Cilt: 09 Sayı: 03 Kategori: 02 Sayfa: 46-53 ISSN No: 1303-6319 Volume: 09 Number: 03 Category: 02 Page: 46-53

QUANTILE REGRESYON PROBLEMİ OLARAK MODELLENEN ÇOK DEĞİŞKENLİ ÇOKLU REGRESYON MODELİNİN ÇOK AMAÇLI PROGRAMLAMA YÖNTEMLERİ İLE ÇÖZÜMÜ

İlkay ALTINDAĞ* **Nimet YAPICI PEHLİVAN****

ÖZET

Quantile Regresyon, basit doğrusal ve çoklu doğrusal regresyon analizinde kullanılan en küçük kareler yöntemine bir alternatif olarak geliştirilen ve daha kapsamlı bir regresyon görüntüsü sunmak amacıyla önerilen bir yöntemdir. Koenker ve Basett (1978) tarafından sunulan Quantile Regresyon, koşullu quantile fonksiyonlarının tahmin modeli için uygun bir yöntem sağlar. Bu çalışmada temel olarak çok değişkenli çoklu regresyon modeli Quantile Regresyon Problemi olarak ele alınmış ve çok amaçlı programlama problemi elde edilmiştir. Elde edilen problemin, çeşitli quantile değerleri için çok amaçlı programlama yöntemlerinden Global Kriter ve Stem Yöntemleri ile çözülmesi amaçlanmıştır.

Anahtar Kelimeler: Çok amaçlı programlama, Çok değişkenli çoklu regresyon, Global kriter yöntemi, Quantile regresyon, Stem yöntemi.

1. GİRİŞ

Regresyon Analizi bağımlı ve bağımsız değişkenler arasındaki fonksiyonel yapıyı oluşturmada kullanılır. Bu yapıyı oluşturmada regresyon modelindeki bilinmeyen parametrelerin tahmini oldukça önemlidir. Regresyon analizinde etkin olarak kullanılan En Küçük Kareler Yöntemi (EKK), hataların bağımsız olması, sıfır ortalamalı ve eşit varyanslı (σ^2) normal dağılıma uygunluk göstermesi durumunda optimal sonuçlar vermektedir. Özelikle normal dağılıma uygunluğun olmadığı durumlarda ve uç değerler olduğunda EKK yöntemi optimallikten çok uzaklaşmaktadır. Koenker ve Basett tarafından 1978 yılında önerilen Quantile Regresyon, koşullu quantile fonksiyonlarının tahmin modeli için uygun bir yöntem sağlar ve bu yöntem quantillere bağlı olarak regresyon katsayılarını belirler.

m bağımlı değişken (*Y*) ve *p* bağımsız değişken (*X*) arasındaki ilişkiyi inceleyen çok değişkenli çoklu regresyon modeli,

$$Y = X\beta + \varepsilon \tag{1}$$

biçiminde matris formunda yazılabilir. Eşitlik (1)'de, $Y: n\times m$ boyutlu bağımlı değişken matrisini, $X: n\times(p+1)$ boyutlu tasarım matrisini, $\beta:(p+1)\times m$ boyutlu regresyon katsayıları vektörünü ve $\varepsilon: n\times 1$ boyutlu hata matrisini göstermektedir. ε hata matrisinin, $E(\varepsilon)=0$ ve $Var(\varepsilon)=\Sigma$ ile normal dağıldığı varsayılmaktadır (Johnson ve Wichern 1988).

* **Arş. Gör., Selçuk Üniversitesi, Fen Fakültesi,** İstatistik Bölümü, e-posta: altindag@selcuk.edu.tr
** **Yrd. Doç. Dr., Selçuk Üniversitesi, Fen Fakültesi,** İstatistik Bölümü, e-posta: nimet@selcuk.edu.tr

2. ÇOK DEĞİŞKENLİ ÇOKLU REGRESYON MODELİ İÇİN QUANTILE PROBLEMİ

Koenker ve Basett (1978) tarafından EKK yöntemine alternatif olarak önerilen Quantile Regresyon yöntemi, koşullu quantile fonksiyonlarının tahmin modeli için uygun bir yöntem sağlamaktadır. Quantile Regresyon, özellikle koşullu quantillerin değişkenlik gösterdiği durumlarda kullanışlıdır ve bu yöntem quantillere bağlı olarak regresyon katsayılarını belirler.

