



DENGESİZ PANEL VERİ MODELİ İLE EM ALGORİTMASI SONUÇLARININ KARŞILAŞTIRILMASI

Selahattin GÜRİŞ*
Şaban KIZILARSLAN**

Öz

Eksik veri problemi, veri setinden elde edilecek bilginin ve yapılacak analizlerin kısıtlanmasına yol açmaktadır. Bu çalışmada eksik veri ile panel veri modellerinin tahmini ele alınmıştır. Panel verinin eksik gözlemler içermesi durumunda model tahmini için alternatif iki yöntem uygulanmış ve sonuçlar karşılaştırılmıştır. İlk olarak dengesiz panel veri modeli tahmin edilmiştir. İkinci olarak eksik veriler EM Algoritması kullanılarak tamamlanmış ve dengeli panel veri modeli tahmin edilmiştir. Eksik verilere atama yapmanın parametre tahminlerinde sapmalara neden olabileceği bilinmektedir. Çalışmanın sonucunda iki tahmin yöntemine ait model sonuçları karşılaştırılarak parametre tahminlerinin nasıl etkilendiği analiz edilmiştir. Çalışmada uygulama konusu olarak eğitim ile ekonomik büyüme arasındaki ilişki ele alınmıştır. Böylece ekonomik büyüme ile eğitim arasındaki ilişki de ortaya konmuştur.

Anahtar Kelimeler: Eksik Veri, Dengesiz Panel, EM Algoritması, Eğitim, Ekonomik Büyüme

COMPARISON OF UNBALANCED PANEL DATA MODEL AND EM ALGORITHM RESULTS

Abstract

The problem of missing data leads to be restricted information and analyses derived from the data set. This study focused on estimating panel data models with missing data. Two alternative estimation methods in case of missing values in panel data were applied and the results were compared. Firstly, it was estimated unbalanced panel data model. Secondly, missing values were completed using EM algorithm and then balanced panel data model was estimated. It is known that imputing the missing data

* Marmara Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü, Öğretim Üyesi, Prof. Dr.

** Yüzüncü Yıl Üniversitesi, İİBF, Ekonometri Bölümü, Araştırma Görevlisi.

may cause deviations in the parameter estimations. As a result of study, it has been analyzed how the parameter estimations were affected by comparing two estimation results. The relationship between education and economic growth were examined as implementation of the present study. Thereby, it was also demonstrated the relation of education and economic growth.

Keywords: Missing Data, Unbalanced Panel, EM Algorithm, Education, Economic Growth

I. GİRİŞ

İktisadî ve ekonometrik araştırmaların temelinde araştırılacak değişkenlere ve birimlere ait veriler yer almaktadır. Belirli bir veri seti ile çalışmak istatistiksel ve ekonometrik analizlerin yapılmasına imkân tanımaktadır. Ancak çalışmada seçilen örnekleme ait verilere ulaşmak her zaman mümkün olmamaktadır. Bazı durumlarda değişkenlere veya birimlere ait hiç veri bulunamazken, bazen de ilgili verilerin belirli dönemlerindeki gözlemlerine ulaşmak mümkün olmamaktadır. Bu problem literatürde eksik veri olarak tanımlanmaktadır. Eksik veri problemi her türde veri setinde sıklıkla karşılaşılan bir durumdur. Bu durumun ortaya çıkması çeşitli nedenlerden kaynaklanabilir. Anket gibi veri toplama araçları yardımıyla elde edilen verilerde katılımcıların cevaplamakta isteksiz oldukları soruların bulunması, ölçeğin uzunluğu, uygulanan örneklemin seçimi vb. nedenlerden dolayı toplanan anketlerde eksik gözlemler ortaya çıkabilmektedir [18], [2]. Zaman serisi verileri ile çalışıldığında ise, çalışmada yer alan değişkenin istenen frekansta gözlemlenmemiş olması veya belirli dönemlerde değişkene ait verilerin bulunmaması gibi nedenlerle eksik veri problemi ortaya çıkabilmektedir [13].

Ekonometrik analizlerde ön plana çıkan bir diğer veri türü ise panel veridir. Yatay – kesit verileri ile zaman serisi verilerinin bir arada kullanıldığı panel verilerde de çeşitli nedenlerden dolayı eksik gözlemler ortaya çıkmaktadır. Düzenli aralıklarla yapılan anketler yoluyla toplanan panel verilerde, panel birimleri bir süre sonra ankete katılmayı bırakabilmekte veya aynı yerde ikamet etmeme gibi nedenlerden ötürü bu birimler için ankete devam edilememektedir. Bu problem “aşınma” olarak adlandırılmaktadır ve anketle elde edilen panel veri setindeki eksikliklerin temel sebeplerindendir [28], [13]. Benzer bir durum makro veriler için de söz konusudur. Çalışmada yer alan birimler ülkeler veya şirketler vb. olduğunda bu birimlere ait bazı değişkenler düzenli olarak gözlemlenemeyebilir veya her birim için her dönemde veri bulunamayabilir. Ayrıca belirli bir dönemden sonra bazı birimler örneklemeden çıkarken, bazı yeni birimlerin örnekleme dahil olması söz konusu olabilir. Genel olarak panel verinin hem kesit hem de zaman boyutundan kaynaklanabilen eksik veri probleminin nedenleri çoğaltılabilir [27].

