif etki ederken enflasyonun ise DYY'nı negatif etkilediği sonucuna varılmıştır. Kalyoncu vd. (2015), Johansen eşbütünlüşme analiziyle 1975-2012 dönemi Türkiye'ye gelen DYY'nı analiz etmişlerdir. GSYİH, ticari açıklık, enerji üretimi ve emek verimliliğinin üstünde pozitif etkisi bulunurken; TÜFE ile ölçülen enflasyonun ise üstünde negatif etkisi bulunmuştur.

Özen ve Kıdemli (2020) 1988-2018 yılları arasında Türkiye'ye gelen DYY'nın belirleyicilerini incelemişlerdir. Yöntem olarak ARDL ve hata düzeltme modelinin kullanıldığı çalışmada uzun dönem için dışı açıklık ve belirsizlik endeksi ile DYY arasında pozitif; saatlik kazanç yani emek ücreti ile uluslararası rezerv göstergeleri

arasında negatif bir ilişki bulunmuştur. Hata düzeltme modeli sonuçlarına göre kısa dönemde uluslararası rezervler ile DYY arasında negatif; dışa açıklık, küresel belirsizlik indeksi ve çalışma çağındaki nüfus ile DYY arasında pozitif ilişkili bulunmuştur.

1996-2020 dönemi için Beşoluk ve Kesin (2023), Hacker ve Hatemi-J nedensellik testini kullanarak DYY’nın belirleyicilerini analiz etmişlerdir. Dışa açıklık, kurumlar vergisi, reel döviz kuru ve enflasyon ile DYY arasında tek taraflı bir nedensellik tespit edilmiş, DYY ile iktisadi büyüklük arasında ise çift yönlü nedensellik bulunmuştur.

Batmaz ve Yürük (2023), 1990-2020 periyodunda ARDL testi sonuçlarına göre kısa dönemde herhangi bir değişkenle DYY arasında anlamlı bir ilişki bulunmazken uzun dönemde DYY ile kişi başına düşen gelir arasında pozitif; enflasyonla DYY arasında ise negatif ilişki bulunmuştur. Yomamoto nedensellik analizi sonuçlarına göre kişi başına gelir ve enflasyondan DYY’na doğru tek taraflı bir nedensellik bulunmuşken, işgücü ile DYY arasında nedensellik bulunamamıştır.

Borensztein ve diğerleri (1998) panel veri regresyonu kullanarak, DYY’nın teknoloji transferi yolu ile ev sahibi ülkenin ekonomik büyümesine katkısının olduğunu ve o ülkenin beşerî sermayesi seviyesine bağlı olduğunu vurgulayan çalışmalarıyla literatüre bu alanda önemli bir bakış açısı kazandırmışlardır. Çalışmada, yeterli insan sermayesine sahip olmayan ülkelerin, DYY’nın faydalarından etkin olarak yararlanamayabileceği belirtilmiştir. Benzer şekilde, Alfaro ve diğerleri (2004) panel regresyon analizi ve sınır testi yaklaşımı ile yaptıkları çalışmalarında, DYY’nın ekonomik büyüme üzerindeki etkisinin yüksek oranda finansal piyasaların gelişmişliğine bağlı olduğunu göstermişlerdir. Ayrıca, finansal sistemi zayıf olan ülkelerde DYY’nın büyümeye olan katkısının zayıf olduğunu belirtmişlerdir.

Çeştepe ve Çapcı (2021) Johansen eşbütünleşme testi ve Hatemi-J nedensellik testi kullanarak yaptıkları çalışmada; Türkiye’de GSYİH ve DYY arasındaki ilişkiyi incelemiş ve ekonomik büyümenin hem pozitif hem de negatif olarak DYY’ni etkileyebileceğini ortaya koymuştur. Ayrıca, negatif şokların da iki değişken arasında nedensellik oluşturduğu tespit edilmiştir. Aydemir ve Genç (2015) ise çoklu eşbütünleşme testi ve hata düzeltme modeli (ECM) ile yaptığı analizde ticari açıklık, enerji tüketimi ve iş gücü verimliliği gibi faktörlerin, DYY üzerinde pozitif ve anlamlı bir etki yarattığını ve uzun vadede DYY’nın ekonomik büyümeye olumlu katkı sağladığını tespit etmişlerdir. Demir vd. (2021) eşbütünleşme analizi ve granger nedensellik testi kullanarak yaptıkları çalışmada; Türkiye’de GSYİH ile DYY arasındaki uzun vadeli ilişkiyi değerlendirirken, politik riskin DYY üzerindeki negatif etkisine işaret etmiş ve büyüme ile yatırım ilişkilerinin politik istikrarsızlıktan etkilenebileceğini belirtmiştir. Enflasyon ve DYY arasında uzun dönemde çift yönlü bir nedensellik ilişkisi tespit edilmiştir. Ayrıca, ekonomik büyümeden DYY’na doğru tek yönlü bir nedensellik ilişkisi bulunmuştur. Kar ve Tatlısöz (2008) EKK ve VAR analizi uyguladıkları çalışmalarında; Türkiye’de GSYİH, ticari açıklık ve enerji üretimi gibi faktörlerin, DYY üzerinde pozitif etki yaptığını, reel döviz kuru ve iş gücü maliyetlerinin ise negatif etki yaptığını bulmuştur.

Carkovic ve Levine (2005) ekonomik büyüme ve DYY arasındaki ilişkinin yönünü belirlemek için dinamik bir panel veri modeli kullanmıştır. Çalışmada, DYY’nın doğrudan ekonomik büyümeyi artırdığı iddiasına diğer çalışmalara nazaran daha temkinli yaklaşmış ve bu etkinin; kurumsal altyapı ve ekonomi politikalarına bağlı olduğu belirtilmiştir.

Gelişmekte olan ülkeler üzerinde yapılan analizlerde ise, Asghar ve Hussain (2014) GSYİH büyümesi ile DYY arasında çift yönlü bir nedensellik ilişkisi bulunduğunu vurgulamış, ayrıca ekonomik istikrarın DYY üzerindeki etkisine dikkat çekmiştir. Adams (2009) panel veri analizini kullanarak DYY’nın özellikle altyapı yatırımları ve üretim kapasitesine etkileri üzerinde durarak, bu tür yatırımların uzun vadede ekonomik büyümeyi desteklediğini belirtmiştir.

Grubaugh (2013) panel veri analizi kullandığı ve 74 ülkeyi kapsayan çalışmasında, pazar büyüklüğü ve büyüme oranının DYY üzerindeki etkisinin doğrusal olmadığını göstermiştir. Bu çalışmada, özellikle düşük gelirli ülkelerin DYY’ni çekme potansiyelinin sınırlı olduğu belirtilmiştir. Benzer şekilde, Blomström ve Kokko (1998) DYY’nın etkilerinin ülkelerin rekabetçi yapıları ve politik istikrar düzeyleri ile yakından ilişkili olduğunu belirtmişlerdir. Çok uluslu şirketlerin yarattığı pozitif dışsallıklar, DYY’nın oluşturduğu bilgi transferi ve teknolojik katkı yoluyla ekonomik büyümeye pozitif etki ettiğini göstermişlerdir. Ancak, etkilerin ev sahibi ülkenin ekonomik yapısına bağlı olarak değiştiğini vurgulamışlardır. Dunning (1988) ise OLI (Ownership-Location-Internalization) paradigması ile DYY’nın ekonomik büyümeye nasıl katkıda bulunabileceğini açıklayarak teorik bir temel sağlamıştır. Sahiplik avantajları, lokasyon cazibesi ve içselleştirme, DYY’nın temel unsurları olarak tanımlanmıştır. Ranjan ve Agrawal (2011) BRICS ülkelerinin verisini kullanarak yaptıkları çalışmada DYY’nın ekonomik büyümeye katkısının pozitif olduğunu, ancak etkinin boyutunun ülkelerin açık ekonomi politikalarına ve makroekonomik istikrara bağlı olduğunu vurgulamıştır.

Herzer ve diğerleri (2008) gelişmekte olan ülkelerde DYY ve ekonomik büyüme arasındaki ilişkiyi panel nedensellik analizi ile incelemiş ve DYY’nın ülkeler arasında homojen dağılmadığını ve bu ilişkinin ülkeye özgü farklılıklar gösterebileceğini tespit etmişlerdir. Ayrıca, çalışma bulgularına dayanarak, DYY’nın büyüme üzerindeki etkisi ile kişi başına düşen gelir seviyesi, eğitim düzeyi, açıklık derecesi veya finansal piyasa

gelişmişliği arasında net bir ilişki olmadığını ortaya koymuşlardır. Moudatsou ve diğerleri (2011) 1970-2003 döneminde Avrupa Birliği (AB) ve Güneydoğu Asya Uluslar Birliği (ASEAN) gibi iki farklı ekonomik birliğe yönelik doğrudan yabancı yatırımlar ile ekonomik büyüme arasındaki nedensellik sırasını ele almışlardır. Ekonomik büyüme ile DYY arasında çift taraflı nedensellik tespit etmişlerdir. Barrell ve Pain (1997) doğrudan yabancı yatırımları Avrupa’da belirleyen faktörleri incelemiş ve DYY üzerinde işgücü maliyetlerinin, pazar büyüklüklerinin ve teşvik politikalarının önemini vurgulamıştır. Özellikle, daha düşük işgücü maliyetlerine sahip ülkelerin daha fazla DYY çektiği gösterilmiştir. Campos ve Kinoshita (2003) Avrupa’da DYY’nın temel itici güçlerini analiz etmişler, kurumsal yapının ve ekonomik istikrarın DYY üzerinde önemli bir etkisi olduğunu ortaya koymuşlardır. Ayrıca, ekonomik reformların ve altyapı yatırımlarının, özellikle Orta ve Doğu Avrupa ülkelerinde DYY’ni arttırdığı sonucuna ulaşmıştır. Bevan ve Estrin (2004) Orta ve Doğu Avrupa ülkelerinin Avrupa Birliği’ne entegrasyonu sırasında DYY akışlarının etkilerini incelemişlerdir. Düşük maliyetli işgücü ve AB üyeliği beklentisinin DYY üzerinde olumlu etkiler yarattığını belirtmişlerdir.

### 3. Doğrudan Yabancı Yatırımlar ve Gayri Safi Yurt İçi Hasıla: Panel Veri Nedensellik Analizi

Küresel ekonomik ilişkilerde doğrudan yabancı sermaye yatırımları ekonomik büyüklük ve kalkınma açısından kritik bir rol oynamaktadır. Bu kısımda, Avrupa Birliği (AB) ülkeleri ve Türkiye özelinde 2010-2022 döneminde DYY ile gayrisafi yurt içi hasıla (GSYİH) büyüklükleri arasındaki nedensellik ilişkisi panel veri tahmin yöntemleri ile ele alınmıştır. Son yıllarda, sürdürülebilir bir ekonomik büyümenin üzerine yapılan çalışmalarda DYY’nın rolünün giderek daha fazla ön plana çıktığı görülmektedir. DYY özellikle AB ülkeleri gibi gelişmiş ekonomilerde, teknoloji transferinin ve yenilikçilik kapasitesinin artmasını sağlamakta, Türkiye gibi gelişmekte olan ekonomilerde sermaye açığını kapatma ve istihdam yaratma gibi önemli etkiler üstlenmektedir. Söz konusu etkilerin büyüklüğü ve yönü, ülkelerin ekonomik yapıları, kurumsal altyapıları ve politika çerçevelerine bağlı olarak farklılık gösterebilmektedir. Bu nedenle DYY ile GSYİH arasındaki ilişkinin panel veri ekonometrisi yöntemleri ile analizi önemli yol gösterici bulgular sunabilmektedir.

#### 3.1. Veri Seti ve Yöntem

Çalışmada, 27 AB ülkesi ve Türkiye’yi kapsayan toplam 28 birim için 2010-2022 dönemine ait panel veri seti kullanılmıştır. 28 ülke (Almanya, Avusturya, Belçika, Bulgaristan, Kıbrıs, Hırvatistan, Danimarka, İspanya, Estonya, Finlandiya, Fransa, Yunanistan, Macaristan, İrlanda, İtalya, Letonya, Litvanya, Lüksemburg, Malta, Hollanda, Polonya, Portekiz, Çek Cumhuriyeti, Romanya, Slovakya, Slovenya, İsveç ve Türkiye) ve 13 yıl panel veri analizinin birim-zaman boyutlarını oluşturmaktadır. Doğrudan yabancı sermaye yatırımları, net girişler (Milyon, USD) olarak alınırken, farklı ülkelerin ekonomik büyüklüklerini karşılaştırmak için sıklıkla kullanılan ve bir ülkenin ekonomik faaliyetlerinin güncel piyasa değerini gösteren GSYİH (Milyon, USD) kullanılmıştır.

**Tablo 1.** Panel Veri Seti İçin Tanımlayıcı İstatistikler\*

Değişkenler	NxT	Ortalama	Std. Sap.	Min	Max
DYY	364	17760.137	56332.096	-330338.47	332926.41
GSYİH	364	577491.35	878144.91	9035.824	4278504

Dünya Bankası, Dünya Kalkınma Göstergeleri kaynağından elde edilmiştir.

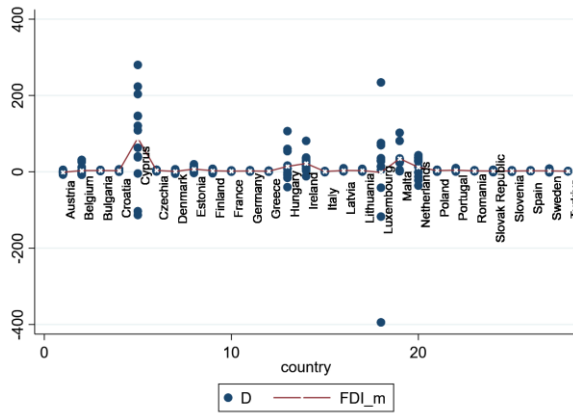
17760 milyon USD, ülkeler genelinde ortalama bir yatırım giriş düzeyini vermektedir. 56332 milyon USD standart sapma DYY değerlerinin ülkeler arasında ciddi farklılıklar gösterdiğine işaret etmektedir. Maksimum ve minimum değerler bazı ülkelerde önemli miktarda sermaye çıkışları gerçekleşirken bazı ülkelerde ise yüksek düzeyde yatırım girişleri olduğunu göstermektedir. Yatırım ortamlarının çeşitliliği ve ülkeler arasında yatırım çekme açısından diğerlerine göre daha başarılı ülkelerin varlığı izlenmektedir. Yabancı sermaye çıkışlarının varlığı, veri setinde negatif DYY büyüklükler ile gözlemlenmektedir. Yüksek standart sapma ve geniş değer aralıkları, panel veri modellemesi sürecinde değişkenlik ve yatay kesit bağımlılığı gibi konulara dikkat edilmesi gerektiğini vurgulamaktadır.

577491 milyon USD GSYİH ortalama, paneldeki veri setindeki ülkelerin ortalama ekonomik büyüklüğünü temsil etmektedir. 878144 milyon USD standart sapma, GSYİH değerlerinin de panel veri ülkeleri arasında önemli farklılık gösterdiğine işaret etmektedir. En düşük GSYİH değerleri diğer ülkelere göre daha küçük ekonomik yapıları en yüksek 4,278504 milyon USD ise daha büyük ekonomileri temsil etmektedir. Tablo 1 hem büyük hem

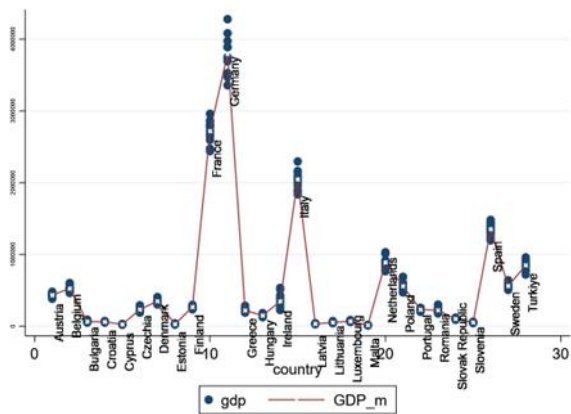


de küçük ekonomilerin aynı panel veri setinde yer aldığını ve analizin hem gelişmiş hem de gelişmekte olan ekonomiler için önemli sonuçlar üretebilmesine olanak sağlayabileceğini göstermektedir. AB ülkeleri arasında ekonomik entegrasyona rağmen, GSYİH seviyelerindeki büyük farklar ülkeler arasındaki kalkınma düzeyinin heterojen olduğunu göstermektedir. Bu durum, AB’nin ekonomik büyümeyi sağlarken kullanacağı dengeleme politikalarının etkilerini ve Türkiye’nin AB içindeki pozisyonunu incelemek için önemli bir bağlam sunar.

Negatif ortalama değerlere sahip ülkelerin, sermaye çıkışı yaşadığını ve bu ülkelerdeki şirketlerin yabancı ülkelerdeki yatırımlara odaklandığını ya da belirli dönemlerde büyük sermaye çıkışları olduğuna işaret etmektedir. Bu ülkeler sermaye ihracatçısı ülkeler olarak da düşünülebilir. Yüksek DYY ortalamalara sahip ülkelerin, güçlü ekonomik yapıları, cazip yatırım ortamları gibi özellikleri ile yabancı yatırımları çektiği söylenebilir. Yüksek standart sapma değerlerine sahip panel veri ülkelerinin, DYY akışlarının oldukça değişken olması ve bazı dönemlerde büyük girişler ya da çıkışlar yaşaması söz konusudur. Düşük DYY ortalamalarına ve nispeten küçük standart sapma değerlerine sahip ülkelerin ise yatırım girişlerinde istikrarlı küçük ekonomilere sahip oldukları görülmektedir.



Şekil 1. Panel Veri Ülkeleri DYY Verisi.



Şekil 2. Panel Veri Ülkeleri GSYİH Verisi.

Şekil 1’de DYY, Şekil 2’de GSYİH veri setlerinin ülkeler bazında birim ortalama değerleri (DYY\_m, GSYİH\_m) bir arada verilmiştir. Şekiller tanımlayıcı istatistikler kısmında yapılan açıklamalara değişkenlik bilgilerini de barındırarak farklı bir bakış sağlamaktadır.

