Ç.Ü. Sosyal Bilimler Enstitüsü Dergisi, Cilt 26, Sayı 3, 2017, Sayfa 13-25

ÇOKLU İÇ İLİŞKİ SORUNU OLAN REGRESYON MODELİNİN HKO ÖLÇÜTÜ İLE BİR ETKİN TAHMİN EDİCİSİ*

Hasan Altan ÇABUK[1]
Sibel ÖRK ÖZEL[2]

ÖZ

Bu çalışmada $y = X\beta + u$ şeklinde ifade edilen genel lineer regresyon modeli ele alınmış olup Kaliforniya yoksulluk veri kümesi (Ramanathan, 2002) üzerinde analiz yapılmıştır. Açıklayıcı değişkenler matrisi X çoklu iç ilişkiye sahip olduğundan bu iç ilişkinin varlığı durumunda modeli en küçük kareler (EKK) ile tahmin e*tmek hatalı sonuçlar verir. Bu problemin çözümü için yanlı fakat kararlı tahmin ediciler kullanılabilir. Çalışmada bu tahmin edicilerden genelleştirilmiş maksimum entropi (GME) tahmin edicisi ve ridge tahmin edici kullanılmış ve bu tahmin ediciler hata kareler ortalamasına (HKO) göre karşılaştırılmıştır. Uygulamada HKO'ların elde edilmesinde bootstrap yöntemi kullanılmıştır. Sonuçta, Kaliforniya yoksulluk veri kümesi için en iyi tahmin edicinin GME olduğuna karar verilmiştir.

Anahtar Kelimeler: Genel Lineer Model, Çoklu İç İlişki, Genelleştirilmiş Maksimum Entropi, Bootstrap

AN EFFECTIVE ESTIMATOR OF LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY PROBLEM ACCORDING TO MSE****

ABSTRACT

In this study, the linear regression model $y = X\beta + u$ is considered and California poverty dataset (Ramanathan, 2002) is analysed. Explanatory variables matrix X may subject to multicollinearity in applications. In the presence of multicollinearity, ordinary least squares (OLS) produce incorrect results. Biased but stable estimators can be used to overcome this problem. In this study, generalized maximum entropy (GME) estimator is compared with the ridge estimator according to the mean squared error (MSE) criteria. In the application section, the bootstrap method is used to obtain MSE values. In conclusion, GME estimator is decided as the best estimator for the California poverty dataset.

Keywords: General Linear Model, Multicollinearity, Generalized Maximum Entropy, Bootstrap

---

***Bu çalışma yüksek lisans tezinden faydalanarak hazırlanmış olup TUBİTAK BİDEB 2211-A Genel Yurtiçi Doktora Burs Programı kapsamında desteklenmiştir. This research is prepared from master thesis and is funded by TUBITAK BIDEB within the scope of 2211-A General Domestic PhD Scholarship**

[1]Prof.Dr., Çukurova Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, haltan@cu.edu.tr [2]Arş.Gör., Çukurova Üniversitesi, İ.İ.B.F., Ekonometri Bölümü, sork@cu.edu.tr

Araştırma, Gönderim Tarihi: 29.11.2017 Kabul Tarihi:: 14.12.2017

1.Giriş

Klasik tahmin yöntemlerini genel lineer modelde (GLM) uygulayabilmek için çeşitli varsayımlar bulunmaktadır. Ancak uygulamada bu varsayımların sağlanmadığı (yanlış örneklem seçimi, veri kümesinin hatalı bir şekilde oluşturulması gibi) durumlarla sıklıkla karşılaşılabilir.

Çalışmada bu problemlerden birisi olan çoklu iç ilişki problemi ele alınmıştır. Lineer regresyon modelinin önemli varsayımlarından biri olan açıklayıcı değişkenler arasında tam çoklu iç ilişki yoktur varsayımından sapma durumunda çoklu iç ilişki sorunu meydana gelir. Bu durum parametre tahminlerinin oldukça kararsız olmasına neden olur. Fakat ekonometrik bir çalışmada kararsız tahmin ediciler değil kararlı tahmin ediciler elde edilmek istenir. Bu nedenle çoklu iç ilişki olması durumunda yanlı tahmin edicilere başvurulur.