Bir veri setinde var olan eksikliklerin yapılacak analizlere etkisi, bu eksik verilerin ortaya çıkış süreci ile doğrudan ilişkilidir [24]. Rubin, eksik verileri ortaya çıkış nedenlerine ve hangi değişkende ortaya çıktıklarına göre üç temel sınıfa ayırmıştır [22]. Buna göre, bir

değişkendeki kayıp gözlemlerin varlığı, değişkenin kendisiyle ve veri setindeki diğer değişkenlerle ilişkili değilse, bu duruma Tamamen Rassal Eksiklik (MCAR) adı verilmektedir. Bir değişkendeki kayıp gözlemlerin varlığı, değişkenin kendisiyle ilişkili değilken, diğer değişkenlerle ilişkili ise bu duruma Rassal Olarak Eksiklik (MAR) adı verilmektedir. Bu iki yapının geçerli olmadığı duruma, yani değişkende kayıp gözlemlerin varlığının değişkenin kendisi ile ilişkili olması durumuna ise Rassal Olmayan Eksiklik (MNAR) adı verilmektedir. MNAR yapısındaki eksik veriler aynı zamanda “ihmal edilemez” kayıp olarak nitelendirilmektedir. Çünkü bu verilerde sistematik bir eksikliğin varlığı söz konusudur ve bu eksiklikler ihmal edilerek yapılan tahminler sapmalı olacaktır [23]. Bu yapılar için genel bir örnek, Little ve Rubin tarafından yaş ve gelir değişkenleri üzerinden şu şekilde verilmiştir; gelir değişkeninde eksikliklerin olması kişinin yaşı veya geliri ile ilişkili değilse MCAR, kişinin yaşı ile ilişkili, ancak geliri ile ilişkili değilse MAR, kişinin geliri ile ilişkili ise MNAR mekanizması geçerlidir [16].

Bir çalışmada kullanılacak veri seti eksik gözlemler içerdiğinde, bu eksik gözlemler için bazı alternatif çözüm yöntemleri önerilmiştir. Bu yöntemler temel olarak iki sınıfta toplanabilir. Bunlardan ilki silme yöntemleridir. Veri setindeki eksik gözlemlerin silinmesine dayanan bu yöntemler, eksik veriler için en pratik çözüm yöntemleridir. Ancak bu yöntemlerin uygulanabilmesi için bazı kısıtlar söz konusudur. Silme yöntemlerinin uygulanması için verilerdeki eksikliklerin tamamen rassal olarak ortaya çıkmış olması gerekir. Bir başka deyişle MCAR mekanizmasının geçerli olması gerekir. Böylece eksik veriler silindiğinde geriye kalan alt örnek, ana kütlemin tesadüfi bir alt örneği olacaktır. Aksi takdirde bu yöntemler yardımıyla elde edilen tam veri seti üzerinden yapılan tahminler sapmalı olacaktır. Bununla birlikte, bu koşulun sağlanması pratikte genellikle oldukça zordur [4].

Silme yöntemlerinin sağlanması kolay olmayan bir ön koşula bağlı olmaları ve veri azaltma sonucunda bilgi kaybına yol açmaları sebebiyle bu yöntemlere alternatif olarak verilere atama yapıp veri setindeki kayıpları tamamlama fikri ortaya çıkmıştır. Atama işlemi, eksik gözlemleri tamamlamak için var olan gözlemler üzerinden yapılan bir hesaplama veya tahmin işlemidir [9]. Atama yapmak için çeşitli yöntemler mevcuttur. Bunlardan daha eski olanlar, herhangi bir modele dayanmayan istatistiksel yöntemlerdir. Bu yöntemler için; serinin ortalamasının atanması, son gözlemin ileri taşınması, benzer değer atama ve en yakın komşu ataması gibi yöntemler örnek olarak verilebilir. Daha moderne atama yöntemleri ise var olan gözlemlerin modellenmesi yardımıyla eksik gözlemler için tahmini değerlerin elde edilmesi temeline dayanmaktadır. Daha karmaşık olmakla birlikte, bu yöntemler geleneksel yöntemlerden daha başarılı sonuçlar vermektedir. Modern yöntemlere örnek olarak regresyon ataması, çoklu atama, en çok olabilirlik yöntemi ve beklenti maksimizasyonu algoritması verilebilir.

Bu çalışmada, panel verilerde eksik gözlemler var olduğunda, bu eksik veriler dikkate alınarak yapılacak iki alternatif tahmin yöntemin sonuçlarının karşılaştırılması amaçlanmıştır.

Bu yöntemlerin ilki dengesiz panel veri modelidir. İkincisi ise beklenti maksimizasyonu algoritması yardımıyla verilere atama yapılarak elde edilecek tam veri seti için dengeli panel veri modelinin tahmin edilmesidir. Bu amaçla iktisadî literatürde var olan eğitim ile ekonomik büyüme ilişkisi uygulama konusu olarak seçilmiştir.

Eğitim ile ekonomik büyüme arasındaki ilişki eskiden beri ilgi çeken bir araştırma konusudur. Ekonomik büyüme üzerinde beşeri sermayenin önemli bir düzeyde etkiye sahip olduğu kabul edildikten sonra, beşeri sermaye kavramı derinliğine ele alınmaya başlanmıştır. Bu kavramın temelde bireylerin gelişimi üzerine kurulu olması nedeniyle, bireylerin gelişiminde kilit bir role sahip olan eğitim faktörü beşeri sermayenin temel göstergelerinden biri olarak kabul edilmektedir. Çünkü eğitim, bireylerden oluşan işgücünün nitelikli hale gelmesini, dolayısıyla verimliliklerinin artmasını sağlayan temel faktördür. Ek olarak, yeni nesil teknoloji ve üretim süreçlerine hakimiyet için eğitim bir ön koşul niteliği taşımaktadır [26]. Ayrıca, gelişmiş ve gelişmekte olan ülkelerin ekonomik performansları arasındaki farkların temel bir nedeni eğitim düzeyleri arasındaki farklılıklardır. Eğitimin istihdam olanakları üzerindeki pozitif etkisi de ekonomik büyümeyi etkileyen sonuçlarından bir tanesidir [20], [11].

