



**T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İSTATİSTİK ANA BİLİM DALI
İSTATİSTİK PROGRAMI**

**İKİ AŞAMALI BOOTSTRAP VERİ ZARFLAMA ANALİZİ İLE
TIMSS VERİLERİ KULLANILARAK EĞİTİM SİSTEMİNDE
ETKİNLİKLERİN DEĞERLENDİRİLMESİ**

Yüksek Lisans Tezi

Firdevs CEVHEROĞLU EREN

**DANIŞMAN
Dr. Öğr. Üyesi Serpil AYDIN**

**SAMSUN
2023**

T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
İSTATİSTİK ANA BİLİM DALI
İSTATİSTİK PROGRAMI



İKİ AŞAMALI BOOTSTRAP VERİ ZARFLAMA ANALİZİ İLE TIMSS VERİLERİ KULLANILARAK EĞİTİM SİSTEMİNDE ETKİNLİKLERİN DEĞERLENDİRİLMESİ

Yüksek Lisans Tezi

Firdevs CEVHEROĞLU EREN

Danışman
Dr. Öğr. Üyesi Serpil AYDIN

SAMSUN
2023

TEZ KABUL VE ONAYI

Firdevs CEVHEROĞLU EREN tarafından, **Dr. Öğr. Üyesi Serpil AYDIN** danışmanlığında hazırlanan '**İKİ AŞAMALI BOOTSTRAP VERİ ZARFLAMA ANALİZİ İLE TIMSS VERİLERİ KULLANILARAK EĞİTİM SİSTEMİNDE ETKİNLİKLERİN DEĞERLENDİRİLMESİ**' başlıklı bu çalışma, jürimiz tarafından 2.10.2023 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak Yüksek Lisans Tezi olarak kabul edilmiştir.

	Unvanı Adı Soyadı Üniversitesi Ana Bilim/Ana Sanat Dalı	Sonuç
Başkan	Dr. Öğr. Üyesi Serpil Aydın Ondokuz Mayıs Üniversitesi İstatistik Anabilim Dalı	☒ Kabul ☐ Ret
Üye	Doç. Dr. Hasan Bulut Ondokuz Mayıs Üniversitesi İstatistik Anabilim Dalı	☒ Kabul ☐ Ret
Üye	Doç. Dr. Tuba Koç Çankırı Karatekin Üniversitesi İstatistik Anabilim Dalı	☒ Kabul ☐ Ret

Bu tez, Enstitü Yönetim Kurulunca belirlenen ve yukarıda adları yazılı jüri üyeleri tarafından uygun görülmüştür.

Prof. Dr. Ahmet TABAK
Enstitü Müdürü

BİLİMSEL ETİĞE UYGUNLUK BEYANI

Hazırladığım Yüksek Lisans tezinin bütün aşamalarında bilimsel etiğe ve akademik kurallara riayet ettiğimi, çalışmada doğrudan veya dolaylı olarak kullandığım her alıntıya kaynak gösterdiğimi ve yararlandığım eserlerin Kaynaklar'da gösterilenlerden oluştuğunu, her unsurun enstitü yazım kılavuzuna uygun yazıldığını ve TÜBİTAK Araştırma ve Yayın Etiği Kurulu Yönetmeliği'nin 3. bölüm 9. maddesinde belirtilen durumlara aykırı davranılmadığını taahhüt ve beyan ederim.

Etik Kurul Gerekli mi ?

Evet ☐ (Gerekli ise ekler kısmına ekleyiniz)

Hayır ☒

02.10.2023

Firdevs Cevheroğlu Eren

TEZ ÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI

Tez Başlığı: İKİ AŞAMALI BOOTSTRAP VERİ ZARFLAMA ANALİZİ İLE TIMSS VERİLERİ KULLANILARAK EĞİTİM SİSTEMİNDE ETKİNLİKLERİN DEĞERLENDİRİLMESİ

Yukarıda başlığı belirtilen tez çalışması için şahsım tarafından 03/07/2023 tarihinde intihal tespit programından alınmış olan özgünlük raporu sonucunda;

Benzerlik oranı : % 18

Tek kaynak oranı : % 5 çıkmıştır.

03 /07 / 2023

Dr. Öğr. Üye. Serpil AYDIN

ÖZET

İKİ AŞAMALI BOOTSTRAP VERİ ZARFLAMA ANALİZİ İLE TIMSS VERİLERİ KULLANILARAK EĞİTİM SİSTEMİNDE ETKİNLİKLERİN DEĞERLENDİRİLMESİ

Firdevs CEVHEROĞLU EREN
Ondokuz Mayıs Üniversitesi
Lisansüstü Eğitim Enstitüsü
İstatistik Ana Bilim Dalı
İstatistik Programı
Yüksek Lisans, Ekim/2023
Danışman: Dr. Öğr. Üyesi Serpil AYDIN

Okul düzeyindeki etkin olmayan kaynak kullanımı ve kötü yönetim, insan kaynakları gelişimi ve işgücü piyasası sonuçları üzerinde ciddi olumsuz etkilere neden olmaktadır. Dolayısıyla okul girdilerinin daha yüksek öğrenme çıktılarına dönüştürülmesi aşamasında daha fazla verim elde etmek için insan kaynaklarının yanı sıra fiziksel kaynakların da etkin yönetimiyle tamamlanması gerekliliğini ortaya koymuştur. Bu nedenle bu tez çalışması küresel bağlamda eğitim sistemlerinin etkinliğini ölçmek ile ülkeler arası düzeyde etkinsizliğin kaynak ve insan yönetimi boyutlarıyla kapsamını ve sonuçlarını araştırmayı amaçlamıştır. Bu kapsamda 4. ve 8. sınıf düzeyindeki öğrencilere dört yılda bir yapılan dünyadaki en büyük ve en kapsamlı uluslararası öğrenci başarılarını değerlendirme çalışması olan TIMSS sınavının yalnızca eğitimin temelini oluşturan 4. sınıf düzeyindeki öğrencilere yapılan kısmı dikkate alınmıştır. Böylece eğitim seviyesinin en başında öğrenmeyi etkileyen bağlamsal faktörlerdeki boşlukların belirlenmesi ve herhangi bir zayıflık alanını saptaması hedeflenmiştir. Araştırma kapsamında kullanılan veriler; öğrencilerin, öğretmenlerin ve okul müdürlerinin doldurduğu anketlerden elde edilen değişkenler ile öğrencilerin matematik ve fen başarılarını temel alarak elde edilmiştir. Bu amaç doğrultusunda ülkelerin göreceli toplam etkinlik skorlarını tahmin etmek için parametrik olmayan yeniden örnekleme veri zarflama (bootstrap VZA) analizinden yararlanılmıştır. Burada bootstrap yaklaşımının kullanılmasının nedeni, tahmin edilen sınırın örnekleme varyasyonlarına göre etkinlik puanlarının duyarlılığını analiz etmenin kolay bir yolu olmasındandır. Matematik ve fen başarısına etki eden eğitim girdilerini göreceli olarak etkin kullanan ülkeler sırasıyla Danimarka, Almanya, Macaristan bulunmuş olup, bu bölgelerin neredeyse optimal düzeyde faaliyet gösterdiği ve istenen sabit çıktıyı üretmek için girdi tüketimleri de neredeyse optimal kullandıkları gözlemlenmiştir. En düşük etkinliğe sahip olan ülkeler ise sırasıyla; Karadağ, Birleşik Arap Emirlikleri, Bosna Hersek olarak belirlenmiştir. İkinci aşamada teknik etkinliği regresyona tabi tutarak okulların okul yönetimi değişkenleri üzerindeki etkilerini Tobit modeli kullanarak incelemiştir. Tobit regresyon analizi kaynak kullanımı ve yönetiminin özellikle ülkelerdeki okulların girdi odaklı etkinliklerini artırabileceğini ortaya koymuştur. Bu sonuçlar göstermektedir ki araştırmaya konu olan ülkelerdeki okulların etkisizliğinin hem matematik ve fen öğretimine özgü kaynakların yetersizliği hem de aynı kaynakların en iyi şekilde kullanılmaması ve yönetilmemesi olduğudur.

Anahtar Sözcükler: TIMSS, Sınav başarısı, Etkinlik, Yeniden örnekleme veri Zarflama, Tobit model.

ABSTRACT

EVALUATION OF EFFICIENCIES IN THE EDUCATION SYSTEM USING TIMSS DATA WITH TWO-STAGE BOOTSTRAP DATA ENVELOPMENT ANALYSIS

Firdevs CEVHEROĞLU EREN

Ondokuz Mayıs University

Institute of Graduate Studies

Department of Statistics

Statistics Programme

Master, October/2023

Supervisor: Assist. Prof. Dr. Serpil AYDIN

Inefficient resource use and mismanagement at the school level have serious negative impacts on human resource development and labor market outcomes. Hence, in order to achieve greater efficiency in transforming school inputs into higher learning outcomes, it is necessary to complement the efficient management of physical as well as human resources. Therefore, this thesis aims to measure the effectiveness of education systems in a global context and to investigate the scope and consequences of inefficiency at the cross-country level in terms of resource and human management dimensions. In this context, only the part of the TIMSS exam, which is the largest and most comprehensive international student achievement assessment study in the world, administered every four years to students at the 4th and 8th grade level, which constitutes the basis of education, was taken into consideration. Thus, it was aimed to identify gaps in the contextual factors affecting learning at the very beginning of the education level and to identify any areas of weakness. The data used in the study were based on students' achievement in mathematics and science, together with variables obtained from questionnaires completed by students, teachers and school principals. For this purpose, nonparametric resampling data envelopment analysis (bootstrap VZA) was used to estimate the relative aggregate efficiency scores of countries. The bootstrap approach is used here because it is an easy way to analyze the sensitivity of efficiency scores to sampling variations of the estimated frontier. Denmark, Germany, and Hungary were found to be relatively efficient in the use of education inputs that affect math and science achievement, respectively, and these regions were found to operate at near-optimal levels and to use input consumption at near-optimal levels to produce the desired constant output. The six countries with the lowest efficiency are Montenegro, United Arab Emirates, Bosnia and Herzegovina, respectively. In the second stage, technical efficiency was regressed and the effects of schools on school management variables were analyzed using a Tobit model. Tobit regression analysis revealed that resource utilization and management can increase the input-oriented efficiency of schools in particular countries. These results suggest that the ineffectiveness of schools in the countries under study is both a lack of resources specific to mathematics and science teaching and a lack of optimal use and management of these same resources.

Keywords: TIMSS, Exam Success, Efficiency, Bootstrap Data Envelopment, Tobit model

ÖN SÖZ VE TEŞEKKÜR

Bu çalışma, Ondokuz Mayıs Üniversitesi Yüksek Lisans Programı kapsamında hazırlanmıştır. Tez konum, “İki Aşamalı Bootstrap Veri Zarflama Analizi ile Timss Verileri Kullanılarak Eğitim Sisteminde Etkinliklerin Değerlendirilmesi “ şeklindedir. Bu çalışma, IEA’nın çeşitli kaynaklarından elde edilen verilerin analiz edilmesiyle oluşturulmuştur. Verilerin toplanması ve analizi sırasında çok çeşitli zorluklarla karşılaştım. Ancak sabırlı ve istekli çalışmam sayesinde sonuçta bu tezi tamamladım. Bu tezi hazırlamam sırasında öncelikle danışman hocam Dr. Öğr. Üyesi Serpil Aydın'a teşekkür etmek istiyorum. Tezimin hazırlık aşamasında kendisinin yönlendirmeleri ve önerileri sayesinde çalışmamın kalitesi arttı. Kendisi, benim sorularıma her zaman sabırla cevap verdi ve beni teşvik etti. Ayrıca çalışmamda kullanmış olduğum verileri temin etmemde yardımcı olan Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu'na (IEA) da teşekkür etmek istiyorum. Verilerin elde edilmesi ve analizi çalışmamın temelini oluşturdu. Bu nedenle bu kurumun desteği olmadan bu çalışmayı tamamlamam mümkün olmayacaktı. Son olarak aileme ve arkadaşlarıma da teşekkür etmek istiyorum. Onların desteği çalışmamı sürdürmemde ve tamamlamamda büyük bir motivasyon kaynağı oldu. Bu süreçte bana her zaman destek oldukları için minnettarım.

Firdevs CEVHEROĞLU EREN

İÇİNDEKİLER

TEZ KABUL VE ONAYI	i
BİLİMSELETİĞE UYGUNLUK BEYANI.....	ii
TEZÇALIŞMASI ÖZGÜNLÜK RAPORU BEYANI.....	ii
ÖZET.....	iii
ABSTRACT.....	iv
ÖNSÖZ VE TEŞEKKÜR.....	v
İÇİNDEKİLER.....	vi
SİMGELER VE KISALTMALAR.....	vii
ŞEKİLLER DİZİNİ.....	viii
TABLolar DİZİNİ.....	ix
1. GİRİŞ.....	1
2. EĞİTİMDE ULUSLARARASI ETKİNLİK ÖLÇÜMÜNE DAYALI LİTERATÜRE GENEL BAKIŞ.....	4
3. TIMSS SKORLARINA GÖRE ÜLKELERİN GÖRECELİ ETKİNLİK HESABINDA KULLANILAN TEMEL DEĞİŞKENLER VE HESAPLAMA YÖNTEMLERİ.....	8
3.1. TIMSS Sınavı Hakkında	8
3.2. TIMSS 2019'a Katılan Ülkeler	9
3.3. TIMSS 2019 Değerlendirme Boyutları	9
3.4. TIMSS 2019 Yeterlik Düzeyleri	10
4. MATERYAL VE METOD	11
4.1. Veri Zarflama Analizi (VZA) Metodolojisi	13
4.2. Ölçeğe Göre Sabit Getiri ve Ölçeğe Göre Değişken Getiri modelleri	16
4.3. Örneklem (Bootstrap) Veri Zarflama Metodolojisi.....	18
4.4. Sınırlı bağımlı değişkenli regresyon (Limited Dependent Variable Regression): Tobit Regresyon Modeli	22
5. ARAŞTIRMANIN KAPSAMI VE VERİSETİ.....	28
6. ARAŞTIRMANIN AMPİRİK SONUÇLARI VE TARTIŞMA	38
7. SONUÇ	47

SİMGELER VE KISALTMALAR

VZA	:Veri Zarlama Analizi (Data Envelopment Analysis)
IEA	:Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (International Association for the Evaluation of Educational Achievement)
TIMSS	:Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS- Trends in International Mathematics and Science Study)



ŞEKİLLER DİZİNİ

Şekil 4.1. Bootstrapped VZA skorları.....	21
Şekil 6.1. Ortalama Etkinlik Skorlarına göre sıralanmış ülkelerin VZA skor değerleri	41



TABLÖLAR DİZİNİ

Tablo 3.1. TIMSS 2019 Uygulamasına Katılan Ülkeler.....	9
Tablo 3.2. Öğrenme Alanlarına Göre TIMSS 2019 Matematik 4. Sınıf Değerlendirmesinin Yüzdelik Dağılımları.....	10
Tablo 3.2. (Devam)	Hata! Yer işareti tanımlanmamış.
Tablo 3.3. Öğrenme Alanlarına Göre TIMSS 2019 Fen Bilimleri 4. Sınıf Değerlendirmesinin Yüzdelik Dağılımları.....	10
Tablo 4.1. Girdi/Çıktı Odaklı-Primal/Dual VZA Modelleri	17
Tablo 4.1. (Devam)	17
Tablo 5.1. Bootstrap (VZA) modelinin analizi için kullanılan girdiler ve çıktılar. ...	30
Tablo 5.2. Bu çalışmada kullanılan TIMSS 2019 veri kümesindeki değişkenlerin ayrıntıları	32
Tablo 5.2. (Devam)	33
Tablo 5.2. (Devam)	34
Tablo 5.2. (Devam)	35
Tablo 5.2. (Devam)	36
Tablo 5.2. (Devam)	37
Tablo 6.1. Model Özet Bilgileri	38
Tablo 6.2. Etkinlik Skorlarının Duyarlılık Sonuçları.....	39
Tablo 6.2. (Devam)	40
Tablo 6.3. Tobit Model Sonuçları	44

1. GİRİŞ

Etkili bir değerlendirme için hem eğitim girdilerinin hem de sonuçlarının (bilgi, tutum ve katılım gibi) incelenmesini gerektirdiği görüşü zaman içerisinde hakimiyet kazanmıştır. Böylece eğitim çıktıları üzerinde anlamlı ve tutarlı etkileri olacak faktörleri belirleme amacı doğmuştur. Bu nedenle 1958 yılında Eğitim Enstitüsü'nde okul ve öğrenci değerlendirmesi sorunlarını tartışmak üzere bir grup akademisyen, eğitim psikoloğu, sosyolog ve psikometristin Almanya'nın Hamburg kentindeki UNESCO Eğitim Enstitüsü'nde bir araya gelmiştir. Tüm bu gelişmelerin sonucunda eğitim başarısı ve eğitimin diğer yönleri hakkında büyük ölçekli karşılaştırmalı çalışmalar yürütmek amacıyla ulusal araştırma kurumları ve devlet araştırma ajanslarından oluşan bağımsız, uluslararası bir kooperatif olan *Uluslararası Eğitim Başarısını Değerlendirme Derneği* (*International Association for the Evaluation of Educational Achievement-IEA*) 1967 yılında tüzel kişilik kazanmıştır (IEA).

1958'de kuruluşundan bu yana IEA, uluslararası başarıya ilişkin 30'dan fazla araştırma çalışması yürütmüştür. IEA çalışmaları matematik, bilim, okuma, yurttaşlık ve vatandaşlık eğitimi, bilgisayar ve bilgi okuryazarlığı ve öğretmen eğitimi ile ilgili konulara odaklanmıştır. Ayrıca eğitimcilere, politika yapıcılara ve ebeveynlere öğrencilerin nasıl performans gösterdiğine dair iç görüş sağlamak için dünya çapında yüksek kaliteli, büyük ölçekli karşılaştırmalı eğitim çalışmaları yürütmüştür.

Boston College'ın Lynch Eğitim ve İnsani Gelişme Okulu'nda bulunan IEA'nın, 60'tan fazla ülkede Uluslararası Matematik ve Bilim Çalışmasındaki Eğilimler (the Trends in International Mathematics and Science Study-TIMSS) ve Uluslararası Okuma Okuryazarlığında İlerleme Çalışması (the Progress in International Reading Literacy Study- PIRLS) alanlarındaki öğrenci başarısının uluslararası karşılaştırmalı değerlendirmelerini düzenli olarak yürütmektedir. TIMSS & PIRLS Uluslararası Çalışma Merkezi IEA'nın temel çalışma döngüsünü oluşturmaktadır (International Study Center). Merkezi Amsterdam'da bulunan ve Hamburg'da büyük bir veri işleme ve araştırma merkezi bulunan IEA, 1959'dan beri öğrenci başarısıyla ilgili uluslararası karşılaştırmalı araştırmalar yürütmeye devam etmektedir. Böylece TIMSS ve PIRLS, katılımcı ülkelerin eğitim politikalarını iyileştirmek için kanıta dayalı kararlar almalarını sağlamaktadır.