Quantile Regresyon aynı zamanda hem alt hem üst quantillerle ya da tüm quantillerle ilgilendiğinden $X = x$ verildiğinde Y'nin koşullu dağılımının tam bir gösterimini sağlamaktadır (Koenker ve Hallock 2001, Chen 2005, Altındağ 2010).

Quantile Regresyon Yönteminde τ. regresyon quantili $(0 < \tau < 1)$;

$$\underset{\beta \in \mathbb{R}^k}{Minimize} \left[\sum_{t \in \{t : y_t \geq x_t \boldsymbol{b}\}} \tau \left| y_t - x_t \boldsymbol{\beta} \right| + \sum_{t \in \{t : y_t < x_t \boldsymbol{b}\}} (1-\tau) \left| y_t - x_t \boldsymbol{\beta} \right| \right] \tag{2}$$

eşitliğinin minimize edilmesiyle elde edilir. Eşitlik (2)'de, $\{y_t,\ t = 1,\ldots,T\}$ bağımlı değişken, $\{x_t, t = 1,\ldots,T\}$, $t \times k$ boyutlu tasarım matrisini ve $\boldsymbol{\beta}$ tahmin edilecek katsayı vektörünü göstermektedir (Koenker ve Basett, 1978).

Quantile Regresyon'da $\tau = 0.5$ alınması durumunda MINMAD problemi elde edilir. Yapıcı Pehlivan ve Apaydın (2003), çok değişkeni çoklu regresyon modelini MINMAD olarak modellemiş ve çok amaçlı programlama problemi elde etmişlerdir. Bu çalışmada ise, çok değişkenli çoklu MINMAD problemi temel alınarak çok değişkenli çoklu Quantile Regresyon problemi modellenerek,

$$\begin{aligned} & Minimize\, Z_i = \tau \sum_{j=1}^{n} u_{ij} + (1-\tau) \sum_{j=1}^{n} v_{ij} \\ & \qquad X\beta_i + u_i - v_i = Y_i \\ & \qquad u_i, v_i \geq 0, \qquad i = 1, 2, \ldots, m \\ & \qquad 0 \leq \tau \leq 1 \\ & \qquad \beta \text{ işareti belirtilmemiş.} \end{aligned} \tag{3}$$

çok amaçlı programlama (ÇAP) problemi elde edilmiştir. Eşitlik (3)'te, β işareti belirtilmemiş değişkenleri $\beta' \geq 0$ ve $\beta'' \geq 0$ alınarak $(\beta' - \beta'')$ durumuna dönüştürülür. Çözüm sonucunda β' ve β'' değişkenlerinin optimum çözümü ve $u_1, u_2, \ldots, u_m$ ve $v_1, v_2, \ldots, v_m$ sapma değişkenleri elde edilecektir. u_i ve v_j sapma değişkenleri,

$$u_i = \begin{cases} Y_i - \hat{Y}_i, & Y_i - \hat{Y}_i \geq 0 \\ 0, & diğer\ durumlarda \end{cases} \qquad v_i = \begin{cases} -\left(Y_i - \hat{Y}_i\right), & Y_i - \hat{Y}_i < 0 \\ 0, & diğer\ durumlarda \end{cases}$$

biçiminde ifade edilmektedir.

3. ÇOK AMAÇLI QUANTILE PROBLEMİ İÇİN GLOBAL KRİTER ve STEM YÖNTEMİ

Çok amaçlı programlama (ÇAP) probleminin matematiksel gösterimi,

$$Max\left[\sum_{i=1}^{n} c_{1i}x_i, \sum_{i=1}^{n} c_{2i}x_i, \ldots, \sum_{i=1}^{n} c_{ki}x_i,\right]$$
$$\sum_{i=1}^{n} a_{ij}x_j \le b_i, \quad j=1,2,\ldots,m \tag{4}$$
$$xi \ge 0, \qquad i=1,2,\ldots,n$$

biçimindedir.

3.1 Global Kriter Yöntemi

Global kriter yöntemi, problemle ilgili kısıtlar ve amaçlar tanımlandıktan sonra, karar vericinin tercihleriyle ilgili bilgisine ihtiyaç duymayan yöntemlerden birisidir. Karar verici, klasik optimizasyon yöntemlerinde olduğu gibi yöntemin bulduğu çözümün kabul edilebilir olduğunu varsayar.

