Süleyman Demirel Üniversitesi
Fen Bilimleri Enstitüsü Dergisi
Cilt 26, Sayı 3, 552-559, 2022

Süleyman Demirel University
Journal of Natural and Applied Sciences
Volume 26, Issue 3, 552-559, 2022

DOI: 10.19113/sdufenbed.1119370

Bell Regresyon Modelinde Liu tipi Tahmin Edici

Melike IŞILAR[1], Yakup Murat BULUT*[2]

[1]Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, 26040, Eskişehir, Türkiye
[2]Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 26040, Eskişehir, Türkiye

(Alınış / Received: 20.05.2022, Kabul/Accepted: 28.06.2022, Online Yayınlanma/Published Online: 20.12.2022)

Anahtar Kelimeler
Bell Regresyon,
Sayım verileri,
Aşırı yayılım,
Çoklu iç ilişki,
Liu tipi tahmin edici

Öz: Bu çalışmada, sayım verilerini modellemek için kullanılan Poisson regresyon modeline alternatif olarak tanımlanan Bell regresyon modelinde çoklu iç ilişki olması durumunda kullanılan yanlı tahmin edicilere alternatif bir tahmin edici önerilmiştir. Bell regresyon modeli aşırı yayılım probleminin çözümü için kullanılan bir modeldir. Bell regresyon modelinin parametreleri genellikle en çok olabilirlik (EÇO) tahmin edicisi kullanılarak tahmin edilmektedir. Fakat, çoklu iç ilişki problemi olması durumunda EÇO tahmin edicisinin performansı düşmektedir. Bu sebeple, Bell Liu-tipi tahmin edicisi önerilmiştir. Önerilen Bell Liu tipi tahmin edicinin performansı Bell Ridge ve Bell Liu tahmin edicileri ile Monte Carlo simülasyon çalışması yardımıyla karşılaştırılmıştır. Ayrıca, simülasyon çalışmasına desteklemek için gerçek veri örneği verilmiştir.

Liu-type Estimator in the Bell Regression Model

Keywords
Bell regression,
Count Data,
Overdispersion,
Multicollinearity,
Liu type estimator

Abstract: This study proposes a new estimator used in the case of multicollinearity problems in the Bell regression model that is an alternative model for the Poisson regression model. The Bell regression model is used to solve the overdispersion problem. Generally, the maximum likelihood estimation (MLE) method is used to estimate the parameters of the Bell regression model. But, the performance of the MLE decreases when the multicollinearity problem occurs. Therefore, the Bell liu-type estimator is proposed. Monte Carlo simulation study is conducted to compare the performance of the proposed estimator with Bell Ridge and Bell Liu estimators. Additionally, the real data example is given to support the simulation study.

1. Giriş

Regresyon analizi bağımsız değişken ve bağımlı değişken arasındaki ilişkiyi matematiksel olarak modelleyerek gerekli çıkarımların yapılmasında yaygın olarak kullanılan istatistiksel analizdir. Regresyon analizinde bağımlı değişken her zaman sürekli değişkenlerden oluşmaz. Çoğu zaman kesikli değerler de alabilmektedir. Sayma verisi olarak adlandırılan bu veri tipleri ile uygulamada yaygın olarak karşılaşılmaktadır. Sayma verileri, kategorik veriler ve sürekli çarpık dağılım gösteren veriler için lineer regresyon modeli kullanmak doğru olmaz. Bu tür durumlarda genelleştirilmiş regresyon modellerinden veri yapısına uygun olan modelin kullanması doğru sonuçların elde edilebilmesi için uygun olmaktadır.

Sayma verilerinin modellenmesinde yaygın olarak Poisson regresyon modeli (PRM) kullanılmaktadır. Poisson dağılımın beklenen değer ve varyans eşit olarak elde edilmektedir. Bu nedenle PRM' nin varsayımlarından biri beklenen değer ve varyansın birbirine eşit olmasıdır. Uygulamada elde edilen verilerin çoğunda bu varsayım sağlanamamaktadır. Genellikle varyans değeri beklenen değerden büyük çıkmaktadır. Bu durum ise aşırı yayılım varsayımı olarak adlandırılmaktadır. Aşırı yayılıma sahip sayma verilerinin modellenmesinde PRM' nin kullanılması elverişli olmamaktadır [1]. Aşırı yayılım gösteren sayma verilerinin modellenmesinde karma dağılımlara bağlı ya da alternatif dağılımlara ait modellerinin kullanılması önerilmiştir. Negatif binom regresyon modeli (NBRM) bunlardan biridir. Karma bir dağılım olan negatif binom dağılımı aşırı yayılım probleminin bir çözümüdür. PRM' nin yaygın olarak tercih edilme sebeplerinden birisi Poisson dağılımın tek parametresi olmasından dolayı işlem kolaylığı sağlamasıdır. Bu nedenle sayma verilerinin modellenmesinde Poisson ve negatif binom dağılımına alternatif olarak Castellares vd. tarafından Bell dağımı (BD) ve bu dağılım temelinde Bell

*İlgili yazar: ymbulut@ogu.edu.tr

regresyon modeli (BRM) önerilmiştir [2]. PRM' nin tek parametresi olması ve NBRM nin aşırı yayılımlı sayma verilerini modelleme avantajlarının ikisine de sahiptir. Bu nedenle son yıllarda yaygın olarak bu modellere alternatif olarak BRM kullanılmaya başlanmıştır. BRM genelleştirilmiş lineer modellerden (GLM) biridir. GLM parametrelerin tahmin edilmesinde yaygın olarak En Çok Olabilirlik (EÇO) tahmin edicisi kullanılmaktadır. Bunun nedeni regresyon analizinin varsayımlarının sağlanması durumunda EÇO tahmin edicisi en iyi tahmin edici olma özelliğine sahiptir.