Çoklu iç ilişkinin varlığı durumunda regresyon katsayılarının en küçük kareler (EKK) tahmin edicileri yansız olmaya devam etmekle birlikte, bu tahmin edicilerin varyans ve kovaryansları büyük çıkma eğilimindedir. Bu durumda güven aralıkları olması gerektiğinden daha geniş çıkabilecek ve hipotez testlerinde güven eksikliği olabilecektir. Ayrıca parametre tahminleri beklenen işaretlere sahip olmayabilecek ve parametre tahminlerinde büyük sapmalar meydana gelebilecektir.

Bu nedenle çoklu iç ilişkinin parametre tahminleri üzerindeki etkisi kontrol edilmelidir. EKK tahminleri modeldeki çoklu iç ilişkiden olumsuz yönde etkileneceğinden, modelde çoklu iç ilişki olması durumunda yanlı fakat kararlı tahmin edicilere başvurulur. Bu çalışmada yanlı tahmin edicilerden literatürde 1970 yılından günümüze kadar önemli bir paya sahip olan Ridge tahmin edici ile son yıllarda sıkça kullanılan genelleştirilmiş maksimum entropi (GME) tahmin edicileri incelenecektir. Çalışmada amaç, GLM'de çoklu iç ilişki problemi olması durumunda EKK yönteminin iyi sonuçlar vermeyebileceğini ve alternatif olarak geliştirilen yanlı tahmin edicilerin çoklu iç ilişkiden daha az etkilendiğini bir veri kümesi üzerinde göstermektir.

Çalışma dört bölümden oluşmaktadır. Giriş bölümünü takip eden ikinci bölümde GLM için GME tahmin edicisi incelenmiştir. Uygulama kısmının yer aldığı üçüncü bölümde çoklu iç ilişkinin tespit edildiği Kaliforniya şehirlerine ait yoksulluk veri kümesi (Ramanathan, 2002) incelenmiştir. Bu veri kümesine EKK, Ridge ve GME tahmin edicileri uygulanarak parametre tahminleri elde edilmiş, tahmin edicilerin etkinlikleri hata kareler ortalaması (HKO) ölçütüne göre karşılaştırılmıştır. Çalışmanın son bölümünde sonuç ve tartışmalara yer verilmiştir.

GENELLEŞTİRİLMİŞ MAKSİMUM ENTROPİ TAHMİN EDİCİSİ

1996 yılında Golan, Judge ve Miller tarafından önerilen GME tahmin edicisi Jaynes (1957) tarafından sunulan maksimum entropi (ME) yönteminin GLM için genelleştirilmiş halidir. ME prensibi Shannon (1948) tarafından geliştirilen bilgi kuramı kavramından yararlanılarak inverse problemlerinin çözümü için kullanılmaya başlanmıştır. ME prensibi tüm olasılık dağılımları içerisinde entropisi en yüksek olan dağılımı seçmeyi ve verinin modelde söz sahibi olmasını hedefler. ME yöntemi sadece eksik-sunumlu inverse problemler için kullanılırken GME hem eksik-sunumlu hem de kötü-koşullu inverse problemler için kullanılır.

$\mathbf{y}: T \times 1$ boyutlu açıklanan değişken vektörü,

$\mathbf{X}: T \times K$ boyutlu açıklayıcı değişkenler matrisi (tasarım matrisi),

$\boldsymbol{\beta}: K \times 1$ boyutlu bilinmeyen (tahmin edilecek) parametreler vektörü

$\mathbf{u}: T \times 1$ boyutlu hata vektörü olmak üzere

GLM (1) numaralı eşitlikteki gibi ifade edilebilir:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \tag{1}$$

Golan, Judge ve Miller (1996) bilinmeyen parametreleri ve hataları GME ile tahmin etmek üzere; kompakt (tıkız) desteklerle birlikte kesikli rassal olasılık değişkenleri kullanarak, (1) numaralı denklemde verilen model parametrelerini ve hata terimlerini yeniden tanımlamışlardır.

Kompakt destek kümesi $M \geq 2$ sayıda destek değerini içerir. $\mathbf{p_k} = \left[p_{k1}, \ldots, p_{kM} \right]'$ toplamları bir olan pozitif ağırlıklı M boyutlu kesikli rassal olasılık değişkenleri ve $\mathbf{z_k} = \left[z_{k1}, \ldots, z_{kM} \right]'$ parametre destek vektörü olmak üzere bilinmeyen parametreler vektörü;

$$\boldsymbol{\beta} = \mathbf{Zp} \tag{2}$$

olarak ifade edilir. Burada;

$\mathbf{Z}: K \times KM$ boyutlu kompakt destek matrisi ve

$\mathbf{p}: KM \times 1$ boyutlu ağırlıklar vektörüdür.