Adım 1.

$$Max\, f_j(x) = c^t x, \quad j=1,2,\ldots,k$$
$$Ax \le b, \quad x \ge 0 \tag{5}$$

problemlerinin mümkün ideal çözümleri $f_j^*, \quad j=1,2,\ldots,k$ elde edilir.

2.Adım: Bulunan ideal çözümler yardımıyla,

$$Minimize \sum_{j=1}^{k}\left[\frac{f_j(x^*) - f_j(x)}{f_j(x^*)}\right]^a$$
$$Ax \le b, \quad x \ge 0 \tag{6}$$

Eşitlik (6)'da, işlem kolaylığı sağlaması ve bir doğrusal programlama problemi elde etmek istenmesi nedeniyle *a=1* alınmıştır. *a=2* alınması durumunda, problem karesel programlama problemine dönüşmektedir. Problemlerinin mümkün ideal çözümleri $f_j^*, \quad j=1,2,\ldots,k$ elde edilir (Evren ve Ülengin, 1992).

3.2 Stem Yöntemi

STEP-metodu adıyla da bilinen bu yöntem iteratif bir çözüm sürecine dayanır. Her iterasyonda hesaplama aşaması, karar verme aşaması ve karar verici (KV) ile analist arasında bir diyalog aşaması vardır. STEM yönteminin adımları aşağıdaki gibidir:

Adım 1.

$$\begin{aligned} & Max\, f_j\left(x\right)=c^t x, \quad j=1,2,\ldots,k \\ & \qquad Ax\le b, \quad x\ge 0 \end{aligned} \tag{7}$$

problemlerinin mümkün ideal çözümleri $f_j^*,\ \ j=1,2,\ldots,k$ elde edilir ve Tablo 1'de verilen alternatif çözümler tablosu oluşturulur.

Tablo 1. Alternatif Çözümler Tablosu

	f_1	f_2	$\cdots$	f_j	$\cdots$	f_k
f_1				z_{j1}		
f_2				z_{j2}		
$\vdots$				$\vdots$		
f_j				f_j^*		z_{jk}
$\vdots$				$\vdots$		
f_k				z_{jk}		

f_j^*: j.sütunun maksimum değerini, $f_j^{\min}$: j.sütunun minimum değerini göstermektedir. Lineer Programlama (LP) probleminin çözümü, ideal çözüm f_j^* en yakın gözükmektedir. Burada $X^m, Ax\le b, x\ge 0$ ve $(m\text{-}1)$. iterasyonda ilave edilen kısıtları içermektedir.

Adım 2. *m*. iterasyonda,

$$\begin{aligned} & \underset{(x,\lambda)}{Min}\,\lambda \\ & \lambda\ge f_j^*-f_j\left(x\right)\pi_j, \quad j=1,2,\ldots,k \\ & x\in X^M \\ & \lambda\ge 0 \end{aligned} \tag{8}$$

LP probleminin çözümü, ideal çözüm f_j^* 'a en yakın gözükmektedir. π_j optimuma olan uzaklığın göreceli önemini verir ve

$$\pi_j=\frac{\alpha_j}{\sum_j \alpha_j} \tag{9}$$

eşitliğinden hesaplanır. Burada α_j,

$$\alpha_j = \begin{cases} \dfrac{f_j^* - f_j^{\min}}{f_j^*} \left[\dfrac{1}{\sqrt{\sum_{i=1}^{n} \left(c_{ji}\right)^2}} \right] & eğer \;\; f_j^* > 0 \; ise \\ \dfrac{f_j^{\min} - f_j^*}{f_j^{\min}} \left[\dfrac{1}{\sqrt{\sum_{i=1}^{n} \left(c_{ji}\right)^2}} \right] & eğer \;\; f_j^* \le 0 \; ise \end{cases} \tag{10}$$

dir.

<u>Adım 3:</u> Uzlaşık çözüm X^m, amaç vektörü f^m ile f_j^* karşılaştırmak üzere KV'ye sunulur. Eğer bazı amaçlar tatminkar ise ve diğerleri değilse, KV tatmin olunan amaç f_j^mden tatmin olunmayan amacın lehine fedakarlık yapmak zorundadır. Böylece, sonraki iterasyon için uygun bölge,

$$X^{M+1} = \begin{cases} X^M \\ f_j\left(x\right) \ge f_j\left(x^m\right) - \Delta f_j \\ f_j\left(x\right) \ge f_j\left(x^m\right); i \ne m, \; i = 1, 2, ... \end{cases} \tag{11}$$

biçiminde oluşturulur. Tatmin olan amaca ait π_j ağırlığı sıfır alınır ve *(m+1)*. iterasyonun hesaplama aşaması başlar ve tatmin edici çözüme ulaşılana kadar işleme devam edilir (Evren ve Ülengin 1992).