Regresyon analizinin genel olarak varsayımlarından birisi bağımsız değişkenlerin lineer olarak bağımsız olmasıdır. Doğadan elde edilen veri setlerinin genel olarak çoğunda bağımsız değişkenler arasında yüksek bir lineer ilişki bulunmaktadır. Bu durum çoklu iç ilişki problemi olarak adlandırılmaktadır. Çoklu iç ilişki problemi olması durumunda EÇO tahmin edicisinin varyansı artmaktadır. Bu durumda parametreye ait güven aralığı genişler, hipotez testlerinde yanlış karar verilebilmektedir. Parametrelerin tahminleri ise beklenen durum ile ters işaretli çıkabilmektedir. Kısacası çoklu iç ilişki problemi olması durumunda EÇO tahmin edicisi güvenilir olmayan sonuçlar vermektedir [3].

Çoklu iç ilişki probleminin bazı araştırmacılar tarafından küçük veri problemi olarak değerlendirilmektedir [4]. Bu nedenle bu problemin çözümü olarak ek veri toplamak önerilmiştir [5]. Fakat araştırmacılar için ek veri toplamak maliyet ve zaman açısından problem yaratabilmektedir. Ayrıca ek veri toplanması durumunda da elde edilen yeni gözlemlerde aynı ilişki yapısında olabilmektedir. Bu durumda ek veri toplamak çoklu iç ilişki problemini çözmemektedir. Bu problemin çözümü için alternatif olarak yüksek ilişki olan değişkenlerin modelden çıkarılması önerilmiştir. Bu durum ise modelden çıkartılan değişkenden elde edilebilecek bilgi kaybına neden olmaktadır. Ayrıca Lipovetsky ve Conklin ise çalışmasında iki değişken arasında ne kadar yüksek ilişki bulunursa bulunsun bu değişkenlerin birbirinin ikamesi olamayacağını söylemektedir [6]. Elde mevcut veri yapısını koruyarak çoklu iç ilişki probleminin çözüm yöntemi için yanlı tahmin ediciler üzerine çalışmalar yapılmaktadır. EÇO tahmin edicisi yansız bir tahmin edicidir. Bu problem söz konusu olduğunda yansız olan EÇO tahmin edicisinin varyansı artmaktadır. Parametrelerin tahmini için çoklu iç ilişki olması durumunda yanlı fakat varyansı daha küçük alternatif tahmin ediciler mevcuttur. Bu yanlı tahmin edicilerin hata kareler ortalaması (MSE) değeri EÇO tahmin edicisinin MSE değerinden daha küçük elde edilmektedir. Bu amaçla literatürde farklı regresyon modelleri için birçok yanlı tahmin ediciler önerilmiştir.

Literatürde yaygın olarak kullanılan yanlı tahmin edicilerden biri Hoerl ve Kennard tarafından önerilen Ridge tahmin edicisidir [7]. İlk olarak lineer model için önerilen bu tahmin edici daha sonra pek çok GLM üyesi modeller için tanımlanmıştır. PRM için Månsson ve Shukur, NBRM için Månsson, BRM için ise Amin vd. tarafından tanımlanmıştır [8-10].

Ridge tahmin edicisi yanlılık parametresine bağlı olan bir tahmin edicidir. Çoklu iç ilişki durumunda $X'X$ matrisinin tersinin alınamaması problemini bu matrise bir k sabiti ekleyerek çözmektedir. Kısacası eklenen yan değeri ile varyans değerini küçültmektedir. Kısacası yan değeri arttıkça varyans değeri azalmaktadır. Bu tahmin edici, varyans değerindeki azalış yanın karesi değerindeki artıştan fazla olduğu sürece EÇO tahmin edicisinden daha iyi sonuç vermektedir.

Ridge tahmin edicisinde k parametresinin tahmin edilmesi problemi ile karşılaşılmaktadır. k parametresi tahmin edicinin lineer bir fonksiyonu olarak elde edilememektedir. Bu nedenle günümüzde hala pek çok k parametresi için tahmin ediciler önerilmektedir. Tahmin edicinin performansı veriye bağlı olarak parametrenin tahmin edicisine göre değişmektedir. Bu nedenle Liu tarafından parametresi tahmin edicinin lineer bir fonksiyonu olarak elde edilebilen yeni bir tahmin edici önerilmiştir [11]. Önerilen bu tahmin edici Akdeniz ve Kaçıranlar tarafından Liu (Lineer Unified) tahmin edicisi (LE) olarak adlandırılmıştır [12]. LE d büzülme parametresine bağlı bir tahmin edicidir. Parametre uzayını büzerek çoklu iç ilişki problemini çözmektedir. NBRM için LE Månsson tarafından önerilmiştir [13]. Kurtoğlu ve Özkale GLM için genel olarak Liu tahmin edicisini tanımlamışlardır [14]. LE' nin RE' ye avantajı parametrenin tahmin edicinin doğrusal bir fonksiyonu olarak elde edilebilmesidir. Bu amaçla Liu tarafından d büzülme parametresinin tahmini için MSE' yi minimum yapan en uygun tahmin edici önerilmiştir [11]. Daha sonra RE' nin k parametresinin tahmin edicilerine benzer yapıda en uygun d parametresine alternatif tahmin ediciler önerilmiştir. Qasim vd. önce Gamma regresyon modelinde alternatif d parametreleri önermiş ve daha sonra bu tahmin edicilerin performanslarını PRM' de karşılaştırmıştır [15-16]. Majid vd. ise çalışmasında farklı d parametresinin performanslarını BRM için karşılaştırmışlardır [17].