Ekonomik büyüme ile ilişkisi araştırılırken, eğitimi temsilen hangi değişkenlerin kullanılacağı net olarak belirli olmamakla birlikte, yapılan çalışmalarda öne çıkan bazı değişkenler bulunmaktadır. Bunlardan bir tanesi eğitim harcamalarıdır. Eğitimin bireylerin gelişimi üzerindeki aktif rolü nedeniyle, eğitime yapılan harcamaların uzun dönemde ülke ekonomisine olumlu şekilde yansıtacağı düşünülmektedir. Eğitimi temsilen kullanılan bir diğer değişken ise okul kayıt oranlarıdır. Yapılan çalışmalar farklı düzeylerdeki okul kayıt oranlarının ekonomik büyüme üzerinde farklı etkilere sahip olduğunu ortaya koymaktadır. Bunlara ek olarak okullaşma ve okur - yazarlık oranları eğitimi temsil eden değişkenler olarak çalışmalarda yer bulmaktadır [8], [10]. Ekonomik büyüme ile eğitim arasındaki ilişkiyi ortaya koyan geniş bir iktisadî literatür mevcuttur. Konu hakkında panel veri analizi kullanılarak yapılan çalışmalar incelendiğinde, ön plana çıkan ilk çalışmaların Barro'ya ait olduğu görülmektedir. Barro, ilk çalışmasında 98 ülkeye ait 1960 - 1985 dönemi verilerini kesit veri olarak incelemiş, daha sonraki çalışmalarında ise sırasıyla 91 ve 84 ülke için panel veri çalışması yapmıştır [5], [6], [7]. Çalışmaların öne çıkan sonuçları olarak, okul kayıt oranlarının ve okullaşma oranlarının ekonomik büyüme üzerinde olumlu etkisi olduğunu ortaya koymuştur. Keller, ekonomik büyüme üzerindeki etkilerini analiz etmek üzere her öğretim düzeyi için yapılan kamu harcamaları ve bu öğretim düzeylerine ait kayıt oranları değişkenlerini kullanmıştır [15]. Sonuçta farklı düzeyler için eğitim harcamalarının ve kayıt oranlarının ekonomik büyüme üzerinde anlamlı etkilere sahip olduğu tespit edilmiştir. Suri vd., 79 ülkeye ait 1960 - 2001 dönemi verilerini incelemiş ve ortaöğretim kayıt oranları ile ekonomik büyüme arasında anlamlı bir ilişki olduğu sonucuna ulaşmışlardır [25]. Oztunc, Chi Oo ve Serin, 11 Asya - Pasifik ülkesi için 1990 - 2010

dönemini inceledikleri çalışmalarında, kadın işgücü katılım oranı ve kadın temel eğitim kayıt oranının ekonomik büyüme üzerinde anlamlı etkilere sahip olduğu sonucuna ulaşmışlardır [21].

II. METODOLOJİ

Bu çalışmada eksik gözlem içeren panel veriler için alternatif iki tahmin yönteminin sonuçlarının karşılaştırılması amacıyla eğitim ile ekonomik büyüme arasındaki ilişki analiz edilmiştir. Çalışmada 34 OECD ülkesine ait 2003 – 2013 dönemi yıllık verileri kullanılmıştır. Veriler Dünya Bankası veri tabanından elde edilmiştir. Bağımlı değişken olarak kişi başına gayri safi yurtiçi hasıla değişkeninin logaritmik değeri kullanılmıştır. Açıklayıcı değişkenler olarak ilköğretim, ortaöğretim ve yükseköğretim düzeyleri için eğitim harcamaları, okul kayıt oranları ve bu düzeylerin her biriyle işgücüne katılım oranları alınmıştır. Bu değişkenlerin tamamı ile modeller kurulduktan sonra, anlamsız değişkenler çıkartılarak en uygun model elde edilmiştir. Sonuçta modelde dört adet açıklayıcı değişken yer almıştır. Bunlar yükseköğretim düzeyi için eğitim harcaması, kayıt oranı, yükseköğretimle işgücüne katılım oranı ve ilköğretimle işgücüne katılım oranıdır. Burada eğitim harcaması, eğitime yapılan hükümet harcaması içinde yükseköğrenimin yüzde olarak payıdır. Kayıt oranı, toplam kayıt miktarının ilgili eğitim düzeyinin gösterildiği yaş grubundaki nüfusa yüzde olarak oranıdır. İşgücüne katılım oranı ise, toplam işgücü içerisinde, en yüksek eğitim düzeyi olarak ilgili eğitim düzeyini bitirmiş olanların yüzde olarak payıdır. Değişkenlere ait bu tanımlamalar, verinin alındığı Dünya Bankası veri tabanında bu değişkenler için yapılmış olan tanımlamalardır. Bağımlı değişken tüm ülkeler için tüm zaman dönemlerinde gözlemlenmiştir ancak açıklayıcı değişkenlerde gözlemlerde eksiklikler bulunmaktadır.

Bu çalışma iki aşamadan oluşmaktadır. İlk aşamada modeli tahmin etmek için dengersiz panel veri modeli kullanılmıştır. Ardından varsayım testleri uygulanmış ve varsayımlardan sapmalara uygun olan robust tahminci yardımıyla model tekrar tahmin edilmiştir. Çalışmada kullanılan verinin zaman boyutu birim boyutuna göre daha kısadır ve yatay kesit boyutu ön plandadır. Ayrıca kullanılan değişkenler oransal değişkenlerdir. Bu nedenle durağanlık incelemesi yapılmamıştır. İkinci aşamada ise veri setindeki eksik gözlemler, model tabanlı bir atama yöntemi olan Beklenti Maksimizasyonu (EM) Algoritması yöntemi ile doldurulmuş ve dengeli panel veri modeli tahmin edilmiştir. Ardından model için varsayım testleri uygulanmış ve varsayımlardan sapmalar için uygun robust tahminci kullanılarak model tekrar tahmin edilmiştir. Daha sonra elde edilen modellere ait sonuçlar karşılaştırılmıştır. Model tahminin de kullanılan panel veri modellerine ve eksik veri atamasında kullanılan EM Algoritması yöntemine bu bölümün devamında değinilmiştir.

