ISSN:1306-3111
e-Journal of New World Sciences Academy
2009, Volume: 4, Number: 4, Article Number: 3A0013

PHYSICAL SCIENCES
Received: November 2008
Accepted: September 2009
Series : 3A
ISSN : 1308-7304

Erol Terzi[1], Talat Şenel[2]
Yüksel Terzi[1], M.Ali Cengiz[1]
Ondokuz Mayıs University[1]
TÜİK Bölge Müdürlüğü[2]
eroltrz@omu.edu.tr
Samsun-Turkey

LATENT REGRESYON MODELLERİN GERÇEK VERİ KÜMESİNE UYGULAMASI

ÖZET

Latent Regresyon modelleri gözlenen açıklayıcı değişkenlerle bir Latent bağımlı değişkeni arasındaki ilişkiyi açıklamada kullanılan güçlü yöntemlerdir. Ancak Latent değişkenlerinin gözlenemeyen yapısı nedeniyle model seçimi standart regresyon yöntemleri kullanılarak yapılamamaktadır. Bu çalışmada Latent Regresyon modeli kullanılarak Karadeniz Bölgesi’ndeki Tütün üretimini etkileyen değişkenlerin seçimi üzerinde duruldu.

Anahtar Kelimeler: Latent Root Regresyon, En İyi Model Seçimi, Geliştirilmiş En Küçük Kareler, Tütün, Regresyon Analizi

AN APPLICATION OF LATENT ROOT REGRESSION MODELS TO REAL DATA SET

ABSTRACT

Latent regression models are powerful tools for studying the relation between a latent outcome variable and observed covariates. However, model selection can not be performed using methods from standard regression models due to the unobserved nature of latent outcome variables. In this study, selection of variables that have effect on the tobacco growth in black sea area was performed using latent regression model.

Keywords: Latent Root Regression, The Best Model Selection, Ordinary Least Square (OLS), Tobacco, Regression Analysis

1. GİRİŞ (INTRODUCTION)

Çoklu regresyon analizinde, bağımlı değişkendeki değişimi açıklayacak bir regresyon eşitliğine girecek bağımsız değişken sayısı ne kadar çok olursa, model o kadar küçük hata taşır. Ancak gerek bağımsız değişkenlerin her birisiyle ilgili sonuçları elde etmenin getireceği maddi yük, gerekse böyle gözlemleri belirli bir zaman içinde elde etme mecburiyetinin getirebileceği zorluklar ve muhtemel hatalar bağımsız değişken sayısını azaltmayı zorunlu kılar. Bu bakımdan tahmin denklemleri, akla gelen her bağımsız değişkeni modele ilave ederek kurulmamalıdır. Araştırmacı, tahminin isabetini mümkün olduğu kadar yüksek tutacak, fakat ekonomik yük ve zorlukları ve çok değişkenle ilgili gözlem elde etmenin getireceği sistematik hataları mümkün olduğu kadar azaltacak sayıda bağımsız değişkenle çalışmaya gayret etmelidir. Bu işlem yapılırken, değişken seçim yöntemleri kullanılarak sonuç değişkenini en iyi açıklayan model seçilir.

Son yıllarda en iyi model seçimi için çok farklı seçim kriterleri üzerinde çalışmalar yapılmaktadır. Aşağıda değişken seçimi için kullanılan belli başlı kriterler verilmiştir:

- En iyi alt regresyonlar kullanımı R^2 ve C_p.
- Adımsal regresyon.
- Ridge regresyon.
- Başlıca regresyon modelleri.
- Latent Root regresyon.

Bu yöntemlerin, özellikle tarım ve sanayi sektöründe son yıllarda yaygın kullanım alanları vardır. Bunlardan yalnızca Latent Root regresyon yöntemi üzerinde durulacaktır. Bu yöntemin temeli, değişkenlere ait korelasyon matrisini kullanarak elde edilen özdeğer ve özvektörlere dayanır. Bu çalışmada, yöntemin teorik yapısı kısaca incelendikten sonra gerçek bir veri kümesi üzerinde uygulamalı değerlendirme yapılmıştır.

Latent değişkenli regresyon modelleri üzerine son yıllarda Box et. Al. (1962), Freund & Minton (1979), Everitt (1984), Draper & Smith (1988), Delbance & Jobson (1993), Elizabeth & Dauglas (1993) önemli çalışmalar yapmışlardır.