RE ve LE' nin farklı avantajları ve dezavantajları vardır. RE çoklu iç ilişki durumunda $(X'X)$ matrisinin tersinin alınamamasından dolayı bu matrisin köşegen elemanlarına k yanlılık değeri ekleyerek matrisin tersinin alınabilmesine olanak sağlamaktadır. LE is parametre uzayının büzerek tahmin edicinin varyansının küçültülmesine olanak sağlamaktadır. RE aslında yan miktarını arttırarak varyansı azaltan bir tahmin edicidir. Kısacası çoklu iç ilişkinin şiddeti ne kadar yüksek ise o kadar yanlı bir tahmin değeri ile bu problem çözülebilmektedir. Her ne kadar varyansı büyük bir tahmin edici yerine yanlı fakat MSE değeri daha küçük bir tahmin edici bu problemin çözümü olsa da uygulama da araştırmacılar yan değerinin yüksek olmasını istememektedir. Bu nedenle Liu lineer regresyon modeli için RE ve LE tahmin edicilerinin avantajlarını birleştirerek Liu tipi tahmin edicisini (LTE) önermiştir [18]. PRM ve NBRM için LTE ise sırasıyla Asar vd. (2015) ve Asar (2018) tarafından önerilmiştir [19-20].

Bu çalışmanın amacı BRM için çoklu iç ilişki probleminin çözümü için alternatif bir tahmin edici önermektir. Bell Ridge tahmin edicisi (BRE) ve Bell Liu tahmin edicisi (BLE) tahmin edicilerinden daha etkin sonuçlar veren Bell Liu tipi tahmin edicisi (BLTE) önerilmiştir.

Bu çalışma kapsamında 2. Bölümde Bell Regresyon modeli anlatılmıştır. BRE' de parametreleri tahmin etmek için literatürde var EÇO, BRE ve BLE verilmiştir. Bu tahmin edicilere alternatif önerdiğimiz Liu tipi tahmin edicisi tanımlanmıştır. Önerilen tahmin edicinin literatürde var olan tahmin edicilere göre performansı teorik olarak karşılaştırılmıştır. 3. Bölümde ele alınan tahmin edicilerin performansları simülasyon çalışması ve gerçek veri örneği üzerinde MSE kriterine göre karşılaştırılmıştır. 4. Bölümde ise yapılan çalışmada elde edilen sonuçlar verilmiştir.

2. Materyal ve Metot

Bu bölümde öncelikle BRM ve parametre tahmini için EÇO tahmin edicisi incelenmiştir. Daha sonra BRE ve BLE tahmin edicileri incelenerek, BLTE önerilmiştir. BLTE nin performansı EÇO, BRE ve BLE ile teorik olarak karşılaştırılmıştır.

2.1. Bell regresyon modeli ve EÇO tahmin edicisi

BD'nin olasılık yoğunluk fonksiyonu $\nu > 0$ olmak üzere,

$$\Pr(Y = y) = \frac{\nu^y e^\nu B_y}{y!}, \qquad y = 0,1,2,\dots \tag{1}$$

şeklindedir. Burada $\nu > 0$ ve B_y , Bell (1934) tarafından tanımlanan Bell sayısıdır. Bell sayısı Eşitlik (2)' de verildiği gibi hesaplanmaktadır [21].

$$B_n = \frac{1}{n}\sum_{i=1}^{\infty}\frac{k^n}{k!} \tag{2}$$

BD' nin beklenen değer ve varyansı ise sırasıyla Eşitlik (3) ve (4)' te verilmiştir.

$$E(y) = \nu e^\nu \tag{3}$$
$$Var(y) = \nu(1+\nu)e^\nu \tag{4}$$

Regresyon modellerinde bağımlı değişkenin ortalaması modellendiği için BRM tanımlanırken BD yeniden parametrize edilerek Eşitlik (5) kullanılmaktadır.

$$\Pr(Y = y) = \exp\{1 - e^{W_0(\theta)}\}\frac{W_0(\theta)\theta B_y}{y!} \tag{5}$$

Burada $\theta = \nu e^\nu$ ve $W_0(.)$, Lambert fonksiyonu olmak üzere $\nu = W_0(\theta)$' dır.

GLM' de parametre tahminleri için yaygın olarak EÇO tahmin edicisi kullanılmaktadır. Bu amaçla BD' nin log-olabilirlik fonksiyonu Eşitlik (6)' da verildiği gibidir.

$$\ell = \sum_{i=1}^{n} y_i log\theta_i + \sum_{i=1}^{n}\left(1 - e^{\theta_i}\right) + logB_{y_i} - \log\left(\prod_{i=1}^{n} y!\right) \tag{6}$$

Eşitlik (6)' da verilen denklemin maksimizasyonu için β' ya göre türevinin alınıp sıfıra eşitlenmiş ve Eşitlik (7) elde edilmiştir.

$$\frac{\partial}{\partial\beta}\ell = \sum_{i=1}^{n}[x_i(1 + \exp\{x_i'\beta\})(y_i - \mu_i)] = 0 \tag{7}$$

Log-olabilirlik fonksiyonunun sıfıra eşitlenmesi ile elde edilen β kapalı bir formda elde edilememektedir. Bu nedenle β ' nın çözümü için iteratif yöntemler kullanılmaktadır. GLM' de EÇO tahmin edicisi için yaygın olarak ağırlıklandırılmış en küçük kareler (IRWLS) yöntemi kullanılmaktadır. β parametresinin EÇO tahmin edicisi Eşitlik (8)' de verilmiştir.

$$\hat{\beta}_{EÇO} = \left(X'\widehat{W}X\right)^{-1}X\widehat{W}\hat{z} \tag{8}$$

Burada $\widehat{W} = diag\{w_1, w_2, \dots, w_3\}$ ve $\widehat{V} = diag\{v_1, v_2, \dots, v_n\}$ olmak üzere $w_i = \left(\frac{d\mu_i}{d\eta}\right)^2 / v_i$ ve $v_i = \mu_i[1 + W_0(\mu_i)]$ olarak hesaplanmaktadır.