II.1. Dengesiz Panel Veri Modelleri

Panel veri setinde her birim için eşit sayıda zaman dönemi olmaması durumu dengesiz panel olarak adlandırılır. Bir başka deyişle dengesiz panelde her birim için her zaman dönemine ait gözlemler mevcut değildir, dolayısıyla dengesiz panel, eksik veri probleminin özel bir durumudur. Bir dengesiz panel veri modeli genel olarak aşağıdaki gibi gösterilebilir:

$$y_{it} = \alpha + \beta x_{it} + u_{it}; \quad u_{it} = \mu_i + \varepsilon_{it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T_i$$

Burada i alt indisi birim sayısını, t alt indisi ise zaman dönemini ifade etmektedir. T_i , i . birim için gözlemlenen zaman dönemi sayısıdır. Böylelikle dengeli panelden farklı olarak, dengesiz panel veri modellerinde zaman döneminin birimlere göre değişmesi modele yansıtılmış olur. y_{it} bağımlı değişkeni, x_{it} açıklayıcı değişkenleri, u_{it} ise hata terimini ifade etmektedir. μ_i terimi i . birime ait birim etkisini, ε_{it} ise panelin genel hata terimini ifade etmektedir. Panel veri modelleri, varsa birim veya zaman etkilerinin modelde nasıl dikkate alındığına göre iki gruba ayrılır. Eğer her bir birimin etkisi modelde sabit bir terim olarak kabul edilirse, bu tür modellere Sabit Etkili Panel Veri Modelleri adı verilir. Bu durumda μ_i terimi modelde sabit bir katsayı gibi ele alınacak ve modelin sabit terimine eklenerek, her bir birime ait sabit terimin ortak sabitten ne kadar farklılaştığını gösterecektir. Eğer birim etkisi sabit bir terim olarak değil de, hata teriminin rassal bir bileşeni olan bir değişken olarak ele alınırsa, bu tür modellere ise Rassal Etkili Panel Veri Modelleri adı verilir. Bu durumda μ_i terimi rassal bir değişken olarak ele alınacak ve dışsallık koşulu gereği, birim etki ile açıklayıcı değişkenler arasında korelasyona izin verilmeyecektir [3], [14]. Yapılan analizde modelin sabit etkili olarak mı yoksa rassal etkili olarak mı ele alınacağı Hausman Testi yardımıyla belirlenmiştir.

II.2. Beklenti Maksimizasyonu (EM) Algoritması

EM Algoritması, eksik verilere atama yapmakta kullanılan, iteratif bir sürece sahip ve model tabanlı bir yöntemdir. Bu yöntem iki basamaktan oluşur. İlk basamak “Beklenti (E)” adımı, ikinci basamak ise “Maksimizasyon (M)” adımıdır. Algoritmanın çalışması için öncelikle verilere ait ortalama vektörü ve kovaryans matrisinin başlangıç değerlerinin var olması gerekmektedir. Bu nedenle, öncelikle eksik veri silme yöntemleri yardımıyla bu iki parametre elde edilir. Bu iki parametre yardımıyla, var olan veriler üzerinden eksik verileri tahmin etmek amacıyla regresyon denklemi kurulur. Daha sonra regresyon denklemi yardımıyla eksik veriler için tahmini değerler bulunur. Böylece tam bir veri seti elde edilir ve E adımı tamamlanmış olur. M adımında ise elde edilen tam veri seti üzerinden ortalama vektörü ve kovaryans matrisi için En Çok Olabilirlik tahminleri elde edilir ve bu iki değer başlangıçta kullanılan değerlerin yerine konarak süreç tekrarlanır. Bu proses, son adımda elde edilen ortalama vektörü ve kovaryans matrisi değerleri ile bir önceki adımda elde edilen

değerler arasında anlamlı bir fark kalmayınca kadar tekrarlanır. Her iterasyon, bir önceki iterasyondan elde edilen bilgi kullanıldığından daha fazla bilgi içermektedir [19], [12], [1], [2], [17].

III. BULGULAR

Bu bölümde model tahmin sonuçları sunulmuştur. Kurulan model aşağıdaki gibidir:

$$\ln gsyh_{it} = \beta_0 + \beta_1 y_harc_{it} + \beta_2 y_kayit_{it} + \beta_3 y_iko_{it} + \beta_4 i_iko_{it} + d9 + \varepsilon_{it}$$

$$i : 1, 2, \dots, 34 \quad t = 1, 2, \dots, T_i \quad T_i \leq 11$$

Burada;

- . $\ln gsyh$: Reel kişi başına gayri safi yurtiçi hasılanın logaritmik değerini,
- . y_harc : Yükseköğretim için yapılan eğitim harcaması oranını,
- . y_kayit : Yükseköğretim kayıt oranını,
- . y_iko : Yükseköğretimle işgücüne katılım oranını,
- . i_iko : İlköğrenimle işgücüne katılım oranını
- . $d9$: 2009 yılına ait kukla değişkeni temsil etmektedir.

Serilerde ortaya çıkan yapısal kırılmayı test etmek amacıyla, her yıl için ayrı ayrı kukla değişken eklenerek modeller tahmin edilmiş ve bu modellerden en düşük toplam hata kare değerini 2009 yılına ait kukla değişkenli model vermiştir. Bu durumda 2009 yılında yapısal kırılma olduğu ortaya çıkmıştır. Bu durum, 2008 yılında yaşanan küresel ekonomik krizin bir sonraki yıldan itibaren ekonomik verilere yansımaları olarak açıklanabilir.