Hükümetlerin ve bakanlıkların TIMSS ve PIRLS sonuçlarını kullanabilme yollarından bazıları şöyle sıralanabilir:

- Küresel bağlamda eğitim sistemlerinin etkinliğini ölçmek
- Öğrenme kaynakları ve fırsatlarındaki boşlukları belirleme
- Herhangi bir zayıflık alanını saptamak ve müfredat reformunu teşvik etmek
- Yeni eğitim girişimlerinin etkisinin ölçülmesi
- Araştırmacıları ve öğretmenleri ölçme ve değerlendirme konusunda eğitmek

TIMSS ve PIRLS ayrıca okul kaynakları, öğrenci tutumları, öğretim uygulamaları ve evde destek dahil olmak üzere öğrenmeyi etkileyen bağlamsal faktörler hakkında kapsamlı veriler toplamaktadır. Bu bilgi, akademik başarıya katkıda bulunan faktörleri keşfetmek için başarı ile ilişkili olarak incelenmesi de yine mümkündür (PIRLS, 2021).

Bu çalışmanın da temel amacı küresel bağlamda eğitim sistemlerinin etkinliğini ölçmek ve bu konuda politik karar alıcılara politikalarını iyileştirmesi yönünde yol göstermektir. Bu amaç kapsamında 4. ve 8. sınıf düzeyindeki öğrencilere dört yılda bir yapılan dünyadaki en büyük ve en kapsamlı uluslararası öğrenci başarılarını değerlendirme çalışması olan TIMSS sınavının yalnızca eğitimin temelini oluşturan 4. sınıf düzeyindeki öğrencilere yapılan kısmi dikkate alınmıştır. Böylece eğitim seviyesinin en başında öğrenmeyi etkileyen bağlamsal faktörlerdeki boşlukların belirlenmesi ve herhangi bir zayıflık alanının saptaması hedeflenmiştir. Sonuç olarak katılımcı ülkelerin eğitim politikalarını iyileştirmek için kanıta dayalı kararlar almalarını sağlanmaya çalışılmıştır. Bu amaç doğrultusunda ülkelerin göreceli toplam etkinlik skorlarını tahmin etmek için parametrik olmayan yeniden örnekleme veri zarflama (bootstrap VZA) analizinden yararlanılmıştır. Burada bootstrap yaklaşımının kullanılmasının nedeni, tahmin edilen sınırın örnekleme varyasyonlarına göre etkinlik puanlarının duyarlılığını analiz etmenin kolay bir yolu olmasındandır. Bu çalışmada öğrencilerin, öğretmenlerin ve okul müdürlerinin doldurduğu anketler kullanılarak IEA IDB Analyzer programıyla öğrencilerin matematik ve fen başarılarına dayanarak elde edilen veriler kullanılmıştır.

Bu tez çalışması bu alanda yapılan literatürün genel bakışı, göreceli etkinlik

hesabında kullanılan temel deęişkenlerin hesaplama yöntemleri, bootstrap VZA metodolojisi, bulgular ve sonuç başlıkları ile devam etmektedir.



2. EĞİTİMDE ULUSLARARI ETKİNLİK ÖLÇÜMÜNE DAYALI LİTERATÜRE GENEL BAKIŞ

Öğrenme sürecine etki eden faktörlerin neler olabileceği ile ilgili literatür genel olarak incelendiğinde; Fuller, Young, Baker (2011) çalışmalarında, okul müdürlerinin nitelikli öğretmen takımlarının oluşturulmasının yanı sıra öğretmenlerin de öğretim becerileri açısından kendilerini geliştirilmesinin öğrenci başarısını artırdığı sonucuna varmışlardır. Benzer şekilde Caputo ve Rastelli (2014)'nin de öğrencinin matematik başarısını etkileyen faktörleri inceledikleri araştırmada yine okul geliştirme ve iyileştirme sürecinin, bu başarıyı arttırdığı sonucuna varmışlardır. Givens Rolland (2012) sınıf yapısının ve öğretmen desteğinin öğrenci motivasyonu üzerinde olumlu etkisinin olduğunu bunun sonucunda da öğrenci başarısının etkilendiği sonucunu ortaya koymuşlardır. Bu temel araştırmaların ardından öğrenme ortamının etkililiğinin başarıyı öğrendi başarı ile ilişkisi sonucu farklı yazarlar tarafından da teyit edilmiştir (Akyüz, 2013; Çobanoğlu ve Badavan, 2017; Carlson ve Lavertu, 2018). Ayrıca Tsui (2005)'nin yansısı Jay, Rose, Simmons (2018) çalışmalarında, ailenin evde sunduğu olanakların (aile geliri, ebeveynlik, ev ortamı) ve öğrenimini destekleyen uygulamaların (örneğin çocukla okul hakkında konuşmalar), çocuğun öğrenme ve ders başarısını arttırdığı sonucuna varmışlardır. Sonuç olarak öğrenme sürecine ve buna bağlı olarak öğrenci başarısına etki eden faktörleri temelde okul ve aile bir bütününde değerlendirilmesi gerektiği sonucuna varılmaktadır. Buna göre öğrenme sürecine etki eden faktörleri ise genel olarak aşağıdaki başlıklarda toplamak mümkündür. Bu faktörler şöyle sıralanabilir:

- Öğrencinin bireysel özelliklerinin yansısı etkileşimde bulunduğu okul ve aile çevresine ilişkin genel yapı
- Okulun ve sınıf ortamının öğrenciye sunduğu imkân ve olanaklar, (örneğin güncel kaynak kullanımı, öğretmen niteliği, okul güvenliği, etkili ve verimli öğrenme süreci, öğretmenin alan bilgisindeki deneyimi ve davranışı....gibi)
- Ailenin desteği ve çocuğa sunduğu öğrenme olanakları

Öğrenme sürecinin ölçümünün sonucu olan öğrenci başarısını standart düzeyde ölçmek için uluslararası alanda uygulanan PISA ve TIMSS sınavları ülkelerin akademik gelişim açısından durumlarını karşılaştırmalı olarak görmelerine olanak

tanınmaktadır. Bu sınavlarda başarı testlerinin yanı sıra okul (öğrenci, öğretmen, okul yöneticisi) ve ebeveynlere anketler uygulayarak yukarıda bahsedilen öğrenci başarısını etkileyen faktörler ile ilgili de bilgiler toplanmaktadır. Böylece öğrencilerin başarı veya başarısızlıklarının nedenlerinin belirlenmesi ve başarıyı etkileyen faktörlerin göreceli olarak etkin kullanıp kullanılmadığını araştırma imkânı sunulmaktadır. Nitekim söz konusu sınavlarla ilgili başarıyı etkileyen faktörlerin etkin kullanılıp kullanılmadığına yönelik birtakım uluslararası kıyaslamalar için yapılan birtakım araştırmalar söz konusudur. Bu araştırmalar ağırlıklı Veri zarflama Analizi (Data Envelopment Analysis-VZA)’nden faydalanmıştır. Bu araştırmalardan bir kısmı PISA sınavı etkinliğini (Afonso ve Aubyn, 2005; Giambona, Vassallo ve Vassiliadis, 2011; Agasisti, 2014; Lorcu ve Bolat, 2015; Gavurova ve diğ., 2017; Erdem-Kara ve Osman, 2019; Ilgaz, Eskici ve Vural, 2019; Mazurek ve Mielcová, 2019) ve bir kısmı TIMSS sınavı (Giménez, Prior ve Thieme, 2007; Arshad, 2014; Koyuncu ve Ilgaz, 2019; Bhutoria ve Aljabri, 2022a) ve bazıları da hem PISA hem de TIMSS sınavı etkinliğini (Stumbrienė ve diğ., 2022) ülkeler bazında incelemiştir.

Araştırmanın kapsamı gereği özellikle TIMSS sınavının başarısında kaynak kullanım etkinliği yapan çalışmalar incelendiğinde; Giménez ve diğ. (2007) TIMSS 1999 sınavının sekizinci sınıftaki öğrencilerine uygulanan matematik ve fen bilgisi sonuçlarını 31 ülkedeki eğitim sisteminin teknik etkinliğini analiz etmek için değerlendirmiştir. Öğrencilerin performansını şu anda eğitim sistemlerine tahsis edilenden daha az kaynakla artırabilecek bir dizi gelişmiş ülke olduğu sonucuna varmışlardır. Arshad (2014), seçilmiş 16 İslam Konferansı Örgütü (Organization of Islamic Conference-OIC) ülkesinde (Batı Şeria ve Gazze dahil) orta öğretimin teknik verimlilik düzeyini araştırmışlardır. Çalışma, 40 ülkeyi kapsayan TIMSS 2011 verilerini kullanmıştır. Seçilen 16 OIC üyesinin neredeyse tamamı, OIC üyesi olmayan ülkelere kıyasla daha iyi TIMSS sonuçları elde etmek için eğitim kaynaklarını kullanma konusunda teknik olarak yetersiz olduğu sonucuna varılmıştır. Çevresel faktörleri kontrol ettikten sonra bile, OIC ülkelerinde orta öğretim teknik olarak yetersiz kaldığı gözlemlenmiştir. Koyuncu ve Ilgaz (2019) TIMSS 2015 kapsamında 8. sınıf matematik çalışmasına katılan 41 ülkenin matematik başarısıyla ilişkili olduğu düşünülen girdilerini etkin şekilde yönetip yönetmediklerini belirlemeyi amaçlamıştır. Verilerin analizinde, veri zarflama analiz tekniği kullanılmıştır. Bu analiz, matematik başarısını etkileyen eğitim girdilerine

odaklanarak Çin-Taipei (Tayvan)'nın birinci, Japonya'nın ikinci ve Kore'nin üçüncü sırada yer aldığını tespit etmiştir. En düşük etkinliğe sahip olan altı ülke ise sırasıyla Kuveyt, Güney Afrika, Arjantin-Buenos Aires, Ürdün, Umman ve Suudi Arabistan olarak belirlenmiştir. Türkiye ise yukarıda belirtilen en yakın etkinliğe sahip ülke olarak gözlemlenmiştir. Düşük etkinliğe sahip olan ülkelerin kaynaklarını verimli bir şekilde kullanmadığı ve harcamalarının karşılığını tam olarak alamadığı sonucuna varılmıştır. Bhutoria ve Aljabri (2022a), Ekonomik İş birliği ve Kalkınma Teşkilatı (Organization for Economic Cooperation and Development-OECD) ve Orta Doğu ve Kuzey Afrika'da (the Middle East and North Africa-MENA) bölgelerindeki ülkeler arası düzeyde kaynak ve insan yönetimi yönleriyle ilgili olarak okulların mevcut teknik verimsizliğinin kapsamını ve sonuçlarını araştırmıştır. (TIMSS) 2019'daki veri setlerini kullanan ikinci aşama Tobit regresyon modelinin yanı sıra radyal olmayan bir Veri Zarflama Analizi (Data Envelopment Analysis-VZA) kullanmıştır. Analiz, 26 ülkedeki 5164 okulu kapsamaktadır. Ampirik bulgular aynı zamanda okuldaki öğrenci topluluğu içinde sürdürülen disiplinin hem MENA hem de OECD bölgelerinde daha yüksek okul düzeyinde girdi ve çıktı verimliliği ile ilişkili en önemli ve anlamlı faktörlerden biri olduğunu ortaya koymaktadır. Ayrıca, insan yönetiminin farklı yönlerinin, özellikle hedef belirleme, öğrenci ve öğretmen motivasyonu ve okul yönetimine veli katılımının, değişen derecelerde de olsa, iki bölgede okul düzeyinde teknik verimlilikle olumlu bir şekilde ilişkili olduğu bulunmuştur. Genel olarak, eğitim yönetimi politikaları, odağı yalnızca daha yüksek miktarda kaynak sağlamaktan, disiplin cezasını teşvik etmenin yanı sıra motivasyon ve katılımı teşvik eden paydaş teşviklerini destekleyerek gelişmiş okul düzeyinde yönetim yoluyla okulların teknik verimliliğini geliştirmeye kaydırmalı sonucuna varılmıştır. Stumbrienė ve diğ. (2022), IEA TIMSS 2015 ve PIRLS 2016 uluslararası öğrenci başarı çalışmalarına katılan eski demokrasi ve post-sosyalist Avrupa Birliği (European Union-EU) üyesi ülkelerin eğitim sistemlerini incelemişlerdir. Araştırma, ilköğretimin etkililiğini ve etkinliği hakkaniyet (sınav sonuçlarındaki varyasyon) açısından değerlendirmeyi amaçlamaktadır. Sosyalizm sonrası ülkelerin oldukça etkili oldukları, ancak eğitimde eşitlik için çabalamada çok etkili olmadıkları sonucuna varılmıştır.

VZA yöntemleri avantajlarından dolayı eğitimde etkinlik ölçümü alanında yaygınca uygulanmıştır. Bilimsel literatürde önerilen ilk model, bir girdi odaklı ve

ölçeğe göre sabit getiri (Constant Return to Scale-CRS) varsayımını dikkate almıştır (Charnes, Cooper ve Rhodes, 1978). Ölçeğe göre değişken getiri (Variable Return to Scale-VRS) koşullarını hesaba katmak için, Banker, Charnes, Cooper (1984) CRS VZA modelinin ötesine geçmiştir. Ancak klasik VZA modellerindeki ana sınırlaması, deterministik bir modele dayandığından, gerçek dünyayı karakterize eden belirsizliği (stokastik hata olarak adlandırılır) dikkate almamasıdır. Aslında, klasik VZA tekniği, güven aralıklarının oluşturulmasına veya tahmin edilen değerler üzerinde testlerin yapılmasına izin vermemektedir (Seiford ve Thrall, 1990).

Güven aralıklarının oluşturulmasındaki sınırlamanın üstesinden gelmek için Leopold Simar ve Wilson (1998); Léopold Simar ve Paul W Wilson (2000); Leopold Simar ve Paul W Wilson (2000) güven aralıkları ve ayarlanmış etkinlik puanları elde ederek sonuçların doğrulanmasına izin veren bir Bootstrap VZA yöntemini önermişlerdir. VZA, kullanılan verilerin kalitesine bağlıdır ve örnekleme yanlılığına, hata ölçümüne ve aykırı değerlere karşı çok hassastır. Bu nedenle Çarpık sonuçlar ve fazla tahmin edilen verimlilik puanları oluşturur (Rouse, Putterill ve Ryan, 2002; Deville, 2009). Bootstrap VZA Prosedürü ise tahmini verimlilik puanlarının korelasyonu nedeniyle yanlılıktan etkilenmeyen tahmini standart hatalar ve güven aralıkları sağlar. Böylece VZA metodolojisinden kaynaklanan bazı zorlukların üstesinden gelinmiş olunur (López-Penabad ve diğ., 2020). Bootstrap VZA'nın standart VZA'ya göre birçok avantajı olmasına rağmen, eğitim ile ilgili uygulamaları sınırlı kalmıştır. Yalnızca Agasisti (2014), çıktı olarak OECD-PISA test puanlarını, girdi olarak ise öğrenci başına harcama değişkenlerini kullanmıştır. Etkinlik sonuçları Bootstrap veri zarflama analizi kullanılarak elde edilmiştir.

Bu tez araştırması hem temel alınan hedef kitle hem de metodolojik anlamda literatürden ayrılmaktadır. Şimdiye kadarki çalışmalar ikinci öğretim düzeyindeki öğrenci kitlesini hedef alırken bu araştırma eğitimin daha başta şekillenmesi gerektiği görüşü ile birinci öğretim düzeyindeki öğrenci kitlesini hedef almıştır. İlgili vaka çalışmasını metodolojik olarak da klasik VZA'nın eksikliğinden dolayı Bootstrap VZA yaklaşımı ile desteklemiştir.

3. TIMSS SKORLARINA GÖRE ÜLKELERİN GÖRECELİ ETKİNLİK HESABINDA KULLANILAN TEMEL DEĞİŞKENLER VE HESAPLAMA YÖNTEMLERİ

3.1. TIMSS Sınavı Hakkında

Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA-International Association for the Evaluation of Educational Achievement), öğrencilerin matematik ve fen bilimleri alanlarında kazandıkları bilgi ve becerilerin değerlendirilmesini amacıyla Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS-Trends in International Mathematics and Science Study) sınavını organize etmektedir. Dördüncü sınıf ve sekizinci sınıf düzeyindeki öğrencilerin matematik ve fen bilimleri alanlarındaki performanslarını TIMSS ile ölçülmektedir. Söz konusu araştırma dört yıllık periyotlar halinde gerçekleştirilmektedir (Suna ve diğ., 2020).

TIMSS’te yalnızca öğrencilerin başarı puanı değil aynı zamanda uygulamaya katılan öğrenciler, bu öğrencilerin öğretmenleri, velileri ve okul idarecilerine birtakım anketler uygulanmakta ve böylece öğrenci başarısı üzerinde etkili olan değişkenlere ilişkin de veriler derlenmektedir. Bu sayede, ülke bazında ve uluslararası düzeyde hem karşılaştırmalı çalışmalar yapılmasına hem de eğitim sisteminin değerlendirmesine imkân vermektedir (Suna ve diğ., 2020).

İlk olarak 1995 yılında düzenlenmiş olan TIMSS, takriben her 4 yılda bir uygulanmaya başlamıştır (Suna ve diğ., 2020). Türkiye 1999 yılında ilk defa TIMSS’e 8.sınıf düzeyinde katıldı.

Türkiye 2019’daki sınava 4.sınıf yaş ortalaması düşük olduğu için 5.sınıf düzeyinde katılmıştır. TIMSS’ e dördüncü sınıf düzeyinde katılan ülkelerdeki öğrenci yaş ortalamaları 10,2 olarak gözlemlenmektedir (I. V. Mullis ve diğ., 2020). Ancak IEA, TIMSS katılımcılarına değerlendirme çerçeveleri, öğrenci yaşları ve benzeri ölçütleri dikkate alarak farklı sınıf düzeylerinde katılması fırsatı da vermektedir. Öyle ki Türkiye dördüncü sınıf öğrencilerinin yaş ortalaması 9,7, beşinci sınıf öğrencilerinin yaş ortalaması 10,6’dır. Bu nedenle TIMSS 2019 döngüsünün dördüncü sınıf uygulamasına Norveç ve Güney Afrika yanı sıra Türkiye de beşinci sınıf düzeyinde katılmıştır (Suna ve diğ., 2020). Öğrenciler her ülkenin nüfus bölgesel yoğunluğuna bağlı olarak tamamen bağımsız olarak seçilir. Örneklem seçme işi tabakalı örnekleme yöntemi ile rastgele seçilir. TIMSS ve PIRLS’ ün veri

setleri birbiriyle bağlantılıdır.

3.2. TIMSS 2019’a Katılan Ülkeler

Yirmi yılı aşkın pek çok ülkenin katılımı ile gerçekleştirilen TIMSS sınavını 2019 döngüsüne ayrı olarak katılan bazı eğitim sistemleri (örneğin, Çin Halk Cumhuriyeti Hong Kong Özel İdari Bölgesi ve Belçika’nın Flaman bölgesi) ile 64 ülke katılmıştır. Bunun haricinde TIMSS 2019 döngüsüne eyaletler ve belediyeler gibi bölgesel birimlerin olduğu 8 karşılaştırma birimi “benchmark participants” de yer almıştır (I. V. Mullis ve diğ., 2020).