Her bir hata için $J \geq 2$ destek noktaları kümesi tanımlansın. $\Pr\left[v_{t1} < u_t < v_{tJ} \right]$ keyfi olarak seçilen ve yeterince küçük olacak şekilde her u_t için v_{t1} ve v_{tJ} hata sınırlarını oluşturmak mümkündür (Golan, Judge ve Miller, 1996).

v_t hata terimleri için destek vektörü ve w_t destek noktaları için ağırlıklar vektörü olmak üzere

$\mathbf{V}: T \times TJ$ boyutlu destek noktaları matrisi,

$\mathbf{w}: TJ \times 1$ boyutlu bilinmeyen ağırlıklar vektörü kullanılarak bilinmeyen hata vektörü matris gösterimiyle

$$\mathbf{u} = \mathbf{Vw} \tag{3}$$

şeklinde yazılabilir.

Golan, Judge ve Miller (1996) hata bileşenlerinin sınırlarının Pukelsheim (1994) tarafından önerilen üç sigma kuralı ile bulunmasını önermişlerdir. Bu çalışmada da hata sınırları için Pukelsheim (1994) tarafından önerilen $\pm 3\sigma$ kuralı uygulanmıştır.

Sonuç olarak (1) nolu eşitlikle ifade edilen GLM yeniden parametrelendirilmiş olarak (4) nolu eşitlikteki gibi ifade edilebilir.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} = \mathbf{XZp} + \mathbf{Vw} \tag{4}$$

Yeniden parametrelendirilen model GME problemi olarak;

$$\max H(\mathbf{p},\mathbf{w}) = -\sum_{k=1}^{K}\sum_{m=1}^{M} p_{km}\ln p_{km} - \sum_{t=1}^{T}\sum_{j=1}^{J} w_{tj}\ln w_{tj} \tag{5}$$

şeklinde ifade edilebilir. Lagrange fonksiyonu kurulup gerekli işlemler yapıldıktan sonra;

$$\Omega_k^p\left(\hat{\boldsymbol{\lambda}}\right) = \sum_{m=1}^{M}\exp\left(-\sum_{t=1}^{T}\hat{\lambda}_t z_{km} x_{tk}\right) \tag{6}$$

ve

$$\Psi\left(\hat{\boldsymbol{\lambda}}\right) = \sum_{j=1}^{J}\exp\left(-\hat{\lambda}_t v_{tj}\right) \tag{7}$$

olmak üzere

$$\hat{p}_{km} = \frac{\exp\left(-\sum_{t=1}^{T}\hat{\lambda}_t z_{km} x_{tk}\right)}{\Omega_k^p\left(\hat{\boldsymbol{\lambda}}\right)} \tag{8}$$

ve

$$\hat{w}_{tj} = \frac{\exp\left(-\hat{\lambda}_t v_{tj}\right)}{\Psi\left(\hat{\boldsymbol{\lambda}}\right)} \tag{9}$$

olduğu gösterilebilir. Elde edilen tahmin ediciler yerlerine yazıldığında;

$$\hat{\boldsymbol{\beta}} = \mathbf{Z}\hat{\mathbf{p}} \tag{10}$$

$$\hat{\mathbf{u}} = \mathbf{V}\hat{\mathbf{w}} \tag{11}$$

GME tahmin edicilerine ulaşılır (Campbell ve Hill, 2006).

UYGULAMA

Bu bölümde, Ramanathan'da (2002, s. 653) yer alan ve Campbell ve Hill'in (2001) ele aldığı Kaliforniya'ya ait yoksulluk veri kümesi kullanılmıştır. Bu veri kümesi 1980 ve 1990 yılları için 58 vilayete ait verilerden oluşmaktadır. Çalışmada 58 gözlemden oluşan 1990 yılına ait veri seti ele alınmıştır. Bu veri setine EKK, Ridge ve GME tahmin edicileri kullanılarak parametre tahminleri elde edilmiştir. Analizler SPSS ve Gauss 10 paket programı kullanılarak gerçekleştirilmiştir. Çalışmada kullanılan değişkenler Tablo 1'de ifade edildiği gibidir. Bu veri kümesine ait özet istatistikler Tablo 2'de, değişkenler arası korelasyonlar Tablo 3'te sunulmaktadır.

