

DOĞRUSAL OLMAYAN TİP II REGRESYON ANALİZİ ÜZERİNE BİR SİMÜLASYON ÇALIŞMASI

Cengiz Gazeloğlu¹, Yrd.Doç.Dr. Sinan Saraçlı²

¹Afyon Kocatepe Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Afyonkarahisar, Türkiye
e-posta: cengiz_gazeloglu@hotmail.com

²Afyon Kocatepe Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Afyonkarahisar, Türkiye
e-posta: ssaracli@aku.edu.tr

ÖZET

Bağımlı ve bağımsız değişkenler arasındaki doğrusal ya da doğrusal olmayan bir ilişkiyi modellemede, regresyon analizi yaygın olarak kullanılan istatistiksel yöntemlerden biridir. Klasik regresyon çözümlemesi, modelde yer alan bağımsız değişkenlerin herhangi bir ölçüm hatası içermediği varsayımı altında çözümleme yapmaktadır. Tip II regresyon ise bu varsayımın sağlanmadığı durumlarda kullanılacak bir regresyon tekniğidir. Bağımsız değişken(ler)'in de hata içerdiği durumlarda, uygun model parametrelerini tahmin etmek için benimsenebilecek bir yol, değişkenleri sırası ile bağımlı ve bağımsız olarak ele almak ve elde edilen iki regresyon denklemini beraber dikkate alan yeni bir model belirlemek olabilir.

Bu çalışmanın amacı, simülasyon uygulamasıyla farklı örneklem hacimleri ve farklı dağılım biçimlerinde doğrusal olmayan Tip II regresyon tekniğinin performansını klasik doğrusal olmayan regresyon tekniği ile karşılaştırmaktır. Çalışmada, değişkenler sırasıyla bağımlı ve bağımsız olarak dikkate alınarak klasik regresyon analizi yardımıyla iki farklı model elde edilmiştir. Daha sonra bağımsız değişkendeki ölçüm hatasını da dikkate almak için elde edilen her bir gözlem değerinin ortalaması alınmıştır. Böylece daha önceki değerlerden farklı ve bağımsız değişkendeki hatayı da dikkate alan yeni değerler elde edilmiştir. Model parametreleri elde edilen bu yeni gözlem değerleri üzerinden tahmin edilmiştir. Doğrusal olmayan Tip II regresyonun performansı HKO (Hata Kareler Ortalaması) değeri ile belirlenmiştir. Simülasyon çalışması olarak yapılan tüm bu analizler MATLAB paket programı ile gerçekleştirilmiştir.

Anahtar Kelimeler: Doğrusal Olmayan Tip II Regresyon, Ölçüm Hatası, Tip II Regresyon

A SIMULATION STUDY ON NONLINEAR TYPE II REGRESSION ANALYSIS

ABSTRACT

Regression analysis is one of the widely used statistical methods to model the linear or nonlinear relation between dependent and independent variables. Classical regression analysis, work under the assumption that the independent variables do not include any measurement error. Type II regression can be used when this assumption is not met. When the independent variables include some error, to estimate the appropriate model parameters one way can be considering each of the variables dependent and independent respectively and obtain a new model which considers the errors of two regression equations.

The aim of this study is to compare the performance of classical nonlinear regression analysis and nonlinear Type II regression analysis for different sample sizes and distribution types via a simulation application. In study, by the help of classical regression analysis, two models are obtained by considering the variables dependent and independent respectively. Then to consider the measurement errors of independent variable, the means of the each new observations are calculated. Thus, new observations which also include the error terms of independent variables and different than the first values are obtained. Model parameters are estimated via these new values. The performance of nonlinear Type II regression is designated by MSE (Mean Square Error) value. These analysis are performed via MATLAB software as a simulation study.

Keywords: Nonlinear Type II Regression, Measurement Error, Type II Regression

1. Giriş

Regresyon analizinin ilk kullanım alanı astronomi olmuştur. Lengendre ve Gauss gezegenlerin yörüngelerini belirlemek amacıyla EKK olarak bilinen tekniği oluşturmuşlardır. Konu ile ilgili değişkenlerden yola çıkarak bu değişkenler için bir regresyon modeli geliştirmişlerdir (Ergül, 2006).