### 3.2. Panel Veri Kümeleme Analizi

Panel Kümeleme, DYY ve GSYİH değişkenleri arasındaki nedensellik ilişkisini incelerken önemli avantajlar sağlayabilir. AB ülkeleri ve Türkiye gibi birbirine ekonomik, politik ve ticari olarak bağlı birimler arasında yatay kesit bağımlılığı olması muhtemeldir. Yatay kesit bağımlılığını hesaba katmak standart hata tahminlerini daha doğru hale getirir ve sonuçların sağlamlığını artırır. Bunun yanı sıra panel veri ülkelerinin benzer ekonomik yapılarına göre gruplandırılması heterojenliği dikkate alan daha spesifik sonuçlar elde edilebilmesini sağlayabilmektedir. Politik veya kurumsal farklılıklar nedeniyle benzer özelliklere sahip ülkelerin gruplandırılması daha hassas analizler yapılmasını sağlar.

Panel veri modeli tahmin sürecinde yer verilen 28 ülkenin Sarafidis ve Weber (2015) tarafından geliştirilen regresyon kümeleme yaklaşımına göre panel veri birimleri farklı kümeler halinde gruplandırılmıştır. Yöntem algoritmasına bağlı olarak kaç kümeye bölündüğü ve ülkelerin hangi kümeler içinde yer aldığını belirlediği çalışma gerçekleştirilmiştir. Farklı küme birliklilikleri içinde en uygun küme sayısı tahmini değeri, model bilgi kriterini (MIC) minimize eden büyüklüktür ve çalışmada bu büyüklük 2’dir. 28 ülke 2 farklı kümeye ayrılabilir.

$$\text{MIC} = N \ln \left( \frac{\text{KKT}}{\text{NT}} \right) + \Omega \theta_n \quad (1)$$

**Tablo 2.** Optimum Küme Sayısı.

<b>Omega</b>	<b>Total RSS</b>	<b>MIC</b>
<b>1</b>	<b>9.154e+23</b>	<b>1384.380</b>
<b>2</b>	<b>6.530e+23</b>	<b>1379.561</b>
<b>3</b>	<b>5.760e+23</b>	<b>1380.686</b>
<b>4</b>	<b>5.670e+23</b>	<b>1384.883</b>
<b>5</b>	<b>5.600e+23</b>	<b>1389.174</b>
<b>6</b>	<b>5.560e+23</b>	<b>1393.612</b>
<b>7</b>	<b>5.550e+23</b>	<b>1398.200</b>
<b>8</b>	<b>5.490e+23</b>	<b>1402.534</b>
<b>9</b>	<b>5.490e+23</b>	<b>1407.172</b>
<b>10</b>	<b>5.490e+23</b>	<b>1411.811</b>

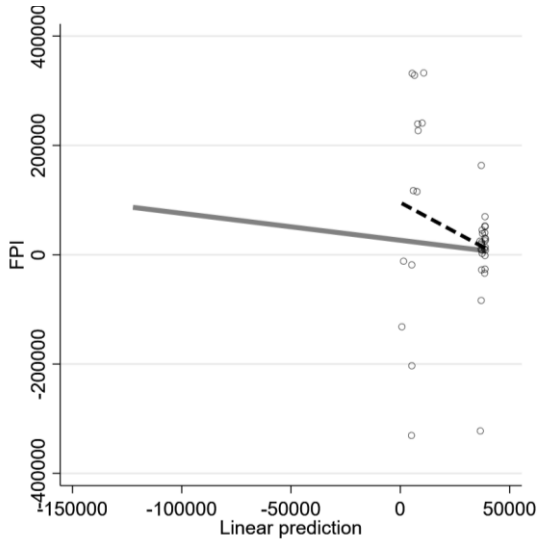
**Tablo 3.** Kümelerin Frekans Değerleri.

<b>Kümeler</b>	<b>Frekans</b>	<b>%</b>	<b>Kümülatif</b>
1	39	10.71	10.71
2	325	89.29	100.00
Toplam	364	100.00	

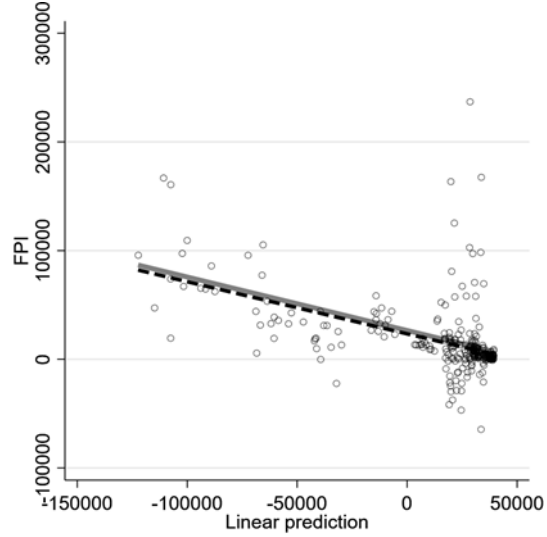
Tablo 2 minimum MIC değerini veren küme sayısını, Tablo 3 ise kümelerin frekans değerlerini vermektedir. 28 ülke 2 farklı kümeye ayrılabilir. Nedensellik ilişkisinin muhtemel yönü itibarıyla DYY ve GSYİH değişkenleri için farklı regresyonlara dayalı küme belirleme çalışmalarından DYY değişkeni ile gerçekleştirilen çalışma, panel veri ülkelerini iki farklı kümeye ayırma noktasında yeterli bilgiyi sağlamıştır. Kümeler,

Küme 1: Almanya, Avusturya, Belçika, Bulgaristan, Hırvatistan, Danimarka, İspanya, Estonya, Finlandiya, Fransa, Yunanistan, Macaristan, İrlanda, İtalya, Letonya, Litvanya, Malta, Polonya, Portekiz, Çek Cumhuriyeti, Romanya, Slovakya, Slovenya ve İsveç.

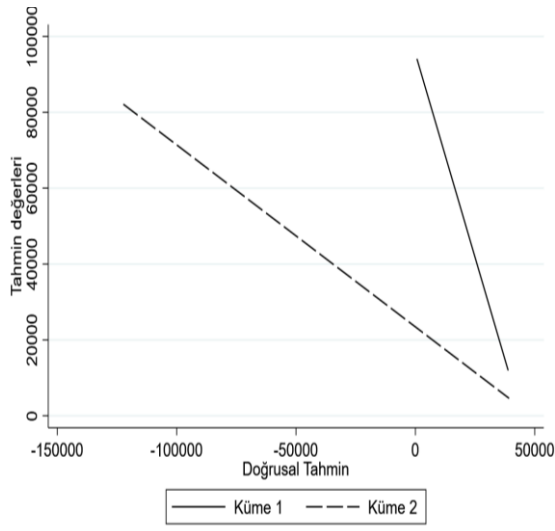
Küme 2: Kıbrıs, Lüksemburg, Hollanda şeklinde belirlenmiştir. Devam eden çalışmalar her iki küme üzerinden ayrı ayrı sürdürülmüştür.



Şekil 3: Küme 1 DYY Tahmini.



Şekil 4: Küme 2 DYY Tahmini.



Şekil 5: Küme 1-2 Doğrusal Tahminleri.

Grafik 3 birinci küme için, Grafik 4 ikinci Küme için DYY ile GSYİH verilerinin gözlem değerleri ile doğrusal tahminlerini göstermektedir. Grafik 5 ise iki farklı küme için tahminlerin hem kesme hem eğim için farklılığını tahmin değerleri ve doğrusal tahminler üzerinden göstermektedir.

### 3.3. Panel Veri Durağanlık Analizi

Dumitrescu & Hurlin testi, heterojen panellerde Granger nedenselliğini analiz etmek için güçlü ve esnek bir araçtır. Bu test Granger nedensellik testinin panel veriler için genişletilmiş halidir (İşcanoğlu-Çekiç ve Gültekin 2019). Panel veri analizlerinde birimlerin farklı yapısını ve nedensellik ilişkilerinin değişkenliğini dikkate alarak daha güvenilir ve kapsamlı sonuçlar elde edilmesini sağlar. Yöntemin güvenilir sonuçlar verebilmesi için kullanılan verilerin durağanlığının test edilmesi gerekir. Panel veri modellerinde durağanlığın testinde kullanılacak yöntemin belirlenebilmesi için öncelikle yatay kesit bağımlılığının varlığını inceleyen testler yapılmalıdır.

Test farklı panel dönemlerinde farklı birimlerin hata terimlerinin ilişkili olmadığı sıfır hipotezini sınamaktadır. Hipotezler,

$H_0: \rho_{ij} = \rho_{ji} = 0$   $i \neq j$  için , yatay kesit bağımlılık yoktur.

$H_1: \rho_{ij} \neq \rho_{ji} \neq 0$   $j \neq i$  için , yatay kesit bağımlılık vardır.

**Tablo 4.** Yatay Kesit Bağımlılığının Testi (Değişkenler için, Pesaran 2015).

Değişken	CD-test	p-değeri
DYY	5.372	0.000*
GSYİH	46.307	0.000*

Not: \* 0.01 anlam düzeyinde testin sıfır hipotezinin reddedildiğini göstermektedir.

Tablo 4’ de katı yatay kesit bağımsızlığı yerine zayıf yatay kesit bağımsızlığını ortaya koyan sıfır hipotezinin sınındığı Pesaran (2015) testi sonuçları yer almaktadır. Test istatistiği,

$$CD = \sqrt{\frac{2}{N(N-1)} \left( \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sqrt{T_{ij} \hat{\rho}_{ij}} \right)} \quad (2)$$

eşitliği ile hesaplanmaktadır. Sonuçlar, DYY ve GSYİH değişkeni için yatay kesit bağımlılığının olduğunu göstermektedir. Bu sonuç ikinci nesil birim kök testlerinin uygulanması gerektiğini ortaya koymaktadır.

**Tablo 5.** Panel Birim Kök Testleri (Düzy Seriler İçin).

Değişkenler / Test	CIPS	CIPS Kritik Değeri Pesaran (2007)
DYY	-1.516	
GSYİH	-0.299	% 5: -1.58
$\Delta$ DYY	-3.821*	% 1: -1.77
$\Delta$ GSYİH	-1.804*	

Not: \* 0.01 anlam düzeyinde testin sıfır hipotezinin reddedildiğini göstermektedir.

Tablo 5’te ikinci nesil birim kök testlerinden CIPS test (Cross-sectionally Augmented IPS Test) sonuçları verilmiştir. Bu test, Pesaran (2007) tarafından geliştirilmiş ve yatay kesit bağımlılığını dikkate alarak panel veri setlerinde birim kök testleri yapmak için kullanılmaktadır. Test, CADF (Cross-sectionally Augmented Dickey-Fuller) istatistiğini hesaplamakta ve tüm birimler için istatistiklerin ortalamasını almaktadır. Otokorelasyon durumunu dikkate alan bir test istatistiği hesaplanmaktadır. DYY ve GSYİH serileri düzeyde durağan değilken ilk fark serileri durağan serilerdir.

### 3.4. Panel Veri Homojenlik Analizi

Eğimlerin homojen özellikte olup olmadığını sınamak amacıyla Pesaran Yamagata (Delta, 2008) testi uygulanmıştır. Küçük ve büyük örnekler için delta istatistiği hesaplanmıştır.

Küçük ve büyük örnekler için delta istatistiği,

$$\hat{\Delta} = \sqrt{N} \left( \frac{N^{-1}\bar{S} - k}{2k} \right) \sim \chi_k^2 \quad (3)$$

$$\hat{\Delta}_{adj} = \sqrt{N} \left( \frac{N^{-1}\bar{S} - k}{v(T, k)} \right) \sim N(0,1) \quad (4)$$

şeklinde ve Swamy (1970) örneklemden hesaplanan test istatistiği ve bağımsız değişken sayısından yararlanılarak Delta İstatistiği hesaplanmaktadır.

**Tablo 6.** Küme 1- Eğitim Homojenliği Testi.

	Delta	p-value
$\hat{\Delta}$	0.624	0.533
$\hat{\Delta}_{adj}$	0.712	0.477

**Tablo 7.** Küme 2- Eğitim Homojenliği Testi.

	Delta	p-value
$\hat{\Delta}$	-1.119	0.263
$\hat{\Delta}_{adj}$	-1.292	0.196

Tablo 6 ve 7’ de yer alan sonuçlar incelendiğinde, küme 1 ve küme 2’nin eğitim homojenliğine sahip oldukları görülmektedir. Eğitim homojenliğini ortaya koyan sıfır hipotezi reddedilememektedir. Kümelerin kendi içlerinde homojenliği söz konusudur.

İki farklı panel kümesi için nedensellik analizi Dumitrescu-Hurlin (2012) testine dayalı olarak gerçekleştirilmiştir. Bu test Granger nedenselliğini panel verilerde analiz etmek için geliştirilmiştir. Testin alternatif hipotezi en az bir panel birimi için nedenselliğin varlığını ortaya koymaktadır. Test sürecinde kullanılan denklemler,

$$\Delta DYY_{it} = \alpha_i \sum_{k=1}^K \gamma_i^{(k)} \Delta DYY_{it-k} + \sum_{k=1}^K \beta_i^{(k)} \Delta GSYİH_{it-k} + \varepsilon_{it} \quad (5)$$

$$\Delta GSYİH_{it} = \theta_i \sum_{k=1}^K \theta_i^{(k)} \Delta GSYİH_{it-k} + \sum_{k=1}^K \pi_i^{(k)} \Delta DYY_{it-k} + \varepsilon_{it} \quad (6)$$

şeklinde. Testte gecikme mertebeleri Akaike Bilgi Kriteri (AIC) ile belirlenmiştir. W-bar istatistiği ortalama bireysel test istatistiklerinden elde edilir ve paneldeki bireysel eğilimleri değerlendirmek için kullanılır. Z-bar istatistiği, ise panel veri analizlerinde güçlü ve pratik bir araçtır. Standardizasyon ve asimptotik normal dağılıma uyması sayesinde anlamlılık testlerinde güvenilir sonuçlar sağlar.

**Tablo 8.** Küme 1- Dumitrescu ve Hurlin (2012) Nedensellik Testi Sonuçları.

Sıfır Hipotezi	Test İstatistiği		p-değeri
	W-bar	Z-bar	
GSYİH DYY'nın nedeni değildir	1.8287	2.9299	0.0034*
DYY GSYİH'nın nedeni değildir	1.3831	1.3546	0.1755

Not: \*, 0.01 anlam düzeylerinde testin sıfır hipotezinin reddedildiğini göstermektedir.

**Tablo 9.** Küme 2- Dumitrescu ve Hurlin (2012) Nedensellik Testi Sonuçları.

Sıfır Hipotezi	Test İstatistiği		p-değeri
	W-bar	Z-bar	
GSYİH DYY'nın nedeni değildir	4.6910	2.3305	0.0198**
DYY GSYİH'nın nedeni değildir	3.4718	3.0273	0.0025*

Not: \* ve \*\* sırasıyla, 0.01 ve 0.05 anlam düzeylerinde testin sıfır hipotezinin reddedildiğini göstermektedir.

Tablo 8’ de yer alan sonuçlar, birinci panel veri kümesindeki ülkelerde GSYİH değişkeni DYY büyüklüğünün nedenidir. DYY değişkeninin ise GSYİH değişkenindeki değişmelere neden olduğu söylenememektedir. Tablo 9 ikinci panel kümesi için nedensellik analizi sonuçlarını barındırmaktadır. Sonuçlar, GSYİH DYY değişkenlerinin karşılıklı olarak birbirinin nedeni olduğunu göstermektedir.

#### 4. Sonuç

Türkiye’nin de dahil edildiği AB ülkeleri ile bir panel veri analizi gerçekleştirilmiştir. Bahsedilen avantajları sebebiyle bir panel kümeleme analizi yapılmıştır. Kıbrıs, Lüksemburg ve Hollanda diğer ülkelerden kümeleme kriterleri sonucu olarak ayrı bir gruba dahil edilmiştir. Bu üç ülkenin bir kümede bir araya gelmesi bir takım ülke özelliklerine dayandırılabilir. Öncelikle Hollanda, Lüksemburg ve Kıbrıs için DYY verilerinin önemli bir kısmı reel sektörden çok finansal işlemleri temsil etmektedir. Başka bir deyişle yatırımların reel ekonomiye etkisi sınırlıdır ve yatırımların çoğu finansal sektörde yoğunlaşmaktadır. Söz konusu üç ülke de yatırımcılar için düşük vergi oranları veya vergi teşvikleri sunmaktadır. Yüksek standart sapma ve yüksek minimum negatif DYY değerleri, bu ülkelerdeki DYY giriş ve çıkışlarının uluslararası şirketlerin sermaye hareketlerinden büyük ölçüde etkilendiğini ve ekonomik dalgalanmalara açık olduğunu göstermektedir. Son olarak bu ülkelerin genellikle uluslararası yatırımların geçiş noktaları olarak kullanıldığı, yani, bu ülkelerin yabancı yatırımların gerçek anlamda bir üretim ya da altyapı geliştirme yerine, daha çok finansal yeniden yapılandırma ve vergi optimizasyonu amaçlı yatırımların kaydedildiği yerler olduğu söylenebilir.

Kümeleme analizi sonuçlarına bağlı kalınarak her iki küme için DYY ve GSYİH değişkenleri arasında nedensellik analizi yapılmıştır. Sonuçlar panel veri birinci küme ülkelerinde (Kıbrıs, Lüksemburg ve Hollanda dışında kalan) GSYİH değişkeni DYY büyüklüğünün nedeni iken DYY değişkeninin ise GSYİH değişkenindeki değişmelere neden olduğu söylenememektedir. Ülkelerin daha cazip yatırım ortamları sunması ve sahip olunan büyük pazar büyüklüğünün, yatırımcılar için talep ve büyüme için potansiyel yaratması GSYİH’nın DYY değişikliklerine neden olmasına gerekçe olarak düşünülebilir. GSYİH’nın artışının, ülkenin ekonomik istikrarına dair bir işaret olarak görülmesine paralel olarak yatırımcıların, istikrarlı büyüyen ekonomilere daha fazla yatırım yapma eğiliminde oldukları söylenebilir. DYY büyüklüğünün GSYİH’ya göre küçük bir oran oluşturması ve/veya

etkisinin yerel yatırım, tüketim, ihracat gibi diğer faktörler tarafından gölgelenmiş olması DYY değişkeninin GSYİH değişkeninin bir nedeni olmamasına sebep olarak gösterilebilir. Yine bu sonuç DYY’nın yöneldiği sektörlerin, GSYİH’nın artması için doğrudan bir katkı yapmamaktadır düşüncesini ortaya koyabilmektedir. İkinci Panel veri kümesi için DYY ve GSYİH değişkenleri için karşılıklı nedenselliğin varlığı tespit edilmiştir. Bu iki yönlü nedensellik bu ülkelerin yatırımcı dostu politikalara ve cazip bir yatırım ortamına sahip olduklarını gösterir. Bu durum, DYY ve ekonomik büyüklük arasında dinamik bir döngü yaratarak bu ülkelerin küresel sermaye hareketlerinden avantaj sağlamalarına olanak tanır. Bu ülkelerin sundukları vergi avantajları ve yatırım dostu politikalar sayesinde, yabancı yatırımların doğrudan, artan bir ekonomik büyüklüğe sebep olan döngüye kavuştukları söylenebilir.