Model öncelikle dengesiz panel veri modeli olarak tahmin edilmiştir. Sabit etkili ve rassal etkili model tahmin edildikten sonra, hangi modelin uygun olduğuna Hausman Testi yardımıyla karar verilmiştir. Hausman Testi sonucu aşağıdaki gibidir:

Tablo 1. Hausman Testi Sonucu

H_0 : Katsayılar arasındaki fark sistematik değildir.	
Test İstatistiği: chi2(5)	Prob>chi2
4.06	0.5402

Test sonucu incelendiğinde, modelin rassal etkili olduğunu ifade eden sıfır hipotezi reddedilememiştir, dolayısıyla rassal etkili model uygun model olarak seçilmiştir. Rassal etkili model tahmin sonucu aşağıdaki gibidir:

Tablo 2. Rassal Etkili Dengesiz Panel Veri Modeli Tahmin Sonuçları

Bağımlı Değişken: lngsyh		Wald chi2(5)= 149.00		
Gözlem Sayısı: 230		Prob > chi2 = 0.0000		
Değişkenler	Katsayılar	Std. Hata	z	P> z
sabit	9.2421	0.1248	74.08	0.000
y_harc	0.0055	0.0024	2.27	0.023
y_kayit	0.0064	0.0008	8.40	0.000
y_iko	0.0100	0.0017	5.90	0.000
i_iko	0.0028	0.0009	2.94	0.003
d9	-0.0463	0.0097	-4.79	0.000
R ² : Grup içi = 0.4297 Gruplar arası = 0.3313 Toplam = 0.3361				

Modele ait tahmin sonuçları incelendiğinde, öncelikle gözlem sayısının dengeli panelde olması gerektiği gibi 34x11 değil, eksik verilerden dolayı 230 olarak verildiği görülmektedir. Modelin genel anlamlılığı için yapılan Wald Testi ve katsayıların anlamlılığı için yapılan test sonuçları incelendiğinde, tüm katsayıların ve genel olarak modelin istatistiksel olarak anlamlı çıktığı görülmektedir. Tüm katsayılar literatüre uygun olarak pozitif bulunmuştur. Yalnızca 2009 yılına ait kukla değişken negatif işaretli bulunmuştur. Katsayılar yorumlanırken bağımlı değişkenin logaritmik olduğu göz önünde bulundurulmalıdır. Buna göre diğer değişkenlerin etkisi sabitken, yükseköğretime yapılan harcama oranındaki 1 birimlik artış, kişi başına GSYH'yı %0.55 oranında arttıracaktır. Benzer şekilde yükseköğretim kayıt oranındaki 1 birimlik artış, kişi başına GSYH'yı %0.64 oranında arttıracaktır. Açıklayıcı değişkenler içerisinde en büyük etkinin yükseköğretimle işgücüne katılım oranına ait olduğu gözlenmektedir. Yükseköğretimle işgücüne katılım oranındaki 1 birimlik artış, kişi başına GSYH'yı %1 oranında arttıracaktır. İlköğrenimle işgücüne katılım oranındaki 1 birimlik artış ise, kişi başına GSYH'yı %0.28 oranında arttıracaktır. Sabit ve rassal etkili panel veri modellerinde üç ayrı belirlilik katsayısı (R²) hesaplanmaktadır. Model çıktısı incelendiğinde, toplam R² %33,61 olarak bulunmuştur. Model için varsayımlardan sapmalar test edilmiştir. Otokorelasyon ve değişen varyans için test sonuçları elde edilmiştir. Ancak yatay kesit bağımlılığı testleri mevcut veri seti için çalışmadığından test edilememiştir. Otokorelasyon için yapılan Benzerlik Oranı (LR) Testi ve değişen varyans için yapılan Levene, Brown & Forsythe Testi sonuçları aşağıdaki gibidir:

Tablo 3. Model İçin Varsayımdan Sapma Testlerinin Sonuçları

H_0 : Otokorelasyon yoktur	
Test İstatistiği	Prob>chi2(1)
210.41	0.0000
7.83	0.0051
H_0 : Değişen varyans yoktur	
Test İstatistiği	Prob>F
W0 = 5.5432664 df(29, 200)	0.0000
W50 = 2.5315577 df(29, 200)	0.0001
W10 = 5.2185301 df(29, 200)	0.0000

Test sonuçları incelendiğinde, hem otokorelasyon hem de değişen varyans testi sonucunda sıfır hipotezlerinin reddildiği görülmektedir, yani modelde otokorelasyon ve değişen varyans bulunmaktadır. Bu nedenle bu iki durum için dirençli olan bir robust tahminci yardımıyla model tekrar tahmin edilmiştir. Bu tahminci Arellano, Froot & Roger tahmincisidir. Robust tahmin sonuçlarında, katsayıların anlamlılığı için z Testi istatistikleri verilmektedir. Yükseköğrenime yapılan harcama oranı değişkeninin katsayısına ait z test istatistiğinin olasılık değerine göre, bu değişken istatistiksel olarak anlamsız çıkmıştır. Bu nedenle aynı model, bu değişken çıkartılarak tekrar kurulmuştur. Modelin robust tahmin sonuçları Tablo 4'de verilmiştir.

Tablo 4. Arellano, Froot & Roger Robust Dengesiz Panel Veri Modeli Tahmin Sonucu

Bağımlı Değişken: lngsyh				Wald chi2(4)= 19.89
Gözlem Sayısı: 230				Prob > chi2 = 0.0005
Değişkenler	Katsayılar	Robust Std. Hata	z	P> z
sabit	9.5342	0.2322	41.06	0.000
y_kayit	0.0045	0.0018	2.52	0.012
y_iko	0.0096	0.0035	2.76	0.006
i_iko	0.0032	0.0014	2.21	0.027
d9	-0.0418	0.0101	-4.13	0.000
R ² : Grup içi = 0.3015 Gruplar arası = 0.1200 Toplam = 0.1693				