Değişkenler arasında istatistiksel bir ilişkiden bahsedilebilmesi, Y bağımlı değişkeninin X bağımsız değişkeniyle sistematik olarak değişme eğiliminde olması, diğer bir ifade ile X ve Y değişkenleri arasında ilişkinin doğrusal/eğrisel bir fonksiyon ile ifade edilebilmesi ve X bağımsız değişkeninin Y bağımlı değişkenini etkilemesini gerektirmektedir (Çankal, 2010).

İki değişken arasında doğrusal bir ilişki olduğunda, bu ilişki dağılım grafiğindeki noktalar arasından geçen uygun bir doğru ile tanımlanabilir. Bu doğruya regresyon doğrusu denir ve matematiksel olarak bir denklem ile gösterilebilir. Bu denkleme de regresyon denklemi denir (Sümbüloğlu, 2002).

Regresyon (bağlanım), sözlük anlamı ile bir şeyi başka bir şeye bağlama işi ve biçimidir. Bilimsel olarak regresyon terimi, bir değişken ile başka bir ya da birden çok değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır (Şıklar, 2000). Basit doğrusal regresyon denklemi denkem 1.1'de verildiği gibidir.

$$Y_i = \beta_0 + \beta_1 X_i + e_i \quad 1.1$$

Bu modelde yer alan β_0 sabit katsayısı, β_1 ise eğim katsayısını ifade etmektedir. X_i değerleri, herhangi bir ölçüm hatası içermeyen stokastik olmayan gözlem değerleridir. Modelde yer alan e_i hata terimleri ise stokastik, bağımsız değerlerdir (Tiku and Akkaya, 2004). EKK tekniğinde aşağıdaki varsayımların (Şıklar, 2000) gerçekleşmesi gerekmektedir.

- 1) Hata terimi e_i ortalaması sıfıra eşit rassal bir değişkendir.
- 2) Hata terimi e_i nin dağılımı normaldir.
- 3) Hata terimi e_i nin değerleri arasında ilişki yoktur.
- 4) Hata terimi e_i nin varyansı her X_i değeri için eşittir.
- 5) X değişkeni hata terimi e ile ilişkili olmayıp, stokastik değildir.
- 6) X değişkeni tekrar eden örnek değerlerine göre sabittir.
- 7) Model belirleme hatası taşımamaktadır.
- 8) Bağımsız değişkenler arasında tam veya kuvvetli bir ilişki yoktur. (Çoklu doğrusal regresyon modellerinde geçerlidir).

Sağlık, fen ve sosyal bilimlerdeki hemen hemen tüm disiplinlerde kullanıma sahip olan regresyon analizi, bağımlı değişken ile bağımsız değişken(ler) arasındaki ilişkiyi modellemede sıklıkla tercih edilen önemli bir istatistiksel yöntemdir. EKK tekniği, regresyon analizi içerisinde yaygın olarak kullanılan bir teknik olup, bu tekniğin kullanılabilmesi için bazı varsayımların gerçekleşmesi gerekmektedir. Bu varsayımlardan birisi, bağımsız değişkenlerin herhangi bir ölçüm hatası içermediğidir. Bu doğrultuda, bağımsız değişken(ler)in ölçüm hatası içermediği varsayılan tekniklere Tip I regresyon teknikleri adı verilmektedir. Söz konusu varsayımların sağlanamadığı durumlarda bağımlı ve bağımsız değişken(ler) arasındaki ilişkiyi modellemek için alternatif teknikler geliştirilmektedir (Gazeloğlu ve Saraçlı, 2011).

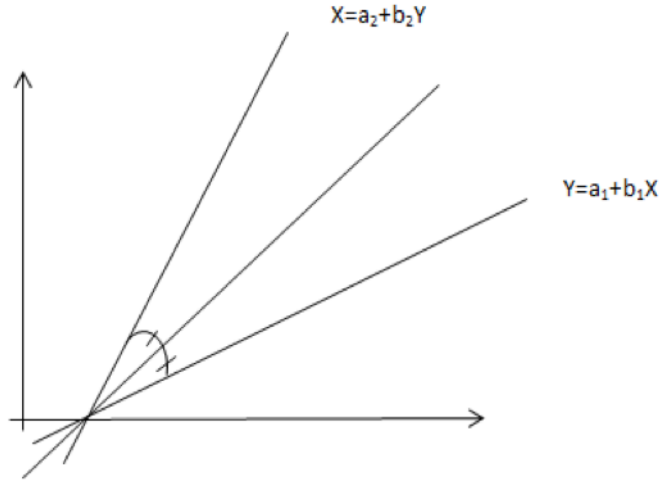
1.1. Tip II regresyon

Klasik regresyonda yaygın olarak kullanılan En Küçük Kareler (EKK) yaklaşımı, hata teriminin kaynağının sadece bağımlı değişkenden olduğunu varsayar. Modeldeki bağımsız değişkenlerinde ölçüm hatası içerebileceğini varsayan regresyon teknikleri, Tip II regresyon teknikleri olarak adlandırılabilir.