Yanlı tahmin edicilerde yanlılık parametrelerinin hesaplanmasında kanonik formlar kullanılmaktadır. β parametresini kanonik formda yazabilmek için $\alpha = Q'\beta$ şeklinde dönüşüm uygulanmaktadır. Burada Q, herhangi bir ortogonal matris olarak alınabilir. Literatürdeki çalışmalar incelendiğinde genel olarak Q matrisi $X'\widehat{W}X$ matrisinin öz vektörlerinden oluşan matris olarak alınmaktadır. Kanonik dönüşüm yapıldığında EÇO tahmin edicisinin matris hata kareler ortalaması (MMSE) ve skaler hata kareler ortalaması (MSE) denklemleri sırasıyla Eşitlik (9) ve (10)' da verilmiştir.

$$MMSE\left(\hat{\beta}_{EÇO}\right) = T^{-1} \tag{9}$$

$$MSE\left(\hat{\beta}_{EÇO}\right) = tr\{Q\Lambda^{-1}Q^{-1}\} = \sum_{j=1}^{l}\frac{1}{\lambda_j} \tag{10}$$

Burada $\mathrm{T} = X'\widehat{W}X$ ve l toplam parametre sayısı olmak üzere λ_j, $X'\widehat{W}X$ matrisinin j. özdeğeridir.

2.2. Ridge ve Liu tahmin edicileri

Hoerl ve Kennard tarafından lineer model için önerilen RE daha sonra pek çok GLM modeli için özel olarak tanımlanmıştır. Amin vd. tarafından tanımlanan Bell Ridge tahmin edicisi (BRE)

$$\hat{\beta}(k) = \left(X'\widehat{W}X + kI\right)^{-1} X'\widehat{W}X\hat{\beta}_{EÇO} \qquad (11)$$

şeklindedir [10]. Burada $k > 0$ olmak üzere yanlılık parametresidir. k ' nın tahmini için birçok farklı çalışmada farklı tahmin ediciler mevcuttur. Simülasyon çalışması ve gerçek veri uygulamasında kullanılan tahmin edici Eşitlik (13)' te verilmiştir.

$$\hat{k} = \frac{1}{\hat{\alpha}'\hat{\alpha}} \qquad (12)$$

BRE' nin MMSE ve MSE denklemleri Eşitlik (13) ve (14)' te verilmiştir.

$$MMSE\left(\hat{\beta}_{BRE}\right) = T_k^{-1}TT_k^{-1} + k^2T_k^{-1}\beta\beta'T_k^{-1} \qquad (13)$$

$$MSE\left(\hat{\beta}_{BRE}\right) = \sum_{j=1}^{l} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + k^2 \sum_{j=1}^{l} \frac{\alpha_j^2}{\left(\lambda_j + k\right)^2} \qquad (14)$$

Burada $T_k = T + kI$ dır.

RE tahmin edicisine alternatif olarak önerilen LE' nin Kurtoğlu ve Özkale tarafından GLM için genel tanımlamasının BRM için gösterimi Eşitlik (15)'te verilmiştir [14].

$$\hat{\beta}(d) = \left(X'\widehat{W}X + I\right)^{-1}\left(X'\widehat{W}X + dI\right)\hat{\beta}_{EÇO} \qquad (15)$$

Burada $0 < d < 1$ olmak üzere büzülme parametresidir.

$$MMSE\left(\hat{\beta}_{BLE}\right) = T_1^{-1}T_{d^+}T^{-1}T_{d^+}T_1^{-1} + (1-d)^2T_1^{-1}\beta\beta'T_1^{-1} \qquad (16)$$

$$MSE\left(\hat{\beta}_{BLE}\right) = \sum_{j=1}^{l} \frac{\left(\lambda_j + d\right)^2}{\left(\lambda_j + 1\right)^2} + (1-d)^2 \sum_{j=1}^{l} \frac{\alpha_j^2}{\left(\lambda_j + 1\right)^2} \qquad (17)$$

Burada $T_1 = T + I$ ve $T_{d^+} = T + dI$ dır.