Ülkeler ve benchmark katılımcılar; dördüncü, sekizinci sınıf düzeylerinde ayrı ayrı katılabilecekleri gibi her iki sınıf düzeyinde TIMSS’e katılma imkanına sahiplerdir. TIMSS 2019 uygulamasına 58 ülke ve 6 benchmark katılımcı dördüncü sınıf düzeyinde, 39 ülke ve 7 benchmark katılımcı ise sekizinci sınıf düzeyinde katılmıştır (I. V. Mullis ve diğ., 2020). TIMSS 2019 döngüsüne katılan ülkeler Tablo 3.1’de gösterilmektedir.

Tablo 3.1. TIMSS 2019 Uygulamasına Katılan Ülkeler

Almanya*	Bulgaristan	Güney Kıbrıs	İsveç*	Kuzey İrlanda	Pakistan	Türkiye*	Batı Kap, GAC
ABD*	Çek Cumhuriyeti*	Gürcistan*	İtalya*	Kuzey Makedonya	Polonya	Umman	Abu Dabi, BAE*
Arnavutluk	Tayvan*	Hırvatistan*	Japonya	Letonya	Portekiz*	Ürdün	Dubai, BAE*
Avustralya	Danimarka*	Hollanda*	Kanada*	Litvanya*	Romanya	Yeni Zelanda	
Avusturya*	Ermenistan	Hong Kong*	Karadağ	Lübnan	Rusya*	Ontario, Kanada*	
Azerbaycan	Fas	İngiltere*	Katar*	Macaristan*	Sırbistan	Quebec, Kanada*	
Bahreyn	Filipinler	İran	Kazakistan	Malezya*	Singapur*	Moskova, Rusya*	
Belçika (Flaman Bölgesi)	Finlandiya*	İrlanda	Güney Kore*	Malta*	Slovakya*	Madrid, İspanya*	
Bleşik Arap Emirlikleri*	Fransa*	İspanya*	Kosova	Mısır	Suudi Arabistan	Gauteng, GAC	
Bosna Hersek	Güney Afrika	İsrail*	Kuveyt	Norveç*	Şili*		

(* eTIMSS yani bilgisayar tabanlı test uygulama sınavını uygulayan ülkeleri göstermektedir)

Kaynak: TIMSS 2019 Türkiye Ön Raporu

3.3. TIMSS 2019 Değerlendirme Boyutları

TIMSS 2019 dördüncü sınıf düzeyinde matematik öğrenme alanlarının yüzdelik dağılımları Tablo 3.2’deki gibidir.

Tablo 3.2. Öğrenme Alanlarına Göre TIMSS 2019 Matematik 4. Sınıf Değerlendirmesinin Yüzdelik Dağılımları

Sayılar %50	• Doğal sayılar (%25) • İfadeler, basit denklemler ve ilişkiler (%15) • Kesirler ve ondalık sayılar (%10)
Ölçme ve Geometri %30	• Ölçme (%15) • Geometri (%15)
Veri %20	• Veri okuma, yorumlama ve gösterimi (%15) • Problemleri çözmek için veriyi kullanma (%5)

TIMSS 2019 dördüncü sınıf düzeyinde fen bilimleri öğrenme alanlarının yüzdelik dağılımları Tablo 3.3'teki gibidir.

Tablo 3.3. Öğrenme Alanlarına Göre TIMSS 2019 Fen Bilimleri 4. Sınıf Değerlendirmesinin Yüzdelik Dağılımları

Canlı Bilimleri %45	• Organizmaların özellikleri ve yaşam süreçleri • Yaşam döngüleri, üreme ve kalıtım • Organizma, çevre ve birbirleriyle etkileşimleri • Ekosistemler • İnsan Sağlığı
Fiziksel Bilimler %35	• Maddenin özellikleri ve sınıflandırılması ve maddedeki değişiklikler • Enerji türleri ve enerji aktarımı • Kuvvetler ve hareket
Yer Bilimi %20	• Yerkürenin fiziksel özellikleri, kaynakları ve tarihi • Yerkürenin hava durumu ve iklimler • Güneş sisteminde yerküre

3.4. TIMSS 2019 Yeterlik Düzeyleri

TIMSS uygulamasında öğrencilerin başarılarının davranış göstergelerini oluşturmak üzere dört farklı yeterlik düzeyi tanımlanmıştır. Bu yeterlik düzeyleri sırasıyla İleri (625 puan), Üst (550 puan), Orta (475 puan) ve Alt (400 puan) olarak sınıflandırılmaktadır. Öğrencilerin aldıkları puana göre matematik ve fen alanlarında neler yapabilecekleri bu yeterlik düzeyleri ile somut bir şekilde görülebilmesi sağlanmaktadır. Puanı 400'ün altında olan öğrenciler ise alt düzeye erişemeyen öğrenciler olarak tanımlanmaktadır ve Birleşmiş Milletler'in "Sürdürülebilir Kalkınma Hedefleri" ne göre ulaşılması gereken en düşük yeterlilik seviyesi olarak kabul edilmektedir (I. Mullis ve diğ., 2020).

4. MATERYAL VE METOD

Bir yönetim birimi, karar alma sürecinde kontrolü sağlamak için performans ölçümü gibi önemli bir araç kullanabilir. Performans ölçümü, karar biriminin faaliyetleri hakkında geri bildirim sağlayarak sorumlulukları yerine getirmesine yardımcı olur. Bu süreçte, karar biriminin farklı kademe ve seviyelerinde etkinlik, etkililik, verimlilik, kârlılık ve performans gibi farklı göstergeler kullanılır. Ancak bu kavramlar sık sık birbirinin yerine kullanılır ve karıştırılır (Cavlak, 2021). Buna göre:

Etkinlik (Efficiency): Kaynakları mümkün olduğunca tasarruflu kullanarak bir işi yapma yeteneğidir. Başka bir deyişle, verilen kaynaklarla mümkün olan en iyi sonucu elde etme kabiliyetidir. Örneğin, bir üretim sürecinde, üretilen birim başına kullanılan kaynakların maliyeti.

Etkililik (Effectiveness): Bir hedefi gerçekleştirmek için ne kadar başarılı olduğumuzu gösteren bir ölçüttür. Bir işin etkililiği, belirlenen hedefe ne kadar yakın bir şekilde tamamlandığına bağlıdır. Örneğin, bir müşteri memnuniyeti hedefi belirlendiğinde, müşterilerin memnuniyetini arttırmak için yapılan çalışmaların etkililiği, müşteri memnuniyeti ölçümleriyle belirlenir.

Verimlilik (Productivity): Belirli bir zaman aralığında gerçekleştirilen iş miktarıdır. Başka bir deyişle verimlilik, kaynakların belirli bir süre içinde ürettiği çıktının miktarıdır. Örneğin, bir saat içinde üretilen ürün miktarı veya bir çalışanın belirli bir süre içinde tamamladığı iş miktarı.

Performans (Performance): Bir görevi yerine getirme yeteneği veya başarılı bir sonuç elde etme yeteneği ile ilgilidir. Performans, bir kişinin ya da bir organizasyonun hedeflerini ne kadar başarılı bir şekilde yerine getirdiğini ölçer. Örneğin, bir sporcu ya da bir takımın maç performansı, belirli bir görevi yerine getirme yeteneğiyle ölçülür.

Etkinlik ölçümü, birimin nerede olduğunun belirlenmesine olanak sağlamakta ve eldeki girdilerden ne denli iyi bir biçimde çıktı üretebileceğini göstermektedir. Etkinlik ölçme yöntemleri, rasyo analizi ve sınır etkinliği analizi olmak üzere iki grupta değerlendirilebilmektedir (Charnes ve diğ., 1997) .

Rasyo (Oran) Analizi etkinlik ölçümünde yaygın kullanılan etkinlik ölçme yöntemidir. Bu yöntem bir tek girdi ile bir tek çıktının birbirleriyle oranlanması sonucu oluşan bir rasyonun zaman içinde izlenmesi şeklinde ifade edilmektedir (Cooper, Seiford ve Tone, 2007).

Sınır Etkinliği Analizi ise, bir birimin en iyi performans gösteren birimlerle karşılaştırılması sonucu elde edilen bir etkinlik ölçütüdür. Bu yöntemde, birimin girdileri ve çıktıları belirlenerek, bu girdi ve çıktılar arasındaki ilişkiler matematiksel olarak modellenir ve işletmenin performansı bu modele göre değerlendirilir (Zhu, 2009). Sınır etkinliği analizi, verimlilik ve etkinlik analizinde kullanılan çok kriterli karar verme yöntemidir. Bu yöntem, girdi ve çıktıları matematiksel olarak modelleyerek bir karar birimlerinin performansını ölçer. VZA, genellikle aşağıdaki dört yöntemle gerçekleştirilir (Hwang, Lee ve Zhu, 2016):

- Veri Zarflama Analizi (Data Envelopment Analysis- VZA): Karar biriminin birim girdi ile birim çıktı arasındaki ilişkilerini modelleyen bir yöntemdir.
- Stokastik Sınır Analizi (Stochastic Frontier Analysis- SFA): Karar biriminin performansını, rastgele hatalar içeren bir üretim fonksiyonu ile modelleyen bir yöntemdir.
- Parametrik Olmayan Sınır Analizi (Non-Parametric Frontier Analysis- NPF): Karar biriminin performansını, parametrelerini önceden belirlemeden modellenen bir yöntemdir.
- Meta-Sınır Analizi (Metafrontier Analysis- MFA): Farklı alanlardaki karar birimlerin (örneğin farklı sektörlerdeki işletmelerin) performanslarını karşılaştıran bir yöntemdir.

VZA yöntemi, en yaygın kullanılan sınır etkinliği analizi yöntemidir. Ancak, her yöntemin kendine özgü avantajları ve dezavantajları vardır. Bu nedenle seçilen yöntem; ölçülen verimlilik türüne, veri setinin büyüklüğüne, ölçülen girdi ve çıktıların doğasına ve diğer faktörlere bağlı olarak değişebilmektedir (Olesen ve Petersen, 2016).

VZA, çok sayıda girdi ve çıktı değişkeni olan karmaşık veri setleriyle çalışmak için daha uygundur. Bu nedenle genellikle çok boyutlu karar verme problemlerinin

çözümü için kullanılmaktadır. Rastgele hatalar içeren bir üretim fonksiyonu ile karar biriminin performansını modelleyen SFA ise ölçülen performansı daha doğru bir şekilde hesaplayabilmektedir. Ancak SFA, çok boyutlu veri setleriyle çalışmak için uygun değildir. Bu nedenle genellikle tek boyutlu veri setleri için kullanılmaktadır. VZA, çıktıların önceden belirlenmiş bir hedef olmadığı durumlarda kullanılabilir. Bunun nedeni, VZA'nın girdi ve çıktıları birbirine göre orantılı olan birimler için etkinlikleri ölçebilmesidir. VZA, karar birimleri arasında kaynak paylaşımının olması durumunda da kullanılabilir. Bu nedenle, özellikle kamu hizmetleri gibi faaliyetlerde yaygın olarak kullanılmaktadır. Ancak SFA, çıktıların önceden belirlenmiş bir hedef olması gerektiği durumlarda kullanılabilmesi söz konusudur. Bu nedenle, özellikle işletmeler gibi önceden belirlenmiş hedefleri olan faaliyetler için daha uygundur (Olesen ve Petersen, 2016). Bu tez araştırmasının kapsamı uluslararası düzeyde ve kamu hizmetlerini ilgilendiren bir kıyaslama üzerine olduğu için veri zarflama metodolojisi tercih edilmiştir.

4.1. Veri Zarflama Analizi (VZA) Metodolojisi

Veri zarflama analizi, birimlerin (örneğin işletme, ülke, kurum, kuruluş, vs.) etkinliğini ölçmek için kullanılan çok değişkenli analiz yöntemidir. VZA, çok sayıda girdi ve çıktı değişkenini kullanarak, verilen kaynaklarla en yüksek çıktıyı sağlayan birimlerin belirlenmesini sağlar. Bu yöntemde, diğer birimlerle karşılaştırılabilir olan birimlerin oluşturduğu bir referans seti oluşturulur ve bu referans setinin dışındaki birimlerin etkinliği ölçülmektedir (Emrouznejad ve Yang, 2018). Bu yöntem, özellikle performans değerlendirmesi ve kaynak tahsisinde de kullanılabilir. Örneğin, bir şirketin üretkenliğini artırmak veya bir kamu hizmetinin maliyet etkinliğini ölçmek için VZA yöntemi tercih edilmektedir. Bu yöntem, aynı zamanda çevresel etkilere de odaklanabilmekte ve çevresel sürdürülebilirliği değerlendirmek için kullanılabilir (Kılış ve Kılış, 2016).

VZA, birimlerin performansını ölçmek için çok sayıda kriteri kullanılmaktadır. Bu kriterler birimlerin verimliliği, etkinliği, kalitesi ve maliyeti gibi faktörleri içerebilmektedir. Böylece birimlerin bu faktörlere ne kadar uygun olduğu belirlenerek performanslarını karşılaştırmak için bir ölçüt sağlanmış olunur. Tüm bunların sonucu olarak bu analiz yöntemi, veri setlerini matematiksel bir modele dönüştürür ve birimlerin girdi ve çıktıları kullanarak etkinliklerini hesaplar (T. J. Coelli ve diğ., 2005).

VZA modellerinin temel amacı, tüm birimleri kullanarak en yüksek performansı elde etmek ve etkinlik arayışında olan birimlerin zayıf yönlerini belirleyerek bunların nasıl geliştirilebileceğine dair öneriler sunmaktır. Bu amaçla genellikle iki tür yaklaşım kullanılır: İlk olarak, Veri Zarflama Analizi Girdi-Çıktı Modeli olarak adlandırılan model, birimlerin girdi ve çıktılarını hesaplar ve bu verileri kullanarak bir performans ölçütü belirler. İkinci olarak, Veri Zarflama Analizi İzleme Modeli olarak adlandırılan model, birimlerin performansını belirli aralıklarla izler ve verimliliklerini sürekli olarak ölçer (Zhu, 2009). Sonuç olarak, birimlerin gözlemlenen verilerine dayanarak etkinliği ölçülür ve her birim için bir etkinlik puanı hesaplanır. Bu puanlar birimlerin performansını kıyaslamak için kullanılabilir. VZA yöntemi çoğu kez optimizasyon modelleri ile birlikte kullanılmakta ve karar vericilere kaynak tahsisinde veya performans değerlendirmelerinde kullanılabilecek alternatif çözümler sunmaktadır (Emrouznejad, Parker ve Tavares, 2008).

1970'lerde ortaya çıkan VZA yöntemleri o zamandan beri önemli bir gelişme kaydetti. Kronolojik olarak VZA yöntemlerinin gelişimi şöyledir:

1978: Charnes, Cooper ve Rhodes tarafından geliştirilen ve lineer programlama yaklaşımı temelinde çalışan CCR (Charnes, Cooper, Rhodes) modeli ile VZA yöntemleri başladı. Bu model ölçeğe göre sabit getiri (constant return scale-CRS) varsayımına dayanmaktadır (Charnes ve diğ., 1978).

1984: Banker, Charnes ve Cooper tarafından geliştirilen ve VZA yöntemlerinin daha etkili bir şekilde kullanılmasını sağlayan BCC (Banker, Charnes, Cooper) modeli yayınlandı. Ölçeğe göre değişken getiri (variable return scale-VRS) varsayımını içermektedir (Banker ve diğ., 1984).

1985: Doyle ve Green tarafından geliştirilen non-radial yöntemlerle VZA yöntemleri genişledi (Doyle ve Green, 1994).

1990'lar: VZA yöntemleri, geleneksel lineer programlama yaklaşımından farklı olarak diğer optimizasyon teknikleri (örneğin, ağ tabanlı, tabu arama ve genetik algoritma) kullanılarak uygulanmaya başlandı (Emrouznejad ve De Witte, 2010).

2000'lerin başları: VZA yöntemleri, çoklu hedef optimizasyonu gibi yeni uygulamaların ortaya çıkmasıyla birlikte daha da genişledi (Cook, Tone ve Zhu, 2014).

2010'ların başları: VZA yöntemleri, daha spesifik uygulama alanlarında (örneğin, çevre etkisi ölçümü) kullanılmak üzere geliştirildi (Cooper ve diğ., 2007).

Literatürde yoğun bir şekilde kullanılan VZA yöntemlerinden bazıları ise şunlardır (Zhu, 2009):

- CCR modeli: Bu model ölçeğe göre sabit getiri varsayımına dayanır. Yani karar verme birimlerinin üretim ölçeği değişse bile etkinliklerinin değişmeyeceğini kabul eder. Bu model girdi veya çıktı odaklı olabilir.

- BCC modeli: Bu model ölçeğe göre değişken getiri varsayımına dayanır. Yani karar verme birimlerinin üretim ölçeği değiştikçe etkinliklerinin de değişebileceğini kabul eder. Bu model de girdi veya çıktı odaklı olabilir.

- SBM modeli: Bu model girdi ve çıktılarıdaki fazlalık ve eksiklikleri dikkate alır. Bu model sadece girdi odaklıdır.

- Süper etkinlik modeli: Bu model aşırı değerlerin etkisini azaltmak için kullanılır. Bu modelde etkin olan karar verme birimleri arasında da bir sıralama yapılır.

- Network modeli: Bu model karar verme birimlerinin birden fazla aşamadan oluşan bir üretim sürecine sahip olduğunu varsayar. Bu modelde her aşama için ayrı bir etkinlik hesaplanır.

- Dinamik model: Bu model karar verme birimlerinin zaman içindeki performansını ölçmek için kullanılır. Bu modelde geçmiş dönemlerden gelen stoklar ve gelecek dönemlere aktarılan stoklar dikkate alınır.

VZA, karar birimlerinin (Decision Making Units - DMU) etkinliklerini ölçmeyi amaçlar. Bir karar birimi etkin bir şekilde işletilmezse ya da çalışma koşullarında dezavantajlı bir duruma sahipse etkinsiz hale gelir. (Kutlar ve Babacan, 2008). CRS (Constant Returns to Scale) ve VRS (Variable Returns to Scale) etkinlik skorları arasındaki oran ölçek etkinliğini belirlemek için kullanılır. CRS ve VRS modelleri girdi veya çıktı odaklı olarak kurulabilir. VZA'nın kullanım alanları ve varsayımlarına bağlı olarak birçok VZA modeli oluşturulabilir. Bu modeller genellikle girdiye ve çıktıya odaklı olmak üzere iki ana gruba ayrılır. Girdiye odaklı modeller, etkin olmayan karar birimlerinin girdilerini herhangi bir çıktı düzeyi için ne kadar azaltmaları gerektiğini araştırır. Çıktıya odaklı etkinlik ölçütleri, etkin

olmayan karar birimlerinin çıktılarını herhangi bir girdi bileşimi için ne kadar artırabileceklerini araştırır. Girdi veya çıktı odaklı modeller kurulamazsa, toplamsal modeller kullanılır. Dolayısıyla girdi odaklı modeller, sabit çıktı düzeyinde girdi kullanımını oransal olarak azaltarak teknik etkinsizliği ölçmeye çalışır. Çıktı odaklı modeller ise sabit girdi düzeyinde çıktı kullanımını oransal olarak artırarak teknik etkinsizliği ölçer. Bu iki ölçek değeri CRS altında aynı değeri verirken VRS altında biraz farklılık gösterir. Literatürdeki çalışmaların çoğunda girdi miktarı öncelikli karar değişkeni olduğu için genellikle girdi odaklı modeller tercih edilir. Ancak kaynak miktarının sabit tutulup çıktı miktarının mümkün olduğunca artırılması gereken durumlarda çıktı odaklı modeller daha uygun olabilir. Yönelim seçimi elde edilen skor değerleri üzerinde küçük bir etkiye sahiptir. (T. Coelli ve Perelman, 2000).