## Kaynaklar

- Adams, S. (2009). Foreign Direct Investment, Domestic Investment, and Economic Growth in Sub-Saharan Africa. *Journal of Policy Modeling*, 31(6), 939-949.
- Alfaro, L., Chanda, A., Kalemli-Ozcan, S., & Sayek, S. (2004). FDI and Economic Growth: The Role of Local Financial Markets. *Journal of International Economics*, 64(1), 89-112.
- Asghar, N., & Hussain, Z. (2014). Financial Development, Trade Openness and Economic Growth in Developing Countries: Recent Evidence from Panel Data. *Pakistan Economic and Social Review*, 52, 99-126.
- Aydemir, O., & Genç, E. (2015). Uluslararası Sermaye Hareketlerinin Belirleyicileri, Doğrudan Yabancı Yatırımlar Kapsamında Bir Analiz, Türkiye Örneği. *Optimum Ekonomi ve Yönetim Bilimleri Dergisi*, 2(2), 17-41. <https://doi.org/10.17541/oeymbd.54709>
- Aydemir, O., & Genç, İ. (2015). Türkiye’de Doğrudan Yabancı Yatırımlara Etki Eden Faktörlerin Analizi: Çoklu Eşbütünleşme Yaklaşımı.
- Barrell, R., & Pain, IN. (1997). The Growth of Foreign Direct Investment in Europe. *Economic Journal*, 107 (445), 1770-1786.
- Batmaz, N., & Tunca, H. (2007). Türkiye’de Doğrudan Yabancı Sermaye Yatırımlarının Bölgesel Belirleyicileri Üzerine Bir Eş Bütünleşme Analizi (1992–2003), *Sosyal Bilimler Araştırmaları Dergisi*, [2\(1\),199-224](https://doi.org/10.1501/2007/1199-224).
- Batmaz, T., & Yürük, B. (2023). Türkiye’de Doğrudan Yabancı Yatırımların Belirleyicileri Üzerine Ampirik Bir Analiz: ARDL Sınır Testi (1990-2020). *Optimum Ekonomi ve Yönetim Bilimleri Dergisi*, [10\(1\), 99-124](https://doi.org/10.17541/oeymbd.54709).
- Beşoluk, E., & Keskin, A. (2023). Türkiye’de Doğrudan Yabancı Yatırımların Belirleyicileri: Hacker ve Hatemi-J Nedensellik Analizi. *JOEEP: Journal of Emerging Economies and Policy*, [8\(1\), 266-284](https://doi.org/10.1501/2023/1266-284).
- Bevan, A. A., & Estrin, S. (2004). The Determinants of Foreign Direct Investment into European Transition Economies. *Journal of Comparative Economics*, 32(4), 775–787.
- Bilgili, F., Tülüce, N. S. H., & Doğan, İ. (2012). The Determinants of FDI in Turkey: A Markov Regime-switching Approach Economic Modelling, [29\(4\), 1161-1169](https://doi.org/10.1016/j.econmod.2011.11.019).
- Blomström, M., & Kokko, A. (1998). Multinational Corporations and Spillovers. *Journal of Economic Surveys*, 12(3), 247-277.
- Borensztein, E., De Gregorio, J., & Lee, J. W. (1998). How Does Foreign Direct Investment Affect Economic Growth? *Journal of International Economics*, 45(1), 115-135.
- Bozdağlıoğlu, E. Y., & Özpınar, Ö. (2011). Türkiye’ye Gelen Doğrudan Yabancı Yatırımların Türkiye’nin İhracat Performansına Etkilerinin VAR Yöntemi ile Tahmini. *Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*,13(3),39-63.
- Campos, N. F., & Kinoshita, Y. (2003). Why Does FDI Go Where It Goes? New Evidence from the Transition Economies. *IMF Working Paper*, No. 2003/228.
- Carkovic, M., & Levine, R. (2005). Does Foreign Direct Investment Accelerate Economic Growth? In *Does FDI Promote Development? (195-220)*. Institute for International Economics.
- Çeştepe, H., & Çapcı, T. (2021). Türkiye’de DYY ve Ekonomik Büyüme Arasındaki Dinamik İlişki: Eşbütünleşme Analizi.
- Çetin, R. (2024). Analysis of Foreign Direct Investment Policies and of the Performance of FDI Flows into Türkiye. *Kahramanmaraş Sütçü İmam Üniversitesi Sosyal Bilimler Dergisi*, 21(1), 118-127. <https://doi.org/10.33437/ksusb.1401753>
- Demir, A., et al. (2021). Türkiye’de Makroekonomik Göstergeler ve DYY Arasındaki Nedensellik İlişkisi: VAR Analizi.



- Atf / Citation:** ÜN T., TAŞSEVEN Ö., KARACA D. A. (2024). Avrupa Birliği Ülkeleri ve Türkiye’de Doğrudan Yabancı Sermaye Yatırımları ile Gayri Safı Yurtiçi Hasıla Arasındaki Panel Veri Nedensellik Analizi. *İstatistik Araştırma Dergisi*, 14 (2), 30-45.
- Demirel, O., (2006). Doğrudan Yabancı Yatırımlar, Ekonomik Büyümeye Etkileri ve Türkiye Uygulaması, Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü İktisat Anabilim Dalı Yüksek Lisans Tezi.
- Dumitrescu, E. I., & Hurlin, C. (2012). Testing for Granger non-causality in heterogeneous panels. *Economic Modelling*, 29(4), 1450-1460. <https://doi.org/10.1016/j.econmod.2012.02.014>
- Dumludag D. (2009). An Analysis of the Determinants of Foreign Direct Investment in Turkey: The Role of the Institutional Context, *Journal of Business Economics and Management*, 10(1), 15-30.
- Dunning, J. H. (1988). *Explaining International Production*. London: Unwin Hyman.
- Esiyok, B. (2011). Determinants of Foreign Direct Investment in Turkey: a Panel Study Approach. MPRA Paper No. 36568.
- Grubaugh, S. (2013). Determinants of Foreign Direct Investment and Its Effect on Economic Growth. *Applied Economics*, 45(18), 2519-2528.
- Herzer, D., Klasen, S., & Nowak-Lehmann, F. D. (2008). In Search of FDI-led Growth in Developing Countries: The way forward." *Economic Modelling*, 25(5), 793-810. <https://databank.worldbank.org/source/world-development-indicators>
- İşcanoğlu Çekiç, A. & Gültekin, H. (2019). R Uygulamalı Panel Veri Analizi ve Ampirik Bir Uygulama, Ekin Basım Yayın.
- Joel Deichmann & Socrates Karidis & Selin Sayek, 2003. Foreign Direct Investment in Turkey: Regional Determinants, *Applied Economics*, Taylor & Francis Journals, 35(16), 1767-1778.
- Kalyoncu, H., Tuluçe, N., & Yaprak, Z. (2015). Determinants of Foreign Direct Investment: An Empirical Analysis for Turkey. *Journal of Economic and Social Studies*, 5(2).
- Kar, M., & Tatlısöz, D. (2008). Türkiye’de Doğrudan Yabancı Yatırımların Belirleyicileri: EKK ve VAR Analizleri.
- Özen, A., & Kıdemli, M. (2020). Türkiye’de Doğrudan Yabancı Yatırımlarının Belirleyicileri Üzerine Ekonometrik bir Analiz. *Ekonomi Politika ve Finans Araştırmaları Dergisi*, 5(3), 643-666.
- Moudatsou, A., & Kyrkilis, D. (2011). FDI and Economic Growth: Causality for the EU and ASEAN. *Journal of Economic Integration*, 26(3), 554–577. <http://www.jstor.org/stable/41348800>
- Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. \*CESifo Working Paper Series No. 1229\*. <https://ssrn.com/abstract=572504>
- Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. \*Journal of Applied Econometrics, 22\*(2), 265-312. <https://doi.org/10.1002/jae.951>
- Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. \*Econometric Reviews, 34\*(6-10), 1089-1117. <https://doi.org/10.1080/07474938.2014.956623>
- Pesaran, M. H., & Yamagata, T. (2008). Testing slope homogeneity in large panels. \*Journal of Econometrics, 142\*(1), 50-93. <https://doi.org/10.1016/j.jeconom.2007.05.010>
- Ranjan, V., & Agrawal, G. (2011). FDI and Economic Growth in BRICS Economies: A Panel Data Analysis. *International Business Research*, 4(3), 203-210.
- Sarafidis, V., & Weber, E. (2015). A critical review of panel data models with cross-sectional dependence. \*Journal of Applied Econometrics, 30\*(4), 429-456. <https://doi.org/10.1002/jae.2372>
- Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica*, 38(2), 311–323. <https://doi.org/10.2307/1913012>
- Yapraklı, S. (2006). Türkiye’de doğrudan yabancı yatırımların ekonomik belirleyicileri üzerine ekonometrik bir analiz. *D.E.Ü.İ.B.F. Dergisi*, 21(2), 23-48.

## Ekler

**Ek 1 Tablo 1.** Panel Veri Seti Ülkeleri DYY Verisi Tanımlayıcı İstatistikleri.

Ülke	Ortalama	Std. Sapma	Min	Max
Austria	-4368.197	16177.918	-28939.34	22983.943
Belgium	15639.764	63983.551	-41601.387	163624.38
Bulgaria	2144.322	664.533	1093.6	3594.55
Croatia	1857.663	1490.546	35.178	4660.185
Cyprus	21115.985	30103.866	-33356.59	69972.766
Czechia	8771.864	3034.546	1699.915	12890.95
Denmark	4287.056	11617.015	-16349.826	27041.426
Estonia	2085.959	1910.696	-715.465	7361.224
Finland	7620.775	10905.97	-10571.699	23869.273
France	47429.411	28987.002	5804.878	105424.96
Germany	85955.626	41594.197	19532.082	166915.56
Greece	3203.383	1917.079	532.927	6846.333
Hungary	21546.754	59761.676	-64364.898	167595.02
Ireland	64422.376	65066.217	-46633.227	237060.22
Italy	18044.395	16921.352	-22090.941	44249.715
Latvia	1143.276	846.546	335.421	3721.078
Lithuania	1637.085	1243.412	357.323	4505.744
Luxbourg	-5916.362	109325.48	-322053.78	163717.59
Malta	4180.574	2459.343	155.879	9244.998
Netherl.	95452.166	218556	-330338.47	332926.41
Poland	18404.154	9656.231	1029	37113
Portugal	9722.494	4975.898	1270.014	21396.375
Romania	5754.635	3133.964	2370.097	11882.525
Slovak Rep.	2220.464	1953.034	-1141.703	5431.593
Slovenia	1178.482	780.604	33.538	2219.248
Spain	36215.609	10896.565	20848.479	58626.18
Sweden	14739.422	18626.334	-8624.293	52477.898
Turkiye	12794.692	3019.298	7700	19263

**Ek 2 Tablo 2.** Panel Veri Seti Ülkeleri GSYİH Verisi Tanımlayıcı İstatistikleri.

Country	Mean_GDP	Std_Dev_GDP	Min_GDP	Max_GDP
Austria	432817.219	28755.733	381971.156	479295.375
Belgium	525620.062	39739.839	462335.562	600748.812
Bulgaria	64094.051	12464.735	50811.996	90346.172
Croatia	59979.098	6005.52	50857.355	71600.047
Cyprus	24891.244	2984.012	19842.492	29482.918
Czechia	231853.422	32566.135	188033.047	290527.562
Denmark	348299.844	30440.916	302673.062	405688.0
Estonia	28353.33	5337.142	22890.762	38100.812
Finland	267237.938	17431.016	234534.375	296776.75
France	2719139.0	158384.702	2439188.75	2959355.75
Germany	3794164.0	262921.631	3357585.75	4278504.0
Greece	218841.562	27070.873	188480.344	282995.938
Hungary	148800.953	19165.991	125174.164	182090.047
Ireland	345917.969	106447.944	225118.719	533140.0
Italy	2047265.375	135889.875	1836637.75	2294994.25
Latvia	32207.258	4573.247	27263.09	40932.031
Lithuania	51421.625	9612.271	41435.535	70974.492
Luxembourg	68774.797	8260.131	59776.383	85584.109
Malta	13365.298	3129.243	9461.776	18125.564
Netherlands	889178.812	78626.209	765572.75	1029678.312
Poland	558331.375	73154.254	470024.562	688125.0
Portugal	230576.625	17922.526	199394.062	255534.844
Romania	222230.5	42737.974	177883.891	300691.344
Slovak Republic	101836.789	9216.809	88900.883	118576.484
Slovenia	51436.047	5727.021	43107.508	61832.203
Spain	1352785.125	86724.272	1196157.0	1480710.5
Sweden	560371.125	36920.686	505103.781	639714.938
Türkiye	849696.688	70844.937	720338.5	957799.125



## Web Kazıma Yöntemlerinin Tüketici Fiyat Endeksi Çalışmalarında Kullanılması Türkiye Örneği

Yeliz YALÇIN

Ankara Hacı Bayram Veli Üniversitesi, Ekonometri Bölümü / Prof. Dr.

yeliz.yalcin@hbv.edu.tr

Orcid No: 0000-0002-7141-3086

Burak BAYSAL

Ankara Hacı Bayram Veli Üniversitesi, Ekonometri Bölümü / Lisansüstü Öğrencisi

burak.baysal@hbv.edu.tr

Orcid No: 0009-0008-5431-3172

### Özet

Tüketici Fiyat Endeksleri, hanehalklarının tüketimine konu olan mal ve hizmetlerin fiyatlarının genel düzeyinde zaman içerisinde meydana gelen değişimlerin bir ölçüsü olarak tanımlanmaktadır. Gelişen teknoloji ile birlikte özellikle fiyat istatistikleri bağlamında kullanılacak yeni veri derleme yöntemleri ortaya çıkmış, bu sayede Tüketici Fiyat Endeksi çalışmaları için oldukça hızlı bir biçimde girdi sağlanabilmesi mümkün hale gelmiştir. Bu yöntemlerden biri internet üzerindeki bilgilerin otomatik kazıcılar vasıtası ile yapısal veriye dönüştürülmesini ve veri tabanlarına kaydedilmesini sağlayan web kazıma yöntemidir. Bu sayede geleneksel veri derleme yöntemlerinin kullanımı ile aylık olarak üretilen Tüketici Fiyat Endeksi göstergelerinin günlük veya haftalık olarak daha yüksek frekanslarla hesaplanabilmesi mümkün hale gelmiştir. Çalışmada, web kazıma yöntemleri ile internet üzerinden derlenecek fiyatlar kullanılarak günlük ve aylık Tüketici Fiyat Endeksleri elde edilmesi amaçlanmıştır. Uluslararası anlamda kabul görmüş endeks hesaplama yöntemleri kullanılarak ulaşılan bulgular dikkate alınan dönemde web kazıma yöntemleri ile elde edilen sonuçların resmi istatistiklerle tutarlı sonuçlar sağladığını ortaya koymuştur.

*Anahtar sözcükler: Tüketici Fiyat Endeksi, Web kazıma, Metin madenciliği.*

**Sorumlu Yazar / Corresponding Author:** Burak BAYSAL, Ankara Hacı Bayram Veli Üniversitesi, Ekonometri Bölümü.

**Atıf / Citation:** YALÇIN Y., BAYSAL B. (2024). Web Kazıma Yöntemlerinin Tüketici Fiyat Endeksi Çalışmalarında Kullanılması Türkiye Örneği. *İstatistik Araştırma Dergisi*, 14 (2), 46-58.

Makalede yer alan görüş ve düşünceler yazarlarına ait olup dergiyi, editörlerini ve Türkiye İstatistik Kurumu'nu bağlamamaktadır.

## Using Web Scraping Methods in Consumer Price Index Studies, Türkiye Example

### Abstract

Consumer Price Indices are defined as a measure of the changes in the prices of goods and services that are subject to household consumption over time. With the developing technology, new data collection methods have emerged that can be used especially in the context of price statistics. In this way, it has become possible to provide a quick input to the Consumer Price Index studies. The web scraping method, which enables the information on the Internet to be converted into structured data by means of automatic scrapers and saved in databases, is one of these methods. In this way, it has become possible to calculate the Consumer Price Index indicators, which are produced monthly with traditional data collection methods, on a daily or weekly basis. In the study, it is aimed to obtain daily and monthly Consumer Price Index by using web scraping methods and prices compiled over the internet. Findings obtained by using internationally accepted index calculation methods revealed that the results obtained by web scraping methods provide results consistent with official statistics.

*Keywords: Consumer Price Index, Web scraping, Text mining.*

### 1. Giriş

Endeksler, bir değişkenin zamana veya mekana göre ne şekilde değiştiğinin sayısal bir ölçüsüdür. Bir ekonomide tüketiciler tarafından satın alınan milyonlarca farklı mal ve hizmet yer almaktadır. Bu mal ve hizmet fiyatlarının genel düzeyinde zaman içerisinde meydana gelen değişimler ise Tüketici Fiyat Endeksleri (TÜFE) yardımı ile ölçülmektedir. TÜFE verileri başlı başına uluslararası karşılaştırmalarda kullanılan bir gösterge olmanın yanı sıra ücret ve maaşların güncellenmesi, birçok değişken için reel değerlerin hesaplanması, enflasyon hedeflemesi ve fiyat istikrarının izlenmesi gibi alanlarda kullanılmaktadır.