Liu tarafından LE için önerilen en uygun d tahmin edicisi için Eşitlik (17)' nin türevi alınarak sıfıra eşitlenmesi ile Eşitlik (18) elde edilmektedir [11].

$$\hat{d} = 1 - \frac{\sum_{j=1}^{l} \frac{1}{\lambda_j\left(\lambda_j+1\right)^2}}{\sum_{j=1}^{l} \frac{\hat{\alpha}_j^2}{\left(\lambda_j+1\right)^2}} \qquad (18)$$

2.3. Liu tipi tahmin edicisi

BRE ve BLE nin avantajlarını birleştirerek bu tahmin edicilerden daha etkin sonuçlar elde etmek amacıyla Bell Liu tipi tahmin edici (BLTE) önerilmektedir. Genel formu ile BLTE tahmin edicisi Eşitlik (19)' da verilmiştir.

$$\hat{\beta}^\star(k,d) = \left(X'\widehat{W}X + kI\right)^{-1}\left(X'\widehat{W}\hat{z} - d\hat{\beta}^\star\right) \qquad (19)$$

Burada $\hat{\beta}^\star$, β nın herhangi bir tahmin edicisi olarak alınabilir. Genel olarak $\hat{\beta}^\star$ lineer regresyon modelinde en küçük kareler tahmin edicisi, GLM' de EÇO tahmin edicisi olarak alınmıştır. Bazı çalışmalarda $\hat{\beta}^\star$ için $\hat{\beta}(k)$ da alınmıştır [18,22]. Bu çalışmada $\hat{\beta}^\star = \hat{\beta}_{EÇO}$ olarak alınmıştır. Bu durumda BLTE Eşitlik (20)' de verildiği gibidir.

$$\hat{\beta}(k,d) = \left(X'\widehat{W}X + kI\right)^{-1}\left(X'\widehat{W}X - dI\right)\hat{\beta}_{EÇO} \qquad (20)$$

Burada $k > 0$, yanlılık parametresi ve $-\infty < d < \infty$, büzülme parametresidir.

Önerdiğimiz tahmin edicinin k ve d parametrelerine göre limit durumları ise aşağıdaki gibidir:

- $\lim_{d\to 0} \hat{\beta}(k,d) = \hat{\beta}(k)$
- $\lim_{\substack{d\to 0\\ k\to 0}} \hat{\beta}(k,d) = \hat{\beta}_{EÇO}$

BLTE' nin MMSE ve MSE denklemleri sırasıyla Eşitlik (21) ve (22)'de verilmiştir.

$$MMSE\left(\hat{\beta}_{BLE}\right) = T_k^{-1}T_{d^-}T^{-1}T_{d^-}T_k^{-1} + (d+k)^2T_k^{-1}\beta\beta'T_k^{-1} \qquad (21)$$

$$MSE\left(\hat{\beta}_{BLE}\right) = \sum_{j=1}^{l} \frac{\left(\lambda_j - d\right)^2}{\left(\lambda_j + k\right)^2} + (k+d)^2 \sum_{j=1}^{l} \frac{\alpha_j^2}{\left(\lambda_j + k\right)^2} \qquad (22)$$

Burada $T_{d^-} = T - dI$' dır.

BLTE'd k parametresi için BRE'de kullanılan herhangi bir tahmin edici kullanılabilir. Bu tahmin edicilere alternatif olarak Liu tarafından Eşitlik (23)' te verildiği gibi önerilen tahmin edici de kullanılabilir.

$$\hat{k}^\star = \frac{\lambda_{max} - 100\lambda_{min}}{99} \qquad (23)$$

BLTE'de d parametresi için BLE' ye benzer şekilde optimal tahmin edici kullanılmıştır. Bunun için Eşitlik (22)' de verilen MSE denkleminin d' ye göre türevi alınıp sıfıra eşitlenmiştir. Buradan elde edilen d parametresinin tahmin edicisi Eşitlik (24)' te verilmiştir.

$$\hat{d}^\star = \frac{\sum_{j=1}^{l} \frac{1-\hat{k}\hat{\alpha}_j^2}{\left(\lambda_j+\hat{k}\right)^2}}{\sum_{j=1}^{l} \frac{1+\lambda_j\hat{\alpha}_j^2}{\lambda_j\left(\lambda_j+\hat{k}\right)^2}} \qquad (24)$$

Bu çalışmada $\hat{\beta}_1(k,d)$ olarak Eşitlik (12)' de verilen, $\hat{\beta}_2(k,d)$ olarak Eşitlik (23)' te verilen k tahmin edicileri ve bunlara bağlı tahmin edilen d parametrelerinin kullanıldığı iki tane BLTE tahmin edicisi kullanılmıştır.

3. Bulgular

Bu bölümde farklı örnek hacmi, değişken sayısı ve korelasyon katsayılarına göre tahmin edicilerin performansları simülasyon çalışması ile karşılaştırılmıştır. Ayrıca teorik karşılaştırma ve simülasyon çalışmasındaki sonuçlar gerçek veri seti uygulaması ile desteklenmiştir.

Tablo 1. $p = 3$ için MSE değerleri

		$\hat{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\beta}}(\boldsymbol{k})$	$\hat{\boldsymbol{\beta}}(\boldsymbol{d})$	$\widehat{\boldsymbol{\beta_1}}(\boldsymbol{k},\boldsymbol{d})$	$\widehat{\boldsymbol{\beta_2}}(\boldsymbol{k},\boldsymbol{d})$
	0.90	6.8802	5.9904	5.5255	4.7262	5.0202
50	**0.95**	9.3248	7.6369	7.0127	6.2237	6.6565
	0.99	32.0634	23.6150	24.9113	20.2162	21.4814
	0.90	5.3233	4.8808	4.6213	3.7521	3.9991
100	**0.95**	6.5322	5.7394	5.3188	4.5758	4.9516
	0.99	16.8403	12.7250	12.5056	10.6863	11.4449
	0.90	5.0779	4.7782	4.6111	3.6701	3.9161
150	**0.95**	5.6629	5.1172	4.8216	3.9919	4.3529
	0.99	12.4667	9.8160	9.1608	8.2255	8.8671
	0.90	4.9806	4.7509	4.6095	3.6411	3.8545
200	**0.95**	5.3417	4.9408	4.7085	3.8515	4.2079
	0.99	10.4849	8.4414	7.6911	7.0666	7.6188
	0.90	4.8174	4.6366	4.5317	3.5648	3.7726
250	**0.95**	4.9903	4.6538	4.5552	3.6091	3.9413
	0.99	9.0794	7.3960	6.6436	6.1122	6.6356
	0.90	4.6850	4.5401	4.4569	3.4565	3.7554
300	**0.95**	4.8375	4.5548	4.5064	3.5190	3.8444
	0.99	7.8437	6.4718	5.8674	5.2988	5.7827