4.2. Ölçeğe Göre Sabit Getiri ve Ölçeğe Göre Değişken Getiri modelleri

Charnes ve diğ. (1978) tarafından geliştirilen model yazarların baş harfleriyle CCR modeli olarak adlandırılır. CCR modeli, sabit getirili bir ölçek modelidir. Bu modelde herhangi bir karar birimi için tüm girdilerdeki %1'lik bir artışın çıktılarda aynı oranda bir değişiklik yaptığı durumda etkin olarak ifade edilir. Banker ve diğ. (1984) tarafından geliştirilen model ise yazarların baş harfleriyle BCC modeli olarak adlandırılır. BCC modeli, değişken getirili bir ölçek modelidir. Bu modelde herhangi bir karar biriminin tüm girdilerindeki %1'lik bir artışın çıktılarda diğer karar birimlerine göre en fazla veya en az değişiklik yaptığı durumda etkin olarak ifade edilir. Dolayısıyla girdi (çıktı) odaklı primal (dual) modeller tablodaki gibi gösterilir.

Tablo 4.1. Girdi/Çıktı Odaklı-Primal/Dual VZA Modelleri

	Primal	Dual
Girdi odaklı	$\text{mak } z_o = \sum_{r=1}^s u_r y_{ro} - (u_o)^*$ Kısıtlar: $\sum_{i=1}^m v_i x_{io} = 1$ $-\sum_{i=1}^m v_i x_{ij} + \sum_{r=1}^s u_r y_{rj} - (u_o)^* \leq 0$ $v_i \geq 0 \quad u_r \geq 0 \quad u_o \text{ Sınırsız}$	$\min \theta_o$ Kısıtlar: $\sum_{j=1}^n \lambda_j x_{ij} - \theta_o x_{io} \leq 0 \quad i = 1, \dots, m$ $y_{ro} - \sum_{j=1}^n \lambda_j y_{rj} \leq 0 \quad r = 1, \dots, s$ $\left(\sum_{j=1}^n \lambda_j = 1 \right)^*$

Tablo 4.2. (Devam)

Çıktı odaklı	$\min z_o = \sum_{i=1}^m v_i x_{io} - (v_o)^*$ Kısıtlar: $\sum_{r=1}^s u_r y_{rj} = 1$ $\sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s u_r y_{rj} - (v_o)^* \leq 0$ $v_i \geq 0 \quad u_r \geq 0 \quad v_o \text{ Sınırsız}$	$\text{mak } \phi_o$ Kısıtlar: $\sum_{j=1}^n \lambda_j x_{ij} - \phi_o x_{io} \leq 0 \quad i = 1, \dots, m$ $\phi_o y_{ro} - \sum_{j=1}^n \lambda_j y_{rj} \leq 0 \quad r = 1, \dots, s$ $\left(\sum_{j=1}^n \lambda_j = 1 \right)^*$
---------------------	--	--

()^{*} BCC modelde kısıt olarak eklenir.

Burada x_{ij} , X kümesine ait olan ve $n \times m$ boyutunda tüm karar birimlerinin girdilerini temsil eden bir matristir. X ise tüm karar birimlerinin girdilerini içeren $n \times m$ boyutunda bir matristir. y_{rj} , Y kümesine ait olan ve $n \times s$ boyutunda tüm karar birimlerinin çıktıları temsil eden bir matristir. Y ise tüm karar birimlerinin çıktıları içeren $n \times s$ boyutunda bir matristir. x_{io} ve y_{ro} , bir karar biriminin sırasıyla girdilerini ve çıktıları temsil eden $m \times 1$ ve $s \times 1$ boyutunda vektörlerdir. u_r ve v_i ise çıktı ve girdi ağırlık vektörleridir. θ_o ve ϕ_o ise bir karar biriminin etkinlik skorunu

ifade eden skaler değerlerdir. Varsayılan bu kısıtlar altında, modelde bulunan n adet karar birimi ayrı ayrı çözülerek her bir karar biriminin skor değerleri elde edilir. Primal doğrusal modelin kısıt sayısı, genellikle işlem kolaylığı açısından dual modele göre daha fazla olduğundan dolayı dual model tercih edilir.

4.3. Örnekleme (Bootstrap) Veri Zarflama Metodolojisi

Bootstrap VZA algoritması, Veri Zarflama Analizi yönteminin etkinliğini hesaplamak için kullanılan bir tekniktir. Bu algoritma, VZA yöntemini kullanarak her bir gözlem birimi için bir etkinlik puanı hesaplar ve bu puanlar aracılığıyla birimlerin göreceli etkinliklerini karşılaştırır (Emrouznejad ve Yang, 2018).

Bootstrap VZA algoritması, birçok farklı girdi ve çıktı değişkeni içeren bir veri setinde kullanılabilir. Bu algoritma örnekleme (sampling) yöntemlerini kullanarak orijinal veri setindeki gözlem birimlerinin alt kümelerini oluşturur ve her alt küme üzerinde VZA yöntemini uygular. Bu şekilde her alt kümeden elde edilen etkinlik puanlarının ortalaması hesaplanır ve birimlerin göreceli etkinlikleri belirlenir. Bu işlem orijinal veri seti üzerinde birden fazla kez tekrarlanır ve elde edilen sonuçların ortalaması daha güvenilir bir etkinlik ölçüsü sağlar. Ancak bu algoritma etkinlik analizi için yaygın olarak kullanılan bir teknik olmasına rağmen büyük veri setleri üzerinde hesaplama yaparken hesaplama gücü gerektiren bir işlemdir (Emrouznejad ve Yang, 2018).

Tahmin edicilerin örnekleme dağılım özelliğinden dolayı geleneksel nokta tahminleri VZA'yı tutarlı bir etkinlik tahmincisi olarak görmek için yetersizdir (Leopold Simar, 1992). Bu durum VZA sonuçlarının güvenilirliğini artırmak için bootstrapping teknikleri kullanılarak geliştirilen Bootstrap VZA yöntemlerinin ortaya atılmasını ve geliştirilmesine neden olmuştur. Bu sınırlamanın üstesinden gelmek için Efron (1992)'un yeniden örnekleme (bootstrap) metodolojisi tahmini sınırın örnekleme varyasyonlarına göre ölçülen etkinlik puanlarının duyarlılığını değerlendirmek için uygun bir araç olarak karşımıza çıkmaktadır. Leopold Simar ve Wilson (1998) geleneksel VZA prosedüründen daha teknik etkinlik tahminlerinin güvenilirliğini artırmak için Efron (1979)'un yaklaşımlarına dayanan parametrik olmayan sınır modelleri için genel bir metodoloji ortaya koymuştur. Ayrıca basitleştirilmiş bir şekilde bootstrap'ın tipik olarak orijinal tahmin edicinin örnekleme dağılımını çoğaltmak için yeniden örnekleme ve orijinal tahmin ediciyi her benzetilmiş örneğe ekleyerek veri üretme sürecini (Data Generation Process-

DGP) tekrar tekrar simüle etme ilkesine dayandığını belirtmişlerdir. Bu nedenle nokta tahminlerinin örnekleme dağılımında stokastik hatalar varsa analitik olarak türetilmeyen güven aralıklarını oluşturmak için bootstrap VZA prosedürü gerçekleştirilebilir (Toma ve diğ., 2017).

Bu çalışmada Simar ve Wilson tarafından geliştirilen bootstrapped VZA yöntemi uygulanmaktadır (Leopold Simar ve Wilson, 1998; Léopold Simar ve Paul W Wilson, 2000; Léopold Simar ve Wilson, 2002). Bootstrapping, değişkenlerin dağılımı veya veri oluşturma süreci (DGP) hakkında önceden varsayımlar veya bilgi olmadan etkinlik skorları için çıkarımsal istatistikler geliştirerek VZA'nın parametrik olmayan doğasına iVZA bir çözüm sunmaktadır. Bootstrapped VZA için örnekleme dağılımları Monte Carlo simülasyonu kullanılarak yaklaştırılır.

Örneğin, gerçek veri oluşturma sürecinin bilinmeyen P olduğunu varsayalım. Leopold Simar ve Wilson (1998)'a göre ampirik örnekleme dağılımına dayalı Monte Carlo simülasyonu kullanılarak $\bar{P}$ olarak adlandırılan tutarlı bir P tahmini aşağıdaki ilişki ile $\bar{\theta}^*$ etkinlik skorlarına karşılık gelen bir vektörü verecek şekilde simüle edilebilir:

$$(\bar{\theta}^* - \hat{\theta})|\bar{P} \approx (\hat{\theta} - \theta^*)|P \quad (1)$$

Burada $\hat{\theta}$, VZA dan elde edilen orijinal etkinlik skoru olup θ^* etkinlik skorlarının gerçek ve bilinmeyen vektörüdür. Denklem (1) VZA ($\hat{\theta}$) ile etkinlik puanının yanlılığını tahmin etmemizi sağlar. Yukarıdaki ilişki $\bar{\theta}^* > \hat{\theta} > \theta^*$ nedeniyledir.

Leopold Simar ve Wilson (1998)'a göre, θ^* tahmini şu şekilde hesaplanabilir:

$$\tilde{\theta}^* = 2\hat{\theta} - \bar{\theta}^* \quad (2)$$

Burada

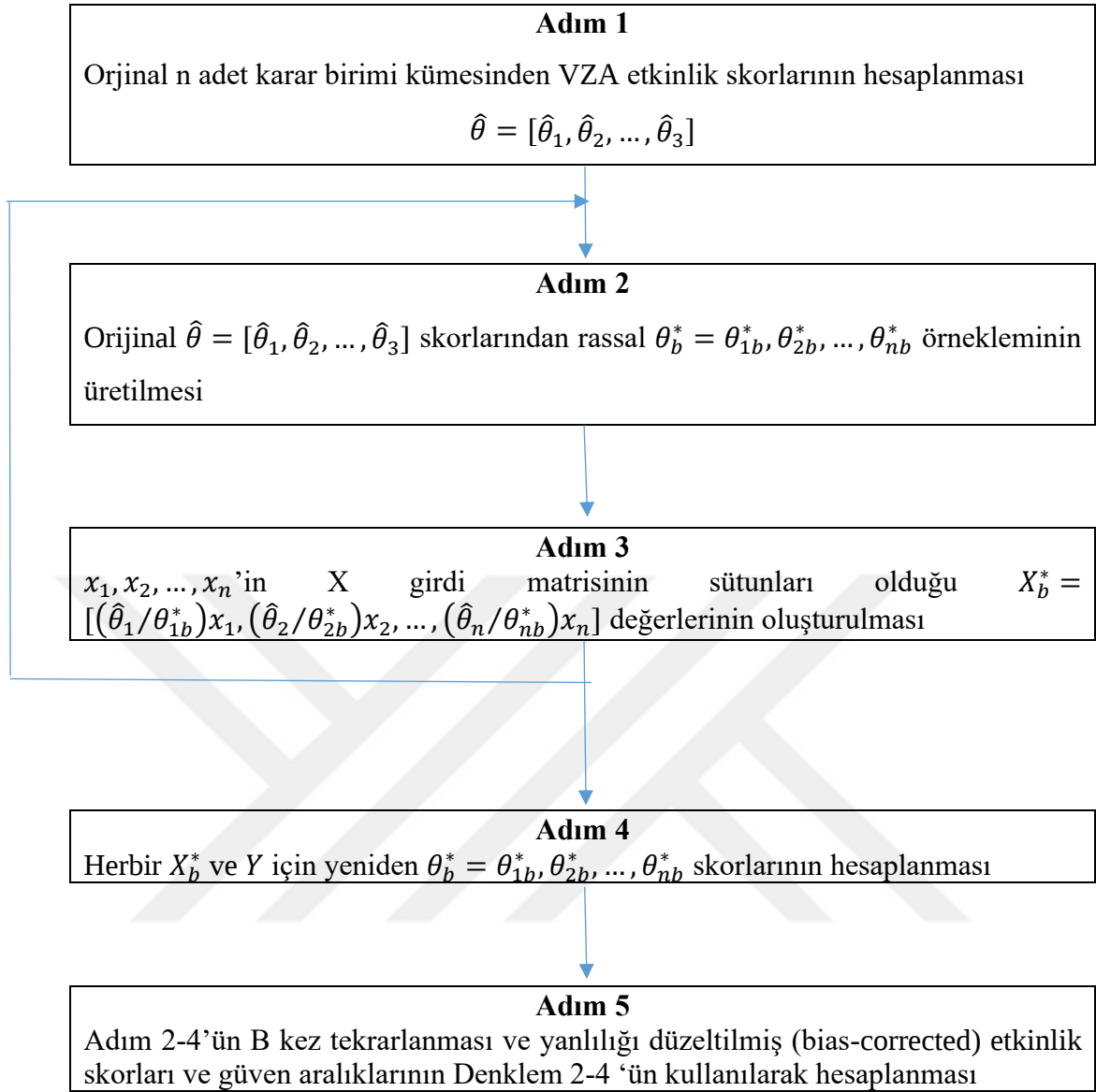
$$\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b \quad (3)$$

$$s_{\hat{\theta}} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b - \bar{\theta}^*)^2} \quad (4)$$

Yeniden örnekleme örneklerinin sayısına göre $b = 1, 2, \dots B$ 'dir.

$\hat{\theta}_b$ 'nin ampirik dağılımı çarpıksa, medyan değer kullanılabilir. Şekil 1, bootstrapped VZA skorlarında kullanılabilecek genel adımları göstermektedir. Adım 2, bootstrapping'in anahtarıdır. Leopold Simar ve Wilson (1998), 1'lik üst sınıra yakın etkinlik puanlarının tahminini iyileştirmek için düzgün yeniden örnekleme (smooth bootstrap) yönteminin kullanılması gerektiğini önermektedir. 1000 ila 2000 kez yeniden örnekleme yapmak yaygın bir uygulamadır. Bu çalışmada veri seti 2000 kez yeniden örneklenmiştir.





Şekil 4.1. Bootstrapped VZA skorları

Kaynak: Leopold Simar ve Wilson (1998) araştırmasından uyarlanmıştır.

Bootstrap veri zarflama analizi, veri zarflama analizinde yaygın olarak kullanılan bir yöntemdir. VZA, veri setinin dağılımı hakkında herhangi bir ön bilgiye sahip olmadan etkin bir şekilde performans değerlendirmesi yapabilir. Bootstrap yöntemi ise örnekleme dağılımını kullanarak veri setinden rastgele örneklemeler çıkararak örnekleme dağılımının istatistiksel özelliklerini hesaplar. Böylece örnekleme dağılımı sayesinde örneklem büyüklüğüne bağlı olarak istatistiksel güçlüklerin üstesinden gelinir (Bogetoft ve Otto, 2010).

Bootstrap veri zarflama analizi yöntemi VZA modelinin klasik yaklaşımlarına kıyasla birçok avantaja sahiptir (Leopold Simar ve Wilson, 2007):

Dağılım Bilgisi: Bootstrap yöntemi VZA modeli için dağılım bilgisi sağlar.

Bu sayede veri setindeki gözlemlerin dağılımını daha iyi anlayabiliriz. Bu VZA modeli için daha güvenilir sonuçlar sağlar.

Veri Dönüştürme Gerekliliği: VZA modeli veri setindeki verilerin oranlarına dayandığından verilerin dönüştürülmesi gerekebilir. Ancak bu dönüşümler sonuçları etkileyebilir. Bootstrap yöntemi verilerin dönüştürülmesi ihtiyacını ortadan kaldırır.

Örneklem Büyüklüğü Sorunu: VZA modeli, örneklem büyüklüğüne duyarlıdır. Örneklem dağılımı sayesinde örneklem büyüklüğüne bağlı istatistiksel güçlüklerin üstesinden gelinir.

Anlamlılık Testleri: Bootstrap yöntemi VZA sonuçları için anlamlılık testleri yapmamıza olanak tanır. Bu, sonuçların güvenilirliğini artırır.

Bu tez çalışmanın temel amacı küresel bağlamda eğitim sistemlerinin etkinliğini ölçmek ve bu konuda politik arar alıcılara politikalarını iyileştirmesi yönünde yol göstermektir. Bu bağlamda tez araştırmasının ikinci aşamasında en son yapılan Uluslararası Matematik ve Bilim Çalışması Eğilimleri sınavı TIMSS 2019'daki 4. Sınıf düzeyindeki öğrencilerin veri kümelerini kullanarak hesaplanan bootstrap VZA modeli ile yapılan etkinlik ölçümünün yanı sıra Tobit regresyon modeli ile yönetsel uygulamaların etkinlik üzerindeki etkisinin incelenmesi amaçlanmıştır. Böylece Tobit regresyonunun yardımıyla daha iyi eğitim yönetim sistemleri yoluyla mevcut kaynakların daha iyi kullanımının ülke düzeyinde daha yüksek bilişsel başarı sağlayıp sağlamadığının araştırılması hedeflenmektedir.

Bu çalışmada Tobit regresyon modelinin tercih edilmesinin nedeni etkinlik ölçüsü olan VZA skorlarının sıfır ve üst sınırlarının olmasıdır. VZA yalnızca verilerin dağılımını kullanarak etkinlik ölçer. Bu nedenle bir okulun etkinlik skoru sadece diğer okullarla göreceli olarak karşılaştırılarak ölçülür. Bu skorlar sıfır ve üst sınırlarla sınırlıdır. Bu da geleneksel regresyon modelleri kullanılarak analiz edilmelerini zorlaştırır. Tobit regresyon modeli sınırlı bağımlı değişkenlerle çalışmak için özel olarak tasarlanmıştır ve sıfır ve üst sınırlarla sınırlı veriler için en uygun modeldir (Bhutoria ve Aljabri, 2022b).

4.4. Sınırlı bağımlı değişkenli regresyon (Limited Dependent Variable Regression): Tobit Regresyon Modeli

Ekonometrik modelleme; ekonomistler tarafından değişkenler arasındaki

ilişkileri açıklamak, tahmin etmek ve analiz etmek için kullanılan istatistiksel bir araçtır. Ekonometrik modeller, tüketici talebi veya finansal piyasalar gibi değişkenlerin davranışını tahmin etmek ya da ekonomik politika kararlarını bilgilendirmek amacıyla birçok ekonomik ilişkiyi analiz etmek için kullanılmaktadır. Sıklıkla tercih edilen ekonometrik modellerden biri olan doğrusal regresyon modelidir. Bununla birlikte doğrusal ekonometrik modeller, üretilen tahminlerin güvenilir olabilmesi için karşılanması gereken bazı varsayımlara dayanmaktadır (Gujarati, Porter ve Gunasekar, 2009).

Doğrusal ekonometrik modellemedeki varsayımlardan biri modelde kullanılan verilerin bağımsız ve özdeş dağılmış (Independent Identity Distribution-IID) olmasıdır. Bu veri setindeki her gözlemin diğerleriyle ilişkisiz olduğu ve verilerin dağılımının her gözlem için aynı olduğu anlamına gelir. Veriler IID değilse yanlış-biased ve yanlış tahminlere yol açabilir. Bu nedenle ekonomistlerin bu varsayımı karşıladığından emin olmak için verilerini dikkatli bir şekilde seçmeleri ve temizlemeleri önemlidir (Wooldridge, 2010).