TÜFE hesaplamalarının gerçekleştirilebilmesi için öncelikle toplumun harcama yapısını temsil edecek bir mal ve hizmet sepeti belirlenmesi, bu sepette yer alan her bir maddenin göreceli ağırlığının hesaplanması, tüketime konu olan mal ve hizmetlerin uygun bir sınıflama sistemine göre tasnif edilmesi ve son olarak baz alınan dönem ve cari dönem fiyatlarının derlenmesi gerekmektedir. Tüm bu girdilerin elde edilmesinin ardından uygun endeks formülü kullanılarak endeks hesaplaması gerçekleştirilmektedir. Bu girdilerin tamamının bir araya getirilmesi ve TÜFE hesaplamalarının yapılmasından ulusal istatistik ofisleri sorumludur. Özellikle geleneksel veri derleme yöntemleri kullanımının ciddi bir zaman ve işgücü maliyeti gerektirmesinden dolayı bireysel olarak bu hesaplamaların yapılması son derece zor görünmektedir. Ancak günümüzde teknolojinin gelişmesi ile alternatif veri derleme yöntemleri ortaya çıkmış, bu sayede daha düşük zaman ve işgücü maliyetleri ile veri derleme uygulamaları gerçekleştirilmesi mümkün hale gelmiştir. Bu çalışmanın da konusunu oluşturan TÜFE çalışmaları açısından başlıca alternatif veri kaynakları olarak internet ve barkod verileri öne çıkmaktadır. Bilindiği üzere son dönemde tüketicilerin satın alma alışkanlıkları önemli ölçüde değişmiş, online alışverişlerin hacmi ciddi oranda artmıştır. Bununla birlikte çok sayıda işyeri geleneksel mağazalarının yanı sıra kendi web siteleri üzerinden de satış yapmaya başlamış, bazı işyerleri ise geleneksel mağazacılık anlayışını terk ederek yalnızca internet üzerinden satış gerçekleştirir hale gelmiştir. İnternet, ekonomi üzerinde son derece etkili olurken sürekli büyüyen bir veri kaynağı haline de gelmiştir. Özellikle TÜFE çalışmaları açısından bu verilerin kullanılması istatistik birimlerin cevaplayıcı yükünün azaltılması ve manuel veri derlemeden kaynaklanan maliyetlerin azaltılması noktalarında faydalı görülmektedir. Ayrıca daha yüksek frekans ve hacimlerde veri derlemenin mümkün hale gelmesi, istatistik üretim süreçlerinin hızlandırılması açısından da faydalı olacaktır. EUROSTAT (2020) çalışmasına göre; artan gözlem sayıları ve web kazıma teknolojilerinin otomasyonu vasıtasıyla yeni göstergelerin geliştirilmesi, veri kalıplarının daha iyi anlaşılması, fiyatların günlük, saatlik ve daha yüksek frekanslarla izlenmesi mümkün hale gelecektir. Ayrıca fiyat ölçümlerini destekleyecek ilave metaveriler toplanabilecektir.

Web kazıma ile elde edilen verilerin fiyat çalışmalarında kullanılmasına yönelik ilk büyük çalışma MIT (Massachusetts Institute of Technology) tarafından yürütülen "Billion Price Project" adlı projedir. Cavallo ve Rigobon (2016) yaptıkları çalışmada çeşitli ülkeler için internet üzerinden fiyat derlemesi gerçekleştirmiş, günlük ve aylık frekanslarla endeks değerleri hesaplamış ve kapsadıkları ürünler özelinde resmi istatistiklerle karşılaştırmalar yapmışlardır. Örneğe seçilen iş yerleri hem çevrimiçi hem de çevrimdışı satış yapan iş yerlerinin

arasından seçilmiştir. Fiyatı derlenen ürünler ise ağırlıkların mevcut olduğu, resmi istatistikler kapsamında fiyat derlemesi yapılan maddeler arasında yer almaktadır. Üçüncü parti uygulamalar yerine iş yerinin kendi web sitesinden veri derlemesi yapılmaktadır. Yaklaşık 25 ülke için TÜFE ağırlığının %70'inden fazlası kapsamaktadır. Çalışma özellikle Arjantin'in enflasyonu üzerine odaklanmış, açıklanan resmi rakamlar ile çalışma kapsamında hesaplanan rakamlar arasında üç kata varan farklar ortaya çıkmıştır.

Billion Price Project, internet üzerinden derlenen fiyatların TÜFE çalışmalarında kullanılabilmesine yönelik önemli bir emsal teşkil etmekle birlikte ulusal istatistik ofisleri de internetten fiyat kazıma yöntemlerini resmi istatistik alanında kullanmaktadır. Bu bağlamda 34 farklı ülkenin veri derleme yöntemleri incelenmiş, bunların 19'unun web kazıma yöntemlerini çeşitli düzeylerde kullandığı görülmüştür. Ülkemizde ise Gençler ve Ulusoy (2021) çalışmasında TÜİK tarafından yayınlanan madde sepeti ve ağırlıkları kullanarak günlük endeks hesaplamaları gerçekleştirmişlerdir. Çalışma sonucunda elde edilen bulgular Enflasyon Araştırma Grubu (ENAG) adı ile yayınlanmaktadır. Bu çalışmanın ENAG enflasyon hesaplamasından farklılıkları şu şekildedir;

- Bu çalışmada sabit sepet yaklaşımına sadık kalınmış, fiyatı derlenen ürünlerin kalitesinin sabit tutulması sağlanmıştır. ENAG çalışmasında ise hesaplamada kullanılan ürünlerin devamlı olarak değiştiği ve herhangi bir kalite düzeltmesi uygulanmadığı anlaşılmaktadır.
- Bu çalışmada literatürde önerilen endeks formülleri olan Jevons endeks formülü, üst düzey endekslerde ise Laspeyres endeks formülü kullanılmaktadır. ENAG çalışmasında ise özellikle günlük zincirlemeden kaynaklı bir zincir sapması olup olmadığına yönelik bulguların paylaşılmasının önemli olduğu düşünülmektedir. EUROSTAT (2022: 9) çalışmasına göre zincirleme sıklığının artırılması endeksin sistematik olarak yanlışlaşmasına, özellikle uzun dönemde makul olmayan sonuçların ortaya çıkmasına neden olmaktadır.
- Bu çalışmada Laspeyres endeks formülünün gereği olarak baz dönemi, ağırlıkların da referans dönemi olan aralık ayı olarak belirlenmiştir. ENAG endeks hesaplamasında ise baz dönemi 26 ağustos olarak belirlenmiş, bunun gerekçesi açıklanmamıştır. Ayrıca ağırlıklarda fiyat güncelleme (price update) işleminin gerçekleştirilip gerçekleştirilmediği de belirtilmemiştir.
- Bu çalışmada fiyatı sürekli olarak derlenebilen ürünlerin (Eşleşen modeller yöntemi) fiyatları kullanılmaktadır. Bu nedenle baz döneminde fiyatının derlenmesi mümkün olmayan ürünler endeks hesaplamasının dışında bırakılmaktadır. ENAG endeksinde özellikle güçlü mevsimsel özellik gösteren ürünlere ilişkin uygulamadan bahsedilmemiştir.
- Son farklılık ise kullanılan ağırlıklar ve sınıflama ile ilgilidir. Bu çalışmada COICOP sınıflaması kamuoyu ile paylaşılan en alt düzeyde kullanılmış olup derlenen tüm fiyatlara metin madenciliği yöntemleri de kullanılarak 7'li düzeyde kod ataması gerçekleştirilmiştir. Ayrıca TÜFE sepetinde yer alan maddelerin tamamının fiyatı derlenemediği için ağırlıklarda yanlışlık oluşmaktadır. Bu nedenle ağırlıklar ana gruplardan başlanarak kapsanan maddelere yeniden dağıtılmıştır. ENAG çalışmasında COICOP sınıflamasının hangi düzeyde kullanıldığı anlaşılamamış olup yalnızca ana gruplar düzeyinde kullanıldığı değerlendirilmiştir. Ağırlıkların doğru şekilde kullanılabilmesi için fiyatların ağırlıklarla en uyumlu şekilde sınıflandırılması gerekmektedir.

İnternet fiyatları kullanılarak günlük TÜFE hesaplanmasına yönelik bir diğer çalışma İsviçre için gerçekleştirilmiştir. İnternet üzerinden derlenen fiyatlar ve banka kartı harcama verileri kullanılarak oluşturulan ağırlıklar vasıtası ile günlük fiyat endeksleri elde edilmiştir. Bu endeksler resmi istatistikler ile karşılaştırılmıştır (Santiago ve Sarah, 2020).

Bununla birlikte ulusal istatistik ofisleri de web kazıma verilerini farklı kapsamlarda endeks hesaplamalarına dahil etmektedir. TÜİK, derlemekte olduğu fiyatların yaklaşık %5'ini web kazıma yöntemi ile derlerken beyaz eşyalar, ev elektroniği, mobilya, otomobil ve otobüs biletleri bu kapsamda yer almaktadır (TÜİK, 2024). Benzer şekilde BLS derlediği fiyatların yaklaşık %8'ini web üzerinden elde etmektedir (U.S Bureau of Labor Statistics, 2020). Avrupa ülkelerindeki durum ele alındığında ise birçok ülkede yine web kazıma yöntemlerinin farklı kapsamlarda kullanıldığı göze çarpmaktadır.

Bu farklılıklar bağlamında çalışmanın literatüre katkılarını aşağıdaki gibi sıralamak mümkün olacaktır:

- i. Çalışma uluslararası TÜFE metodolojisi ile tamamen uyumlu bir hesaplama ortaya koymaktadır.
- ii. TÜFE çalışmalarında web kazıma yöntemi ile derlenmesi uygun olmayan maddeler ortaya konulmuştur.
- iii. TÜFE çalışmaları bakımından web kazıma yöntemi kullanılmasının ortaya çıkaracağı olası zorluklara yer verilmiştir.
- iv. Eşleşen modeller yöntemiyle fiyat derlenmesi sağlanarak kalite faktörünün zaman içerisinde sabit tutulması garanti edilmiştir.
- v. COICOP sınıflaması, alt düzeyde kod atanarak ve metin madenciliğiyle yapılmış; ağırlıklar daha doğru sınıflandırılmıştır.

Çalışmanın ikinci bölümünde hesaplamada kullanılan endeks formüllerinden, üçüncü bölümde veri toplama süreci ve endeks hesaplamasından, dördüncü bölümde çalışmada elde edilen bulgulardan bahsedilecek, tartışma ve sonuç bölümü ile çalışma sonlandırılacaktır.

## 2. Yöntem

### 2.1. Tüketici Fiyat Endeksi

TÜFE hesaplamaları iki aşamalı olarak gerçekleştirilmektedir. İlk aşamada ağırlık verilebilen en alt düzeyde endeksler hesaplanmakta ikinci aşamada ise sınıflamanın hiyerarşik olarak üst basamaklarına çıkılmaktadır.

İlk aşamada ağırlık verilemeyen fiyatlar toplulaştırılarak endekslenmektedir. Literatürde bu işlemde kullanılacak üç temel endeks formülü yer almaktadır. Bunlar Dutot, Carli ve Jevons endeks formülleridir.

Bir endeks formülünün geçerli kabul edilebilmesi için arzu edilen bazı istatistiksel özellikleri ihtiva etmesi gerekmektedir. Bu bağlamda endeks rakamlarının bazı aksiyomları sağlayıp sağlamadığının test edilmesi gerekmektedir. Bu testler literatürde aksiyomatik testler olarak anılmakta olup endeks formüllerinin birbirlerine karşı üstünlüklerinin ve zayıflıklarının anlaşılmasında kullanılmaktadır (Armknrecht ,1996). Bu testlerden en yaygın olarak kullanılanları zaman çevrimi testi, eş ölçülebilirlik testi ve geçişlilik testidir.

Zaman çevrimi testine göre, iki farklı periyotta gözlenen fiyatlar yer değiştirdiğinde endeks de aynı şekilde yer değiştirmelidir. Geçişlilik testi, iki periyot arasındaki zincirleşmiş endeks ile doğrudan endeksin eşit olup olmadığını sınımlamaktadır. Eş ölçülebilirlik testi ise fiyatı ölçülen ürünlerin ölçüm birimleri ile ilgilidir. Bu test endeksin ölçü biriminden etkilenip etkilenmediğini ölçer. Jevons endeksi bu testlerin tamamında başarılı sonuçlar ortaya koyarken Dutot endeksi eş ölçülebilirlik testinde Carli endeksi ise zaman çevrimi ve geçişlilik testlerinde başarısız sonuçlar üretmektedir (Fenwick, 2009: 154,155). Aksiyomatik testlerde en başarılı sonuçları Jevons endeksi sağladığından literatürde alt düzey endekslerde Jevons önerilmektedir. Jevons endeksi iki farklı şekilde hesaplanabilir. Temel toplamlar düzeyinde karşılaştırılan iki dönemde de eşleşen fiyatların geometrik ortalamalarının oranı şeklinde hesaplanabileceği gibi göreceli fiyatların geometrik ortalaması biçiminde de hesaplanabilir (Graf, 2020: 124).

$$I_j^{t/0} = \prod \left( \frac{P_i^t}{P_i^0} \right)^{1/n} = \frac{\prod (P_i^t)^{1/n}}{\prod (P_i^0)^{1/n}} \quad (2.1)$$

Yukarıdaki denklemde  $I_j^{t/0}$  t dönemi endeks rakamını,  $P_i^t$  i. ürünün t dönemindeki fiyatını,  $P_i^0$  i. ürünün baz dönemi fiyatını, n ise ürün sayısını ifade etmektedir. Buna göre hesaplanan endekslerden sınıflamanın üst basamaklarına çıkılırken kullanılacak Laspeyres, Paashe, Fisher, Törnqvist gibi çok sayıda endeks formülü bulunmaktadır. Ancak elde edilebilen en güncel harcama verisi bir önceki yıla ait olduğu için yalnızca Laspeyres formülü kullanılabilir.

$$I_{LA}^{t/0} = \frac{\sum P_i^t Q_i^0}{\sum P_i^0 Q_i^0} \quad (2.2)$$

Yukarıdaki denklemde  $P_i^0$  i. ürünün baz dönem fiyatını,  $P_i^t$  i. ürünün t dönemindeki fiyatını,  $Q_i^0$  i. ürünün baz dönem miktarını,  $Q_i^t$  i. ürünün t dönemindeki miktarını,  $I_{LA}^{t/0}$  ise Laspeyres fiyat endeksini ifade etmektedir (EUROSTAT, 2018: 170). Buna göre ilk aşamada hesaplanan endeks değerlerinin ağırlıklı ortalaması alınarak sınıflamanın üst basamaklarına çıkılmaktadır. Bu işlem yapılırken baz dönemi için hesaplanan ağırlık seti kullanılmaktadır.

### 2.2. Web kazıma

Web kazıma; web sitelerinde yer alan bilgilerin hızlı, verimli ve otomatik bir biçimde çıkarılarak bu bilgilerin kullanılabilir bir formatta sunulmasıdır (Castrillo-Fernández, 2015). Daha açık bir ifade ile web sitelerinde yer alan bilgiler alınarak tanınabilir hale getirilir ve bir veri tabanına kaydedilir. Web kazıma işlemleri Python ve R başta olmak üzere Java, Visual Basic gibi birçok programlama dili kullanılarak yapılabilmektedir. Bu paketlerin yanı sıra programlama bilgisi gerektirmeyen üzerine tıklama (point and click) şeklinde çalışan çeşitli programlar da yer almaktadır.

Web kazıma amacı ile en yaygın olarak kullanılan ve tavsiye edilen programlama dilleri Python ve R dilleridir. Hem Python hem de R web kazıma amacıyla geliştirilmiş kütüphaneler içermektedir. Bu kütüphaneler aracılığı ile HTML sayfasının indirilerek hedef bilgilerin seçilmesi sağlanmaktadır (EUROSTAT, 2020). Web kazıma işlemlerinde Python programlama dili sıklıkla kullanılmakta olup temelde üç paket üzerinden gerçekleştirilmektedir. Bunlar BeautifulSoup, Selenium ve Scrapy'dir. Bu paketlerden BeautifulSoup, HTML ve XML dosyalarından veri çekmek üzere oluşturulmuş bir Python kütüphanesidir (Richardson, 2019). HTML yapılarını anlaşılır hale getirmenin yanı sıra XML yapılarını temsil eden, kolay geçişli Python nesnelere sunarak dağınık web'i biçimlendirmeye ve düzenlemeye yardımcı olur (Mitchell, 2015: 6). Kullanımı kolay olmakla birlikte sunucuya talep yollama işlemi gerçekleştirilememektedir. Ayrıca yalnızca statik web sitelerinde kullanılabilir. Elementleri kullanıcı hareketlerine göre güncellenebilen, dinamik yapıları web sitelerinde kullanılamamaktadır.

Selenium doğrudan web kazıma amacıyla geliştirilmiş bir araç değildir. Asıl amacı web tarayıcılarının otomasyonudur (Selenium, 2022). Selenium, amacı her ne kadar web kazıma olmasa da bu işlemler için de kullanılabilir. Bir web tarayıcısı aracılığı ile insan benzeri bir davranış sergilediğinden tüm web sitelerine erişilebilmekte ve buralardaki bilgiler alınabilmektedir. Selenium doğrudan bir web tarayıcısı üzerinden web sayfalarını taramakta ve buradaki verileri kaydetmektedir. Bu çalışma biçimi nedeniyle hız açısından dezavantajlıdır. Selenium sunucu bazlı olarak çalışan, çoğunlukla JavaScript kodlarının tetiklenmesine olanak sağlayan dinamik web sitelerinden veri alınması için uygun bir çözümdür. Bir sayfa üzerinde tıklama, sayfayı hareket ettirme, form doldurma gibi birçok işlem de gerçekleştirilebildiği için web kazıma kullanıma mükemmel biçimde uyumlu olarak görülmektedir (EUROSTAT, 2020).

Scrapy, yüksek seviyeli bir web tarama ve web kazıma çerçevesidir. Web sayfalarının taranması ve bu sayfalarda yer alan yapısal verilerin çıkarılmasında kullanılır. Bunun yanı sıra otomatik testlerden veri madenciliğine kadar çok geniş bir alanda farklı görevlerde kullanılabilir (Scrapy developers, 2022). Scrapy, diğer araçlara kıyasla çok daha hızlı olmasına karşılık Javascript kullanan web sitelerinde kullanılamamaktadır. İlave bir araç kullanılarak bu sorunun üstesinden gelinebilmektedir.

Yeni bir veri derleme yöntemi olarak web kazıma uygulamasının kalite ve verimlilik bakımlarından TÜFE çalışmalarında kullanılmaya son derece uygundur. Web kazıma yöntemlerinin TÜFE çalışmalarına entegre edilmesi ile birlikte yüksek frekanslı veriler daha düşük maliyetle derlenebilecektir (Griffioen ve Ten Bosch, 2016).

Bu çalışmada; kullanımının kolay olması, bir web tarayıcı aracılığı ile insan benzeri bir davranış göstermesi nedeniyle her türden web sitesinin kazınmasına olanak sağlaması ve yasaklanma olasılığının daha düşük olduğunun değerlendirilmesi nedenleri ile Selenium ve BeautifulSoup paketleri kullanılmıştır.

### **3. Veri Toplama Süreci ve Endeks Hesabı**

TÜFE hesaplamasının gerçekleştirilebilmesi için hanchalkı harcamalarını temsil edici bir mal ve hizmet sepetine, tüketime konu mal ve hizmetlerin tasnif edilmesinde kullanılacak bir sınıflamaya, sepette yer alan mal ve hizmetlerin ağırlıkları ile baz ve cari dönem fiyatlarına ihtiyaç duyulmaktadır.