Bir model için bu robust tahminci uygulandığında standart hatalar yerine robust standart hatalar hesaplanmaktadır. Dolayısıyla eğer yükseköğrenim harcama oranı değişkeni anlamsız çıkmasa ve modelden çıkartılmasaydı, modelin katsayıları Tablo 3'de verilen model sonuçlarındaki katsayılar ile aynı kalacak, yalnızca standart hatalar değişmiş olacaktır. Ancak istatistiksel olarak anlamsız çıkan değişken varken modelin sonuçlarını yorumlamak çok doğru olmayacağından, bu değişken çıkartılarak model yenilenmiştir. Bu nedenle katsayılar ve bu katsayılara ait standart hatalar ve test istatistikleri değişmiştir. Robust modelin son halinde tüm katsayıların istatistiksel olarak anlamlı ve Wald test sonucuna göre modelin de genel olarak anlamlı olduğu görülmektedir. Rassal etkili modelin sonuçları ile karşılaştırıldığında, robust modelde katsayılarda nispeten düşük değişmeler olduğu gözlemlenmektedir. Yükseköğrenime kayıt oranı ve yükseköğrenimle işgücüne katılım oranının etkisi azalırken, ilköğrenimle işgücüne katılım oranının etkisi ise yükselmiştir. Robust standart hataların ise normal standart hatalara nazaran daha yüksek çıktığı dikkat çekmektedir. Ayrıca her üç belirlilik katsayısı da rassal etkili modele kıyasla azalmıştır, ancak bu durum açıklayıcı değişkenlerden biri eksildiği için toplam açıklanma oranının azalmasından kaynaklanmaktadır ve beklenen bir durumdur.

Bu bölümün devamında, veri setindeki eksiklikler EM Algoritması yardımıyla tamamlanmıştır. Ardından atama sonucunda elde edilen dengeli panel veri seti için model tahmin sonuçları sunulmuştur. Sabit ve rassal model arasında karar vermek için yapılan Hausman Testi sonucu aşağıdaki gibidir:

Tablo 5. Hausman Testi Sonucu

H_0 : Katsayılar arasındaki fark sistematik değildir.	
Test İstatistiği: chi2(5)	Prob>chi2
2.82	0.7273

Test sonucu incelendiğinde, modelin rassal etkili olduğunu ifade eden sıfır hipotezi reddedilememiştir, dolayısıyla rassal etkili model uygun model olarak seçilmiştir. Rassal etkili model tahmin sonucu aşağıdaki gibidir:

Tablo 6. EM Algoritması İle Elde Edilen Dengeli Panel İçin Rassal Etkili Model Tahmin Sonuçları

Bağımlı Değişken: lngsyh		Wald chi2(5)= 62.74		
Gözlem Sayısı: 374		Prob > chi2 = 0.0000		
Değişkenler	Katsayılar	Std. Hata	z	P> z
sabit	9.9100	0.0973	101.83	0.000
y_harc	0.0030	0.0013	2.25	0.024
y_kayit	0.0017	0.0004	4.12	0.000
y_iko	0.0019	0.0009	2.04	0.041
i_iko	0.0017	0.0005	3.06	0.002
d9	0.0206	0.0071	2.89	0.004
R ² : Grup içi = 0.1604 Gruplar arası = 0.2563 Toplam = 0.1930				

Model çıktısı incelendiğinde, gözlem sayısının dengeli panele uygun olarak birim sayısı ile zaman dönemi sayısının çarpımı olarak elde edildiği görülmektedir. Tüm değişkenlerin ve genel olarak modelin istatistiksel olarak anlamlı çıktığı görülmektedir. Değişkenlerin etkisi incelendiğinde; diğer değişkenlerin etkisi sabitken, yükseköğrenim yapılan harcama oranındaki 1 birimlik artış, kişi başına GSYH'yı %0.30 oranında arttıracaktır. Yükseköğrenim kayıt oranındaki 1 birimlik artış, kişi başına GSYH'yı %0.17 oranında arttıracaktır. Yükseköğrenimle işgücüne katılım oranındaki 1 birimlik artış, kişi başına GSYH'yı %0.19 oranında arttıracaktır. İlköğrenimle işgücüne katılım oranındaki 1 birimlik artış ise, kişi başına GSYH'yı %0.17 oranında arttıracaktır. Ayrıca kırılma kuklasına ait katsayının işaretinin pozitif çıktığı görülmektedir. Modelin sonucu ile Tablo 2'de verilen dengesiz panel veri için rassal etkili model sonucu karşılaştırıldığında, katsayıların dengesiz panel veri modeline göre daha düşük çıktığı göze çarpmaktadır. Bir başka deyişle atama sonucunda elde edilen dengeli panel veri modelinde değişkenlerin etkisinin dengesiz panel veri modeline göre azaldığı sonucuna ulaşılmıştır. Toplam belirlilik katsayısı %19 civarında ve dengesiz panel veri modeline daha düşük bulunmuştur. Tablo 6'da sonuçları sunulan rassal etkili model için varsayımlardan sapmalar test edilmiştir. Otokorelasyon için yapılan Benzerlik Oranı (LR) Testi, değişen varyans için yapılan Levene, Brown & Forsythe Testi ve yatay kesit bağımlılığı için yapılan Pesaran Testi sonuçları aşağıdaki gibidir:

Tablo 7. Model İçin Varsayımdan Sapma Testlerinin Sonuçları

H₀: Otokorelasyon yoktur	
Test İstatistiği	Prob>chi2(1)
381.47	0.0000
22.75	0.0000
H₀: Değişen varyans yoktur	
Test İstatistiği	Prob>F
W0 = 5.7204816 df(33, 340)	0.0000
W50 = 3.7534622 df(33, 340)	0.0000
W10 = 5.2129960 df(33, 340)	0.0000
H₀: Yatay kesit bağımlılığı yoktur	
Test İstatistiği	Prob
34.648	0.0000

Test sonuçları incelendiğinde, sıfır hipotezlerinin reddildiği görülmektedir ve dolayısıyla modelde otokorelasyon, değişen varyans ve yatay kesit bağımlılığının var olduğu sonucuna ulaşılmıştır. Bu nedenle bu üç durumun varlığına dirençli olan bir robust tahminci kullanılarak model yeniden tahmin edilmiştir. Bu tahminci Parks – Kmenta tahmincisidir. Robust tahmin sonuçları Tablo 8’de görülmektedir.