Ekonometrik modellemedeki diğer bir varsayım da bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin doğrusal olmasıdır. Bu bağımsız değişkendeki bir değişikliğin bağımlı değişkende orantılı bir değişikliğe yol açtığı anlamına gelir. İlişki doğrusal değilse yanlış tahminlere yol açabilir. Ayrıca model bağımsız değişkenler arasında çoklu bağlantı olmadığını varsaymaktadır. Bu yanlış tahminlere de yol açabileceğinden bağımsız değişkenlerin birbiriyle yüksek oranda ilişkili olmadığı anlamına gelir (Wooldridge, 2010).

Genel olarak klasik ekonometrik modellemenin varsayımları güvenilir tahminler ve tahminler üretmek için kritik öneme sahiptir. Ekonomistler modellerini tasarlarlarken, verilerini seçip temizlerken de bu varsayımları dikkatle göz önünde bulundurmalıdır. Bu varsayımlar kısıtlayıcı görünse de ekonometrik modellerin sonuçlarının ekonomik politika kararlarını bilgilendirmek için kullanılabilmesini sağlamak için gereklidir.

Sınırlı bağımlı değişken regresyon modelleri ise bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi incelemek için kullanılan başka bir ekonometrik modelleme türüdür. Bu tür regresyon modelleri bağımlı değişkenin normal dağılmadığı veya sınırlı aralığa sahip olduğu durumlarda yaygın olarak kullanılır. Bu tür regresyon modellerinin amacı da tıpkı diğer klasik regresyon

modellerinde olduğu gibi bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi tahmin etmektir. Birçok durumda bağımlı değişken bir kişinin bir ürünü satın alıp almayacağı bir adaya oy verip vermeyeceği veya belirli bir hastalığa sahip olup olmayacağı gibi bir olayın veya kararın sonucudur. Bu gibi durumlarda bağımlı değişken ikili olabilir veya sınırlı bir aralığa sahip olabilir. Bu da geleneksel regresyon modellerinin kullanılmasını zorlaştırır. Sınırlı bağımlı değişken regresyonu bu tür ilişkileri modellemek ve bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini tahmin etmek için bir yol sağlar (Gujarati ve diğ., 2009).

Her biri kendine özgü özelliklere ve uygulamalara sahip olan çeşitli Sınırlı Bağımlı Değişkenli Regresyon modelleri vardır. Örneğin, en yaygın türlerden biri ikili sonuçları modellemek için kullanılan İkili Logit Regresyonudur. Bu model bir dizi tahmin edici değişken göz önüne alındığında bir olayın gerçekleşme olasılığını tahmin eder. Diğer bir tür ise ikiden fazla olası sonuç olduğunda kullanılan Multinomial Logit Regresyondur. Bu model bir dizi öngörücü değişken göz önüne alındığında her bir sonucun olasılığını tahmin eder. Sınırlı Bağımlı Değişkenli Regresyon modellerinin diğer türleri arasında Probit Regresyonu, Tobit Regresyonu ve Poisson Regresyonu yer alır (Stock ve Watson, 2003).

Sınırlı Bağımlı Değişkenli Regresyon modelleri; ekonomi, finans, pazarlama ve sosyoloji dahil olmak üzere birçok alanda yaygın olarak kullanılmaktadır. Özellikle müşteri davranışları, oy verme kalıpları veya tıbbi teşhisler gibi ikili veya kategorik sonuçlara sahip verileri analiz ederken kullanılırlar. Araştırmacılar bu modelleri kullanarak ilgilenilen sonuçları etkileyen faktörleri daha iyi anlayabilir ve daha doğru tahminlerde bulunabilirler (Maddala, 1987).

Sınırlı bağımlı değişkenli model türlerinden biri olan Tobit modeli, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi tahmin etmek için kullanılan bir tür regresyon analizidir. Spesifik olarak Tobit modeli, bağımlı değişken sansürlendiğinde kullanılır. Yani bazı gözlemler tam olarak gözlemlenmez veya kesilir. Bu model ilk olarak 1950'lerde James Tobin tarafından tanıtılmış ve o zamandan beri ekonometride yaygın olarak kullanılan bir araç haline gelmiştir (W. Greene, 2011).

Tobit model fikri, olabilirlik fonksiyonunu gizli bağımlı değişkenin belirlenen eşik üstünde veya altında olmasına bağlı olarak her gözlem için eşit olmayan örnekleme olasılığını yansıtacak şekilde değiştirmektir (Amemiya, 1984). Tobit

modelde orijinal durumunda olduğu gibi sıfırda aşağıdan sansürlenene bir örnek için limit dışı her gözlem için örnekleme olasılığı basitçe uygun yoğunluk fonksiyonunun yüksekliğidir. Herhangi bir limit gözlem için ise kümülatif dağılım yani uygun yoğunluk fonksiyonunun sıfırın altındaki integralidir. Dolayısıyla tobit olabilirlik fonksiyonu, yoğunlukların ve kümülatif dağılım fonksiyonlarının bir karışımıdır (Kennedy, 2008).

Aşağıda I. tip Tobit için olabilirlik ve log-olabilirlik fonksiyonları verilmiştir. Bu gizli-latent değişken $y_j^* \leq y_L$ olduğunda y_L 'de aşağıdan sansürlenene bir tobittir. Olabilirlik fonksiyonunu yazarken önce bir gösterge fonksiyonu I tanımlarız:

$$I(y) = \begin{cases} 0 & y \leq y_L \\ 1 & y > y_L \end{cases} \quad (5)$$

Daha sonra, Φ standart normal kümülatif dağılım fonksiyonu ve φ standart normal olasılık yoğunluk fonksiyonu olsun. N gözlemlili bir veri seti için I tipi bir tobit için olabilirlik fonksiyonu şöyledir.

$$L(\beta, \sigma) = \prod_{j=1}^N \left(\frac{1}{\sigma} \varphi \left(\frac{y_j - X_j \beta}{\sigma} \right) \right)^{I(y_j)} \left(1 - \Phi \left(\frac{y_j - X_j \beta}{\sigma} \right) \right)^{1-I(y_j)} \quad (6)$$

Bir başka örnek de y_U 'nin üzerindeki değerlerin sansürlenmesidir.

$$y_i = \begin{cases} y_i^* & y_i^* < y_U \\ y_U & y_i^* \geq y_U \end{cases} \quad (7)$$

Tobit modeli, bağımlı değişkenin belirli bir noktada sansürlendiğini varsayar. Bu durum bu noktanın altına veya üstüne düşen herhangi bir gözlemin tam olarak gözlemlenmediği ve bunun yerine sansür noktasının alt veya üst sınırı olarak kaydedildiği anlamına gelir. Sansürlenene değişkenlere örnek olarak genellikle yalnızca geniş aralıklarda rapor edilen gelir veya belirli bir gün sayısı ile sınırlanabilecek bir kişinin hastalık nedeniyle işe gelmediği gün sayısı verilebilir. Tobit modeli ayrıca bağımsız değişkenlerin bağımlı değişkenle doğrusal bir ilişkiye sahip olduğunu varsayar (W. H. Greene, 2003).

Tobit modelinin en önemli avantajlarından biri bağımsız değişkenlerin bağımlı değişken üzerindeki etkisinin bağımlı değişken sansürlenmiş olsa dahi tahmin edilebilmesine imkân vermesidir. Bu özellikle bağımlı değişkenin ilgi çekici olduğu ancak tam olarak gözlemlenmediği durumlarda kullanışlıdır. Örneğin, gelir

eşitsizliğiyle ilgili bir çalışmada gelir yalnızca geniş aralıklarda rapor edildiğinde bile eğitim, yaş ve diğer faktörlerin gelir üzerindeki etkisini tahmin etmek için Tobit modeli kullanılabilir. Genel olarak, Tobit modeli sansürlenmiş verileri analiz etmek için güçlü bir araçtır ve ekonomi, sosyoloji ve halk sağlığı dahil olmak üzere çeşitli alanlarda uygulamaları vardır. Ancak Tobit modelinin geçerli olabilmesi için belirli varsayımların karşılanması gerekir ve model uygulanmadan önce bu varsayımların dikkate alınması gerekir (Gagnon ve Barber, 2018).

Tobit modelinin temel varsayımlarından biri hata teriminin normal bir dağılım izlemesidir. Bu da hata teriminin dağılımının ortalama etrafında simetrik olması gerektiği anlamına gelir. Hata teriminin normal dağılıma uymaması yanlış tahmin ve kestirime yol açabilir. Bu nedenle Tobit modelini kullanmadan önce hata teriminin normalliğini test etmek önemlidir. Çünkü Tobit modeli maksimum olabilirlik tahminine dayanır ve bu da hata teriminin normal bir dağılım izlemesini gerektirir. Hata terimi normal dağılım göstermiyorsa Tobit modelinin sonuçları yanlış veya tutarsız olabilir. Araştırmacılar normallik varsayımını Shapiro-Wilk testi veya Kolmogorov-Smirnov testi gibi tanısal testler kullanarak test etmelidir (W. Greene, 2011).

Tobit modelinin bir diğer önemli varsayımı da sansürleme mekanizmasının dışsal olmasıdır. Bu bir değerin sansürlenme olasılığının değerin kendisiyle veya modeldeki başka herhangi bir değişkenle ilişkili olmaması gerektiği anlamına gelir. Eğer sansürleme mekanizması içsel ise Tobit modeli yanlış veya tutarsız sonuçlar üretebilir. Dışsallığı test etmek için araştırmacılar Hausman testi veya Durbin-Wu-Hausman testi gibi tanısal testleri kullanabilirler (Frey, 2018).

Tobit modelinin üçüncü bir varsayımı hata teriminin varyansının tüm gözlemlerde sabit olmasıdır. Bu varsayım homoscedasticity olarak bilinir. Hata terimi gözlemler arasında eşit olmayan varyanslara sahipse Tobit modeli yanlış veya tutarsız sonuçlar üretebilir. Araştırmacılar, Breusch-Pagan testi veya White testi gibi tanısal testleri kullanarak homoscedasticity'yi test edebilirler. Eşvaryanslılık varsayımı ihlal edilirse araştırmacıların bunun yerine genelleştirilmiş bir Tobit modeli kullanması gerekebilir.

Genellikle bağımlı değişkenin sadece negatif olmayan değerler aldığı ve bir alt sınıra tabi olduğu durumlarda kullanılır. Model; tüketim, harcamalar ve diğer ilgili değişkenlere ilişkin verileri analiz etmek için ekonomi, sağlık bilimleri ve sosyal

bilimler de dahil olmak üzere bir dizi alanda yaygın olarak kullanılmaktadır (Davidson ve MacKinnon, 2004).

Tobit Modelinin bir diğer önemli varsayımı bağımlı ve bağımsız değişkenler arasındaki ilişkinin doğrusal olması gerektiğidir. Bu bağımsız değişkenlerin bağımlı değişken üzerindeki etkisinin bağımlı değişkenin tüm değerleri boyunca sabit olması gerektiği anlamına gelir. İlişkinin doğrusal olmaması yanlış tahmin ve kestirimlere yol açabilir. Bu nedenle Tobit modelini kullanmadan önce doğrusallığın test edilmesi önemlidir (Cameron ve Trivedi, 2010).

Son olarak, Tobit modeli hata teriminin varyansının bağımlı değişkenin tüm değerlerinde sabit olduğunu varsayar. Bu bağımlı değişkendeki değişkenliğin bağımsız değişkenlerin tüm değerlerinde aynı olması gerektiği anlamına gelir. Varyansın sabit olmaması yanlış tahmin ve kestirimlere neden olabilir. Bu nedenle Tobit modelini kullanmadan önce hata teriminin homoscedasticity'sinin test edilmesi önemlidir (W. Greene, 2007).

Sonuç olarak, Tobit modelinin varsayımları tahminlerin ve kestirimlerin doğruluğunu ve güvenilirliğini sağlamada çok önemlidir. Hata teriminin normalliği, bağımlı ve bağımsız değişkenler arasındaki ilişkinin doğrusallığı ve hata teriminin homoscedasticity olması Tobit modeli kullanılmadan önce test edilmesi gereken önemli varsayımlardır. Araştırmacılar ve analistler bu varsayımlara bağlı kalarak verilerinden güvenilir ve geçerli çıkarımlar yapabilirler.

5. ARAŞTIRMANIN KAPSAMI VE VERİSETİ

Öğrencilerin matematik ve fen alanlarında kazandıkları bilgi ve becerilerin değerlendirilmesine yönelik bir tarama yapan Uluslararası Matematik ve Fen Eğilimleri Araştırması (TIMSS- Trends in International Mathematics and Science Study), Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (International Association for the Evaluation of Educational Achievement-IEA)'nın bir projesi olarak sunulmaktadır. IEA'nın Uluslararası Matematik ve Bilim Çalışmalarında Eğilimler (TIMSS), dünyanın dört bir yanındaki öğrencilerin matematik ve fen bilgilerine ilişkin bir dizi uluslararası değerlendirmedir. Katılan öğrenciler ekonomik gelişme, coğrafi konum ve nüfus büyüklüğü açısından çeşitli eğitim sistemlerinden (ülkeler veya ülkelerin bölgesel yetki alanları) gelmektedir. Katılan eğitim sistemlerinin her birinde en az 4.000 ila 5.000 öğrenci değerlendirilir. Katılımcı öğrencilerin matematik ve fen bilgisi öğrenme koşulları hakkında bağlamsal veriler öğrencilerden ve öğretmenlerinden, müdürlerinden ve velilerinden anketler yoluyla toplanmaktadır (IEA, 2021). TIMSS, IEA tarafından dünya çapındaki eğitim sistemlerinin öğrencilerin eğitim başarılarını karşılaştırmasına ve etkili eğitim politikası tasarlamada başkalarının deneyimlerinden öğrenmesine izin vermeyi amaçlayan çalışmalardan biridir. Bu değerlendirme ilk olarak 1995 yılında yapılmış ve her 4 yılda bir uygulanmıştır. Bu nedenle katılan eğitim sistemlerinden bazıları 1995'ten 2019'a kadar olan değerlendirmeler arasında eğilim verilerine sahiptir (PIRLS, 2021). TIMSS 4. ve 8. sınıf öğrencilerini, TIMSS Advanced ise ortaokulun son yılındaki öğrencileri ileri matematik ve fen alanlarında değerlendirmektedir.

IEA'nın uzun tarihi boyunca her zaman ayrı olarak katılan ülkelerdeki bazı farklı eğitim sistemleri de dahil olmak üzere altmış dört ülke TIMSS 2019'a katılmıştır (örneğin, Belçika'nın Felemenkçe konuşulan kısmı ve Çin Halk Cumhuriyeti'nin Hong Kong Özel İdari Bölgesi (SAR- Special Administrative Region)). Ayrıca TIMSS 2019'da sekiz kıyaslama katılımcısı vardı (örneğin, iller ve belediyeler gibi bölgesel kuruluşlar). Ülkeler ve kıyaslama katılımcıları dördüncü sınıf değerlendirmeye, sekizinci sınıf değerlendirmeye veya her ikisine de katılmayı seçebilmektedir. 2019 yılında TIMSS dördüncü sınıf değerlendirmesine 58 ülke ve altı kıyaslama kuruluşu, sekizinci sınıf değerlendirmesine otuz dokuz ülke ve yedi kıyaslama katılımcısı katılmıştır.

TIMSS'in amacı matematik ve fen öğretimi ve öğrenimini geliştirmeye

yardımcı olmak için politikayla ilgili en iyi bilgileri sağlamaktır. Bu kapsamda TIMSS 2019 öğrenci, öğretmen ve okul anketlerini ve öğrencilerin ebeveynleri veya bakıcıları tarafından doldurulan Erken Öğrenme Anketini ("ev anketi" olarak da adlandırılır) içermektedir. Anket maddelerinin çoğu okullaşmanın önemli yönlerine (örneğin; okul güvenliği, öğrencilerin matematik ve fen öğrenmeye yönelik tutumları, erken başlamanın etkisi) ilişkin sağlam göstergeler sağlamak için madde yanıt teorisi (item response theory-IRT) ölçekleri olarak geliştirilmiş ve analiz edilmiştir. Anket sonuçlarıyla birlikte öğrencilerin matematik ve fen öğrendiği ev, okul ve sınıf bağlamları hakkında zengin bilgiler sağlamaktadır.

TIMSS, katılımcı ülkelerdeki öğrencileri sınav sırasındaki ortalama yaşın en az 9,5 olması koşuluyla örgün eğitimin dördüncü yılında ve sınav sırasındaki ortalama yaşın 13,5 olması koşuluyla örgün eğitimin sekizinci yılında değerlendirir. Eğitim sistemleri, okula başlama yaşı, terfi ve elde tutma ile ilgili olarak yapı ve politika ve uygulamalarda farklılık gösterdiğinden hedef notların nasıl etiketlendiği ve öğrencilerin ortalama yaşı konusunda ülkeler arasında farklılıklar vardır. Ayrıca bazı ülkeler TIMSS'i örgün eğitimin dördüncü veya sekizinci yılından farklı bir sınıfa uygulamayı seçmektedir. Norveç, İsveç ve Finlandiya ile daha iyi karşılaştırmalar elde etmek için beşinci ve dokuzuncu sınıf öğrencilerini değerlendirmeyi seçmiştir. Güney Afrika ve kıyaslama sistemleri, beşinci ve dokuzuncu sınıf öğrencilerini müfredatlarına daha iyi uymaları ve trend ölçümünü sürdürmeleri için değerlendirdi. Türkiye de beşinci sınıftaki öğrencileri değerlendirmeyi tercih etmiştir.

TIMSS 2019'a her sınıfta 150'den 200'e kadar okuldan yaklaşık 4.000 öğrenciden oluşan ulusal temsili örneklem katılmıştır. Matematik ve fen değerlendirmeleri ve bağlam anketleri de dahil olmak üzere 330.000'den fazla öğrenci, 310.000 veli, 11.000 okul ve 22.000 öğretmen dördüncüsüne katılmıştır. 8. sınıf değerlendirmesine 8. sınıf değerlendirmesine 8. sınıf değerlendirmesine 8.000 okuldan 250.000 öğrenci ve 30.000 öğretmen katılmıştır (I. Mullis ve diğ., 2020).

Bu tez araştırmasının ilk aşaması, TIMSS 2019 kapsamında 58 ülkenin 4. sınıf matematik ve fen başarısıyla ilişkili olduğu düşünülen girdilerin etkin şekilde yönetilip yönetilmediğini belirlemeyi hedeflemiştir. Bu amaçla nicel araştırma yöntemleri kullanılarak betimsel tarama modeli uygulanmıştır. Verilerin analizinde ise bootstrap veri zarflama analiz tekniği kullanılmıştır. TIMSS çalışmalarında örnekleme 4. sınıf öğrencileri üzerinden yapılmıştır. Bu çalışmada kullanılan veriler

öğrencilerin, öğretmenlerin ve okul müdürlerinin doldurduğu anketlerin IEA IDB Analyzer programıyla öğrencilerin matematik ve fen sınav başarılarını temel alarak elde edilmiştir. Bu çalışma bootstrap VZA yöntemini uygulamış ve sonuçları standart VZA yöntemiyle üretilen sonuçlarla karşılaştırmıştır.