**Tablo 3.1.** Ana gruplara göre fiyat derlenen işyeri sayıları.

Ana Grup	İşyeri Sayısı*
Genel	53
Gıda ve alkolsüz içecekler	4
Giyim ve ayakkabı	11
Konut	8
Ev eşyası	11
Ulaştırma	20
Haberleşme	6
Eğlence ve kültür	6
Lokanta ve oteller	4
Çeşitli mal ve hizmetler	7

\* Web sitelerinde birden fazla ana grup verisi derlenebildiğinden ana gruplar toplamı geneli vermemektedir.

Çalışmada kullanılan işyerlerinin COICOP ana gruplarına göre sayıları Tablo 3.1’de yer almaktadır. Madde sepeti ve ağırlıklar TÜİK web sitesinden alınmış, sınıflama olarak COICOP sınıflaması kullanılmıştır. Fiyatların derleneceği internet siteleri ulusal düzeyde faaliyet gösteren, satış düzeyi itibariyle temsil edici olduğu düşünülen web siteleri arasından karar örnekleme yönteminde seçilmiştir.

**Tablo 3.2.** Kapsanan maddelerin ana grup içerisindeki ağırlıkları.

Ana Grup	Kapsama Oranı
Gıda ve alkolsüz içecekler	89.02
Alkollü içecekler ve tütün	93.74
Giyim ve ayakkabı	82.4
Konut	77.97
Ev eşyası	72.43
Sağlık	0
Ulaştırma	78.69
Haberleşme	80.16
Eğlence ve kültür	52.94
Eğitim	0
Lokanta ve oteller	69.48
Çeşitli mal ve hizmetler	63.51
<b>Toplam</b>	<b>75.25</b>

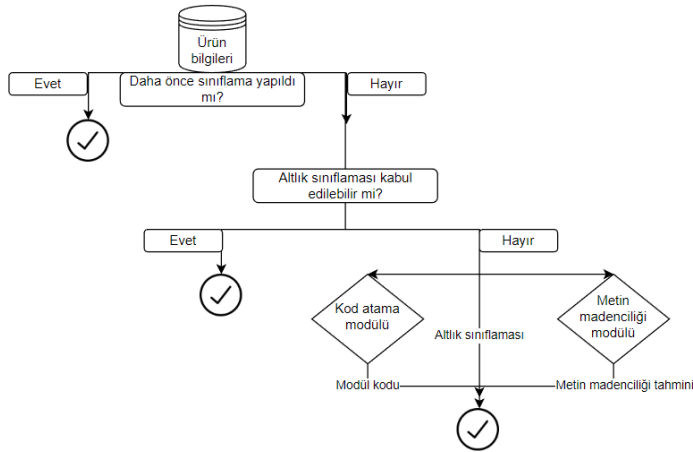
2022 TÜFE madde sepetinde yer alan 409 maddenin 233 tanesinin fiyatı hesaplamaya dahil edilmiştir. Kapsanan maddelerin ilgili ana grup içerisindeki ağırlıkları Tablo 3.2’de yer almaktadır. Burada yer alan değerler ana grup toplamının yüze eşit olduğu duruma göre hesaplanmıştır. Ana grup içerisindeki ağırlığı yüksek maddelerin fiyatlarının web üzerinden derlenmesi mümkün olmayan sağlık ve eğitim ana grupları çalışma kapsamı dışında bırakılmıştır. Bu nedenle ağırlıkları da sıfır olarak ele alınmıştır.

**Tablo 3.3.** Ana grup bazında ağırlıklar.

Ana Grup	Resmi Ağırlık	Yeni Ağırlık
Gıda ve alkolsüz içecekler	25.32	26.73
Alkollü içecekler ve tütün	4.31	4.55
Giyim ve ayakkabı	6.42	6.78
Konut	14.12	14.91
Ev eşyası	8.85	9.34
Sağlık	3.24	0
Ulaştırma	16.8	17.73
Haberleşme	3.78	3.99
Eğlence ve kültür	3.06	3.23
Eğitim	2.03	0
Lokanta ve oteller	7.11	7.51
Çeşitli mal ve hizmetler	4.96	5.24

Çalışmada kullanılan ağırlık seti üzerinde ise birtakım işlemler gerçekleştirilmiştir. TÜİK’in TÜFE madde sepetinde yer alan tüm ürünlerin fiyatlarının derlenememesi nedeniyle ağırlık yapısı bozulmakta ve yeniden dağıtılma ihtiyacı doğmaktadır. Ağırlıklar ana gruplardan başlanarak her bir düzeyde yeniden dağıtılmış, bu sayede ağırlık yapısı korunmuştur. Ayrıca fiyat derlenemeyen ana grupların ağırlıkları diğer gruplara kendi ağırlıkları oranında dağıtılmıştır. Ana gruplar düzeyindeki TÜİK resmi ağırlıkları ve hesaplanan yeni ağırlıklar Tablo 3.3’te yer almaktadır.

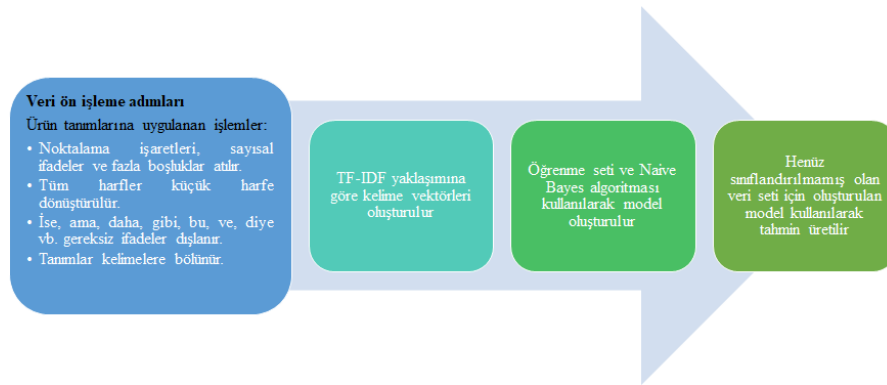
Fiyat derleme uygulamasına başlamadan önce web sayfalarının linkleri ve bu linklerde yer alan ürünlere atanacak COICOP kodları bir tabloda kayıtlanmıştır. Oluşturulan bu tablo çalışmaya altlık teşkil etmektedir ve izleyen bölümde fiyatı derlenen ürünlerin sınıflandırılmasında birincil olarak bu tabloda yer alan kodlamalar kullanılmıştır. Bu yöntem otomobiller, akaryakıtlar, cep telefonu görüşme ücretleri, giyim ürünleri ve ayakkabılar gibi birçok madde için uygun sonuçlar sağlamakla birlikte her ürün için aynı durum geçerli değildir. Özellikle market ürünlerinde web sayfalarının COICOP sınıflamasına uygun biçimde bölünmesi mümkün olmamaktadır. Örneğin nohut, mercimek, pirinç, bulgur gibi bakliyat ürünlerine bazı web sitelerinde tek bir link üzerinden bağlantı yapılmaktadır. Bu ve benzeri durumlar için ayrıca bir kod atama çalışması yapılması gerekmektedir. Çalışmada fiyatı kazınan her ürüne COICOP 7’li düzeyde kod ataması yapılmıştır. Bu düzey kamuoyu ile paylaşılan en detaylı sınıflama düzeyidir.



Şekil 3.1 Ürünlerin sınıflanma süreci.

Ürünlerin kod atama süreci Şekil 3.1’de özetlenmiştir. Kod atama uygulaması geliştirilmeden önce manuel olarak yaklaşık yetmiş bin ürüne kod atanması işlemi gerçekleştirilmiş, “if else” bloklarından oluşan kod atama modülü oluşturulmuştur. Bu modülde yalnızca market ve teknoloji market ürünleri kapsama alınmıştır. Ayrıca altlık tablosunda yer alan linklere karşılık tanımlanmış olan kodlardan güvenilir olanlar tespit edilerek ilgili kodların doğrudan atanması sağlanmıştır. Ürünler sınıflama işlemine sokulmadan önce hali hazırda kodlanmış ürünler arasında yer alıp almadığına bakılmaktadır. Daha önce sınıflama işlemi yapılmışsa aynı sınıf etiketi atanarak sınıflandırma işlemi tamamlanmaktadır. Daha önce sınıflama işlemi yapılmamış ise altlık tablosunda ürün linkine karşılık gelen sınıf etiketinin doğru kabul edilen sınıflar arasında yer alıp alınmadığı incelenmektedir. Ürünün kabul sınıflar arasında yer alması halinde bu sınıf etiketi doğrudan atanmakta aksi halde yazar tarafından oluşturulan kod atama modülü ve metin madenciliği uygulamasının yer aldığı metin madenciliği modülüne sokularak birer sınıf etiketi atanmaktadır. Son olarak altlık tablosunda yer alan sınıf etiketi de dikkate alınarak çoğunluk oyu yöntemi ile sınıflama işlemi tamamlanmaktadır. Çoğunluk oyu yöntemi kod ataması için yeterli değil ise yani atanan üç sınıf etiketi de birbirinden farklı ise kod atama işlemi manuel olarak gerçekleştirilmektedir.

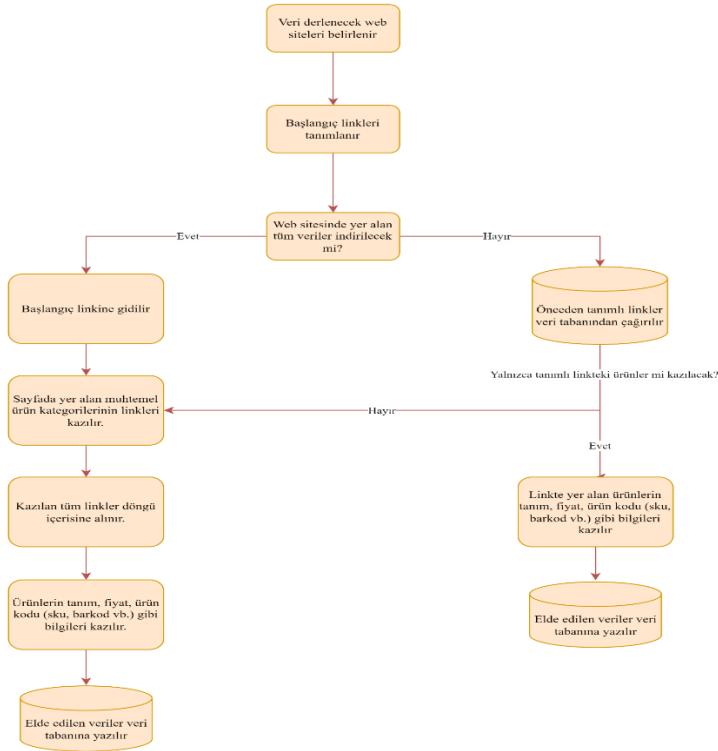
Metin madenciliği uygulamasına ilişkin yürütülen süreç ise Şekil 3.2’de yer almaktadır.



Şekil 3.2 Metin madenciliği uygulaması.

Buna göre önce metinlere ön işleme yapılarak metinler standart bir formata dönüştürülmüş devamında kelime vektörleri oluşturulmuştur. Tahminde Naive Bayes algoritması kullanılmıştır.

Çalışmada kullanılan fiyatlar ise web kazıma yöntemleri kullanılarak internet üzerinden derlenmiştir. Uygulamada Python dili kullanılmış Selenium ve BeautifulSoup paketlerinden yararlanılmıştır.



Şekil 3.3 Metin madenciliği uygulaması.

Çalışmada öncelikle karar örneklemesine göre fiyatı derlenecek işyerleri seçilmiştir. Fiyatların derlenmesi noktasında iki tip uygulama yapılabilmektedir. İlk olarak web sitesinde yer alan tüm ürünlerin fiyatları derlenebilir (Toplu kazıma). Bu yöntemde TÜFE sepetinde yer almayan ürünlerin fiyatları da derleneceğinden bunların ayıklanması ve COICOP kodu atanması zorlaşacaktır. Diğer yöntem ise hedeflenmiş kazımadır. Bu yöntemde ise önceden tanımlanmış başlangıç linkleri ve altında yer alan linklerdeki veriler kazılmaktadır. İkinci yöntemin dezavantajı ise takip edilecek link sayısı çoğalacağından bunların sağlıklı olup olmadığının takip edilmesinden kaynaklanan zaman maliyetidir. Bu çalışmada hibrit bir yaklaşım kullanılmış olup hem toplu kazıma hem de hedeflenmiş kazıma yöntemleri birlikte kullanılmıştır. Çalışmada kullanılan fiyat derleme süreci Şekil 3.3'te yer almaktadır. Buna göre hangi web sitelerinden hangi yöntemle veri derleneceği kararlaştırılır. Toplu kazıma yapılacak web sitesinin başlangıç linkine gidilir ve sitede yer alan sayfalar gezilerek fiyat kazılacak linkler taranır. Bu linkler döngü içerisine alınarak endeks hesaplamasında ihtiyaç duyulan ürün kodu, barkod, ürün tanımı, fiyat gibi veriler derlenir. Hedeflenmiş kazıma yapılacak ise zaten başlangıç linkleri önceden tanımlanmış olduğu için bu linklere gidilerek varsa alt linkler de toplu kazıma yöntemine benzer şekilde taranır yoksa doğrudan tanımlı linklerdeki veriler derlenerek süreç tamamlanır.

TÜFE çalışmalarında en çok satış yapan işyerlerinden en çok satılan ürünlerin fiyatlarının derlenmesi ve mümkün olduğu ölçüde aynı ürünün fiyatının izlenmesi esastır. Fiyat değişimlerinin kalite faktöründen bağımsız olması gerekmektedir. Bu nedenle aynı ürünün fiyatının herhangi bir nedenle izlenememesi durumunda benzer kalitede yeni ürünlerin ikame alınması, fiyat değişiminin ikame alınan ürünün kalitesinden kaynaklanması durumunda ise uygun kalite düzeltme yöntemleri kullanılarak fiyatların kalite etkisinden arındırılması gerekmektedir.

Çalışmada; ulusal düzeyde faaliyet gösteren ve online satış gerçekleştiren işyerlerinden yirmi milyonun üzerinde fiyat derlenmiş, endeks hesaplaması fiyatı sürekli olarak derlenebilen ürünler üzerinden madde düzeyinde Jevons endeks formülü, sınıflamanın hiyerarşik olarak daha üst düzeylerinde ise Laspeyres endeks formülü kullanılarak gerçekleştirilmiştir. Ürün takibi mümkün olduğu ölçüde işyerlerince sağlanan barkod, sku (stock keeping unit) ve ürün kodları ile yapılmıştır. Barkod, sku veya ürün kodlarına ulaşamadığı hallerde ürün tanımlarından faydalanılmıştır. Endeks hesaplamasında fiyatları devamlı olarak takip edilebilen ürünler kullanıldığından herhangi bir kalite düzeltmesi ihtiyacı ortaya çıkmamıştır.

Genel uygulama fiyatları sürekli derlenen ürünlerin hesaplamada kullanılması olmakla birlikte bazı istisnai uygulamalar da kullanılmıştır. Bu uygulamalar şunlardır:

- Gerçek kiraların derlenmesinde sürekli aynı konutun fiyatının izlenmesi mümkün olmamakla birlikte değişen konutların aynı kalitede olup olmadığı da bilinmemektedir zira kiraları etkileyen çok sayıda etken bulunmaktadır. Bunun yanı sıra gerçek kiraların takip edilmesinde stoktaki kiracıların başka bir deyişle

mevcut kontratı devam eden kiralardan fiyat gelişmelerinin de izlenmesi gerekmektedir. İnternet üzerinden bu fiyatlar derlenmemektedir. Bu nedenlerle gerçek kiralardan metodolojiye uygun bir biçimde tamamen internet üzerinden derlenmesi uygun değildir. Ancak gerçek kiralardan ağırlığının çok yüksek olması nedeniyle çalışmada konutların yaş aralıkları, oda sayıları, lokasyonları ve metre kareleri sınırlanarak bu değişkenlerin belirlenen aralıklarda tutulması sağlanmış ve endeks hesaplamasında kullanılmıştır.

- Paket tur fiyatlarının derlenmesinde ulaşım araçları ve konaklama koşullarının sabit tutulması mümkün olmamıştır.
- Otel fiyatlarının derlenmesinde tek kişi için gecelik fiyatlar derlenmiştir. Ancak standart oda, aile odası, süit oda vb. ayrımına gidilmemiştir.
- TÜFE metodolojisi gereği hanehalkı tüketimine konu mal ve hizmet fiyatları endeks hesaplamasında kullanılmalıdır. Bu bağlamda altın fiyatlarının derlenmesinde gram altın fiyatları yatırım harcaması olarak değerlendirilmiş altın bileklik fiyatları derlenerek kullanılmıştır.
- İnternet üzerinden satışı yasaklı olan bütün ürünlerinin fiyatları manuel olarak girilmiştir.

Çalışmada fiyat derleme periyodu 1 Aralık 2021- 25 Nisan 2022 tarihleridir. Fiyat derleme sürecinin başlangıç tarihinin 1 Aralık olarak belirlenme nedeni resmi istatistik kapsamında kullanılan ağırlık setinin referans döneminin 2021 Aralık ayı olmasıdır. Yeni bir ağırlık setinin hesaplanmasına olanak sağlayacak daha zamanlı bir veri bulunmamaktadır. Gerek günlük endekslerin gerekse aylık endekslerin hesaplanmasında aynı ağırlık seti ve aynı fiyatlar kullanılmaktadır. Endeks hesaplamasının günlük olması bakımından veri derlenen zaman periyodunun yeterli olduğu değerlendirilmektedir.

#### 4. Bulgular

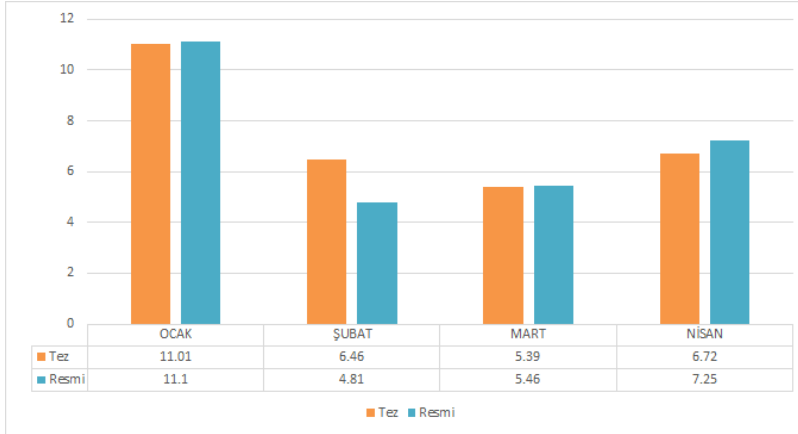
Bu bölümde internet üzerinden derlenen fiyatlar kullanılarak hesaplanan aylık endeks değişim oranları ile resmi istatistik kapsamında hesaplanan TÜFE verilerine karşılaştırmalı olarak yer verilmiştir. Her ne kadar bu bölümde iki verinin karşılaştırması yapılsa da aşağıdaki nedenlerle veriler tam anlamıyla karşılaştırılabilir değildir (TÜİK, 2022):

- Kapsam farklılıkları,
- İşyeri ve ürün örneklemelerinin farklı olması,
- Veri derleme sıklıklarının farklı olması,
- Yerel etkilerin göz ardı edilmesi.