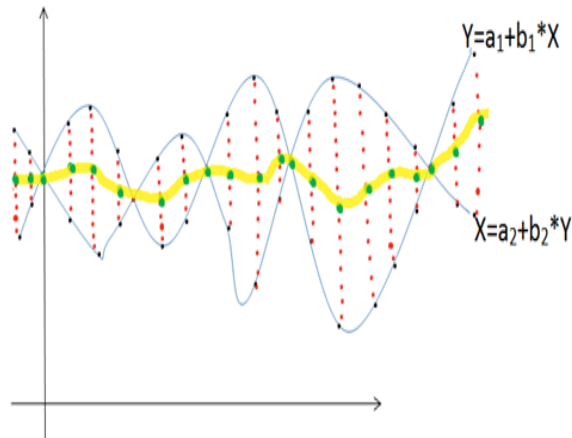
Literatürde birçok farklı Tip II regresyon tekniği vardır. (Saraçlı, Yılmaz, Doğan, 2009, s. 335-342). Genel olarak gözlem değerlerinin elde edilen regresyon denkleminde dik ya da hata miktarına bağlı olarak hesaplanan uzaklıklarının alınması sonucunda her iki değişkendeki hataları da dikkate alma mantığına dayanan bu teknikler, Ortogonal Regresyon, Deming Regresyon, York Regresyon teknikleri ve bunların çeşitli koşullar altında türetilmiş halleridir. Regresyon parametrelerini tahmin etmedeki

hesaplanışları bakımından Ortogonal Regresyon Tekniği; Majör Eksen ve İndirgenmiş Majör Eksen olmak üzere ikiye, Deming Regresyon Tekniği; Deming, Optimal Deming ve Ağırlıklandırılmış Deming olmak üzere üçe, York Regresyon Tekniği ise York ve Optimal York Regresyon Tekniği olmak üzere iki gruba ayrılmaktadır. Passing-Bablok Regresyon Tekniği ise EKK Tekniğine alternatif olan ve parametrik olmayan diğer bir regresyon tekniğidir (Saraçlı 2010, s. 26-35). Bilinmeyen bir dağılım biçimindeki gerçek bir veri seti söz konusu olduğunda, hem bağımlı hem bağımsız değişken ölçüm hatası içeriyorsa Deming ya da EKK-Açıortay tekniğini kullanmak daha sağlıklı sonuçlar elde etmeyi sağlayabilir (Saraçlı ve Türkan, 2012).

Genel olarak yapılan çalışmalarda, değişkenler arasındaki ilişkilerin doğrusal olmadığı durumlarda klasik doğrusal olmayan regresyon tekniği kullanılmaktadır, ancak bağımsız değişken(ler)in ölçüm hatası içermesi durumunda doğrusal olmayan Tip II regresyon tekniğinin kullanılması daha anlamlı ve daha geçerli sonuçlar verecektir. Doğrusal olmayan Tip II regresyon tekniği ile bağımsız değişkendeki ölçüm hatasını dikkate almak için elde edilen her bir gözlem değerinin ortalaması alınmaktadır. Böylece daha önceki değerlerden farklı ve bağımsız değişkendeki hatayı da dikkate alan yeni değerler elde edilmektedir. Bu çerçevede, model parametreleri elde edilen bu yeni gözlem değerleri üzerinden tahmin edilebilmektedir. Ortadaki doğru açıortay doğrusu olmak üzere, Şekil 1'de sunulan Tip II Açıortay regresyon tekniğinden farklı olarak doğrusal olmayan Tip II regresyon tekniği yardımı ile yapılan parametre tahmini Şekil 2'de gösterilen eğri yardımı ile elde edilebilir.



Şekil 1. Doğrusal Tip II Açıortay regresyon modeline ilişkin grafik



Şekil 2. Doğrusal olmayan Tip II regresyon modeline ilişkin grafik

Bu çalışmanın amacı, simülasyon uygulamasıyla farklı örneklem hacimleri ve farklı dağılım biçimlerinde doğrusal olmayan Tip II regresyon tekniğinin performansının klasik doğrusal olmayan regresyon tekniği ile karşılaştırılmasıdır.