VZA metodolojisi için girdi ve çıktıların seçimi büyük önem taşımaktadır. Önemli girdi ve çıktılarını seçmek için standart araçlar yoktur. Literatürdeki araştırmamalardan Koyuncu ve Ilgaz (2019)'nın Tablo 5.1.'deki etkinlik göstergeleri önerilmiştir.

Tablo 5.1. Bootstrap (VZA) modelinin analizi için kullanılan girdiler ve çıktılar.

Girdiler	Çıktılar
• Öğretmenlik kıdemi (years been teaching) • Akademik başarıda okulun etkisi (School Emphasis on Academic Success) • Okulun güvenli ve düzenli olması (Safe and Orderly Schools) • Öğrenci ihtiyaçları ile sınırlandırılmış öğretim (teaching limited by student needs) • Evdeki eğitim kaynakları (home educational resources) • Matematik öğretim kaynağı ile öğretim sıkıntısı (instruction affected by mathematics resource shortages) • Fen bilimleri öğretim kaynağı ile öğretim sıkıntısı (instruction affected by mathematics resource shortages) • Akademik başarıda okulun etkisi (school emph on acad success) • Okul disiplin problemleri (school discipline problems)	• Ortalama Matematik Başarısı (Average Mathematics Achievement) • Ortalama Bilim Başarısı (Average Mathematics Achievement)

Yukarıda belirtilen çalışmada bootstrap VZA metodolojisi içerisinde girdi odaklı CCR modelinin kullanılması tercih edilmiştir. CCR tekniği, toplam etkinliği ölçerek bir ülkenin teknik ve ölçek açısından etkin olup olmadığını belirlemektedir. Girdi temelli yaklaşımın seçilmesinin nedeni veri zarflamanın doğasından kaynaklanmaktadır. Eğer bir karar verme birimi etkin değilse girdilerinde azaltma

yapılması önerilmektedir. Bu çalışma bağlamında değişkenlerle ilgili olarak öğretmen veya okul müdürlerinin algılarını azaltmaları önerilmiştir. Ancak girdi azaltma önerisiyle puanın düşürülmesinin doğru şekilde yorumlanması gerektiğine dikkat edilmelidir. Örneğin okul kaynaklarıyla ilgili algı yüksek olsa da azaltma önerisi alınmışsa ya kaynaklar gerçekten yeterli değildir ancak yeterli gibi algılanmaktadır ya da kaynaklar yeterli olsa da verimli bir şekilde kullanılamamaktadır.

Tez araştırmasının ikinci aşamasında etkinlik ve okul yönetimi arasındaki ilişki ülke düzeyinde Tobit modeli ile açıklanmaya çalışılmıştır. Teknik etkinlik puanları bize ülkelerin eğitim girdilerini öğrenme çıktılarına dönüştürmede ne ölçüde etkin olduklarını göstermektedir. Bu aşamada herhangi bir etkinliğe katkıda bulunabilecek farklı ülkelerin okul yönetimi uygulamalarının rolünü ve yönetim uygulamalarının okul etkinliğini ne ölçüde etkilediğini araştırılmaktadır. Okul yönetimi uygulamaları, kaynak yönetimi ve insan yönetimi faktörlerini içerir. Bu faktörler geleneksel anlamda girdi olmasalar da bir okulun genel işleyişinde kritik öneme sahiptirler. Özellikle bu faktörler okul girdilerinin istenen çıktılara dönüştürülmesinde kritik öneme sahiptir. Bu nedenle ilk aşamadan elde edilen etkinlik puanlarının Tobit regresyon yöntemi kullanılarak Eşitlik (8)'i tahmin etmek için kullanıldığı iki aşamalı bir yaklaşım benimsenmiştir:

$$e_i = \beta_0 + \beta M_i + \epsilon_i \quad (8)$$

Burada e_i , i ülkesindeki okulların teknik etkinlik skorlarıdır ve 0 ile 1 arasındadır. M_i ise, i ülkesindeki okulların yönetim değişkenlerinin ortalama vektörüdür. Yönetim değişkenleri ise aşağıdaki gibidir.

Tablo 5.2. Bu çalışmada kullanılan TIMSS 2019 veri kümesindeki değişkenlerin ayrıntıları

DEĞİŞKENLER	AÇIKLAMALAR
Öğrencilerin testi hangi dillerde yaptıklarının yüzdesi	Öğrencilere,"Testi hangi dillerde yaptınız?" sorusu yöneltilmiştir. ACBG04 değişkeni, öğrencilerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken yüzde cinsinden ifade edilir ve öğrencilerin hangi dillerde testi yaptıklarını yansıtır. Örneğin; ACBG04 değişkeni değeri '%80' olan bir veri seti öğrencilerin %80'inin testi ana dili olarak bildikleri dilde yaptığını gösterir.
Okuldaki toplam bilgisayar sayısı	Okul müdürlerine, "Bu okulda toplam kaç bilgisayar bulunmaktadır?" sorusu yöneltilmiştir. ACBG07 değişkeni, müdürlerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken tam sayı değerlerini alır ve okuldaki toplam bilgisayar sayısını gösterir. Örneğin; ACBG07 değişkeni değeri "50" olan bir okul toplam 50 bilgisayarın bulunduğunu yansıtır.

Tablo 5.3. (Devam)

Okulun velilerin öğrencilerin eğitime katılım düzeyine verdiği önemi	Okul müdürlerine, "Okulunuz velilerin öğrencilerin eğitime katılım düzeyine ne kadar önem vermektedir?" sorusu yöneltilmiştir. ACBG14E değişkeni müdürlerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken 5 puanlık bir ölçekte ifade edilir. 1 puan en düşük önemi, 5 puan en yüksek önemi temsil eder. Örneğin; ACBG14E değişkeni değeri "4" olan bir okul velilerin öğrencilerin eğitime katılım düzeyine oldukça önem verdiğini yansıtır.
Okul müdürünün toplam yöneticilik deneyimi	Müdürlük yapan okul yöneticilerine, "Toplam kaç yıl müdürlük yaptınız?" sorusu yöneltilmiştir. ACBG18 değişkeni yöneticilerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken tam sayı değerlerini alır ve müdürün toplam yöneticilik deneyimini yıllar cinsinden gösterir. Örneğin; ACBG18 değişkeni değeri "10" olan bir okul müdürü toplam 10 yıl müdürlük yapmıştır.

Tablo 5.4. (Devam)

Matematik Kaynaklarının Yetersizliğinden Etkilenen Öğretim	Bu değişken matematik kaynaklarındaki eksikliklerin eğitimin kalitesini nasıl etkilediğine dair okul müdürlerinin görüşlerini yansıtır. Okul müdürlerine, "Son bir yıl içinde okulunuzdaki matematik öğretimi matematik kaynaklarında yaşanan eksikliklerden ne kadar etkilendi?" sorusu yöneltilmiştir. ACBGMRS değişkeni müdürlerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken 4 puanlık bir ölçekte ifade edilir. 1 puan "hiç etkililenmedi" ve 4 puan "çok etkilendi" anlamına gelir. Örneğin; ACBGMRS değişkeni değeri "3" olan bir okul müdürü matematik kaynaklarındaki eksikliklerin okulun matematik öğretimini oldukça etkilediğini düşünmektedir.
--	---

Tablo 5.5. (Devam)

Fen Kaynaklarının Yetersizliğinden Etkilenen Öğretim	Bu değişken fen bilimleri kaynaklarındaki eksikliklerin eğitimin kalitesini nasıl etkilediğine dair okul müdürlerinin görüşlerini yansıtır. Okul müdürlerine, "Son bir yıl içinde okulunuzdaki fen bilimleri öğretimi fen bilimleri kaynaklarında yaşanan eksikliklerden ne kadar etkilendi?" sorusu yöneltmiştir. ACBGSRS değişkeni müdürlerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken 4 puanlık bir ölçekte ifade edilir. 1 puan "hiç etkilendi" ve 4 puan "çok etkilendi" anlamına gelir. Örneğin; ACBGSRS değişkeni değeri "2" olan bir okul müdürü fen bilimleri kaynaklarındaki eksikliklerin okulun fen bilimleri öğretimini çok az etkilediğini düşünmektedir.
---	--

Tablo 5.6. (Devam)

Okul müdürlerinin akademik başarıya verilen önem	Okul müdürlerine, "Okulunuzda akademik başarı ne kadar önemlidir?" sorusu yöneltilmiştir. ACBGEAS değişkeni müdürlerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken 4 puanlık bir ölçekte ifade edilir. 1 puan "hiç önemli değil" ve 4 puan "çok önemli" anlamına gelir. Örneğin, ACBGEAS değişkeni değeri "3" olan bir okul müdürü okulun akademik başarısına oldukça önem verdiğini düşünmektedir. Bu veri TIMSS 2019 sonuçlarına ilişkin bir veri setinde yer alır ve farklı analizler için kullanılabilir.
Okul müdürlerinin disiplin konusuna verdiği önem	Okul müdürlerine, "Okulunuzda disiplin ne kadar önemlidir?" sorusu yöneltilmiştir. ACBGDAS değişkeni müdürlerin bu soruya verdikleri yanıtları temel alarak hesaplanır. Değişken 4 puanlık bir ölçekte ifade edilir. 1 puan "hiç önemli değil" ve 4 puan "çok önemli" anlamına gelir. Örneğin; ACBGDAS değişkeni değeri "2" olan bir okul müdürü okul disiplinine bir dereceye kadar önem verdiğini düşünmektedir.

Tablo 5.7. (Devam)

Öğrencilerin yıllık ders saati sayısı	Anket sorusuna göre öğrencilerin bir yıl içinde aldığı tüm ders saatleri toplanarak bu değişken hesaplanmıştır. Değişkenin birimi öğrencilerin bir yılda aldığı toplam ders saati sayısını yansıtır. Örneğin; ACDGTIHY değişkeni değeri "1000" olan bir okulda öğrenciler bir yıl içinde toplam 1000 saat ders almıştır.
Öğrencilerin sınıflarında tablet bilgisayarların bulunup bulunmaması durumu	Anket sorusuna göre öğrencilere, "Sınıfınızda tablet bilgisayar var mı?" sorusu yöneltilmiştir. ATBS03A değişkeni öğrencilerin bu soruya verdikleri yanıtlar temel alınarak hesaplanır. Değişken evet/hayır şeklindeki bir nominal ölçekte ifade edilir. Örneğin; ATBS03A değişkeni değeri "Evet" olan bir öğrenci sınıfında tablet bilgisayarın bulunduğunu düşünmektedir. Bu değişken öğrencilerin bilgisayar teknolojileriyle olan tanışıklıklarını ve bu teknolojilerin eğitimde kullanımına yönelik politikaların etkilerini incelemek için kullanılabilir.

Genel olarak bu değişkenler öğrencilerin matematik ve fen bilimlerindeki akademik başarılarını etkileyebilecek eğitim kaynakları ve ortamına ilişkin içgörü sağlamaktadır.

6. ARAŞTIRMANIN AMPİRİK SONUÇLARI VE TARTIŞMA

Bu tez araştırmasında “Materyal ve Metot” bölümünde belirtildiği gibi bootstrap VZA metodolojisiyle radyal temelli girdi odaklı CCR modelinin uygulanması tercih edilmiştir. CCR tekniği; Charnes, Cooper ve Rhodes tarafından geliştirilen bir yöntemdir ve toplam etkinliği ölçerek bir ülkenin teknik ve ölçek açısından etkin olup olmadığını belirlemektedir. Girdi temelli yaklaşımın tercih edilmesinin nedeni veri zarflamanın doğasından kaynaklanmaktadır. Eğer bir karar birimi etkin değilse girdilerinde azaltma yapılması önerisi alır. Bu çalışmada değişkenlerle ilgili olarak öğretmen veya okul müdürlerinin algılarını azaltmaları önerilmektedir. Ancak dikkat edilmesi gereken nokta girdi azaltma önerisiyle puanın düşürülmesinin doğru bir şekilde yorumlanmasıdır. Örneğin, okul kaynaklarıyla ilgili algı yüksek olsa da azaltma önerisi alınmışsa bu durumda ya kaynaklar gerçekten yeterli değildir ancak yeterli gibi algılanmaktadır ya da kaynaklar yeterli olsa da verimli bir şekilde kullanılamamaktadır. Söz konusu problemin çözümünde MaxVZA 8 yazılımından faydalanılmıştır. Model ile ilgili bilgiler Tablo 6.1 ‘de etkinlik skorları Tablo 6.2’de gösterilmektedir.

Tablo 6.1. Model Özet Bilgileri

Özellik	Değerler
Model türü	Zarf Modeli
Karar verme birimi KVB sayısı	58
Girdi Sayısı	8
Çıktı Sayısı	2
Mesafe	Radyal
Oryantasyon	Girdi odaklı
Ölçeğe göre getiri	Sabit
Aylak Hesaplama	1 Aşama
Bootstrap Yöntemi	Bootstrap Skoru (Simar ve Wilson 1998'e göre)
Yüzdeliklerine Dayalı Güven Aralığı Yöntemi	Bootstrapped Puanları- 2* Bootstrapped Bias Ortalaması (Simar ve Wilson 1998)
Tekrar Sayısı	2000
Geçen zaman	26 Dakika 21,68 Saniye

Tablo 6.2. Etkinlik Skorlarının Duyarlılık Sonuçları

Ülkeler	VZA	Bootstrap VZA	Yanlılık	Alt sınır	Üst sınır
Almanya	1.000	0.984	0.016	0.964	0.999
Arnavutluk	0.924	0.908	0.016	0.869	0.923
Avustralya	0.947	0.932	0.015	0.907	0.946
Avusturya	0.958	0.942	0.016	0.912	0.957
Azerbaycan	0.970	0.955	0.015	0.918	0.969
Bahreyn	0.928	0.914	0.013	0.882	0.927
Belçika (Flaman Bölgesi)	1.000	0.976	0.024	0.946	0.999
Birleşik Arap Emirlikleri	0.908	0.895	0.013	0.864	0.908
Birleşik Devletler	1.000	0.980	0.020	0.957	0.999
Bosna Hersek	0.914	0.900	0.014	0.879	0.913
Bulgaristan	0.918	0.907	0.011	0.892	0.917
Çek Cumhuriyeti	1.000	0.983	0.017	0.962	0.999
Çin Taipei	1.000	0.971	0.029	0.922	0.999
Danimarka	0.998	0.985	0.014	0.968	0.997
Ermenistan	0.921	0.909	0.012	0.894	0.920
Fas	1.000	0.968	0.032	0.909	0.999
Filipinler	0.965	0.956	0.010	0.941	0.965
Finlandiya	1.000	0.979	0.021	0.952	0.999
Fransa	0.991	0.976	0.015	0.958	0.990
Güney Afrika	1.000	0.974	0.026	0.941	0.999
Gürcistan	0.904	0.894	0.011	0.881	0.904
Hırvatistan	0.934	0.920	0.015	0.899	0.933
Hollanda	1.000	0.979	0.021	0.953	0.999
Hong Kong	1.000	0.968	0.032	0.909	0.999
İngiltere	1.000	0.970	0.030	0.918	0.999
İran, İslam Cumhuriyeti	0.990	0.979	0.012	0.958	0.989
İrlanda	0.911	0.895	0.016	0.864	0.910
İspanya	0.884	0.872	0.012	0.850	0.883
İsveç	1.000	0.978	0.022	0.941	0.999
İtalya	0.992	0.979	0.013	0.967	0.991
Japonya	1.000	0.968	0.032	0.908	0.999
Kanada	0.968	0.957	0.010	0.941	0.967
Karadağ	0.899	0.886	0.013	0.861	0.899
Katar	0.943	0.929	0.014	0.895	0.942
Kazakistan	0.977	0.967	0.010	0.952	0.976
Kıbrıs	0.986	0.972	0.014	0.953	0.985
Kore Cumhuriyeti	1.000	0.967	0.033	0.907	0.999
Kosova	1.000	0.969	0.031	0.924	0.999
Kuveyt	0.963	0.948	0.015	0.917	0.962
Kuzey İrlanda	0.978	0.962	0.016	0.935	0.977
Kuzey Makedonya	0.956	0.945	0.011	0.927	0.955

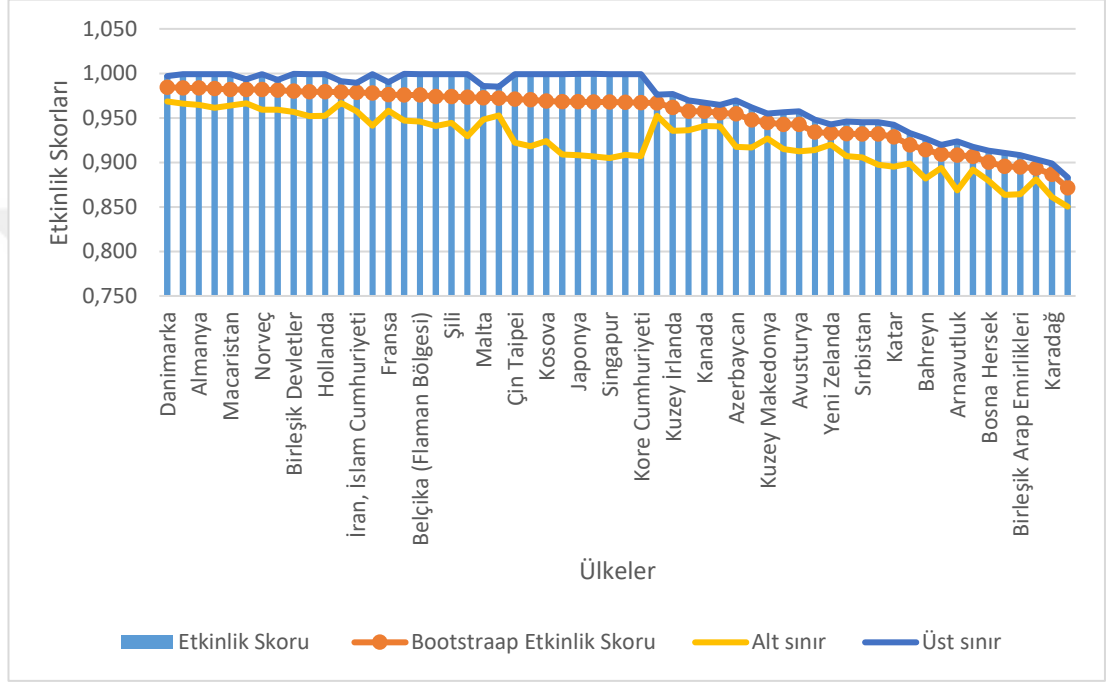
Tablo 6.3. (Devam)

Letonya	0.949	0.934	0.015	0.914	0.948
Litvanya	1.000	0.976	0.024	0.947	0.999
Macaristan	1.000	0.982	0.018	0.964	0.999
Malta	0.987	0.972	0.015	0.948	0.986
Norveç	1.000	0.982	0.018	0.959	0.999
Pakistan	0.946	0.932	0.014	0.897	0.945
Polonya	0.970	0.957	0.013	0.936	0.969
Portekiz	0.994	0.982	0.013	0.966	0.993
Rusya Federasyonu	1.000	0.973	0.027	0.930	0.999
Şili	1.000	0.974	0.026	0.944	0.999
Singapur	1.000	0.968	0.032	0.905	0.999
Sırbistan	0.946	0.932	0.014	0.906	0.945
Slovak Cumhuriyeti	1.000	0.984	0.016	0.966	0.999
Suudi Arabistan	0.957	0.943	0.014	0.915	0.956
Türkiye	1.000	0.968	0.032	0.907	0.999
Umman	0.993	0.981	0.012	0.959	0.993
Yeni Zelanda	0.943	0.933	0.011	0.920	0.943

Tablo 6.2. Bootstrap VZA sonuçlarını göstermektedir. Sütun 1 etkinliği incelenen karar birimleri olan ülkeleri göstermektedir. Sütun 2-4, sırasıyla ülkelerin girdi odaklı ölçeğe göre sabit getirili orijinal VZA etkinlik skorlarını sapma düzeltmeli ortalama tahmini (Sapma düzeltmesi, veri setindeki gürültüden ve sapmalardan kaynaklanan yanıltıcı etkileri ortadan kaldırmayı amaçlar. Bu nedenle sapma düzeltmeli etkinlik skorları daha güvenilirdir ve daha doğru bir gösterge olarak kabul edilir.) bootstrap sapması (Bu skorlar sapma düzeltmeli etkinlik skorlarına benzerdir. ancak bootstrap yöntemi kullanılarak elde edilirler. Bootstrap sapması, veri setinin tekrar örnekleme yapılarak orijinal veri setinin çeşitli alt kümelerinin oluşturulması ve her alt kümenin üzerinde ayrı ayrı hesaplama yapılmasıdır. Bu yöntem örnekleme hatası ve veri setindeki değişkenlik nedeniyle ortaya çıkabilecek sapmaları düzeltmeyi amaçlar.) vermektedir. Son iki sütun yanlılık-düzeltilmiş (bias-correction) %95 güven aralıklarını göstermektedir. (Yanlılık düzeltilmiş güven aralıkları, sapma düzeltmeli etkinlik skorlarından hesaplanır ve bir ülkenin gerçek etkinlik skorunun %95 güvenle hangi aralıkta olabileceğini gösterir.)