Tablo 1. Çalışmada Kullanılan Değişkenler

Değişken	**Tanımı**
Y	Geliri yoksulluk düzeyinin altında kalan ailelerin oranı
X_2	Ortalama hanehalkı büyüklüğü

X_3	İşsizlik oranı
X_4	Lise diplomasına sahip 25 yaş ve üzeri kişilerin yüzdesi
X_5	En az 4 yıllık üniversite diplomasına sahip 25 yaş ve üzeri kişilerin yüzdesi
X_6	Medyan gelir
X_7	Şehirleşme oranı
X_8	Şehirleşme oranının karesi

Çalışmada kullanılan model (12) nolu eşitlikte verildiği gibidir:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + u \qquad (12)$$

Vazquez, Panudulkitti ve Timofeev (2009) şehirleşmenin yoksulluk eğrisi üzerindeki etkisinin U biçiminde olduğunu belirtmiştir. Buna göre başlangıçta şehirleşme oranının artması yoksulluğu azaltırken, belirli bir noktadan sonra şehirleşmedeki artış yoksulluk seviyesinin artmasına neden olmaktadır. Bu nedenle yoksulluk seviyesi incelenirken şehirleşme oranının karesinin de modele eklenmesi uygun görülmüştür.

(12) nolu modelde ortalama hanehalkı büyüklüğü (X_2) ve işsizlik oranı (X_3) arttıkça yoksulluğun artması beklenmektedir. Öte yandan lise okullaşma oranı (X_4), üniversite okullaşma oranı (X_5) ve medyan gelir (X_6) değişkenlerinin yoksulluk oranı üzerinde negatif bir etkisi beklenmektedir. Şehirleşme ve yoksulluk arasındaki U şeklindeki ilişki nedeniyle şehirleşme oranının (X_7) yoksulluk üzerinde negatif, karesininse (X_8) pozitif etkisinin olması beklenmektedir. Bu durumda parametrelerinin işaretleri hakkındaki önsel bilginin $\beta_2, \beta_3, \beta_8 > 0$ ve $\beta_4, \beta_5, \beta_6, \beta_7 < 0$ şeklinde olduğu söylenebilir.

Tablo 2. Değişkenlere ait özet istatistikler

Değişken	**Ortalama**	**Standart Sapma**
Y	9,9034	3,95545
X_2	2,6909	0,24054
X_3	9,9517	3,93323
X_4	57,5655	6,21601
X_5	18,7914	7,69945
X_6	35,3377	8,26424
X_7	34,1017	19,48169
X_8	1535,9202	1635,35641

Tablo 3. Örneklem korelasyon matrisi

	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_2	1	0,485**	-0,508*	-0,300 *	-0,035	0,350**	0,236
X_3	0,485**	1	-0,109	-0,757**	-0,714**	0,110	0,054

X_4	-0,508**	-0,109	1	-0,358**	-0,280*	0,211	0,260*
X_5	-0,300	-0,757**	-0,358**	1	0,848**	-0,358**	-0,299*
X_6	-0,035	-0,714**	-0,280*	0,848**	1	-0,084	-0,045
X_7	0,350**	0,110	0,211	-0,358**	-0,084	1	0,951**
X_8	0,236	0,054	0,260*	-0,299*	-0,045	0,951**	1

* : p < 0,055, ** : p < 0,01

Tablo 3 incelendiğinde açıklayıcı değişkenler arasındaki ikili korelasyonların oldukça yüksek oldukları görülmektedir. Bu durumda değişkenler arasında yüksek korelasyonlar olduğu ve kurulacak olan regresyon modelinin çoklu iç ilişkiden etkilenebileceği söylenebilir.

(12) nolu modelin parametreleri EKK yöntemiyle tahmin edildiğinde elde edilen sonuçlar Tablo 4'teki gibidir.

Tablo 4. EKK Tahminleri

Değişken	**Tahmin**	**Standart Hata**
Sabit	16,852	8,518
X_2	6,509	1,941
X_3	-0,028	0,121
X_4	-0,121	0,068
X_5	0,170	0,098
X_6	-0,544	0,071
X_7	-0,055	0,043
X_8	0,000	0,000