Tablo 8. Parks - Kmenta Robust Dengeli Panel Veri Modeli Tahmin Sonucu

Bağımlı Değişken: lngsyh				Wald chi2(4)= 339.31
Gözlem Sayısı: 374				Prob > chi2 = 0.0000
Değişkenler	Katsayılar	Std. Hata	z	P> z
sabit	9.0048	0.0790	114.03	0.000
y_harc	0.0082	0.0012	6.84	0.000
y_iko	0.0261	0.0015	16.83	0.000
i_iko	0.0040	0.0007	5.87	0.000
d9	0.0322	0.0070	4.60	0.000

Değişkenlerin anlamlılığı için yapılan test sonucunda yükseköğrenime kayıt oranı değişkeni istatistiksel olarak anlamsız bulunmuş ve bu değişken çıkartılarak model yeniden tahmin edilmiştir. Model çıktısı incelendiğinde, modelin anlamlılığı için yapılan Wald Testi sonucunda model genel olarak anlamlı bulunmuştur. Kalan tüm değişkenlerin istatistiksel olarak anlamlı olduğu modelin son halinde, yükseköğretimle işgücüne katılım oranı değişkeninin katsayısının rassal etkili modele göre oldukça yüksek çıktığı dikkat çekmektedir. Buna göre, diğer değişkenlerin etkisi sabitken yükseköğretimle işgücüne katılım oranı 1 birim arttığında kişi başına GSYH %2.6 birim artmaktadır. Aynı değişken dengesiz panel için yapılan model tahminleri sonucunda da en yüksek katsayıya sahip değişkendir. Diğer değişkenlerin de etkisi rassal etkili modele oranla daha yüksek bulunmuştur. Bu robust tahminci için belirlilik katsayısının hesaplanmadığı görülmektedir.

IV. SONUÇ

Eksik veri problemi, özellikle daha geniş bir veri seti ile çalışma imkanı veren panel verilerde ortaya çıktığında, panel verinin bu avantajını kısıtlamaktadır. Bu çalışmada panel verilerde eksik gözlemlerin bulunması durumunda, model tahmini için iki alternatif çözüm ele alınmıştır. Bu amaçla uygulama konusu olarak eğitim ile ekonomik büyüme arasındaki ilişki incelenmiştir. Alternatif çözümlerden ilki, eksik verilerin dikkate alınarak modelin tahmin edildiği dengesiz panel veri modelidir. Bu model sonucunda eğitimi temsilen alınan yükseköğretime yapılan harcama oranı, yükseköğretim kayıt oranı, yükseköğretimle işgücüne katılım oranı ve ilköğretimle işgücüne katılım oranı değişkenlerinin ekonomik büyüme üzerinde anlamlı ve pozitif bir etkiye sahip oldukları ortaya çıkmıştır. Bu sonuçtan hareketle, eğitim düzeyleri arasında ekonomik büyüme üzerinde en fazla etkisi olanın yükseköğretim olduğu göze çarpmaktadır. Model için varsayımlardan sapmalar dikkate alınarak uygulanan robust tahmin sonucunda yükseköğretime yapılan harcama oranı değişkeni anlamsız hale gelmiş, diğer yükseköğretim değişkenlerinin etkileri ise düşmüştür. İlköğretimle işgücüne katılım oranı değişkeninin etkisi ise nispeten yükselmiştir.

İkinci alternatif çözüm olarak eksik verilere EM Algoritması yardımıyla atama yapılmış ve elde edilen tam veri seti ile dengeli panel veri modeli tahmin edilmiştir. Bu tahmin sonucunda aynı değişkenler istatistiksel olarak anlamlı çıkmakla birlikte, katsayılar dengesiz modele göre düşük bulunmuştur. Bir başka deyişle, değişkenlerin büyüme üzerindeki etkisi dengesiz panel veri modeline göre daha düşük düzeyde çıkmıştır. Öte yandan 2009 yılına ait kukla değişken dengesiz panel veri modelinde negatif katsayılı iken, atama sonucu elde edilen modelde pozitif katsayılı bulunmuştur. Modelin açıklama gücünü temsil eden belirlilik katsayılarının da dengesiz panel veri modeline göre daha düşük çıktığı gözükmektedir. Öte yandan, varsayımlardan sapmalar dikkate alınarak uygulanan robust tahmin yöntemi sonucunda daha önceki model tahmin sonuçlarında anlamlı olarak bulunmuş olan

yükseköğretime kayıt oranı değişkeni anlamsız çıkarken, dengesiz panel veri modelinin robust tahmininde anlamsız bulunmuş olan yükseköğretime yapılan harcama oranı değişkeni ise anlamlı bulunmuştur. Kırılma kuklasının katsayısı ise robust tahmin sonucunda pozitif bulunmuştur. Öte yandan her iki modelde de işgücüne katılım oranı değişkenleri anlamlı ve özellikle yükseköğretimle işgücüne katılım oranı değişkeni oldukça yüksek katsayıya sahip bulunmuştur.

İki yöntemle ait tahmin sonuçları karşılaştırıldığında, ilk modellerde değişkenlerin yalnızca etkileri farklılaşırken, robust tahmin sonuçlarında değişkenlerin anlamlılıklarında da farklılaşmalar olduğu dikkat çekmektedir. Bu durumun, eksik verilere atama yapmanın parametre tahminlerinde sapmalara neden olmasından kaynaklanması olasıdır. Çünkü dengesiz panel veri modeli daha az gözlem olmasına rağmen, var olan veriler üzerinden sapmasız tahminler yapmaktadır. Buna karşılık elde edilen tahmin sonuçları eksik verilere atama yapmanın parametreleri etkilediğini göstermektedir. Bu nedenle eksik verilerle çalışırken bir çözüm olarak atama yöntemlerinin kullanılması durumunda bu sakıncaların göz önünde bulundurulmasında yarar vardır. Buna karşılık eksik gözlemlere atama yapmanın, eksik olmayan veriler ile çalışılırsa elde edilecek sonuçlar hakkında yaklaşık bir fikir vermesi açısından avantajlı olduğu söylenebilir.