Tablo 6.2'deki sonuçlar, etkinlik ölçümlerinin örnekleme varyasyonuna karşı duyarlılığını ortaya koymaktadır. Sonuçlar, orijinal VZA etkinlik skorlarına dayalı ülkelerarası göreceli performans karşılaştırmasında dikkatli olunması gerektiğini

göstermektedir. Örneğin; Danimarka'nın orijinal VZA etkinlik skoru 0,998'dir. İngiltere ise 1,000 etkinlik skoru ile görünüşte etkin gözükmektedir. Ancak, 3. sütundaki yanlışlığı düzeltilmiş-bias corrected ölçüye göre Danimarka'nın az farkla da olsa kaynak kullanımını açısından İngiltere'den daha etkin olduğu gözlemlenmektedir. Bununla birlikte son iki sütun iki ülkenin etkinliği için güven aralıklarının büyük ölçüde örtüştüğünü göstermektedir. Bu nedenle iki ülkenin toplam etkinlikleri açısından önemli ölçüde fark olduğunu söyleyemeyiz.



Şekil 6.1. Ortalama Etkinlik Skorlarına göre sıralanmış ülkelerin VZA skor değerleri

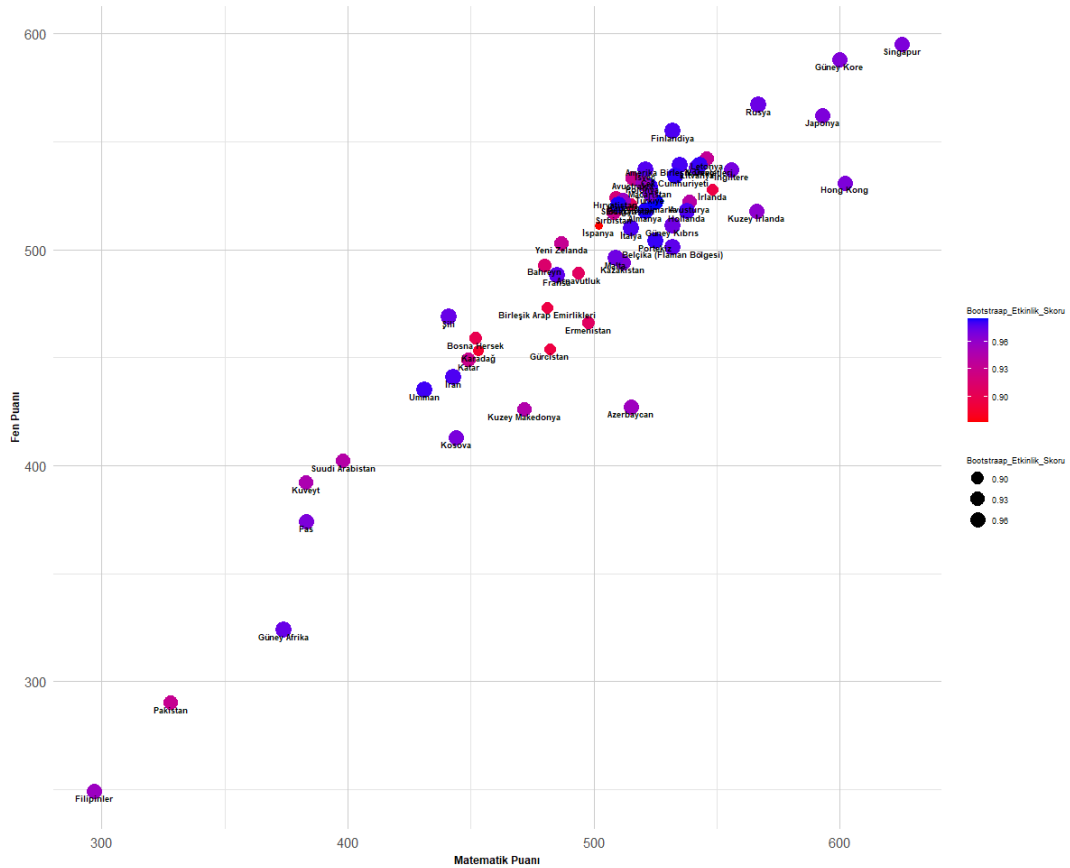
Tablo 6.1, Tablo 6.2 (Devam) ve Şekil 6.1'de gösterildiği gibi analiz edilen dönemde ortalama etkinlik skorları sıralamasına göre göreceli olarak en etkin ülke Danimarka (0,985), onu Slovak Cumhuriyeti (0,984) ve Almanya (0,984) takip etmekte olup bu bölgelerin neredeyse optimal düzeyde faaliyet gösterdiği ve istenen sabit çıktıyı üretmek için girdi tüketimleri de neredeyse optimal olduğu anlamına gelmektedir. Ayrıca bu sonuçlar bu üç bölgenin kaynak tasarrufu yaklaşımını izleyerek arzu edilen sabit çıktı düzeyine ulaşmak için mevcut kaynaklarını etkin bir şekilde yönettiklerinin varsayıldığını göstermektedir. Bununla birlikte geleneksel VZA skorlarına göre sadece Danimarka etkinsiz olup diğer ülkeler etkin çıkmıştır.

Benzer şekilde söz konusu geleneksel VZA skoruna göre, Türkiye Danimarka'ya göre etkin çıkmışken Bootstrapped VZA yöntemine göre

Danimarka'nın Türkiye'ye göre kaynaklarını daha etkin kullandığını göstermektedir.

Ortalama etkinlik skorunun (uluslararası dağılımda) yaklaşık olarak 0,95 olduğu hesaplanmaktadır. Buna göre ülkeler okul girdilerini ortalama olarak %5 civarında azaltsaydı bile aynı TIMSS puanlarını yakalayabileceği sonucu çıkmaktadır. Bu sonuç Danimarka okulları için (ortalama) %1,5'ten, İspanya okulları için %12,8'e kadar değişen oranlarda ülkeler arasında büyük farklılıklar göstermektedir. Başka bir perspektifle değerlendirildiğinde bootstrap etkinlik skorlarına göre en yakın etkin ülke Danimarka'dır ve 0,985 puan almaktadır. En uzak ülke ise İspanya'dır ve 0,872 puan almaktadır. Ülkelerin teknik etkinlik puanlarının ortalaması 0,953'tür. Bu çalışmada incelenen ülkeler matematik ve fen başarısına yönelik kaynaklarını veya çabalarını yaklaşık olarak %5 ($100-95,03=4,97$) oranında hatalı bir şekilde kullanmaktadırlar.

Araştırma kapsamında etkinlik skorlarının matematik ve fen puanları ile arasındaki ilişkisi de incelenmiştir. İlgili sonuçlar Şekil 6.2'de gösterilmektedir.



Şekil 6.2’den gözlemlendiği üzere ülkelerin matematik ve fen puanları arasında doğrusal bir artış bulunmaktadır. Yani aralarında pozitif yönlü bir ilişki mevcuttur. Genel olarak kaynaklarını etkin kullanan ülkelerin matematik ve fen skorlarının da yüksek olduğunu söyleyebiliriz. Ancak bu genel yargının dışında kalan ülkeler de bulunmaktadır. Örneğin Avustralya, Avusturya, Bulgaristan, Hırvatistan, İrlanda, İspanya, Letonya ve Sırbistan ortalamasının altında bir etkinlik skoru göstermelerine karşın matematik ve fen puanlarında ülkelerin ortalamasının üzerindedirler. Tersine Fas, Filipinler, Fransa, Güney Amerika, İran, Kosova ve Şili ülkelerin ortalamasının üstünde bir etkinlik skoru elde etmelerine rağmen matematik ve fen puanlarında ortalamasının altındadırlar. Bu tezatlığı sergileyen ülkelerin durumlarının ayrıca incelenmesi ve nedenlerinin araştırılması bu tezin kapsamı dışında ele alınması gereken başka bir konudur.

Bu tez araştırmasının ikinci aşamasında okul yönetiminin okul bağlamında ülke düzeyindeki etkinlik skorları üzerinde etkisi var mıdır, varsa ne şekilde etkilidir? Sorusunu yanıtlamak için Tobit modelini kullanarak okul bağlamında ülke düzeyindeki yönetim değişkenlerini diğer açıklayıcı değişkenlerle birlikte birinci aşamada elde edilen teknik etkinlik skorları üzerinde regresyona tabi tutulmuştur. Tablo 6.3 bu ikinci aşama Tobit model regresyonundan elde edilen sonuçları göstermektedir.

Tablo 6.4. Tobit Model Sonuçları

		Gözlem Sayısı:	Toplam	Soldan sansürlü	Sansürlü	Sağdan sansürlü
			58	0	28	0
		Bağımlı Değişken:	Bootstrapped Etkinlik Skorları			
Bağımsız Değişken Kodları	Bağımsız Değişken Açıklamaları	Tahmin	Standart hata	Z değeri	P değeri	
Sabit		1.700	0.171	9.924	0.000	***
ACBG04	Öğrencilerin testi hangi dillerde yaptıklarının yüzdesi	-0.006	0.004	-1.463	0.143	
ACBG07	Okuldaki toplam bilgisayar sayısı	0.000	0.000	2.582	0.010	**
ACBG14E	Okulun velilerin öğrencilerin eğitime katılım düzeyine verdiği önemi	-0.049	0.020	-2.420	0.016	*
ACBG18	Okul müdürünün toplam yöneticilik deneyimi	0.005	0.002	2.900	0.004	**
ACBGMRS	Matematik Kaynaklarının Yetersizliğinden Etkilenen Öğretim	-0.020	0.026	-0.755	0.450	
ACBGSRS	Fen Kaynaklarının Yetersizliğinden Etkilenen Öğretim	0.012	0.025	0.463	0.643	
ACBGEAS	Okul müdürlerinin akademik başarıya verilen önem	-0.026	0.007	-3.694	0.000	***
ACBGDAS	Okul müdürlerinin disiplin konusuna verdiği önem	-0.027	0.007	-3.689	0.000	***
ACDGTIHY	Öğrencilerin yıllık ders saati sayısı	0.000	0.000	1.021	0.307	
ATBS03A	Öğrencilerin sınıflarında tablet bilgisayarların bulunup bulunmaması durumu	-0.039	0.019	-2.054	0.040	*
Log(ölçek-scale)		-3.684	0.093	-39.676	0.000	***
Anlamlılık kodları:		0 "****"	0.001 "****"	0.01 "****"	0.05 "****"	0.1 "****"
Ölçek-scale: 0.02513						
Gaussian Dağılım						
Newton-Raphson						
İterasyon Sayısı: 5						
Log-Olabilirlik: 12 Serbestlik Derecesinde 131.4						
Wald-İstatistik: 10 Serbestlik Derecesinde 46.83						
p değeri: 0						

Model, bağımlı değişken (öğretim sürecinde ülkelerin eğitim girdilerini ne kadar etkin kullandıklarının ölçüsü olan) Bootstrapped Etkinlik Skorlarındaki değişimi çeşitli bağımsız değişkenler ile açıklamak için kullanılmıştır. Modelin açıklama gücünü ölçmek için, modelin sabiti 1.700 ve diğer tüm bağımsız değişkenlerin katsayıları kullanılarak "Bootstrapped Etkinlik Skorları"nın tahmin edilmesi sağlanmıştır. Bu tahmin edilen değerler, gerçek "Bootstrapped Etkinlik Skorları" ile karşılaştırılmış ve modelin doğruluğunu ölçmek için "Standart hata", "Z değeri" ve "P değeri" hesaplanmıştır.

Modelin bağımsız değişkenleri arasında okuldaki toplam bilgisayar sayısı (ACBG07), okulun velilerin öğrencilerin eğitime katılım düzeyine verdiği önem (ACBG14E), okul müdürünün toplam yöneticilik deneyimi (ACBG18), öğrencilerin yıllık ders saati sayısı (ACDGTIHY), öğrencilerin sınıflarında tablet bilgisayarların bulunup bulunmaması durumu (ATBS03A), okul müdürlerinin akademik başarıya ve disiplin konusuna verilen önem (ACBGEAS ve ACBGDAS) yer almaktadır. Bu bağımsız değişkenlerin tümü "Bootstrapped Etkinlik Skorları" üzerinde istatistiksel olarak anlamlı etkilere sahip olduğu görülmektedir. Ancak öğrencilerin testi hangi dillerde yaptıklarının yüzdesi (ACBG04), matematik kaynaklarının yetersizliğinden etkilenen öğretim (ACBGMRS) ve fen kaynaklarının yetersizliğinden etkilenen öğretim (ACBGSRS) bağımsız değişkenleri, "Bootstrapped Etkinlik Skorları" üzerinde anlamlı bir etkiye sahip değildir.

Modeldeki diğer bağımsız değişkenlerin, "Bootstrapped Etkinlik Skorları" üzerinde anlamlı etkileri olduğu görülmektedir. Özellikle okul müdürlerinin akademik başarıya ve disiplin konusuna verilen önem (ACBGEAS ve ACBGDAS) ve okul müdürünün toplam yöneticilik deneyimi (ACBG18) gibi değişkenler "Bootstrapped Etkinlik Skorları" üzerinde az da olsa anlamlı etkilere sahip olduğu görülmektedir.

Model sonuçlarına göre, okuldaki toplam bilgisayar sayısı ve okul müdürünün toplam yöneticilik deneyimi eğitim etkinliğinin artmasında pozitif bir etkiye sahiptir. Ayrıca öğretmenlerin akademik başarıya ve disipline verdiği önem de pozitif etki göstermektedir.

Öğrencilerin testleri hangi dillerde yaptıklarının yüzdesi, matematik ve fen kaynaklarının yetersizliğinden etkilenen öğretim, öğrencilerin yıllık ders saati sayısı ve öğrencilerin sınıflarında tablet bilgisayarların bulunup bulunmaması ise eğitim etkinliğine anlamlı bir etki yapmamaktadır.

Ayrıca ölçek-scale olarak verilen 0.02513 değeri, modeldeki tahminlerin yaklaşık olarak 2.5% oranında hata ile tahmin edildiğini göstermektedir.

Sonu olarak bulgular; okulun velilerin ğrencilerin eēitimine katılım düzeyine verdiēi nem, okul mdrlерinin akademik bařarıya ve disiplin konusuna verdiēi nem, ğrencilerin sınıflarında tablet bilgisayarların bulunup bulunmaması durumu, katsayılarının tm rneklem iin hem negatif hem de anlamlı olduēunu gstermektedir. Bu da sz konusu kaynakların tm lkelerde yeterince kullanılmadıēı anlamına gelmektedir. Bařka bir deyiřle girdi odaklı etkinlik, belirli bir ıktı düzeyi gz nne alındıēında girdilerin kullanımındaki etkinliēi deēerlendirdiēinden sonular genel olarak gzlemlenen ıktıların daha az kaynak kullanarak elde edilebileceēini gstermektedir. Yetersiz kullanım ve optimal altı kaynakların ynetimi lke dzeyinde genel etkisizliēe katkıda bulunmaktadır. Ayrıca okul mdrnn toplam yneticilik deneyimi ve (az da olsa) okuldaki toplam bilgisayar sayısındaki artışı tm rneklemde girdi odaklı etkinliēi artırdıēını gstermektedir.

7. SONUÇ

Hem kaynakların hem de insanların yönetimi mevcut kaynakların ve girdilerin daha iyi öğrenci öğrenimine dönüştürülmesinin ayrılmaz bir bileşenini oluşturmaktadır. Okullarda ve eğitim kurumlarında kaynakların kötü yönetimi eğitimin kötü bir şekilde sunulmasına ve düşük kaliteli sonuçlara olumsuz katkıda bulunabilir. Dolayısıyla yönetsel etkinlik, mümkün olan en yüksek çıktıları elde etmek için hem fiziksel hem de insan kaynaklarının en iyi şekilde kullanılması anlamına gelir. Bu çalışma, TIMSS 2019'a katılan 58 ülkedeki 5164 okuldan alınan okul düzeyindeki verilerin ortalaması kullanmış ve iki aşamalı bir metodoloji kullanarak okulların teknik etkinliği ve bunların yönetsel uygulamalarıyla ilişkisini incelemiştir.

Bu araştırma kapsamında parametrik olmayan Bootstrap VZA yöntemini kullanarak sınav başarısına yönelik kaynak verimliliği yaklaşımını değerlendirmiştir. Bu durum değerlendirmesi politika yapıcılara ve planlayıcılara bu politika yaklaşımının güçlü ve zayıf yönlerini ve bu maliyetli politikayı sürdürmenin önemini belirlemede yardımcı olabilmelerini sağlayacak bir tavsiye niteliğindedir. Üstelik Bootstrap VZA yaklaşımı geleneksel VZA prosedürü kullanılarak elde edilenlerden istatistiksel olarak daha doğru verimlilik puanları ile sonuçlanan geleneksel etkinlik puanlarını düzeltmemize ve güven aralıkları elde etmemize izin vermiştir.

Türkiye dördüncü sınıflarda 58 ülke arasından fen sınavı bakımından 36'ncı ve matematik sınavı bakımından 22'inci her iki sınavın ortalamaları bakımından 26.'dır. Türkiye, toplam etkinlik açısından diğer ülkelerle benzer düzeyde olmasına rağmen sınav başarısı açısından alt sıralarda yer almaktadır.