Modelin belirleyicilik katsayısı $R^2 = 0,916$ olup, parametrelere ilişkin t değerleri ve modelin F değeri istatistiksel olarak anlamlıdır. EKK tahminlerine bakıldığında parametre tahminlerinden $\hat{\beta}_2$'nın pozitif işaretli olması beklenirken negatif işaretli olduğu ve $\hat{\beta}_5$'nın negatif işaretli olması gerekirken pozitif işaretli olduğu, diğer parametrelerin ise beklenen işaretleri taşıdığı görülmektedir. Bu durumun çoklu iç ilişkiden kaynaklandığı düşünülmektedir. Çoklu iç ilişkinin derecesini ölçmek için kullanılabilecek bir diğer ölçüt Belsley, Kuh ve Welsch (1980) tarafından önerilen $\kappa = \sqrt{\lambda_{\max} / \lambda_{\min}}$ şeklindeki koşul sayısıdır. Tablo 5'te verilen özdeğerler kullanıldığında koşul sayısı için elde edilen 123,760 değeri oldukça yüksektir. Tüm bu göstergeler ilgili veri seti için yüksek düzeyde çoklu iç ilişki olduğu ve EKK tahminlerinin bu durumdan olumsuz yönde etkilediği anlamına gelmektedir.

Tablo 5. $X'X$ matrisi için özdeğerler ve koşul indisleri

Boyut	Özdeğer	Koşul İndisleri
1	7,019	1,000
2	0,667	3,245
3	0,255	5,247
4	0,029	15,504
5	0,017	20,045
6	0,009	28,709
7	0,004	43,081
8	0,000	123,760

İncelenen modeldeki çoklu iç ilişki problemini çözmek için Hoerl ve Kennard (1970) tarafından tanımlanan (13) numaralı eşitlikteki Ridge regresyon tahmin edicisi kullanılabilir:

$$\hat{\boldsymbol{\beta}}_{\mathbf{k}} = \left(\mathbf{X'X} + k\mathbf{I}\right)^{-1} \mathbf{X'Y},\ (k > 0) \tag{13}$$

Burada k yanlılık parametresi olup k'nın optimum seçimi için literatürde çeşitli yöntemler yer almaktadır. Hoerl, Kennard ve Baldwin (1975);

$$k_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \tag{14}$$

eşitliğini önermektedir. Burada p parametre sayısı, $\hat{\sigma}^2$ ise EKK ile elde edilen σ^2 tahminidir. Hoerl, Kennard ve Baldwin (1975) yaptıkları simülasyon çalışmasında k için (14) eşitliğini kullandıklarında Ridge tahmin edicinin EKK tahmin ediciden daha küçük HKO değerlerine sahip olduğunu göstermişlerdir. Lawless ve Wang (1976) ise k'nın seçimi için

$$k_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2} \tag{15}$$

eşitliğini önermişlerdir. Burada $\hat{\alpha}_i$ kanonik model parametrelerinin EKK tahminleri, λ_i ise $X'X$ matrisinin büyükten küçüğe sıralanmış özdeğerleridir.

Çalışmada Ridge tahmin edici için optimum k'nın seçiminde (14) ve (15)'de verilen yöntemler kullanılmıştır. Hoerl, Kennard ve Baldwin tarafından önerilen k parametresi ile yapılan Ridge tahminleri Ridge HKB, Lawless ve Wang tarafından önerilen k parametresi ile yapılan Ridge tahmini Ridge LW olarak belirtilmiş olup sonuçlar Tablo 6'da sunulmuştur.

Tablo 6. Ridge tahminleri

Değişken	Ridge HKB	Ridge LW
Sabit	6,98689	15,55809
X_2	7,99289	6,72847
X_3	0,03589	-0,02173
X_4	-0,04417	-0,11078
X_5	0,25013	0,18049
X_6	-0,56337	-0,54735
X_7	-0,05272	-0,05501
X_8	0,00038	0,00043

Tablo 6'da verilen sonuçlar incelendiğinde $\hat{\beta}_5$ hariç tüm Ridge tahminlerinin, $\hat{\beta}_3$ hariç tüm Ridge LW tahminlerinin beklenen işarete sahip olduğu görülmektedir. Zaten Ridge tahmin edici işaret kısıtı olmadan kullanılan bir tahmin edici olduğundan elde edilen tüm tahminlerin beklenen işaretlere sahip olması garanti edilemez. Elde edilen sonuç bu durumu destekler niteliktedir.

Çoklu iç ilişki problemini çözmesi beklenen Ridge tahmin edici için, elde edilen tahminlerin beklenen işaretlere uymadığı görülmektedir. Bu nedenle alternatif bir tahmin yöntemi olan GME tahmin edici kullanılarak parametre tahminleri yapılmıştır. İlk olarak modeldeki parametreler üzerine herhangi bir kısıt konmaksızın parametre tahminleri GME ile elde edilmiştir. Ardından parametreler üzerine işaret ve eşitsizlik kısıtları konularak GME tahminleri yeniden hesaplanmıştır.