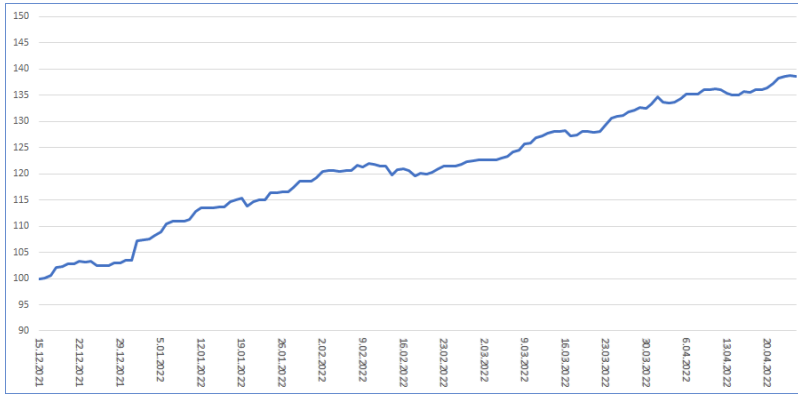
Çalışmada TÜFE genel tahminlerine yönelik hesaplamada kullanılan fiyatlar toplam ağırlığın %75,25'inden oluşmaktadır.

Tablo 4.1'de TÜFE genel endeksine ilişkin aylık değişim oranları yer almaktadır. Buna göre resmi istatistiklerde bir önceki aya göre değişim oranları ocak ayında %11,1, şubat ayında %4,81, mart ayında %5,46 ve nisan ayında %7,25'tir. Bu çalışma bağlamında internet üzerinden derlenen fiyatlarla oluşturulan endekse göre ise ocak ayında %11,01 olan endeks değişim oranı şubat ayında %6,46, mart ayında %5,39 ve nisan ayında %6,72 olmuştur. Buna göre iki hesaplamadan elde edilen sonuçlar ciddi bir farklılık göstermemektedir. Dört aylık kümülatif değişim oranları ise resmi istatistiklere göre %31,71 iken çalışma bulgularında bu oran %32,91 olarak tahmin edilmektedir. Buradan hareketle genel endeks noktasında çok benzer sonuçlar elde edildiği görülmektedir. TÜİK tarafından gerçekleştirilen TÜFE çalışması ile örneklem ve kapsam farklılıkları olmakla birlikte kullanılan endeks formülleri aynıdır. Bunun yanı sıra gerek ağırlıkların gerekse fiyatların referans dönemleri de TÜFE çalışmasına göre belirlenmiştir. İki çalışma her ne kadar tamamen karşılaştırılabilir olmasa da elde edilen sonuçların benzer olması tahmin güvenilirliği açısından olumlu olarak değerlendirilmektedir.

**Tablo 4.1.** Genel endeks aylık değişim oranları, %.



Şekil 4.1'de ise genel endeksin günlük değerleri yer almaktadır. Buna göre endekste dönem dönem düşüşler olmakla birlikte genel trend yukarı yönlüdür. Ocak ayı başında özellikle yönetilen ve yönlendirilen fiyatlara gelen zamları etkisi belirgin biçimde görülmektedir.



**Şekil 4.1.** Genel endeks günlük değerleri.

Çalışma sonucunda elde edilen diğer önemli bulgular ise gelecek çalışmalara da altlık teşkil edecek olan karşılaşılan sorunlar ve dikkat edilmesi gereken hususlardır. Bunlar aşağıda özetlenmiştir:

- Her web sitesinin kendine özgü bir sayfa yapısı bulunması nedeniyle her biri özel bir çalışma gerektirmekte bu web sitelerindeki değişikliklerin izlenerek internet gezgininin güncellenmesi gerekmektedir.
- İnsan tanıma uygulamaları internet gezgininin çalışmasını aksatmaktadır.
- İnternet gezgini, web siteleri tarafından tespit edilmesi durumunda kalıcı veya geçici olarak yasaklanabilmektedir. Bu nedenle tasarım olabildiğince insan davranışını yansıtacak biçimde olmalıdır. Bununla birlikte "crawl delay" benzeri yapılarla kazıma işlemi yavaşlatılabilmektedir.
- Bölgesel fiyatlandırma davranışları endekste yansıtılamamaktadır.
- Hanehalkı tüketiminde önemli paya sahip bazı işyerleri (Pazar, manav veya çevrimiçi satış yapmayan marketler) kapsanamamaktadır.
- Güçlü mevsimsel özellik gösteren ürünler sabit ağırlık sisteminde kapsanamamaktadır. Bir yıllık döngünün tamamlanması durumunda imputasyon yöntemleri uygulanabilir.
- Miktar veya adet değişiminden kaynaklanan kalite değişimleri izlenememektedir (Örn. 10 adet olarak satılan bir paket bisküvinin 8 adete düşürülmesi). Bu sorunun çözümü için kayıtlara miktar ve ölçü birimi (Adet, paket, litre, gram, kg vb.) ataması yapılması gerekmektedir.
- Hizmetler ve çevrimiçi satışı yasak olan ürünlerin fiyatları derlenememektedir.
- Ayrı bir HTML etiketi ile ifade edilen indirimlerle ilgili dikkatli olunmalıdır. Birçok web sitesinde indirimli fiyatlar ayrı olarak gösterilmektedir. Genele yaygın indirimlerin endeks hesabında dikkate alınması gerekmektedir.
- Sepet indirimi olarak tanımlanan ve tüm tüketicilere eşit olarak uygulanan indirimlerin endeks hesabına yansıtılması gerekmektedir. Ancak indirimli fiyat sepete gidildiğinde görülebildiğinden bu fiyatlar derlenememiştir.

#### 4. Sonuç ve Tartışma

TÜFE; hanehalklarının tüketimine konu olan mal ve hizmetlerin fiyatlarının genel düzeyinde zaman içerisinde meydana gelen değişimlerin bir ölçüsü olmakla birlikte maaş ve ücretlerin belirlenmesinden kira artışlarına, makroekonomik göstergelerin deflate edilmesinden parasal istikrarın ölçülmesine kadar çok geniş kullanım alanı bulunan bir göstergedir. Bu bağlamda TÜFE'nin uluslararası standartlara uygun, güvenilir ve zamanlı olması gerekmektedir. Günümüzde teknolojinin gelişmesiyle birlikte bu ihtiyaca cevap verecek yeni veri derleme yöntemleri ortaya çıkmış ve bu yöntemler daha yüksek frekanslarda veri derlenebilmesinin önünü açarak daha yüksek frekanslı TÜFE hesaplamalarının yapılabilmesini sağlamıştır. Web kazıma yöntemi bu yöntemlerin arasında yer almaktadır. Web kazıma yöntemi ile internet üzerinde yer alan bilgilerin otomatik bir şekilde alınması, yapısal veriye dönüştürülerek veri tabanlarına kaydedilmesi sağlanmaktadır. TÜFE çalışmaları bağlamında web kazıma yöntemi ile internet üzerinden fiyatlar, fiyatı derlenen ürünün tanımı ve özellikleri otomatik olarak derlenebilmektedir.

Bu çalışmada uluslararası TÜFE hesaplama yöntemlerine mümkün olan en detaylı biçimde uyulmuş bu yöntemlerden sapmalar objektif biçimde ortaya konulmuştur. Ayrıca web kazıma yöntemi ile fiyatının TÜFE metodolojisine uygun olarak derlenmesi mümkün olmayan ürünlere yer verilmektedir. Çalışma bu özellikleri bakımından gelecek çalışmalara güçlü bir temel oluşturma amacına hizmet etmektedir. Çalışma neticesinde ulaşılan bir diğer sonuç TÜFE bakımından web kazıma ile elde edilen verilerin geleneksel yöntemler kullanılarak derlenen verilere göre üstünlük ve zayıflıklarının ortaya konulmuş olmasıdır. Buna göre yüksek frekans ve düşük maliyetle veri derlenebilmesi, veri giriş hatalarının düşük olması, çok yüksek örnek hacimlerine ulaşılabilmesi web kazımının üstünlükleri iken ikame ve kalite düzeltilmesi süreçlerinin otomasyonundaki zorluklar, bölgesel hareketlerin kapsanamaması, satış verilerinin elde edilememesi nedeniyle ürünlerin hanehalkı tüketimindeki öneminin göz ardı edilmesi, verisi kazınan fiyatların sınıflandırılmasındaki zorluklar ve önceki bölümlerde belirtilen bazı ürünlerin fiyatlarının metodolojiye uygun olarak derlenememesi web kazımının zayıflıkları olarak karşımıza çıkmaktadır. Ayrıca web kazıma işlemi için kullanılan botun web siteleri tarafından tespit edilerek kısıtlanması, sunucu isteklerinin geciktirilmesi uygulamaları, insan tanıma uygulamaları, bazı verilerin PDF formatındaki dosyalar biçiminde sunulması, her bir web sayfasının yapısal özelliklerinin farklı olması nedeniyle özel olarak ele alınması ve bu web sitelerinde meydana gelen değişikliklere göre programın güncellenmesinin gerekmesi gibi başka zorluklar da bulunmaktadır.

Veri derleme süreçlerinde yaşanan zorluklar, gerçek kira başta olmak üzere TÜFE metodolojisine uygun şekilde fiyatı derlenemeyecek ürünlerin varlığı gibi sorunlar olmasına karşılık elde edilen bulguların resmi istatistiklerle tutarlı olması, web kazıma yönteminin TÜFE çalışmasında kullanılabilir anlamı bir alternatif veri derleme metodu olduğunu göstermiştir.

Gelecekte yapılacak çalışmalarda özellikle devir hızı yüksek ürünler başta olmak üzere ürün ikamesi ve kalite düzeltilmesi süreçlerinin otomasyonu, gerçek kira fiyatlarının metodolojiye uygun biçimde hesaplanabilmesi için stok kiracı oranına bağlı bir hesaplama yöntemi geliştirilmesi ve kapsama oranının artırılması hususlarının ele alınması gerekmektedir. Özellikle yüksek frekans ve hacimli verilerin TÜFE çalışmalarında en etkin şekilde kullanılmasını amaçlayan çok boyutlu endeks hesaplama yöntemlerinin kullanımına yönelik çalışmalar yapılması gerekmektedir.

Çalışmanın sonuçları resmi istatistiklerde kullanılan hesaplama yöntemi ile uyumlu bir biçimde kullanılacak web kazıma verilerinin bir hesaplama yanlılığına neden olmadığını göstermektedir. Bu bakımdan istatistik kurumlarının aylık veya yıllık zincirleme şeklinde eşleşen modeller yaklaşımını kullanarak bu verilerin endeks hesaplamalarında kullanılması veya kapsamının genişletilmesinin değerlendirilmesi tavsiye edilmektedir.

**Citation / Atıf:** YALÇIN Y., BAYSAL B. (2024). Web Kazıma Yöntemlerinin Tüketici Fiyat Endeksi Çalışmalarında Kullanılması Türkiye Örneği. *İstatistik Araştırma Dergisi*, 14 (2), 46-58.

## Kaynaklar

- Armknacht, P. (1996). Improving the Efficiency of the U.S. CPI. *IMF Working Papers*, 96(103).
- Castriello-Fernández, O. (2015). *Web Scraping: Applications and Tools*. European Public Sector Information Platform. [https://data.europa.eu/sites/default/files/report/2015\\_web\\_scraping\\_applications\\_and\\_tools.pdf](https://data.europa.eu/sites/default/files/report/2015_web_scraping_applications_and_tools.pdf), Erişim Tarihi:10.12.2021).
- Cavallo, A., and Rigobon, R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. *Journal of Economic Perspectives*, 30(2), 151-178.
- EUROSTAT. (2018). *Harmonized Index of Consumer Prices Methodological Manual*. Luxembourg: Publications office of the European Union.
- EUROSTAT. (2020). Practical Guidelines on Web Scraping for the HICP. (<https://ec.europa.eu/eurostat/web/hicp/methodology>, Erişim Tarihi: 10.12.2021)
- EUROSTAT. (2022). *Guide on Multilateral Methods in the Harmonised Index on Consumer Prices*. Luxembourg: Publications Office of the European Union.
- Fenwick, D. (2009). *Practical Guide to Producing Consumer Price Indices*. New York: UN. (<http://digitallibrary.un.org/record/670125>, Erişim Tarihi: 01.12.2021).
- Gençer, S., ve Ulusoy, V. (2021). Dinamik Ve Yüksek Frekanslı Fiyat Endeksi: Türkiye'nin Günlük Ve Saatlik Bazda Enflasyon Hesaplaması. *Maliye ve Finans Yazıları*, (Özel Sayı 2), s. 173-190.
- Graf, B. (2020). *Consumer Price Index Manual*. USA: International Monetary Fund. (<https://www.elibrary.imf.org/view/book/9781484354841/9781484354841.xml>, Erişim Tarihi: 12.12.2021).
- Griffioen, A. R., & Ten Bosch, O. (2016). On the use of Internet data for the Dutch CPI. CPI. In UNECE Meeting of the Group of Experts on Consumer Price Indices, Geneva.
- Mitchell, R. (2015). *Web Scraping with Python: Collecting Data from Modern Web*. Sebastopol: O'Reilly.
- Richardson, L. (2019). *Beautiful Soup Documentation*. (<https://beautiful-soup-4.readthedocs.io/en/latest/pdf/>, Erişim Tarihi: 04.15.2022).
- Santiago, E. A., and Sarah, M. L. (2020). Tracking inflation on a daily basis. *Swiss Journal of Economics and Statistics*, 156, 1-13.
- Selenium. (2022). Selenium Documentation. (<https://www.selenium.dev/documentation>, Erişim Tarihi: 03.01.2022).
- Scrapy developers. (2022). Scrapy Documentation. (<https://docs.scrapy.org/en/latest/>, Erişim Tarihi: 09.04.2022).
- TÜİK. (2022). Tüketici Fiyat Endeksi Haber Bülteni. (<https://data.tuik.gov.tr/Bulten/Index?p=Tuketici-Fiyat-Endeksi-Ocak-2022-45790>, Erişim Tarihi: 05.04.2022).
- TÜİK. (2024). TÜFE Metodoloji Dokümanı. (<https://data.tuik.gov.tr/Bulten/Index?p=Tuketici-Fiyat-Endeksi-Ekim-2024-5361>, Erişim Tarihi: 12.11.2024).
- U.S Bureau of Labor Statistics. (2020). Handbook of methods. (<https://www.bls.gov/opub/hom/cpi/home.htm>, Erişim Tarihi: 10.10.2022).





## Performance Comparison of Least Squares, Ridge, Lasso and Principal Component Regression for Addressing Multicollinearity in Regression Analysis

Semih ERGİŞİ

Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı / Lisansüstü Öğrencisi  
semihstat@gmail.com

Orcid No: 0009-0007-1364-1252

Beyza DOĞANAY

Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı / Doç. Dr.  
bdoganay@medicine.ankara.edu.tr

Orcid No: 0000-0001-8845-2287

Yasemin YAVUZ

Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı / Prof. Dr.  
dr.yasemin.yavuz@gmail.com

Orcid No: 0000-0003-1661-9468

### Abstract

The purpose of this research was to evaluate the predictive accuracy of various regression methods in the context of multiple linear regression when multicollinearity invalidates the underlying assumptions of the least squares method. These methods included least squares regression (LS), ridge regression (RR), lasso regression (LR), and principal component regression (PCR). For this aim, the dataset including 6 variables simulated from normal with different sample of size from range of 50 to 1000. The performance was assessed using mean square error (MSE) and R square value. Despite the existence of multicollienarity among independent variables, research findings showed that LS method had the smallest MSE in the training dataset but RR had the smallest mse in the test dataset. When the sample size increases, the mse values increase for each methods in the training set but decrease in the test set. They are closer to each other. In terms of R square values, all methods showed similar performance both training and test data set.

*Keywords: Lasso regression, Principal components, Multicollinearity, Least square regression.*

**Sorumlu Yazar / Corresponding Author:** Semih ERGİŞİ, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı.

**Atf / Citation:** ERGİŞİ S., DOĞANAY B., YAVUZ Y. (2024). Performance Comparison of Least Squares, Ridge, Lasso and Principal Component Regression for Addressing Multicollinearity in Regression Analysis. *İstatistik Araştırma Dergisi*, 14 (2), 59-72.

## Regresyon Analizinde Çoklu Doğrusallığın Ele Alınması için En Küçük Kareler, Ridge, Lasso ve Temel Bileşen Regresyonunun Performans Karşılaştırması

### Özet

Bu araştırmanın amacı, çoklu doğrusal regresyon bağlamında çeşitli regresyon yöntemlerinin tahmin doğruluğunu değerlendirmektir. Bu bağlamda, en küçük kareler yönteminin (LS) temel varsayımlarını geçersiz kılabilen çoklu bağlantı sorunları altında çeşitli yöntemler incelendi. İncelenen yöntemler arasında en küçük kareler regresyonu (LS), ridge regresyon (RR), lasso regresyon (LR) ve temel bileşen regresyonu (PCR) yer aldı. Bu amaçla, 50 ila 1000 arasında değişen örneklem boyutlarına sahip, normal dağılımdan simüle edilmiş 6 değişken içeren bir veri seti kullanıldı. Performans, hata kareler ortalaması (hko) ve R kare değeri kullanılarak değerlendirildi. Çoklu bağlantı sorunu olmasına rağmen, araştırma bulguları, LS yönteminin eğitim veri setinde en küçük hata kareler ortalamasına sahip olduğunu, ancak RR'nin test veri setinde en küçük hata kareler ortalamasına sahip olduğunu gösterdi. Örneklem boyutu arttıkça, her yöntem için eğitim setindeki hata kareler ortalaması değerleri arttı ancak test setinde azaldı ve yöntemler birbirine daha yakın hale geldi. R kare değerleri açısından, tüm yöntemler hem eğitim hem de test veri setlerinde benzer performans gösterdi.