4. UYGULAMA

Uygulamada, Rencher (2002)'den alınan kimyasal reaksiyon içeren bir deneye ilişkin sonuçlar uygulama verisi olarak kullanılmıştır. Bu verideki bağımlı değişkenler;

Y_1:Değişmeyen başlangıç maddesinin yüzdesi
Y_2:İstenen ürüne dönüştürme yüzdesi
Y_3:İstenmeyen yan ürünün yüzdesi

Bağımsız değişkenler;

X_1:Sıcaklık
X_2: Konsantrasyon
X_3: Zaman

olarak ele alınmıştır.

Global Kriter, Stem ve EKK yöntemleri ile çözüm sonucunda elde edilen tahmini değerlere ilişkin ortalama mutlak sapma (MAD) ve hata kareler ortalaması (MSE) değerleri karşılaştırılmış ve Tablo 2'de verilmiştir.

Tablo 2. Ortalama Mutlak Sapma ve Hata Kareler Ortalaması Değerleri

Tau (τ)	Global Kriter Yöntemi		Stem Yöntemi		En Küçük Kareler (EKK)		Global Kriter Yöntemi		Stem Yöntemi		En Küçük Kareler (EKK)		Global Kriter Yöntemi		Stem Yöntemi		En Küçük Kareler (EKK)	
	Y_1		Y_1		Y_1		Y_2		Y_2		Y_2		Y_3		Y_3		Y_3	
	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE	MAD	MSE
0.10	4.42	31.77	4.42	31.77			5.11	46.87	3.06	15.62			3.88	26.43	3.14	20.62		
0.20	4.30	32.52	3.44	19.70			5.11	46.87	4.43	36.08			3.43	22.32	3.43	22.32		
0.25	3.84	25.91	1.70	5.35			3.07	16.16	3.09	16.37			3.38	21.63	3.38	21.63		
0.30	1.66	5.14	1.70	5.35			3.05	14.18	3.04	14.14			3.28	23.04	3.28	23.04		
0.40	1.55	4.69	1.70	5.35			2.92	16.73	2.92	14.83			3.11	21.49	3.02	20.79		
0.50	1.52	4.37	1.51	4.36	1.56	4.22	2.99	15.88	2.99	18.33	2.97	13.13	3.37	25.08	3.00	19.65	2.89	12.17
0.60	1.53	4.44	1.53	4.34			3.07	17.47	3.18	20.38			4.07	30.96	4.07	30.96		
0.70	1.85	7.13	1.85	7.13			3.34	23.58	3.09	17.01			4.14	31.78	3.82	25.95		
0.75	2.03	8.32	1.85	7.13			3.61	27.35	3.67	26.03			4.17	32.13	4.17	32.13		
0.80	2.11	9.72	2.21	9.83			3.61	27.35	3.67	25.96			5.18	47.52	3.16	21.95		
0.90	2.95	17.66	2.29	10.54			4.99	44.33	4.60	34.79			7.45	83.53	7.81	87.76		

5. TARTIŞMA ve SONUÇ

Bütün regresyon modellerinde olduğu gibi, çok değişkenli çoklu regresyon modellemesinde amaç, e_i hatalarını minimum yapacak şekilde β_i parametrelerinin tahmin edilmesidir. Bu nedenle, çok değişkenli çoklu regresyon modelinde hataların minimum yapılması amaçlanmış ve çözüm için Quantile Regresyon yaklaşımı önerilmiştir. Önerilen yaklaşımda Quantile Regresyon temeline dayanan çok amaçlı programlama problemi elde edilmiştir. Elde edilen (ÇAP) çözümü için Global Kriter ve Stem Yöntemleri Kullanılmıştır.