Kısıt olmaksızın seçilen parametre destekleri Tablo 7'de ve elde edilen GME tahminleri Tablo 8'de verilmiştir.

Tablo 7. Parametre Destek Vektörleri (Kısıtsız)

Değişken	Parametre Desteği	Önsel Ortalama
Sabit	$z_1 = \{-50, -25, 0, 25, 50\}$	0
X_2	$z_2 = \{-10, -5, 0, 5, 10\}$	0
X_3	$z_3 = \{-2, -1, 0, 1, 2\}$	0
X_4	$z_3 = \{-2, -1, 0, 1, 2\}$	0
X_5	$z_3 = \{-2, -1, 0, 1, 2\}$	0
X_6	$z_2 = \{-10, -5, 0, 5, 10\}$	0
X_7	$z_3 = \{-2, -1, 0, 1, 2\}$	0
X_8	$z_3 = \{-2, -1, 0, 1, 2\}$	0

(12) ile ifade edilen doğrusal regresyon modelinin GME tahminleri incelendiğinde $\hat{\beta}_5, \hat{\beta}_7$ ve $\hat{\beta}_8$'in hala beklenen işaretleri taşımadığı görülmektedir.

Tablo 8. GME Tahminleri (Kısıtsız)

		%95'lik Güven Aralığı	
Değişken	**GME Tahmini**	**Alt Sınır**	**Üst Sınır**
Sabit	17,84319	9,77837	23,39601
X_2	4,20016	2,10547	5,24626
X_3	0,12681	-0,04028	0,34009
X_4	-0,11055	-0,18785	-0,03354
X_5	0,16608	-0,00447	0,31325
X_6	-0,47657	-0,61240	-0,29391
X_7	-0,01684	-0,11596	0,13016
X_8	0,00007	-0,00178	0,00111

Parametrelerle ilgili önsel bilgiler kullanılarak (12) nolu GLM için GME tahmini yeniden yapılmıştır. Bunun için her bir katsayının beklenen işareti düşünülerek işaret kısıtları ve değişkenlerin tanımları düşünülerek büyüklük kısıtları konulmuştur.

(12) nolu model için Campbell ve Hill (2006) tarafından belirtildiği üzere X_2, X_3 değişkenlerine ait katsayıların pozitif; X_4, X_5, X_6 değişkenlerine ait katsayıların negatif işaretli olması beklenmektedir. Ayrıca Vazquez, Panudulkitti ve Timofeev (2009) tarafından belirtildiği üzere şehirleşme oranının yoksulluk oranına etkisi negatif, şehirleşme oranının karesinin etkisi ise pozitiftir. İşaret kısıtlarına ek olarak yine Campbell ve Hill (2006) tarafından uygulanan büyüklük kısıtı da modele dahil edilmiştir. Buna göre üniversite okullaşma oranının yoksulluk oranına etkisinin, lise okullaşma oranının yoksulluk oranına etkisinden daha düşük olması beklenmektedir**.** (12) numaralı model için bu kısıt matematiksel olarak $\beta_5 < \beta_4 < 0$ şeklinde ifade edilebilir. Bu durumda parametreler için;

$$\begin{bmatrix} \beta_4 \\ \beta_5 \end{bmatrix} = Z^* \begin{bmatrix} p_4 \\ p_5 \end{bmatrix} = \begin{bmatrix} z_4' & 0 \\ z_4' & z_5' \end{bmatrix} \begin{bmatrix} p_4 \\ p_5 \end{bmatrix} \tag{16}$$

eşitliği yazılabilir. Burada $Z : 2 \times 2M$ matrisi β_4, β_5 parametreleri ile p_4, p_5 bilinmeyen olasılıkları için destek noktaları matrisleridir.

İşaret ve büyüklük kısıtları için parametre destekleri Tablo 9'da, elde edilen GME tahminleri Tablo 10'da verilmiştir. Tablo 10'da görüldüğü üzere doğrusal regresyon modelinin GME tahminlerinin tamamı beklenen işaretlerle uyumludur. Ayrıca elde edilen güven aralıkları sıfırı içermediğinden tüm parametrelerin %95 güven düzeyinde istatistiksel olarak anlamlı olduğu söylenebilir.