Tablo 2. $p = 5$ için MSE değerleri

		$\hat{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\beta}}(\boldsymbol{k})$	$\hat{\boldsymbol{\beta}}(\boldsymbol{d})$	$\widehat{\boldsymbol{\beta_1}}(\boldsymbol{k},\boldsymbol{d})$	$\widehat{\boldsymbol{\beta_2}}(\boldsymbol{k},\boldsymbol{d})$
	0.90	9.7366	8.6117	7.9871	5.8113	6.8906
50	**0.95**	14.6611	12.1683	11.4927	8.3554	10.0047
	0.99	58.2158	43.6769	49.5360	30.6494	37.0113
	0.90	7.7481	7.2077	6.8588	4.9884	5.8321
100	**0.95**	9.6856	8.5715	8.0077	5.7673	7.0265
	0.99	28.6504	21.8586	23.0618	15.1196	18.3236
	0.90	7.0249	6.6879	6.4508	4.7638	5.5287
150	**0.95**	8.0716	7.3551	6.8868	5.0038	6.0729
	0.99	21.4129	16.9065	16.3678	11.8731	14.3666
	0.90	6.7516	6.5023	6.3321	4.7400	5.4338
200	**0.95**	7.5534	7.0259	6.6856	4.8556	5.8794
	0.99	16.2291	13.0311	12.4108	8.9703	10.9976
	0.90	6.5210	6.3287	6.2116	4.6153	5.2626
250	**0.95**	7.0128	6.6007	6.3242	4.7819	5.5508
	0.99	14.9050	12.1819	11.5252	8.5824	10.3605
	0.90	6.3162	6.1531	6.0519	4.5608	5.1538
300	**0.95**	6.9157	6.2614	6.2272	4.6898	5.1881
	0.99	13.0292	10.8914	10.2808	7.6680	9.2868

Tablo 3. $p = 7$ için MSE değerleri

		$\hat{\boldsymbol{\beta}}$	$\hat{\boldsymbol{\beta}}(\boldsymbol{k})$	$\hat{\boldsymbol{\beta}}(\boldsymbol{d})$	$\widehat{\boldsymbol{\beta_1}}(\boldsymbol{k},\boldsymbol{d})$	$\widehat{\boldsymbol{\beta_2}}(\boldsymbol{k},\boldsymbol{d})$
	0.90	14.1267	12.8730	12.2195	9.0956	11.5774
50	**0.95**	20.9727	18.1303	17.0990	11.6897	15.1683
	0.99	83.0918	63.5483	72.9760	38.8755	52.8858
	0.90	12.9342	12.3180	11.9297	8.8961	10.5664
100	**0.95**	15.7958	14.4914	13.9819	10.0500	12.5013
	0.99	42.4081	33.4052	35.1552	20.8655	28.0555
	0.90	11.7057	11.3180	11.0428	8.4840	9.8822
150	**0.95**	13.1157	12.2855	11.7975	8.5251	10.6248
	0.99	34.2448	28.3299	28.8530	19.4427	24.5913
	0.90	10.7084	10.4381	10.2514	7.8817	9.1668
200	**0.95**	12.5138	11.9080	11.4970	8.4343	10.3915
	0.99	25.9335	21.6665	21.3201	14.3705	18.6911
	0.90	10.0983	9.8916	9.7504	7.5911	8.6945
250	**0.95**	12.1013	11.6411	11.3165	8.4074	10.2616
	0.99	22.4882	19.3318	18.8294	13.1063	16.8073
	0.90	9.9062	9.7361	9.6274	7.5049	8.6064
300	**0.95**	10.9796	10.6037	10.3522	7.6096	9.3357
	0.99	20.3048	17.6517	16.7645	12.1625	15.4642

3.1. Simülasyon çalışması

Simülasyon çalışmasında Bell dağılımına uygun bağımlı değişken ve çoklu iç ilişkili bağımsız değişkenlerden oluşan veri setleri üretilerek BRM kurulmuştur. Üretilen veri setleri için EÇO, BRE, BLE ve BLTE kullanılarak bu tahmin edicilere ait MSE değerleri karşılaştırılmıştır.

İlk olarak bağımsız değişkenler arasında lineer ilişki olması için McDonald ve Galerneau tarafından Eşitlik (25)' da verilen denklem kullanılmıştır [23].

$$x_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho z_{ip} \qquad (25)$$

Burada ρ^2 bağımsız değişkenler arasındaki korelasyon katsayısı ve z_{ij} standart normal dağılımdan üretilmiş rassal sayılardır. Bağımlı değişken vektörü $y_i \sim Bell\big(W_0(\exp\{\beta_0 + x_{i1}\beta_1 + \cdots + x_{il}\beta_l\})\big)$ olacak şekilde üretilmiştir. Simülasyon çalışmasında β vektörü $\beta'\beta = 1$ olacak şekilde seçilmiştir [24].