En küçük kareler yöntemi, $\hat{\beta}_i$ regresyon parametresi tahmin edicilerini, hata kareler toplamını minimum yapacak şekilde hesaplamaya dayanmaktadır. Alternatif olarak önerilen Quantile Regresyon ise, hataların mutlak değerlerinin toplamını minimum yapma esasına dayanmaktadır. Uygulamada, çeşitli quantile değerleri için çok amaçlı programlama yöntemlerinden Global Kriter ve Stem Yöntemleri ile çözüm yapılmıştır.

Global Kriter, Stem, EKK yöntemi sonuçlarına göre Tablo 2 incelendiğinde;

Y_1 bağımlı değişkeni için,

- ✓ MSE kriterine göre en küçük değer (4.360), $\tau = 0.50$ için Stem yönteminden elde edilmiştir.
- ✓ MAD kriterine göre en küçük değer (1.517), $\tau = 0.50$ için Stem yönteminden elde edilmiştir.

Y_2 bağımlı değişkeni için,

- ✓ MSE kriterine göre en küçük değer (13.129), EKK yönteminden elde edilmiştir.
- ✓ MAD kriterine göre en küçük değer (2.917), $\tau = 0.40$ için Stem yönteminden elde edilmiştir.

Y_3 bağımlı değişkeni için,

- ✓ MSE kriterine göre en küçük değer (2.893), EKK yönteminden elde edilmiştir.
- ✓ MAD kriterine göre en küçük değer (12.168), EKK yönteminden elde edilmiştir.

Çözüm sonuçlarına ilişkin Tablo 2'de verilen MSE ve MAD değerleri incelendiğinde, en küçük MSE ve MAD değerlerinin $\tau = 0.50$ değeri civarında elde edildiği görülmektedir.

6. KAYNAKLAR

Altındağ, İ., 2010. Quantile Regresyon ve Bir Uygulama. Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Yüksek Lisans Tezi, Konya.

Arthanari, T. S., Dodge Y., 1981. Mathematical Programming in Statistics. New York.

Chen, L., 2005. An Introduction to Quantile Regression and the QUANTREG Procedure. Statistics and Data Analysis, pp. 213- 230.

Evren R., Ülengin, F., 1992. Yönetimde Çok Amaçlı Karar Verme. İTÜ Yayınları, İstanbul.

Johnson, R. A., Wichern D. W., 1988. Applied Multivariate Statistical Analysis. Prentice-Hall, USA.

Gilchrist, W. G., 2000. Statistical Modelling with Quantile Functions. Florida: Chapman & Hall/Crc.

Koenker R., Bassett, G., 1978. "Regression Quantiles", Econometrica, Vol. 46, No.1.

Koenker R., Hallock, K., F., 2001. Quantile Regression, Journal of Economic Perspecives. Volume 15, Number 4, pp. 143-156.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, USA.

Portnoy, S., Koenker, R., 1989. Adaptive L Estimation for Linear Models. The Annals of Statistics, Vol. 17, No. 1.

Rencher, A, C., 2002. Methods of Multivariate Analysis. John Willey and Sons, Inc.

Saçaklı, İ., 2002. Kantil Regresyon ve Alternatif Modelleri ile Karşılaştırılması. Marmara Üniversitesi, Sosyal Bilimler Enstitüsü,Yüksek Lisans Tezi, İstanbul.

Serper, Ö., 1996. Uygulamalı İstatistik I., 3. Basım, İstanbul: Filiz Kitapevi.

Yapıcı, Pehlivan, N., Apaydın, A., 2003. Çok Değişkenli Regresyon Modelinin Minmad Problemi Olarak Modellenmesi ve Global Kriter Yöntemi ile Çözümü. İstatistik Araştırma Dergisi, Cilt 2, No 2, Ankara.

MULTI OBJECTIVE PROGRAMMING METHODS FOR SOLVING MULTIVARIATE MULTIPLE REGRESSION MODELLED AS QUANTILE REGRESSION PROBLEM

ABSTRACT

Quantile Regression has been advanced as an alternative to the least squares method which is used for simple and multiple linear regression analysis and is proposed in order to present a comprehensive regression image. Quantile Regression, proposed by Koenker and Basett (1978), provides an appropriate method for the estimation model of conditional quantile functions. This study mainly considered multivariate multiple regression model as a quantile regression problem and obtained a multi objective program problem. The obtained problem is aimed to be solved for various quantile values by Global Criteria and Stem Methods, which are among the multi objective program methods.

Keywords: Multi objective program, Multivariate multiple regression, Global Criteria Method, Quantile regression, Stem Method.