*Anahtar sözcükler: Lasso regresyon, Temel bileşenler, Çoklu bağlantı, En küçük kareler regresyon.*

### 1. Introduction

Regression analysis is the frequently utilized statistical method for predicting the quantitative relationship between a dependent variable ( $Z$ ) and one or more independent variables (Draper, 1966). The general applications of regression analysis are stated in the two main parts. One of this is to summarize data, relationship between dependent and independent variables while the other is to estimate the future values of dependent variable by using independent variables (Montgomery et al., 2021).

The most commonly utilized methodology in the regression analysis is the Least Squares (LS) method, which is dependent upon the satisfaction of certain necessary assumptions (Wooldridge, 2016). This approach is founded on the concept of minimizing the total sum of squares as the difference between observed  $y$  values and predicted  $y$  values from the regression equation (Altland, 1999). The dependability of the produced model is related to the fulfillment of the presumptions inherent in the LS method. If there is a considerable amount of multicollinearity exists among the independent variables, the regression coefficients obtained using the Least Squares method may lead to misinterpretation of the results (Alpar, 2017).

Ridge regression (RR) and Lasso regression (LR) are popular methods for handling multicollinearity in regression analysis. They are specifically designed to mitigate the adverse impact of multicollinearity on parameter predictions in MLR analysis. Principal Components Regression (PCR) is a regression technique that aims to elucidate original variables with intercorrelations by utilizing a reduced set of newly derived variables, which are linear combinations of the original variables (Alpar, 2017).

In disciplines like environmental science, public health, engineering, and finance, the utilization of the LS method for making predictions without adhering to its fundamental assumptions can lead to inaccurate deductions. In the finance sector, failing to consider multicollinearity among economic indicators may lead to untrustworthy assessments of investment risk (Gujarati & Porter, 2009). As a result, it is important to thoroughly examine the accuracy of the outcomes derived from the LS method when its assumptions are not sufficiently fulfilled.

Nevertheless, in research conducted within this framework, the existence of multicollinearity, a crucial factor in the field of statistics, and resolution methods are not afforded the requisite attention. In addition, despite the decreasing effectiveness of MLR analysis, the optimal method for addressing multicollinearity remains uncertain. A variety of approaches have been implemented to ensure the prevention of multicollinearity issues. Some examples of these techniques include lasso regression (LR), ridge regression (RR), and principal component regression (PCR). These three approaches are commonly employed to address issues of multicollinearity as they

are capable of producing effective estimators for predictors, making them a focus of interest for many researchers. This research paper examines and compares the performance of four selected regression techniques of LS, Lasso, Ridge, and PCR across different sample sizes. The Mean Squared Error (MSE) and R square metrics are used to evaluate the performance of various regression models.

## 2. Dataset and Method

In this study, Monte Carlo simulation techniques were utilized to create the dataset, which comprised six continuous variables, with one designated as the dependent variable and five as independent variables. To ensure the robustness of our findings across different sample sizes, datasets were simulated for sample sizes of 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, and 1000, with 1000 independent simulations conducted for each sample size.  $X_1$  was generated as random values drawn from a normal distribution with a mean of 3 and a standard deviation of 5.  $X_2$  was derived as a linear transformation of  $X_1$ , specifically calculated as in the Equation 2.1.  $X_3$  was derived as linear combination of  $X_1$  and  $X_2$  and it was given in the Equation 2.2.  $X_4$  was independently generated as random values from a normal distribution with a mean of 3 and a standard deviation 3.  $X_5$  was created as linear combination of  $X_1$ ,  $X_2$ ,  $X_3$  and it was given in the Equation 2.3.

$$X_2 = 4 * X_1 + e_2, \quad e_2 \sim N(0.5, 2) \quad (2.1)$$

$$X_3 = 2 * X_1 + 0.6 * X_2 + e_3, \quad e_3 \sim N(5, 2) \quad (2.2)$$

$$X_5 = X_1 + X_2 - X_3 + e_5, \quad e_5 \sim N(0, 0.1) \quad (2.3)$$

The dependent variable  $Y$  was then created as linear combination of independent variables using specified weights and a random error term was added to introduce variability. The simulation of the  $Y$  variable was given in the Equation 2.4.  $\omega_1 = 1.5$ ,  $\omega_2 = 2$ ,  $\omega_3 = 0.5$ ,  $\omega_4 = 0.8$  and  $\omega_5 = 1.2$  were initial weights.

$$Y = \omega_1 X_1 + \omega_2 X_2 + \omega_3 X_3 + \omega_4 X_4 + \omega_5 X_5 + e, \quad e \sim N(0, 20) \quad (2.4)$$

While MLR models are often seen using matrix notation, Equation 2.5 provides a general outline of the regression analysis framework.

$$\gamma = \alpha_0 + \alpha_1 * \omega_1 + \alpha_2 * \omega_2 + \alpha_3 * \omega_3 + \dots + \alpha_p * \omega_p + \varepsilon \quad (2.5)$$

In Equation 2.5, the dependent variable is on the left side, while the independent variables are on the right side. This equation mathematically expresses the portion of the variation in the dependent variable that can be explained by the independent variables. The error term in Equation 2.5 represents the portion of the variation that is not explained by the independent variables. In order to see the significance of multicollinearity, correlation matrices for selected sample of sizes (50, 200, 500, 900) are given in the Appendix.

### 2.1. Least Square Regression

The method of LS regression is frequently employed in statistical analysis to establish a relationship between the dependent variable and the predictor variables (Miles & Shevlin, 2000). This method is effective unless confronted with the issue of multicollinearity, which occurs when two or more explanatory variables are closely linearly correlated (Altland, 1999). The objective of this approach is to reduce the total of squared error values, under the assumption that the error values adhere to a normal distribution with consistent variance, and to enhance the model based on this (Weisberg, 1985).

By leveraging the power of the least squares regression method, designed to minimize residuals, meticulous researchers and astute analysts can achieve an unquestionably reliable estimation of coefficients. This increases statistical inference which allows more investigation and interpretation of complex data sets (Draper, 1966). Furthermore, the LS regression method stands out as an important aspect of many regression techniques. Simplicity and efficiency of LS regression makes its status as a necessary data analysis tools which allow researchers and analysts to go deeper into the field's ever-increasing depth and complexity (Alpar, 2017). The regression equation is written in the form of weighted averages of independent variables and it is given in Equation 2.6. In the Equation 2.6,  $\varepsilon$  is normally distributed with zero mean and  $\sigma$  standard deviation. In the Equation 2.6,  $\omega_i$ ,  $i = 0, 1, \dots, k$  are weights, and learned from training data. These parameters can be estimated by minimizing the Equation 2.7 that minimizes the sum squared difference between observed  $Y$  values and estimated  $\hat{Y}$  values.

$$Y = \omega_0 + \omega_1 * X_1 + \omega_2 * X_2 + \dots + \omega_k * X_k + \varepsilon, \quad \varepsilon \sim N(0, \sigma) \quad (2.6)$$

$$\hat{Y} = \hat{\omega}_0 + \hat{\omega}_1 * X_1 + \hat{\omega}_2 * X_2 + \dots + \hat{\omega}_k * X_k \quad (2.7)$$

$$\min_{\omega} \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 \quad (2.8)$$

Minimizing the Equation 2.8, the regression parameters can be found by taking derivative with respect to  $\omega$  and equating it zero. In contrast, despite the benefits of the LS regression method, its application may be hampered by failure to meet certain basic assumptions (Kutner et al.,2004). A vital assumption is that the independent variables should not be significantly linearly correlated with each other (Kutner et al.,2004). Under the existence of multicollinearity, this can lead to increased variance in the LS estimates, potentially giving biased results that deviate from the true values (Kutner et al.,2004). Therefore, if there is a multicollinearity in a regression model, using different techniques instead of the LS method can reduce the variance and produce more reliable results (Kutner et al.,2004).

## 2.2. Ridge Regression

Ridge regression (RR) is a technique that aims to apply the L2 norm shrinkage penalty to the regression coefficients and thereby reduce the effects of multicollinearity in regression analysis containing multicollinear data (Hoerl & Kennard, 1970). The negative effects of multicollinearity can be reduced by constraining the coefficients from becoming too large. The penalty term applied in the RR method reduces the coefficient estimates towards zero by penalizing them without making them completely zero. As a result, some of the highly correlated predictors are allowed to remain in the model, albeit with reduced magnitudes. (Zoe & Hastie, 2005). While it is not aimed to select variable in the scope of RR, it is aimed to reduce the effect of coefficients and by this way increase the performance of model on the test data set. In the training data set, the performance of the model is not as good as the LS model (Müller & Guido, 2016).

Ridge regression excels in this aspect by permitting the retention of all predictors in the model without elimination. However, in addition to preserving the relevant predictors, the L2-norm reduction method forces the important coefficients to become smaller, bringing them closer to zero. This algorithm balances the coefficients by adding a penalty term to the residual sum of squares (Müller & Guido, 2016). It has a tuning parameter that provides a balance between reducing the effect of the coefficients and reducing the variance of the model(s) (Tibshirani, 1996). In fact, the form of this penalty term can be expressed in the Equation 2.9.

$$\min_{\omega} \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 + \lambda \sum_{m=1}^k \omega_m^2 \quad (2.9)$$

In the Equation 2.9, the right hand side of the plus sign is a regularization term to the least squares objective function, which serves to get rid of overfitting and enhance the model's generalization performance (Müller & Guido, 2016). The inclusion of this penalty component pushes the model to strike a compromise between precisely fitting the data and keeping coefficients low. The coefficients are biased toward zero by this regularization procedure, which also keeps any one predictor from governing the model (Müller & Guido, 2016). Ridge regression is therefore especially remedy for datasets with multicollinearity. Ridge regression stabilizes the model's performance and produces more accurate coefficient estimates by lessening the effect of multicollinearity. Ridge regression is an effective method for enhancing a linear regression model's stability and predictive precision when multicollinearity is present (Hoerl & Kennard, 1970).

## 2.3. Lasso Regression

Lasso (Least Absolute Shrinkage and Selection Operator) regression (LR) has been widely used to address the problem of multicollinearity in regressors in high-dimensional datasets (Müller & Guido, 2016). This is because the L1 penalty used in lasso regression allows for the automatic reduction and deletion of coefficients associated with weak or irrelevant variables, substantially limiting the influence of multicollinearity (Müller & Guido, 2016). Consequently, lasso regression increases model interpretability while also increasing predicted accuracy by saving only the most significant factors. The feature-selective behavior of Lasso regression is particularly useful when the effects of individual predictors that are truly important to the model are obscured due to multicollinearity. By penalizing the coefficients of variables which leading to multicollinearity, their influences are reduced effectively by Lasso regression, thereby enhancing the interpretability and robustness of the model (Fu ,1998). In the mathematical logic under the lasso regression is that it applies an L1 regularization technique, which encourages sparsity in the model. As a result, some coefficients being shrunk to zero, effectively performing variable selection and allowing for a simpler model that includes only the most important variables. The logic behind LR is that it adds a penalty term to the loss function, which discourages complexity and helps mitigate the impact of

multicollinearity. In Equation 2.10, the mathematical expression of this model is given. In the Equation 2.10, the right hand side of the plus sign is a regularization term to the least squares objective function.

$$\min_{\omega} \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 + \lambda \sum_{m=1}^k |\omega_m| \quad (2.10)$$

By decreasing the detrimental effect of multicollinearity, the coefficients of the regression model are stabilized, resulting in more credible and interpretable results. This strategy essentially decreases some coefficients to zero, thus conducting variable selection and increasing the model's predictive power (Müller & Guido, 2016).

## 2.4. Principal Component Regression

The Principal Component Regression (PCR) method is a statistical technique that allows researchers to estimate the coefficients of independent variables by performing multiple linear regression analysis between the observed dependent variable and the new variables obtained using the PCR method, without having to remove the independent variables in the presence of multicollinearity (Alpar,2017). In this kind of study used to deal with multicollinearity, researchers who choose to utilize this model need first acquire independent components using principal component analysis. They should then retrieve the coefficients of the original variables by applying the back transformation procedure described in Equations 2.11 and 2.12. Unscaled regression coefficients in terms of MLR is given in left hand side of the Equation 2.11,  $P$  is the matrix of Eigen values and  $\psi_{PCR}$  is the coefficients of principal components.

$$\omega' = P * \psi_{PCR} \quad (2.11)$$

$$\omega = \omega' * \frac{\sigma_Y}{\sigma_{x_i}} \quad (2.12)$$

The coefficients of independent variables are given in the left hand side of the Equation 2.12. In fact, this method not only provides a summary of how independent variables are related through linear combinations of a set of separate variables, but also overcomes the misinterpretation of original variables under the effect of multicollinearity by providing the opportunity to return to the original variables (Johnson, 1998). This method converts the original regression problem into a condensed regression structure in which the number of independent variables is reduced to the maximum number of principal components used in the PCA sequence to explain the variance while providing the benefit of dimensionality reduction (Hintze, 2007). The goal of this method is to apply the least squares method to a set of derived variables known as principle components, which are produced from the correlation matrix using a specific process (Göktaş, 2010). In the context of PCR, the multicollinearity situation can be addressed by estimating the regression coefficients through the application of the Least Squares regression method on a new set of variables (Topal, 2010). These fresh variables are acquired through conducting orthogonal transformations on the initial variables in the dataset (Ortabaş, 2001).

## 3. Result and Discussion

In this study, the performance metrics of four different models were compared, presented in four distinct tables and one visual, with separate results for the training and test data. The tables sequentially display the MSE values for various sample sizes across different models. The same structure applies to the R-squared values. Each table consists of 5 columns and 11 rows, where the first column contains the sample sizes, while the remaining columns provide the average MSE and R-squared values for each model.

When examining Table 1, the MSE values obtained from different models for various sample sizes in the training data are presented. According to the results, the MSE values obtained through LS analysis are the lowest across all sample sizes, while the highest average MSE is observed in PCR analysis.

When examining Table 2, the changes in average MSE values in the test set can be observed. In all sample sizes, the lowest average MSE value is obtained from the RR and LR analysis, while the highest average MSE value is obtained from the PCR analysis. The most notable observation in this table is that, for the dataset with a sample size of 50, the highest average MSE is recorded in the LS analysis.

From the results presented in Table 1 and Table 2 for the training and test sets, it is observed that as the sample size increases, the MSE values in the training set increase and converge across the three models, while in the test set, the MSE values decrease and converge. Furthermore, in terms of generalization performance, the RR and LR model demonstrates the best results, though with increasing sample size, the performance of LS, RR, and LR models also converges. Figure 1 visualizes the results provided in Table 1 and Table 2, showing the changes in average MSE values for each model in both the training and test sets as the sample size increases.

**Table 1.** Average MSE for training datasets.

Sample Size	LS	RR	LR	PCR
50	338,32	352,55	352,22	393,02
100	373,91	379,98	380,25	414,02
200	386,52	389,24	389,53	420,01
300	390,08	391,91	392,14	421,7
400	391,8	393,07	393,26	421,65
500	393,95	395,01	395,16	422,77
600	394,2	395,05	395,14	423,36
700	396	396,72	396,82	424,44
800	396,87	397,49	397,56	425,47
900	396,62	397,17	397,24	424,93
1000	397,39	397,88	397,94	425,45

**Table 2.** Average MSE for test datasets.

Sample Size	LS	RR	LR	PCR
50	475,29	446,96	447,2	461,48
100	433,02	424,1	423,46	445,18
200	414,95	411,6	411,37	433,06
300	413,92	411,42	411,53	434,8
400	408,11	406,43	406,48	430,79
500	403,36	402,49	402,61	427,44
600	408,34	407,33	407,36	432,81
700	406,36	405,62	405,7	430,6
800	401,63	400,82	400,81	426,4
900	403,19	402,59	402,62	428,27
1000	402,58	402,03	402,08	427,6

**Table 3.** Average R-squared for training datasets.

Sample Size	LS	RR	LR	PCR
50	0,92	0,91	0,91	0,90
100	0,91	0,91	0,91	0,90
200	0,91	0,91	0,91	0,90
300	0,91	0,91	0,91	0,90
400	0,91	0,91	0,91	0,90
500	0,91	0,91	0,91	0,90
600	0,91	0,91	0,91	0,90
700	0,91	0,91	0,91	0,90
800	0,91	0,91	0,91	0,90
900	0,91	0,91	0,91	0,90
1000	0,91	0,91	0,91	0,90

**Table 4.** Average R-squared for test datasets.

Sample Size	LS	RR	LR	PCR
50	0,85	0,86	0,86	0,85
100	0,88	0,88	0,88	0,88
200	0,90	0,90	0,90	0,89
300	0,90	0,90	0,90	0,89
400	0,90	0,90	0,90	0,90
500	0,90	0,90	0,90	0,90
600	0,90	0,90	0,90	0,90
700	0,90	0,90	0,90	0,90
800	0,91	0,91	0,91	0,90
900	0,91	0,91	0,91	0,90
1000	0,91	0,91	0,91	0,90

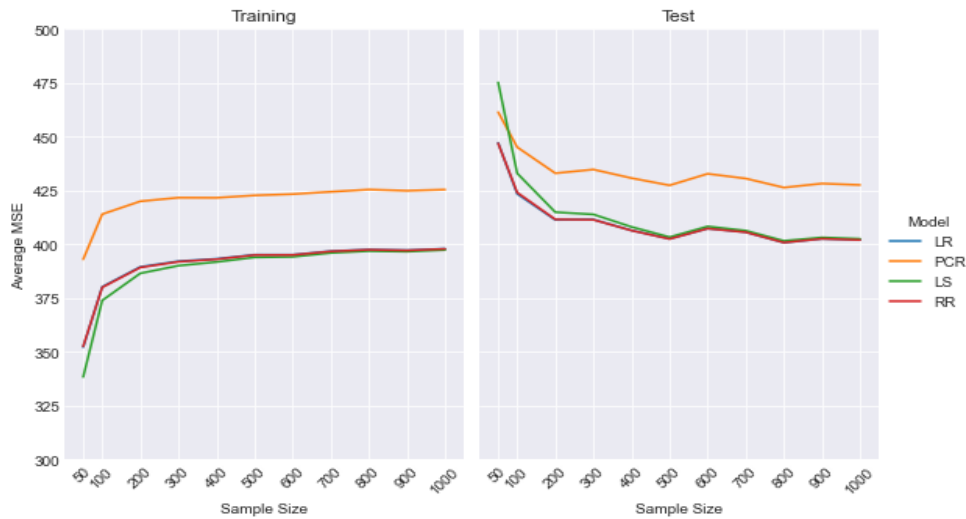


Figure 1. The change of MSE according to the sample size.

There are two different columns in the Figure 1 and they represent the change of MSE values for training and test data set respectively. X axes represent the sample size and y axis represents the average MSE values for this study. Each model is represented by line with different colored. Blue lines represent Lasso regression, orange lines represent Principal Component regression, green lines represent Least square regression and red lines represent Ridge regression. When the sample size increases in the training data, average MSE values increase. However, when the sample size increases in test data, average MSE values decrease.