2. Metot

Çalışmada doğrusal olmayan Tip II regresyon tekniğinin simülasyon yardımı ile modellenmesi için yapılan MATLAB paket programı yardımıyla rassal olarak üretilen 50000 gözlem değeri (Ana kütle 50000 birim olarak belirlenmiştir), basıklıkları farklı üç dağılım biçimi için ($T_{\sim 4}$, $T_{\sim 10}$, $T_{\sim 30}$), farklı örneklem büyüklüklerinde ($n=50, 100, 200$) analiz edilmiştir.

Ele alınan örneklem için simülasyon aracılığı ile yapılan tekrarlar $N=[100000/n]$ alınarak yapılmıştır. Yapılan bu simülasyonlar için kurulan modele ilişkin hata kareler ortalaması eşitlik 2.1'de ki gibi hesaplanmıştır.

$$HKO = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-k} \quad 2.1$$

3. Bulgular

Tablo 1. 4, 10 ve 30 serbestlik dereceli t dağılımında farklı örneklem büyüklükleri için regresyon tekniklerinin performansları

	Regresyon Tekniği	Farklı Örneklem Hacminde ki HKO'sı Değerleri			
		n=25	n=50	n=100	n=200
4 serbestlik dereceli	OLS (y x)	3,9285	3,3897	3,1498	2,5592
	OLS (x y)	3,9475	2,0467	2,0634	1,4243
	Doğrusal Olmayan Tip II	3,9380	2,7182	2,6067	1,9917
10 serbestlik dereceli	OLS (y x)	3,0475	2,4957	2,1127	1,6706
	OLS (x y)	2,1291	1,8852	1,1710	0,6641
	Doğrusal Olmayan Tip II	2,5883	2,1905	1,6419	1,2674

	Regresyon Tekniđi	Farklı Örneklem Hacminde ki HKO'sı Deđerleri			
		n=25	n=50	n=100	n=200
30 serbestlik dereceli	OLS (y x)	2,8577	1,6908	1,6667	1,5683
	OLS (x y)	1,9351	1,6717	1,1121	0,6576
	Dođrusal Olmayan Tip II	2,3964	1,6312	1,3894	1,2629

4, 10 ve 30 serbestlik dereceli farklı örneklem büyüklüğündeki regresyon tekniklerinin performanslarının sunulduğu tablo 1'de 4 serbestlik dereceli t dağılımında örneklem hacmin 25 olması durumunda (n=25), Y değişkeninin bağımlı, X değişkeninin bağımsız olduğu regresyon tekniğinde (OLS (y|x)) HKO değeri 3,9285 olarak tespit edilmiştir. Örneklem hacminin 50, 100 ve 200 olduğu durumlarda ise HKO sırasıyla 3,3987, 3,1498 ve 2,5592 olarak gözlemlenmiştir. X değişkeninin bağımlı Y değişkeninin bağımsız olduğu regresyon tekniğinde, örneklem hacminin 25, 50, 100, 200 olduğu durumlarda HKO sırasıyla 3,9475 2,0467, 2,0634, 1,4243 olarak hesaplanmıştır. Doğrusal olmayan Tip II regresyon tekniğinde de örneklem hacminin 25, 50, 100 ve 200 olduğu durumlarda HKO değerleri sırasıyla 3,9380 2,7182, 2,6067, 1,9917 olarak gözlemlenmiştir.

10 serbestlik derecesinde ise örneklem hacmin 25 olması durumunda (n=25), Y değişkeninin bağımlı, X değişkeninin bağımsız olduğu regresyon tekniğinde (OLS (y|x)) HKO değeri 3,0475 olarak tespit edilmiştir. Örneklem hacminin 50, 100 ve 200 olduğu durumlarda ise HKO sırasıyla 2,4957, 2,1127 ve 1,6706 olarak gözlemlenmiştir. X değişkeninin bağımlı Y değişkeninin bağımsız olduğu regresyon tekniğinde, örneklem hacminin 25, 50, 100, 200 olduğu durumlarda HKO sırasıyla 2,1291, 1,8852, 1,1710, 0,6641 olarak hesaplanmıştır. Doğrusal olmayan Tip II regresyon tekniğinde de örneklem hacminin 25, 50, 100 ve 200 olduğu durumlarda HKO değerleri sırasıyla 2,5883, 2,1905, 1,6419, 1,2674 olarak gözlemlenmiştir.