Tablo 9. Parametre Destek Vektörleri (Kısıtlı)

Değişken	Parametre Desteği	Önsel Ortalama
Sabit	$z_1 = \{-50, -25, 0, 25, 50\}$	0
X_2	$z_2 = \{0, 2.5, 5, 7.5, 10\}$	5
X_3	$z_3 = \{0, 0.5, 1, 1.5, 2\}$	1
X_4	$z_4 = \{-2, -1.5, -1, -0.5, 0\}$	-1
X_5	$z_5 = \{-2, -1.5, -1, -0.5, 0\}$	$\hat{\beta}_4 - 1$
X_6	$z_6 = \{-10, -7.5, -5, -2.5, 0\}$	-5
X_7	$z_5 = \{-2, -1.5, -1, -0.5, 0\}$	0
X_8	$z_3 = \{0, 0.5, 1, 1.5, 2\}$	0

Tablo 10. GME Tahminleri (Kısıtlı)

		%95'lik Güven Aralığı	
Değişken	**GME Tahmini**	**Alt Sınır**	**Üst Sınır**
Sabit	14,80827	9,20682	20,55863
X_2	4,74872	4,03248	5,42809
X_3	0,20454	0,11729	0,35228
X_4	-0,09599	-0,17303	-0,04413
X_5	-0,13533	-0,25004	-0,07917
X_6	-0,23753	-0,32326	-0,10567
X_7	-0,14990	-0,34348	-0,07401
X_8	0,00118	0,00043	0,00354

Elde edilen sonuçlar beklenen işaret bakımından GME tahmin edicinin hem EKK hem de Ridge tahmin ediciden daha iyi sonuçlar verdiğini göstermektedir. Öte yandan tahmin edicileri karşılaştırmada sadece işaretlerin tutarlılığına bakmak yerine, tahmin edicilerin HKO değerlerini kullanmak da gerekir. GME tahmin edici için HKO değeri kapalı formda ifade edilemediğinden, bu tahmin edicinin HKO değeri Efron (1979) tarafından önerilen bootstrap yöntemi ile tahmin edilebilir. Bu yöntemin aşamaları şu şekildedir: Parametrelerin herhangi bir tahmini $\tilde{\beta}$ olmak üzere, çoklu doğrusal regresyon modeli için $y = X\beta + e$ denkleminden rezidüler $e = y - X\tilde{\beta}$ eşitliği ile hesaplanır. Daha sonra j-inci bootstrap denemesinde e'nin elemanlarından n tanesi yerine konularak rasgele çekilir. Çekilen rezidülerin oluşturduğu vektör e_j^*, X ve $\tilde{\beta}$ kullanılarak j-inci deneme için bağımlı değişken vektörü y_j^* hesaplanır. Böylece bağımlı değişkene ait yeni bir örneklem elde edilmiş olur. Bu aşamada y_j^* ve

X kullanılarak j-inci deneme için β 'nın tahmini olan $\tilde{\beta}_j$ hesaplanır. Bu işlem yeterince tekrar edildiğinde $\tilde{\beta}$ 'nın örnekleme dağılımının bir tahmini elde edilmiş olur. Bu ampirik örnekleme dağılımı kullanılarak HKO değerleri tahmin edilebilir (Çabuk ve Akdeniz, 2007 & Akdeniz vd., 2011).

EKK, Ridge ve GME tahmin edicilerin HKO değerlerini elde etmek için bu işlem 400 defa tekrarlanmış ve böylece her tahmin edici için HKO tahminleri elde edilmiştir. Hesaplanan HKO değerleri Tablo 11'de sunulmuştur. Kısıt olmadan yapılan GME tahmini GME1, kısıt konularak yapılan GME tahmini GME2 olarak adlandırılmıştır.

Tablo 11. Bootstrap Sonuçları

Tahmin Yöntemi	k Değeri	HKO Tahminleri
EKK	—	72,07552
Ridge HKB	0,07040	191,05802
Ridge LW	0,00357	61,36033
GME1	—	19,911157
GME2	—	16,01474

Tablo 11 incelendiğinde;

$$HKO_{GME2} < HKO_{GME1} < HKO_{RİDGELW} < HKO_{EKK} < HKO_{RİDGEHKB}$$

sonucu elde edilmiştir. Buna göre kuvvetli çoklu iç ilişkiye sahip bu veri seti için en iyi tahmin edicinin GME tahmin edici olduğu görülmektedir. Hem kısıtsız hem de kısıtlı GME tahmin edicilerin HKO değerleri, diğer tahmin edicilerin HKO değerlerine göre oldukça düşüktür. Bu tahmin edicileri EKK ve Ridge tahmin ediciler izlemektedir. Kısıtsız GME tahmin edicinin HKO değeri en küçük olmasına rağmen, kısıtlı GME ile elde edilen parametre değerlerinin işaretleri beklenen işaretlerle uyumludur. Bu durumda kısıtlı GME'nin kısıtsız GME'ye tercih edilebileceği söylenebilir.