It can be said that PCR method in terms of average MSE values are separated from other methods both in training and test data set. When the sample size is equal and above from 500, average MSE values are overlapped for LR, LS and RR in the training data set. On the other hand, when the sample size is below from 500, average MSE values are lowest for LS method and it is expected in the literature because the performance of LS method under the multicollinearity is not affected for the training data set.

When the sample size is equal 50, average MSE values are highest for LS in the test data set. When the sample size is equal and above from 500, average MSE values of LS are getting closer to LR and RR. On the other hand, when the sample size is below from 500, average MSE values are lowest for LR and RR methods and it is expected in the literature because the generalization performance of RR and LR methods under the multicollinearity is better than LS method.

From the Figure 1, it can be said that average MSE values are getting closer for each model both in training and test data set. The first one increases with increasing sample size and up to certain point and the second one decreases with increasing sample size up to certain point.

When examining Table 3 and Table 4, the R-square values for the models across different sample sizes are presented for both the training and test datasets. In the training set, despite the increase in sample size, all models achieve high R-square values, indicating that the increase in sample size has not led to significant changes in R-square values. On the other hand, in the test set, the R-square values show a slight increase as the sample size grows, but even for smaller sample sizes, the R-square values remain relatively high.



**Table 5.** Changing of coefficients across different sample size and model.

Sample Size	Coefficient	LR	RR	PCR	LS
50	Intercept	37.33 (0.34)	37.33 (0.34)	37.33 (0.34)	37.33 (0.34)
	X1	15.41 (0.7)	11.42 (0.62)	16.26 (0.06)	2.46 (5.98)
	X2	36.31 (0.77)	22.23 (0.41)	16.33 (0.06)	19.2 (23.22)
	X3	5.81 (0.37)	21.7 (0.25)	16.12 (0.06)	32.99 (25.61)
	X4	2.14 (0.1)	2.31 (0.1)	2.27 (0.11)	2.33 (0.11)
	X5	3.81 (0.15)	6.76 (0.17)	14.32 (0.07)	7.96 (4.27)
100	Intercept	37.89 (0.23)	37.89 (0.23)	37.89 (0.23)	37.89 (0.23)
	X1	14.36 (0.53)	10.43 (0.52)	16.51 (0.04)	2.58 (3.77)
	X2	41.28 (0.58)	23.47 (0.34)	16.58 (0.04)	17.68 (14.98)
	X3	3.4 (0.23)	22.58 (0.21)	16.36 (0.04)	34.91 (16.45)
	X4	2.2 (0.07)	2.41 (0.07)	2.39 (0.08)	2.42 (0.07)
	X5	3.23 (0.11)	6.68 (0.12)	14.63 (0.05)	8.33 (2.74)
200	Intercept	37.59 (0.16)	37.59 (0.16)	37.59 (0.16)	37.59 (0.16)
	X1	12.98 (0.42)	8.85 (0.45)	16.55 (0.03)	7.37 (2.59)
	X2	44.96 (0.45)	24.61 (0.29)	16.62 (0.03)	35.95 (10.08)
	X3	1.56 (0.12)	23.17 (0.19)	16.39 (0.03)	14.71 (11.12)
	X4	2.17 (0.05)	2.39 (0.05)	2.36 (0.05)	2.39 (0.05)
	X5	2.87 (0.08)	6.7 (0.09)	14.67 (0.03)	5.06 (1.86)
300	Intercept	37.74 (0.13)	37.74 (0.13)	37.74 (0.13)	37.74 (0.13)
	X1	11.72 (0.34)	7.39 (0.39)	16.54 (0.02)	6.26 (2.14)
	X2	47.06 (0.37)	25.59 (0.26)	16.61 (0.02)	35.37 (8.37)
	X3	0.97 (0.1)	23.73 (0.18)	16.38 (0.02)	16.3 (9.22)
	X4	2.21 (0.04)	2.39 (0.04)	2.4 (0.04)	2.4 (0.04)
	X5	2.64 (0.07)	6.63 (0.07)	14.69 (0.03)	5.15 (1.54)
400	Intercept	37.63 (0.12)	37.63 (0.12)	37.63 (0.12)	37.63 (0.12)
	X1	10.71 (0.31)	6.28 (0.35)	16.54 (0.02)	5.18 (1.79)
	X2	48.47 (0.32)	26.45 (0.27)	16.62 (0.02)	34.42 (7.08)
	X3	0.59 (0.06)	23.92 (0.18)	16.38 (0.02)	18.05 (7.78)
	X4	2.19 (0.04)	2.35 (0.04)	2.33 (0.04)	2.35 (0.04)
	X5	2.66 (0.06)	6.73 (0.06)	14.7 (0.02)	5.52 (1.3)
500	Intercept	37.62 (0.11)	37.62 (0.11)	37.62 (0.11)	37.62 (0.11)
	X1	10.7 (0.29)	6.21 (0.32)	16.58 (0.02)	4.21 (1.66)
	X2	48.8 (0.3)	26.35 (0.25)	16.65 (0.02)	28.72 (6.49)
	X3	0.36 (0.05)	24.13 (0.2)	16.42 (0.02)	23.91 (7.15)
	X4	2.2 (0.03)	2.35 (0.03)	2.35 (0.03)	2.35 (0.03)
	X5	2.73 (0.06)	6.84 (0.06)	14.73 (0.02)	6.66 (1.19)

**Table 5.** Changing of coefficients across different sample size and model(cont.).

Sample Size	Coefficient	LR	RR	PCR	LS
600	Intercept	37.64 (0.1)	37.64 (0.1)	37.64 (0.1)	37.64 (0.1)
	X1	10.89 (0.26)	6.43 (0.29)	16.56 (0.02)	6.85 (1.48)
	X2	48.67 (0.28)	26.39 (0.25)	16.63 (0.02)	37.07 (5.86)
	X3	0.38 (0.05)	23.97 (0.21)	16.39 (0.02)	14.32 (6.44)
	X4	2.25 (0.03)	2.38 (0.03)	2.38 (0.03)	2.38 (0.03)
	X5	2.55 (0.05)	6.62 (0.06)	14.71 (0.02)	4.87 (1.08)
700	Intercept	37.69 (0.09)	37.69 (0.09)	37.69 (0.09)	37.69 (0.09)
	X1	11.09 (0.25)	6.43 (0.28)	16.57 (0.02)	8.04 (1.4)
	X2	48.64 (0.27)	26.58 (0.26)	16.64 (0.02)	40.63 (5.46)
	X3	0.23 (0.04)	23.79 (0.24)	16.4 (0.02)	10.04 (6.02)
	X4	2.23 (0.03)	2.35 (0.03)	2.35 (0.03)	2.35 (0.03)
	X5	2.61 (0.05)	6.66 (0.06)	14.73 (0.02)	4.3 (1.0)
800	Intercept	37.64 (0.08)	37.64 (0.08)	37.64 (0.08)	37.64 (0.08)
	X1	10.8 (0.24)	6.2 (0.26)	16.57 (0.02)	8.54 (1.27)
	X2	49.1 (0.25)	27.16 (0.26)	16.64 (0.02)	43.89 (5.03)
	X3	0.14 (0.03)	23.55 (0.23)	16.4 (0.02)	6.82 (5.52)
	X4	2.32 (0.03)	2.43 (0.02)	2.43 (0.03)	2.43 (0.02)
	X5	2.52 (0.04)	6.53 (0.05)	14.72 (0.02)	3.65 (0.92)
900	Intercept	37.6 (0.08)	37.6 (0.08)	37.6 (0.08)	37.6 (0.08)
	X1	10.52 (0.22)	5.74 (0.25)	16.59 (0.01)	7.24 (1.23)
	X2	49.45 (0.24)	27.09 (0.28)	16.66 (0.01)	39.6 (4.84)
	X3	0.15 (0.03)	24.08 (0.27)	16.42 (0.01)	11.82 (5.33)
	X4	2.29 (0.02)	2.4 (0.02)	2.39 (0.03)	2.4 (0.02)
	X5	2.53 (0.04)	6.63 (0.06)	14.75 (0.02)	4.5 (0.89)
1000	Intercept	37.61 (0.07)	37.61 (0.07)	37.61 (0.07)	37.61 (0.07)
	X1	10.7 (0.21)	5.89 (0.24)	16.59 (0.01)	8.32 (1.18)
	X2	49.39 (0.22)	27.23 (0.28)	16.66 (0.01)	42.78 (4.56)
	X3	0.04 (0.01)	23.82 (0.29)	16.42 (0.01)	8.04 (5.04)
	X4	2.29 (0.02)	2.4 (0.02)	2.38 (0.02)	2.4 (0.02)
	X5	2.53 (0.04)	6.6 (0.06)	14.75 (0.02)	3.88 (0.84)

Note: Estimate (standard error) given in the Table 6.

Tables 5 and 6 present the regression coefficients along with their standard errors. It is important to highlight that, even with varying sample sizes, the coefficients derived from the LR, RR, and PCR methods are fairly consistent. In contrast, the coefficients produced by the LS method show considerable fluctuations in their values as the sample size varies. This indicates that the LS method may not provide reliable coefficient estimates when multicollinearity is present. Also, both multicollinearity exist and small sample size from the Table 5 and Table 6, coefficients have higher standard error and they are not statistically significant.

In addition, an analysis of the standard errors for the coefficients indicates that the LR, RR, and PCR approaches consistently produce exceedingly low standard errors, which exhibit a slight decrease as the dataset's number of observations grows. Conversely, the LS method consistently demonstrates high standard errors across all sample sizes.

According to the findings of this study, the LS regression method yielded the lowest MSE value in the training dataset for various sample sizes. The size of our sample had a significant impact on our results. The Ridge regression and Lasso regression methods performed best in terms of the Mean Squared Error (MSE) in the test dataset. Conversely, the PCR method showed the highest average MSE value in both the training and test datasets. In terms of MSE performance, these findings are consistent with those reported by (Çankaya et al. 2019).

Moreover, as the size of the sample grows, the average mean squared error (MSE) values also rise for the least squares (LS), ridge regression (RR), and Lasso regression (LR) methods in the training data set. In relation to the validity of the models, it was found that the models generated with RR and Lasso regression yielded the most favorable outcomes for the test data especially for the small sample sizes. Additionally, it was noted that as the sample size increases, the average MSE values for LS, RR, and LR methods tend to converge.

There are actually two main goals of the regression analysis in the literature. One of this is to summarize data set for the purposes of finding relationships among the variables, finding which independent variables have highest effect on dependent variable and getting adjusted coefficients of each independent variables. The other goal of this is to predict future value of the dependent variables by using independent variable. These two goals can be used under two main title as explorative model and predictive model respectively.

This study yields two main conclusions. Initially, the performance metrics, such as Mean Squared Error (MSE) and R-squared values, are not able to effectively differentiate between models in the presence of multicollinearity. However, the generalization of the LS method is low in the scope of the predictive model when the sample size small and multicollinearity exists. In simpler terms, the LS method does not exhibit a noticeable decline compared to other models when assessing model performance under the high sample size. On the other hand, when it comes to interpreting coefficients in the scope of summarizing the data set, the LS method clearly falls short, resulting in misleading conclusions and inaccurate interpretations due to high standard error of coefficients. Conversely, the LR, RR, and PCR methods yield reliable outcomes in scenarios where dataset summarization and coefficient interpretation are essential. Despite of the increasing sample size, the stability of the confidents for LS method cannot be guaranteed.

## References

- Alpar, R., (2017). Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. Detay Yayıncılık.
- Altland H. W. (1999). Regression analysis: statistical modeling of a response variable.
- Çankaya, S., Eker, S., & Abacı, S. H. (2019). Comparison of least squares, ridge regression and principal component approaches in the presence of multicollinearity in regression analysis. *Turkish Journal of Agriculture - Food Science and Technology*, 7(3), 180-190. <https://doi.org/10.24925/turjaf.v7i3.180-190.2019>
- Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example*. John Wiley & Sons.
- Draper, N. R. and Smith H. (1998). Applied Regression Analysis. New York, John Wiley and Sons, Inc.
- Fu, W. J. (1998). Penalized regression: the Bridge versus the Lasso, *Journal of Computation and Graphical Statistics*, 7, 397-416
- Göktaş, A., & Öznur, İ. (2010). Türkiye'de işsizlik oranının temel bileşenli regresyon analizi ile belirlenmesi. *Sosyal Ekonomik Araştırmalar Dergisi*, 10, 279-294.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
- Hintze, J. L. (2007). *NCSS User's Guide III - Regression and Curve Fitting, Chapter 340 - Principal Components Regression*. Kaysville/Utah: NCSS Statistical System.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Johnson, R., A. (1998). Applied multivariate statistical analysis, Prentice Hall, 458-498.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2004). *Applied linear statistical models* (5th ed.). McGraw-Hill.
- Miles J. & Shevlin M. (2000). Applying Regression and Correlation-A Guide for Students and Researchers. Sage Publication, London.
- Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. sut.ac.th
- Ortabaş, N. (2001). Principal components in the problem of multicollinearity (Master's thesis). DEÜ Fen Bilimleri Enstitüsü.
- Topal, M., Eydurhan, E., Yağanoğlu, A. M., Sönmez, A., & Keskin, S. (2010). Çoklu doğrusal bağlantı durumunda ridge ve temel bileşenler regresyon analiz yöntemlerinin kullanımı. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi*, 41, 53-57.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267-288.
- Weisberg, S. (1985). Applied Linear Regression, Wiley.
- Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.
- Zou, H., & Hastie, T. (2005). The use of ridge regression in the analysis of high-dimensional data. *Journal of Computational and Graphical Statistics*, 14(3), 814-828.

**Appendix**

**Table 6.** Correlation matrix for selected sample size (50).

		X1	X2	X3	X4	X5	Y
<b>X1</b>	<b>P.C</b>	1	,998**	,997**	0.229	,854**	,965**
	<b>Sig</b>		0.000	0.000	0.109	0.000	0.000
<b>X2</b>	<b>P.C</b>	,998**	1	,997**	0.235	,870**	,967**
	<b>Sig</b>	0.000		0.000	0.100	0.000	0.000
<b>X3</b>	<b>P.C</b>	,997**	,997**	1	0.236	,827**	,962**
	<b>Sig</b>	0.000	0.000		0.099	0.000	0.000
<b>X4</b>	<b>P.C</b>	0.229	0.235	0.236	1	0.182	0.254
	<b>Sig</b>	0.109	0.100	0.099		0.205	0.075
<b>X5</b>	<b>P.C</b>	,854**	,870**	,827**	0.182	1	,855**
	<b>Sig</b>	0.000	0.000	0.000	0.205		0.000
<b>Y</b>	<b>P.C</b>	,965**	,967**	,962**	0.254	,855**	1
	<b>Sig</b>	0.000	0.000	0.000	0.075	0.000	

\*\* Correlation is significant at the 0.01 level (2-tailed).

**Table 7.** Correlation matrix for selected sample size (200).

		X1	X2	X3	X4	X5	Y
<b>X1</b>	<b>P.C</b>	1	,996**	,996**	0.012	,814**	,950**
	<b>Sig</b>		0.000	0.000	0.866	0.000	0.000
<b>X2</b>	<b>P.C</b>	,996**	1	,995**	0.014	,835**	,952**
	<b>Sig</b>	0.000		0.000	0.845	0.000	0.000
<b>X3</b>	<b>P.C</b>	,996**	,995**	1	0.016	,780**	,945**
	<b>Sig</b>	0.000	0.000		0.822	0.000	0.000
<b>X4</b>	<b>P.C</b>	0.012	0.014	0.016	1	-0.006	0.040
	<b>Sig</b>	0.866	0.845	0.822		0.931	0.570
<b>X5</b>	<b>P.C</b>	,814**	,835**	,780**	-0.006	1	,818**
	<b>Sig</b>	0.000	0.000	0.000	0.931		0.000
<b>Y</b>	<b>P.C</b>	,950**	,952**	,945**	0.040	,818**	1
	<b>Sig</b>	0.000	0.000	0.000	0.570	0.000	

\*\* Correlation is significant at the 0.01 level (2-tailed).

**Table 8.** Correlation matrix for selected sample size (500).

		X1	X2	X3	X4	X5	Y
<b>X1</b>	<b>P.C</b>	1	,995**	,995**	0.014	,822**	,956**
	<b>Sig</b>		0.000	0.000	0.756	0.000	0.000
<b>X2</b>	<b>P.C</b>	,995**	1	,995**	0.011	,839**	,958**
	<b>Sig</b>	0.000		0.000	0.803	0.000	0.000
<b>X3</b>	<b>P.C</b>	,995**	,995**	1	0.008	,784**	,953**
	<b>Sig</b>	0.000	0.000		0.854	0.000	0.000
<b>X4</b>	<b>P.C</b>	0.014	0.011	0.008	1	0.030	0.033
	<b>Sig</b>	0.756	0.803	0.854		0.508	0.460
<b>X5</b>	<b>P.C</b>	,822**	,839**	,784**	0.030	1	,815**
	<b>Sig</b>	0.000	0.000	0.000	0.508		0.000
<b>Y</b>	<b>P.C</b>	,956**	,958**	,953**	0.033	,815**	1
	<b>Sig</b>	0.000	0.000	0.000	0.460	0.000	

\*\* Correlation is significant at the 0.01 level (2-tailed).

**Table 9.** Correlation matrix for selected sample size (900).

		X1	X2	X3	X4	X5	Y
<b>X1</b>	<b>P.C</b>	1	,995**	,995**	0.000	,820**	,954**
	<b>Sig</b>		0.000	0.000	0.990	0.000	0.000
<b>X2</b>	<b>P.C</b>	,995**	1	,995**	-0.003	,835**	,958**
	<b>Sig</b>	0.000		0.000	0.924	0.000	0.000
<b>X3</b>	<b>P.C</b>	,995**	,995**	1	-0.004	,779**	,951**
	<b>Sig</b>	0.000	0.000		0.898	0.000	0.000
<b>X4</b>	<b>P.C</b>	0.000	-0.003	-0.004	1	0.009	0.040
	<b>Sig</b>	0.990	0.924	0.898		0.797	0.229
<b>X5</b>	<b>P.C</b>	,820**	,835**	,779**	0.009	1	,814**
	<b>Sig</b>	0.000	0.000	0.000	0.797		0.000
<b>Y</b>	<b>P.C</b>	,954**	,958**	,951**	0.040	,814**	1
	<b>Sig</b>	0.000	0.000	0.000	0.229	0.000	

\*\* . Correlation is significant at the 0.01 level (2-tailed).