Son olarak 30 serbestlik derecesinde, örneklem hacmin 25 olması durumunda (n=25), Y değişkeninin bağımlı, X değişkeninin bağımsız olduğu regresyon tekniğinde (OLS (y|x)) HKO değeri 2,8577 olarak tespit edilmiştir. Örneklem hacminin 50, 100 ve 200 olduğu durumlarda ise HKO sırasıyla 1,6908, 1,6667 ve 1,5683 olarak gözlemlenmiştir. X değişkeninin bağımlı Y değişkeninin bağımsız olduğu regresyon tekniğinde, örneklem hacminin 25, 50, 100, 200 olduğu durumlarda HKO sırasıyla 1,9351, 1,6717, 1,1121, 0,6576 olarak hesaplanmıştır. Doğrusal olmayan Tip II regresyon tekniğinde de örneklem hacminin 25, 50, 100 ve 200 olduğu durumlarda HKO değerleri sırasıyla 2,3964, 1,6312, 1,3894, 1,2629 olarak gözlemlenmiştir.

4. Sonuç

Regresyon çözümlemesi ile gerçekleştirilmiş daha önceki çalışmalar incelendiğinde, Tip I regresyon analizi için birçok doğrusal ve eğrisel çözümlemelere rastlanmaktadır. Tip II regresyon analizi için ise şimdiye kadar yapılan çalışmaların tamamının doğrusal olduğu görülmektedir. Bu çalışmada doğrusal olmayan Tip II regresyon analizinin önemi vurgulanmaya çalışılmıştır.

MATLAB paket programı ile Monte-Carlo simülasyon çalışmasının yapıldığı bu çalışmada, farklı dağılım biçimlerinde, farklı örneklem büyüklüklerindeki veri setleri karşılaştırılmıştır ve sonuç olarak aynı serbestlik derecesine sahip farklı örneklem büyüklüklerinde üretilmiş veri setlerinde örneklem hacmi büyükçe hata kareler ortalamasının da azaldığı gözlemlenmiştir. Ayrıca 4, 10 ve 30 serbestlik dereceli t dağılımlarından üretilen veri setinde yapılan analizler sonucunda dağılımın serbestlik derecesi arttıkça HKO değerlerinin de düştüğü tespit edilmiştir.

Doğrusal olmayan Tip II regresyon tekniğine ait HKO değerinin, sırasıyla bağımlı ve bağımsız değişken olarak ele alınan ve klasik doğrusal olmayan regresyon tekniklerinden elde edilen HKO değerlerinin ortasında bir değer olması her iki değişkendeki hatayı da dikkate alarak çözülemeye gitmesi bakımından anlamlıdır.

Kaynaklar

Çankal, E. (2010). İstatistik, Lisan Yayınları, Ankara.

Ergül, B. (2006). Robust Regresyon ve Uygulamaları. Yüksek Lisans Tezi, Eskişehir Osmangazi Üniversitesi, *Fen Bilimleri Enstitüsü*, Eskişehir.

Gazeloğlu, C. ve Saraçlı, S. (2011). Çoklu Tip II Regresyon Analizi, 7. Uluslar Arası İstatistik Kongresi, Antalya.

Saraçlı S., Türkan A.H. (2012). A Comparison of Linear Regression Techniques In Method Comparison Studies, *Journal of Statistical Computation and Simulation*, doi:10.1080/00949655.2012.673168.

Saraçlı S., Yılmaz V. ve Doğan I., (2009), Simple Linear Regression Techniques in Measurement Error Models (Review), *Anadolu Üniversitesi Bilim ve Teknoloji Dergisi*, 10: 335-342.

Saraçlı S., (2010) Tip II Regresyon Tekniklerinin Monte-Carlo Simülasyonu İle Karşılaştırılması, *E-Journal of New World Sciences Academy*, 6: 26-35.

Sümbüloğlu, K. ve Sümbüloğlu, V. (2002). Bioistatistik, Hatipoğlu Yayınları. Ankara.

Şıklar, E. (2000). Regresyon Analizine Giriş, Anadolu Üniversitesi *Fen-Edebiyat Fakültesi Yayınları*, Eskişehir.

Tiku, M.L. and Akkaya, A.D. (2004). Robust Estimation and Hypothesis Testing, *New Age International, New Delphi*.