Simülasyon çalışmasında bağımsız değişken sayısı, $p = 3,5,7$ örnek hacmi, $n = 50, 100, 150, 200, 250, 300$ ve $\rho = 0.9, 0.95, 0.99$ alınarak simülasyon tekrar sayısı 2000 olarak seçilmiştir. Çalışmada R programının *bellreg* paketi kullanılmıştır [25]. Tahmin edicilere ait MSE değerleri ise Eşitlik (26) kullanılarak hesaplanmıştır.

$$MSE(\hat{\alpha}_i^{\star}) = \frac{1}{2000}\sum_{k=1}^{2000}(\hat{\alpha}_i^{\star} - \alpha)'(\hat{\alpha}_i^{\star} - \alpha) \qquad (26)$$

Simülasyon çalışmasından elde edilen sonuçlar Tablo 1-3' te verilmiştir.

- Simülasyon tablolarından elde edilen sonuçlar aşağıdaki gibi özetlenmiştir.
- Tüm tahmin edicilerin MSE değerleri örnek hacmi arttıkça azalmaktadır.
- Bağımsız değişkenler arasındaki ilişki arttıkça tahmin edicilerin MSE değerleri artmaktadır.
- Bağımsız değişken sayısı arttıkça aynı korelasyon değerleri için ilişkili değişken sayısı arttığından dolayı tahmin edicilerin MSE değerleri artmaktadır.
- Tüm tablolarda aynı korelasyon değeri ve aynı örnek hacmi için tüm BLTE tahmin edicilerinin MSE değerleri BRE ve BLE' den daha küçüktür.

Önerilen BLTE arasında $\hat{\beta}_1(k,d)$, $\hat{\beta}_2(k,d)$' den daha etkin sonuçlar vermektedir.

3.2. Gerçek veri uygulaması

Çalışmanın bu kısmında önerilen tahmin edicinin performansı gerçek veri set sonuçları ile incelenmiştir. Amin vd. nin çalışmalarında kullandıkları Myers vd.' nin kitabından alınan Mine fracture verisi kullanılmıştır [10, 26]. Mine fracture verisi batı Virginia' da Appalachian bölgesindeki kömür yataklarındaki madenlerde meydana gelen yalanma veya kırık sayılarını araştırmak için yapılan gözlemlerden elde edilen bir veri setidir [10]. Veri seti 44 gözlem ve dört bağımsız değişkenden oluşmaktadır. Bağımsız değişkenler sırasıyla ayak cinsinden iç yük kalınlığı, daha önce kanılmış alt dikişin yüzde ekstraksiyonu, alt dikiş yüksekliğini ve madenin açıldığı zamanı göstermektedir. Veri için hesaplanan koşul sayısı 296.5585 olduğu için veri setinde çoklu iç ilişki problemi söz konusudur. Bu veri seti için EÇO, BRE, BLE ve BLTE' ye ait parametre tahminleri, tahminlerin standart hataları ve tahmin edicilerin MSE değerleri Tablo 4' te verilmiştir

Tablo 4 incelendiğinde veri setinde çoklu iç ilişki olduğundan dolayı BRE ve BLE tahmin edicileri ile elde edilen MSE değerleri EÇO tahmin edicisinin MSE değerinden daha küçüktür. Bu çalışmada önerilmiş olan BLTE ile BRE ve BLE' den daha etkin sonuçlar elde edilmiştir.

4. Tartışma ve Sonuç

Uygulamada yaygın olarak karşılaşılan veri tiplerinden birisi sayma verileridir. Sayım verilerinin modellenmesinde yaygın olarak PRM kullanılmaktadır. Fakat aşırı yayılımlı sayma verilerinin modellenmesinde PRM uygun bir model olmamaktadır. Bu modelin alternatifi olarak kullanılan NBRM' ye alternatif olarak son yıllarda

Tablo 4. Mine Fracture veri seti için analiz sonuçları

	$\hat{\beta}$	$\hat{\beta}(k)$	$\hat{\beta}(d)$	$\hat{\beta}_1(k,d)$	$\hat{\beta}_2(k,d)$
$\hat{\beta}_0$	-0.00294	-0.00294	-0.00294	-0.00291	-0.00294
	(1.38907)	(1.24013)	(0.47412)	(0.00291)	(0.00063)
$\hat{\beta}_1$	-0.01126	-0.01126	-0.01126	-0.01001	-0.01126
	(0.00106)	(0.00106)	(0.00106)	(0.00094)	(0.00106)
$\hat{\beta}_2$	0.01819	0.01819	0.01819	0.00740	0.01819
	(0.01684)	(0.01519)	(0.00736)	(0.00239)	(0.00489)
$\hat{\beta}_3$	-0.02384	-0.02384	-0.02383	-0.00197	-0.02384
	(0.00679)	(0.00673)	(0.00653)	(0.00281)	(0.00650)
$\hat{\beta}_4$	-4.00835	-3.57855	-1.36813	-0.00839	-0.00182
	(0.02179)	(0.02177)	(0.02167)	(0.00179)	(0.02166)
	1.93033	1.53868	0.22574	0.00414	0.00108
$MSE(\hat{\beta})$	1.93033	1.53868	0.22574	0.00414	0.00108

yaygın olarak BRM kullanılmaya başlanmıştır. Negatif binom dağılımından daha esnek bir dağılım ve tek parametresi bulunmasından dolayı işlem kolaylığı sağlamasından dolayı daha NBRM' den daha kullanışlı bir model olarak tercih edilebilmektedir.