Değişkenlerden elde edilen bulguların gerçekliğini yansıttığı varsayımı doğrultusunda özellikle tam etkin olmayan ülkelerin çoğunun ya eğitim için gereksiz kaynaklara aşırı yatırım yaptığı ya da girdilerin sonuçlara dönüşümünde etkin olmadığı söylenebilir. Az gelişmiş veya gelişmekte olan ülkeler, harcamalarını sonuçları iyileştirmeye yönelik çabalarla dengelemeli veya eğitime harcadıkları fazla girdileri azaltma yoluna gitmelidir. Bu durum özellikle Ermenistan, Karadağ, Birleşik Arap Emirlikleri (Dubai), Kuzey Makedonya, Kuveyt, Gürcistan, Filipinler gibi ülkeler için geçerlidir.

Genel anlamda, ülkelerin matematik ve fen puanları arasında genel olarak

pozitif bir ilişki bulunmaktadır. Fakat bazı ülkeler bu genellemeye uymamaktadır. Özellikle Avustralya, Avusturya, Bulgaristan ve birkaç diğer ülke ortalamasının altında bir etkinlik skoruyla matematik ve fen puanlarında yüksek skorlar elde ederken; Fas, Filipinler, Fransa ve bazıları ortalamasının üstünde etkinlik skorlarına sahip olmalarına rağmen bu puanlarda düşük kalmışlardır. Bu tezatlık daha derinlemesine incelenmeye değerdir.

İkinci aşamada Teknik etkinliği regresyona tabi tutularak okulların okul yönetimi değişkenleri üzerindeki etkilerini Tobit modeli kullanarak incelemiştir. Tobit regresyon analizi, kaynak kullanımı ve yönetiminin özellikle ülkelerdeki okulların girdi odaklı teknik etkinliklerini artırabileceğini ortaya koymuştur. Bu sonuçlar göstermektedir ki araştırmaya konu olan ülkelerdeki okulların etkisizliğinin hem matematik ve fen öğretimine özgü kaynakların yetersizliği hem de aynı kaynakların en iyi şekilde kullanılmaması ve yönetilmemesi olduğudur. Küresel okul yönetimine ilişkin bulgular bazı yeni perspektifler sunmaktadır.

Özetle bu çalışma, okul girdilerinin daha yüksek öğrenme çıktılarına dönüştürülmesinde daha fazla verim elde etmek için bunların insan kaynaklarının yanı sıra fiziksel kaynakların da etkin yönetimiyle tamamlanması gerektiğini ortaya koymaktadır. Herhangi bir eğitim sisteminin karmaşıklığı göz önüne alındığında bu tez çalışmasında yapılan değişken seçimi sınırlıdır ve eğitim operasyonunun altında yatan girdi-çıkıtı çerçevesinin tamamını kapsamamaktadır. Bu nedenle bir genelleme tartışmalı ve daha ileri bir araştırma konusu olarak kabul edilmelidir. Buna ek olarak VZA'nın girdi ve çıktı seçimine duyarlı olmasının getirdiği bariz sınırlama da burada belirtilmelidir. Ayrıca gözlemlenemeyen girdi ve çıktılarından kaynaklanan ölçüm hatası olasılığı da tamamen göz ardı edilmemelidir. Bununla birlikte VZA parametrik olmayan bir teknik olduğundan üretim sınırının "fonksiyonel formundaki" değişikliklere duyarlılık ekonometrik tahminlerde olduğundan daha az önemli bir sorundur (Stolp, 1990). Ayrıca bu çalışmanın araştırma odağının yönetsel uygulamalar ile ülkelerdeki okulların verimli işleyişi arasındaki ilişkiyle ilgili olduğu ve dolayısıyla VZA tekniğinin Karar Verme Birimi (DMU) olarak okul bağlamındaki değerlerin ülke ortalaması kullanılmaktadır. Bu durum öğrenci düzeyinde heterojen bir analiz yapma veya teknik verimlilikte buna bağlı olarak ortaya çıkan değişim olasılıklarını ele alma kapsamını kısıtlamaktadır. Çalışmanın bir diğer kısıtlılığı da karar verme birimlerinin çıktılarını ifade etmek için sadece iki

dersin, yani matematik ve fen derslerinin öğrenci başarı puanlarını dahil etmiş olmamızdır. Oysa okul bağlamında ülke düzeyindeki performans ancak okulda öğretilen tüm kritik bilişsel ve bilişsel olmayan yönler ölçülüp değerlendirildiğinde çıktı üretme açısından bütünsel olarak değerlendirilebilir. Literatür, okulların çıktılarının sadece test sonuçları veya öğrenci terfisi ile sınırlı olmadığını çok boyutlu bir yapı olabileceğini göstermiştir. Bazı akademik olmayan çıktılar da okulların önemli çıktıları olarak kabul edilebilir (örneğin; öğrenci refahı, sosyo-duygusal beceriler, yaşam becerileri, sivil katılım vb.) ve akademik başarı ile ilgisiz olabilir. Veri kısıtlamaları nedeniyle bu çalışma sadece bilişsel başarıya odaklanmaktadır. Bu tür sınırlamalara rağmen bu makaledeki bulgular, özellikle politika amaçları açısından önemli gözlemler ortaya koymaktadır. Bu gözlemler daha yüksek girdi düzeyinin okullarda daha yüksek etkinlik sağlamak için yeterli olmadığını göstermektedir. Dünya genelinde hükümetler okullar için girdi ve kaynakları artırmaya odaklanırken bu kaynakların etkin yönetimi daha iyi öğrenme çıktıları elde etmek için kritik hale gelmektedir. Ayrıca öğrenci ve öğretmenlerin hedef belirlemesi ve motivasyonu, öğrenci disiplini ve velilerin okul faaliyetlerine katılımı gibi insan kaynakları yönetiminin farklı yönleri, fiziksel kaynakların mevcudiyeti ve kullanımına kıyasla okulların daha verimli olmasına katkıda bulunacağı bu araştırmanın sonucunda da aşıkardır. Okul yönetimini iyileştirmeye yönelik dengeli bir yaklaşım okulun eğitimindeki tüm paydaşları, yani öğrencileri, velileri ve öğretmenleri güven altına almayı içerir. Fiziksel kaynakların ve daha da önemlisi insan kaynaklarının en iyi şekilde kullanılmasına odaklanan okul yönetimi uygulamaları potansiyel olarak öğrenmeye elverişli ve bütüncül bir ortam yaratabilir. Böylece öğrencilerin öğrenme çıktılarını iyileştirilebilir.

KAYNAKLAR

- Afonso, A, Aubyn, M S. (2005). "Non-parametric approaches to education and health efficiency in OECD countries". *Journal of applied economics*, 8(2), 227-246.
- Agasisti, T. (2014). "The efficiency of public spending on education: An empirical comparison of EU countries". *European Journal of Education*, 49(4), 543-557.
- Akyüz, G. (2013). "Öğrencilerin okul dışı etkinliklere ayırdıkları süreler ve matematik başarıları arasındaki ilişkinin incelenmesi". *Elektronik sosyal bilimler dergisi*, 12(46), 112-130.
- Amemiya, T. (1984). "Tobit models: A survey". *Journal of econometrics*, 24(1-2), 3-61.
- Arshad, M N M. (2014). "Efficiency of Secondary Education of a Selected OIC Countries". *Global Education Review*, 1(4).
- Banker, R D, Charnes, A, Cooper, W W. (1984). "Some models for estimating technical and scale inefficiencies in data envelopment analysis". *Management science*, 30(9), 1078-1092.
- Bhutoria, A, Aljabri, N. (2022a). "Managerial practices and school efficiency: a data envelopment analysis across OECD and MENA countries using TIMSS 2019 data". *Large-scale Assessments in Education*, 10(1). doi: 10.1186/s40536-022-00147-3
- Bhutoria, A, Aljabri, N. (2022b). "Managerial practices and school efficiency: a data envelopment analysis across OECD and MENA countries using TIMSS 2019 data". *Large-scale Assessments in Education*, 10(1), 24.
- Bogetoft, P, Otto, L. (2010). *Benchmarking with dea, sfa, and r* (157. bs): Springer Science & Business Media.
- Cameron, A C, Trivedi, P K. (2010). *Microeconometrics using stata* (2. bs): Stata press College Station, TX.
- Caputo, A, Rastelli, V. (2014). "School improvement plans and student achievement: Preliminary evidence from the Quality and Merit Project in Italy". *Improving Schools*, 17(1), 72-98.
- Carlson, D, Lavertu, S. (2018). "School improvement grants in Ohio: Effects on student achievement and school administration". *Educational Evaluation and Policy Analysis*, 40(3), 287-315.
- Cavlak, H. (2021). "Etkinlik, Etkililik, Verimlilik, Kârlılık, Performans: Kavramsal Bir Çerçeve Ve Karşılaştırma". *Journal of Research in Business*, 6(1), 99-126.
- Charnes, A, Cooper, W, Lewin, A Y, Seiford, L M. (1997). "Data envelopment analysis theory, methodology and applications". *Journal of the Operational Research society*, 48(3), 332-333.
- Charnes, A, Cooper, W W, Rhodes, E. (1978). "Measuring the efficiency of decision making units". *European journal of operational research*, 2(6), 429-444.
- Coelli, T, Perelman, S. (2000). "Technical efficiency of European railways: a distance function approach". *Applied economics*, 32(15), 1967-1976.
- Coelli, T J, Rao, D S P, O'Donnell, C J, Battese, G E. (2005). *An introduction to efficiency and productivity analysis*: springer science & business media.
- Cook, W D, Tone, K, Zhu, J. (2014). "Data envelopment analysis: Prior to choosing a model". *Omega*, 44, 1-4.
- Cooper, W W, Seiford, L M, Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software* (2. bs): Springer.

- Çobanoğlu, F, Badavan, Y. (2017). "Başarılı okulların anahtarı: etkili okul değişkenleri". *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*(26), 114-134.
- Davidson, R, MacKinnon, J G. (2004). *Econometric theory and methods* (5. bs): Oxford University Press New York.
- Deville, A. (2009). "Branch banking network assessment using DEA: A benchmarking analysis—A note". *Management Accounting Research*, 20(4), 252-261.
- Doyle, J, Green, R. (1994). "Efficiency and cross-efficiency in DEA: Derivations, meanings and uses". *Journal of the operational research society*, 45, 567-578.
- Efron, B. (1979). "Computers and the theory of statistics: thinking the unthinkable". *SIAM review*, 21(4), 460-480.
- Efron, B. (1992). Bootstrap methods: another look at the jackknife *Breakthroughs in statistics* (pp. 569-593): Springer.
- Emrouznejad, A, De Witte, K. (2010). "COOPER-framework: A unified process for non-parametric projects". *European Journal of Operational Research*, 207(3), 1573-1586.
- Emrouznejad, A, Parker, B R, Tavares, G. (2008). "Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA". *Socio-economic planning sciences*, 42(3), 151-157.
- Emrouznejad, A, Yang, G-l. (2018). "A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016". *Socio-economic planning sciences*, 61, 4-8.
- Erdem-Kara, B, Osman, T. (2019). "Eğitim kaynaklarının kullanım etkinliği üzerine bir uluslararası karşılaştırma". *Eğitimde Kuram ve Uygulama*, 15(2), 153-170.
- Frey, B B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*: Sage Publications.
- Fuller, E, Young, M, Baker, B D. (2011). "Do principal preparation programs influence student achievement through the building of teacher-team qualifications by the principal? An exploratory analysis". *Educational Administration Quarterly*, 47(1), 173-216.
- Gagnon, J C, Barber, B R. (2018). The Sage encyclopedia of educational research, measurement, and evaluation *The SAGE encyclopedia of educational research, measurement and evaluation* (pp. 668): Sage.
- Gavurova, B, Kocisova, K, Belas, L, Krajcik, V. (2017). "Relative efficiency of government expenditure on secondary education". *Journal of International Studies*, 10(2).
- Giambona, F, Vassallo, E, Vassiliadis, E. (2011). "Educational systems efficiency in European Union countries". *Studies in Educational Evaluation*, 37(2-3), 108-122.
- Giménez, V, Prior, D, Thieme, C. (2007). "Technical efficiency, managerial efficiency and objective-setting in the educational system: an international comparison". *Journal of the Operational Research Society*, 58(8), 996-1007.
- Givens Rolland, R. (2012). "Synthesizing the evidence on classroom goal structures in middle and secondary schools: A meta-analysis and narrative review". *Review of Educational Research*, 82(4), 396-435.
- Greene, W. (2007). *Econometric Analysis*, : USA: Pearson Education, Inc., Upper Saddle River.
- Greene, W. (2011). *Econometric Analysis*: Pearson Education.
- Greene, W H. (2003). *Econometric analysis*: Pearson Education India.

- Gujarati, D, Porter, D, Gunasekar, S. (2009). Basic econometrics Tata McGraw-Hill Education.
- Hwang, S-N, Lee, H-S, Zhu, J. (2016). *Handbook of operations analytics using data envelopment analysis*: Springer.
- IEA.). IEA- Main Page. from <https://www.iea.nl/>
- IEA. (2021). International Association for the Evaluation of Educational Achievement. from <https://timss2019.org/reports/about/>
- Ilgaz, G, Eskici, M, Vural, L. (2019). "Which Country Is More Effective in Science Teaching? Evidence from PISA 2015 as a Secondary School Assessment Tool". *Asian Journal of Education and Training*, 5(2), 349-361.
- International Study Center, T P.
- Jay, T, Rose, J, Simmons, B. (2018). "Why is parental involvement in children's mathematics learning hard? Parental perspectives on their role supporting children's learning". *Sage Open*, 8(2), 2158244018775466.
- Kennedy, P. (2008). *A guide to econometrics*: John Wiley & Sons.
- Kılıkış, Ş, Kılıkış, Ş. (2016). "Benchmarking airports based on a sustainability ranking index". *Journal of Cleaner Production*, 130, 248-259.
- Koyuncu, B, Ilgaz, G. (2019). "Matematik Öğretimi Sürecinde Ülkelerin Eğitim Girdilerini Ne Kadar Etkin Kullandıklarının TIMSS 2015 Verilerine Göre İncelenmesi". *İlköğretim Online*, 18(4).
- Kutlar, A, Babacan, A. (2008). "Türkiye'deki kamu üniversitelerinde CCR etkinliği-ölçek etkinliği analizi: DEA tekniği uygulaması". *Kocaeli Üniversitesi Sosyal Bilimler Dergisi*(15), 148-172.
- López-Penabad, M-C, Maside-Sanfiz, J M, Torrelles Manent, J, Iglesias-Casal, A. (2020). "Application of the DEA Double Bootstrap to Analyze Efficiency in Galician Sheltered Workshops". *Sustainability*, 12(16), 6625.
- Lorcu, F, Bolat, B A. (2015). "Comparison of secondary education PISA results in European member states and Turkey via DEA and SEM". *Journal of WEI Business and Economics*, 4(3), 7.
- Maddala, G S. (1987). "Limited dependent variable models using panel data". *Journal of Human resources*, 307-338.
- Mazurek, J, Mielcová, E. (2019). "On the relationship between selected-socio-economic indicators and student performances in the PISA 2015 study".
- Mullis, I, Martin, M, Foy, P, Kelly, D, Fishbein, B. (2020). "TIMSS 2019 international results in mathematics and science". *Boston College, TIMSS & PIRLS International Study Center*.
- Mullis, I V, Martin, M O, Foy, P, Kelly, D L, Fishbein, B. (2020). "TIMSS 2019 international results in mathematics and science". Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-results>.
- Olesen, O B, Petersen, N C. (2016). "Stochastic data envelopment analysis—A review". *European journal of operational research*, 251(1), 2-21.
- PIRLS, T. (2021). Trends in International Mathematics and Science Study. from <https://timssandpirls.bc.edu/>

- Rouse, P, Putterill, M, Ryan, D. (2002). "Integrated performance measurement design: insights from an application in aircraft maintenance". *Management Accounting Research*, 13(2), 229-248.
- Seiford, L M, Thrall, R M. (1990). "Recent developments in DEA: the mathematical programming approach to frontier analysis". *Journal of econometrics*, 46(1-2), 7-38.
- Simar, L. (1992). Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping *International Applications of Productivity and Efficiency Analysis* (pp. 167-199): Springer.
- Simar, L, Wilson, P W. (1998). "Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models". *Management science*, 44(1), 49-61.
- Simar, L, Wilson, P W. (2000). "A general methodology for bootstrapping in non-parametric frontier models". *Journal of applied statistics*, 27(6), 779-802.
- Simar, L, Wilson, P W. (2000). "Statistical inference in nonparametric frontier models: The state of the art". *Journal of productivity analysis*, 13(1), 49-78.
- Simar, L, Wilson, P W. (2002). "Non-parametric tests of returns to scale". *European Journal of Operational Research*, 139(1), 115-132. doi: [https://doi.org/10.1016/S0377-2217\(01\)00167-9](https://doi.org/10.1016/S0377-2217(01)00167-9)
- Simar, L, Wilson, P W. (2007). "Estimation and inference in two-stage, semi-parametric models of production processes". *Journal of econometrics*, 136(1), 31-64.
- Stock, J H, Watson, M W. (2003). *Introduction to econometrics* (104. bs): Addison Wesley Boston.
- Stolp, C. (1990). "Strengths and weaknesses of data envelopment analysis: An urban and regional perspective". *Computers, Environment and Urban Systems*, 14(2), 103-116.
- Stumbrienė, D, Želvys, R, Žilinskas, J, Jakaitienė, A. (2022). "Using PIRLS and TIMSS data for the assessment of efficiency and effectiveness of primary education in post-socialist and old democracy EU member states". *Lietuvos statistikos darbai*, 61, 1-15.
- Suna, D H E, Şensoy, D S, Parlak, D B, Özdemir, E. (2020). TIMSS 2019 Türkiye Ön Raporu. Atatürk Bulvarı No: 98 Bakanlıklar ANKARA: T.C. Millî Eğitim Bakanlığı
- Toma, P, Miglietta, P P, Zurlini, G, Valente, D, Petrosillo, I. (2017). "A non-parametric bootstrap-data envelopment analysis approach for environmental policy planning and management of agricultural efficiency in EU countries". *Ecological indicators*, 83, 132-143.
- Tsui, M-s. (2005). "Functions of social work supervision in Hong Kong". *International social work*, 48(4), 485-493.
- Wooldridge, J M. (2010). *Econometric analysis of cross section and panel data*: MIT press.
- Zhu, J. (2009). *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets* (2. bs): Springer.

ÖZ GEÇMİŞ

Firdevs Cevheroğlu Eren, Erzurum Mehmet Akif Ersoy Lisesi'ni bitirdikten sonra Atatürk Üniversitesi Eğitim Fakültesi'nden 2011 yılında mezun oldu. 2020 yılında OMÜ LEE İstatistik Yüksek Lisans programını girdi. Mezuniyetinden bu yana öğretmen olarak görev yapan Firdevs Cevheroğlu Eren, orta derecede İngilizce bilmektedir. Temel ilgi alanları, Sınıf Eğitimi, Çocuk ve Ergen Psikolojisi (Mayıs 2023).

İletişim Bilgileri

ORCID ID : 0000-0003-1808-0355

Yayınlar:

1. AYDIN, S. G., & EREN, F. C. (2023). Bootstrap Veri Zarflama Analizi ile TIMSS Verileri Kullanılarak Eğitim Sisteminde Teknik ve Yönetsel Etkinliklerin Karşılaştırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 27(2), 197-206.