Yararlanılan Kaynaklar

- [1] Acock, A. (2005). Working With Missing Values. *Journal of Marriage and Family*, 1012–1028.
- [2] Allison, P. (2002). *Missing Data*. Thousand Oaks, CA: Sage Publications, Inc.
- [3] Baltagi, B. (2005). *Econometric Analysis of Panel Data*. 3. Edit. West Sussex: John Wiley & Sons, Ltd.
- [4] Baraldi, A., & Enders, C. (2013). Missing Data Methods. T. D. Little içinde, *The Oxford Handbook of Quantitative Methods* (s. 635 - 664). New York: Oxford University Press.
- [5] Barro, R. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, 407-443.
- [6] Barro, R. (1996). Determinants of Economic Growth: A Cross-Country Empirical Study. *NBER Working Paper Series No: 5698*.
- [7] Barro, R. (2001). Human Capital and Growth. *American Economic Review*, 12-17.
- [8] Breton, T. (2013). The Role of Education in economic Growth: Theory, History and Current Returns. *Educational Research*, 121–138.
- [9] Cameron, A., & Trivedi, P. (2005). *Microeconometrics Methods and Applications*. New York: Cambridge University Press.
- [10] Clarke, W., Jones, A., & Lacy, B. (2015). Education Spending and Workforce Quality as Determinants of Economic Growth. *Journal of Rural and Community Development*, 24-35.

- [11] Doğrul, N. (2009). Gelir Seviyeleri Farklı Döllerde Eğitimin Ekonomik Büyüme Etkisi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 259-267.
- [12] Enders, C. (2010). *Applied Missing Data Analysis*. New York: The Guilford Press.
- [13] Greene, W. (2008). *Econometric Analysis*. 6. Edit. New Jersey: Pearson Education, Inc.
- [14] Güriş, S. (2015). Panel Veri ve Panel Veri Modelleri. S. Güriş içinde, *Stata ile Panel Veri Modelleri* (s. 1-38). İstanbul: Der Yayınları.
- [15] Keller, K. (2006). Education Expansion, Expenditures per Student and the Effects on Growth in Asia. *Global Economic Review*, 21-42.
- [16] Little, R., & Rubin, D. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.
- [17] Madbully, D., Maravelakis, P., & Mahmoud, M. (2013). The Effect of Methods for Handling Missing Values on the Performance of the MEWMA Control Chart. *Communications in Statistics - Simulation and Computation*, 1437-1454.
- [18] McKnight, P. M. (2007). *Missing data : A Gentle Introduction*. New York: The Guilford Press.
- [19] McLachlan, G., & Krishnan, T. (2008). *The EM Algorithm and Extensions*. Second Edition. New Jersey: John Wiley & Sons, Inc.
- [20] Mercan, M., & Sezer, S. (2014). The Effect of Education Expenditure on Economic Growth: The Case of Turkey. *Procedia - Social and Behavioral Sciences*, 925 – 930.
- [21] Oztunc, H., Chi Oo, Z., & Serin, Z. (2015). Effects of Female Education on Economic Growth: A Cross Country Empirical Study. *Educational Sciences: Theory & Practice*, 349-357.
- [22] Rubin, D. (1976). Inference and Missing Data. *Biometrika*, 581 - 592.
- [23] Silva, L., & Zárate, L. (2014). A Brief Review of The Main Approaches for Treatment of Missing Data. *Intelligent Data Analysis*, 1177-1198.
- [24] Stock, J., & Watson, M. (2011). *Introduction to Econometrics*. 3. Edit. Boston: Pearson Education, Inc.
- [25] Suri, T., Boozer, M., Ranis, G., & Stewart, F. (2011). Paths to Success: The Relationship Between Human Development and Economic Growth. *World Development*, 506-522.
- [26] Tsamadias, C., & Prontzas, P. (2012). The Effect of Education on Economic Growth in Greece over the 1960-2000 Period. *Education Economics*, 522-537.
- [27] Verbeek, M. (2004). *A Guide to Modern Econometrics*. 2.Edit. West Sussex: John Wiley & Sons Ltd.
- [28] Young, R., & Johnson, D. (2015). Handling Missing Values in Longitudinal Panel Data with Multiple Imputation. *Journal of Marriage and Family*, 277-294.



Selahattin GÜRIŞ – sguris@marmara.edu.tr

Professor Selahattin GÜRIŞ received his undergraduate degree in Economics from Faculty of Economics and Administrative Sciences at Academy of Economics and Commercial Sciences in 1975. He received his M.A degree in Economics from Institute of Social Sciences at Boğaziçi University in 1980. He received his Ph.D. degree in Statistics Chair of Economics Department from Faculty of Economics at Marmara University in 1984. He is currently a faculty member in the Faculty of Economics at Marmara University. He teaches course on theoretical and applied econometrics. His research interests are econometric theory, time series and panel data econometrics.



Şaban KIZILARSLAN – saban.kizilarслан@marmara.edu.tr

Şaban KIZILARSLAN was born in Van, Turkey. He graduated from Department of Mathematics, Faculty of Science, Akdeniz University in 2011. He is currently a research assistant in Econometrics Department of Faculty of Economics and Administrative Sciences at Yüzüncü Yıl University. He is also a PhD student in Econometrics Department of Faculty of Economics at Marmara University. His research interests are econometric theory and panel data econometrics.