TARTIŞMA VE SONUÇ

Lineer regresyon modellerinde açıklayıcı değişkenler arasında önemli derecede çoklu iç ilişki problemi olduğu zaman parametre tahminlerinde kararsızlıklar oluşabilmektedir. Ancak ekonometrik çalışmalarda kararsız tahminler değil kararlı tahminler elde edilmek istendiğinden, çoklu iç ilişki durumunda yanlı tahmin edicilere başvurulabilir. Bu çalışmada, çoklu iç ilişki probleminden etkilenen bir veri seti için kararlı tahminler elde edilmesi ve çeşitli alternatif tahmin edicilerin karşılaştırılması amaçlamıştır.

Bu amaç doğrultusunda incelenen yoksulluk verisi için EKK tahminleri elde edilmiş ve tahminlerin işaretlerinin önsel bilgilerle uyuşmadığı görülmüştür. Bu nedenle model parametrelerinin tahmininde yanlı tahmin edicilerden Ridge ve GME yöntemlerine başvurulmuş, bu tahmin ediciler HKO ölçütüne göre karşılaştırılmıştır. GME tahminleri elde edilirken işaret kısıtları içermeyen durumun yanı sıra teoriyle uyumlu işaret ve büyüklük kısıtlarını içeren durum da göz önüne alınmıştır.

Elde edilen sonuçlar incelendiğinde önsel bilgiler doğrultusunda beklenen işaretlerin yalnızca kısıtlı GME için elde edildiği görülmüştür. GME tahmin edicisinin bootstrap yöntemi ile tahmin edilen HKO değeri, EKK ve Ridge tahmin edicilerin HKO değerlerinden daha küçük olarak elde edilmiştir. Bu doğrultuda çoklu iç ilişkiden etkilenen Kaliforniya yoksulluk veri kümesi için GME tahmin edicinin Ridge ve EKK tahmin edicilerden daha iyi olduğu sonucuna varılmıştır.

Kaynaklar

Akdeniz, F., Çabuk, A. ve Güler, H., 2011. Generalized Maximum Entropy Estimators: Applications to the Portland Cement Dataset. The Open Statistics and Probability Journal, 3, 13-20

Belsley, D. A., Kuh, E., & Welsch, R. E., 1980. Regression diagnostics, New York, Wiley.

Campbell, R. C., & Hill, C. R. (2001). Maximum Entropy Estimation in Economic Models with Linear Inequality Restrictions. *Deparmental Working Paper*s. Department of Economics, Lousiana State University, 1-29.

Çabuk, A. ve Akdeniz, F., 2007. İçilişki ve genelleştirilmiş maksimum entropi tahmin edicileri, Journal of Statistical Research, 5(2), 1-19.

Efron, B. (1979). Bootstrap methods: another look at the jacknife. The Annals of Statistics, 7(1), 1-26.

Golan, A., Judge, G., & Miller, D. (1996). Maximum entropy econometrics: robust estimation with limited data. John Wiley & Sons, New York, USA.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

Hoerl, A. E., Kennard, R.W., & Baldwin, K. F. (1975). Ridge regression: some simulation. Communication in Statistics, 4, 105-123.

Jaynes, E. T. (1957). Information theory and statistical mechanics II. Physics Rewiev, 108(2), 171-190.

Lawless, J. F., & Wang, P. (1976). A simulation study of ridge and other regression estimators. Communications in Statistics-Theory and Methods, 14, 1589-1604.

Pukelsheim, F. (1994). The three sigma rule. The American Statistician, 48(2), 88-91.

Ramanathan, R. (2002). Introductory Econometrics With Applications. Fifth Edition, Harcourt College Publishers.

Shannon, C. E. (1948). A Mathematical Theory of Communication. The Bell System Technical Journal, 27(3), 379-423.

Vazquez, J. M., Panudulkitti, P., & Timofeev, A. (2009). Urbanization and the poverty level. International Studies Program Working Paper 9-14 (updated), Georgia State University.