Regresyon analizinde sıklıkla karşılaşılan problemlerden biri olan çoklu iç ilişki probleminin çözümü için BRM modeli için BRE ve BLE literatürde önerilmiştir. Bu çalışma da bu tahmin edicileri de kapsayan ve avantajlarını birleştirerek yeni bir tahmin edici önerilmiştir. Önerilen BLTE ile BLE ve BRE' den daha etkin sonuçlar elde edilmiştir. Bu sonuçlar farklı senaryolar altında hazırlanan simülasyon çalışmasından elde edilen tablolarda açıkça gösterilmiştir. MSE kriterine göre karşılaştırılan tahmin ediciler içinde en küçük MSE değeri tarafımızdan önerilen BLTE' ye aittir. Ayrıca çalışmada verilen gerçek veri örneğinde elde edilen sonuçlar da simülasyon sonuçlarını desteklemektedir.

Çoklu iç ilişki bulunan sayım verilerinin modellenmesinde BRM' de BLTE tahmin edicisi ile daha etkin sonuçlar elde edilebileceği ortaya konulmuştur.

Teşekkür

Makalemizin iyileştirilme sürecinde değerli görüş ve önerileri ile katkı veren editörlere ve hakemlere teşekkür ederiz.

Etik Beyanı

Bu çalışmada, "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması gerekli tüm kurallara uyulduğunu, bahsi geçen yönergenin "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbirinin gerçekleştirilmediğini taahhüt ederiz.

Kaynakça

[1] Lemonte, A., Moreno-Arenas, G., & Castellares, F. 2020. Zero-inflated Bell regression models for count data. Journal of Applied Statistics, 47(2), 265-286.

[2] Castellares, F., Ferrari, S., Lemonte, A. 2018. On the Bell distribution and its associated regression model for count data. Applied Mathematical Modelling, 56, 172-185.

[3] Asar, Y., Genç, A. 2016. New shrinkage parameters for the Liu-type logistic estimator . Communications in Statistics - Simulation and Computation, 45(3), 1094-1103.

[4] Farrar, D. E., Glauber, R. R. 1967. Multicollinearity in Regression Analysis: The Problem Revisited, The Review of Economics and Statistics, 49 (1), 92-107.

[5] Silvey S. 1969. Multicollinearity and Imprecise Estimation. Journal of the Royal Statistical Society: Series B (Methodological), 31 (3), 539-552.

[6] Lipovetsky S., Conklin, W. M. 2001. Multiobjective Regression Modifications for Collinearity. Computers and Operations Research, 28, 1333-1345.

[7] Hoerl, A., Kennard, R. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 42(1), 80-86.

[8] Månsson, K., Shukur, G. 2011. A Poisson ridge regression estimator. Econ. Model. 28, 1475-1481.

[9] Månsson, K. 2012. On ridge estimators for the negative binomial regression model. Economic Modelling, 29, 178-184.

[10] Amin, M., Akram, M., Majid, A. 2021. On the estimation of Bell regression model using ridge estimator. Communications in Statistics-Simulation and Computation.

[11] Liu, K. 1993. A new class of blased estimate in linear regression. Communications in Statistics - Theory and Methods, 22(2), 393-402.

[12] Akdeniz, F., Kaçıranlar, S. 1995. On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE. Communications in Statistics-Theory and Methods, 24(7), 1789-1797.

[13] Urgan N., Tez M. 2008. Liu estimator in logistic regression when the data are collinear. The 20th international conference, EURO mini conference, Continuous Optimization and Knowladge-based Technologies, 323-327.

[14] Kurtoğlu F., Özkale M. 2016. Liu estimation in generalized linear models: application on gamma distributed response variable. Statistical Papers, 57(4), 911-928.

[15] Qasim, M., Amin, M., Amanullah, M. 2018. On the performance of some new liu parameters for the gamma regression model. Journal of Statistical Computation and Simulation, 88(16), 3065-3080.

[16] Qasim, M., Kibria, B., Månsson, K., Sjölander, P. 2019. A new Poisson Liu Regression Estimator: method and application. Journal of Applied Statistics.

[17] Majid, A., Amin, M., Akram, M. N. 2021. On the Liu estimation of Bell regression model in the presence of multicollinearity, Journal of Statistical Computation and Simulation.

[18] Liu, K. 2003. Using Liu-type Estimator to Combat Collinearity. Communications in Statistics-Theory and Methods, 32(5), 1009-1020.

[19] Asar, Y., Karaibrahimoğlu, A., Başbozkurt, H. Genç, A. 2015. Developing a Liu-type estimator for the Poisson Regression. 9th International Statistics Congress, Antalya.

[20] Asar, Y. 2018. Liu-Type Negative Binomial Regression: A Comparison of Recent Estimators and Applications. In Trends and Perspectives in Linear Statistical Inference, 23–39. Cham: Springer.

[21] Bell, E. 1934. Exponential numbers. The American Mathematical Monthly, 41(7), 411-419.

[22] Ertaş, H., Kaçıranlar, S., Güler, H. 2017. Robust Liu type estimator for regression based on M-estimator. Communication in Statistics – Simulation and Computation, 46(5), 3907-3932.

[23] McDonald, G., Galarneau, D. 1975. A monte carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association, 70(350), 407-416.

[24] Newhouse, J., Oman, S. (1971). An evaluation of ridge estimators. Rand Corporation (p-716-PR), Santa Monica., 1-16.

[25] Team, R. C. (tarih yok). R: Alanguage and environment for statistical computing. Vienna, Austria: http://www.R-project.org.

[26] Myers, R., Montgomery, D., Vining, G., Robinson, T. 2012. Generalized Linear Models with Applications in Engineering and the Sciences. Second Edition, Wiley, A John Wiley&Sons, Inc., Publication.