2. ÇALIŞMANIN ÖNEMİ (RESEARCH SIGNIFICANCE)

Bu çalışmada, latent regresyon modellerin gerçek veri kümesine ilişkin bir uyguma yer almaktadır. Latent regresyon modeli kullanılarak hazırlanan Karadeniz Bölgesi'ndeki tütün üretimini etkileyen değişkenlerin seçimi ve elde edilen sonuçlar bundan sonra yapılacak çalışmalara ışık tutması açısından önem arzetmektedir.

3. LATENT ROOT REGRESYON VE PARAMETRE TAHMİNİ (LATENT ROOT REGRESSION AND PARAMETER ESTIMATION)

Açıklayıcı değişkenlerin eliminasyonu ve alternatif ön tahmin eşitlikleri ilk olarak Box ve Tidwell tarafından incelenmiştir. Yapılan çalışmada, standartlaştırılmış bağımlı ve bağımsız değişkenlerin

$$Z^* = (y, Z) \tag{1}$$

eşitliğini sağlaması için veri matrisi incelenmişlerdir. Burada Z standartlaştırılmış ve ölçülmüş "X matrisi", $y = (Y - \overline{Y})\big/\sqrt{S_{YY}}$ ve $S_{YY} = \sum (Y_i - \overline{Y})^2$ (Box ve Tidwell, 1962). Çalışmamıza ait doğrusal genel modelimiz aşağıdaki gibi verilir:

$$Y^* = \beta_0^* + \beta_1^* Z_1 + \cdots + \beta_r^* Z_r + \varepsilon \qquad (2)$$

Burada Z açıklayıcı değişkenler matrisi, β^* ise bilinmeyen parametre vektörüdür.

Z veri kümesine ait korelasyon matrisi

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & \cdots & r_{2k} \\ \vdots & & & \\ r_{k1} & r_{k1} & & r_{kk} \end{bmatrix} \qquad (3)$$

ile verilir. (2) ile verilen modeldeki β^* katsayılar vektörü $\hat{\beta}_0^* = \overline{Y}$ olmak üzere $i = 1,2,...,r$ için

$$\hat{\beta}^* = \begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \vdots \\ \hat{\beta}_r^* \end{bmatrix} = c \sum_j {}^* \gamma_{0j} \lambda_j^{-1} \begin{bmatrix} \gamma_{1j} \\ \gamma_{2j} \\ \vdots \\ \gamma_{rj} \end{bmatrix} = c \begin{bmatrix} \sum_j {}^* \gamma_{0j} \lambda_j^{-1} \gamma_{1j} \\ \sum_j {}^* \gamma_{0j} \lambda_j^{-1} \gamma_{2j} \\ \vdots \\ \sum_j {}^* \gamma_{0j} \lambda_j^{-1} \gamma_{rj} \end{bmatrix} \qquad (4)$$

formülü ile verilir. Burada;

$$c = -\left\{ \sum_j {}^* \gamma_{0j} \lambda_j^{-1} \right\}^{-1} \left\{ \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right\}^{1/2} \qquad (5)$$

sabit bir değer, λ_j'ler (3)'de verilen korelasyon matrisinden elde edilen özdeğerler (Latent Root) ve γ_{ij}'ler ise bunlara karşılık gelen özvektörlerdir (Everitt, 1984).

Modele hangi değişkenlerin katılmasına karar vermek için $\lambda_j \le 0.05$ ve $\gamma_{0j} < 0.10$ olan vektörler elimine edilir. Geriye kalan vektörler üzerinden $\sum_j {}^*$ toplam alınarak "Modified Least Square" (MLS) eşitliği tahmin edilir. Buradaki amaç, tahmin edilen MLS eşitliğinin hata kareler toplamında (RSS) meydana getirdiği değişimi ölçmektir. Herhangi bir MLS eşitliği için hata kareler toplamları;

$$RSS_1 = \left\{ \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right\} \left\{ \sum_j {}^* \gamma_{0j}^2 \lambda_j^{-1} \right\}^{-1}$$

$$= \left\{ \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right\}^{1/2} \left\{ \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right\}^{1/2} \left\{ \sum_j {}^* \gamma_{0j}^2 \lambda_j^{-1} \right\}^{-1} = -c \left\{ \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \right\}^{1/2} \qquad (6)$$

formülü ile hesaplanır. Burada c, (5)'de verildiği gibidir (Delbance ve Jobson, 1993).

MLS ile tahmin edilen modelde, değişken seçimi için geriye doğru $\ell = 1,2,\ldots,r$ için X_ℓ açıklayıcı değişkeni çıkarıldıktan sonra hata kareler toplamı;

$$RSS_2 = \left\{\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right\}\left\{t_{00} - \frac{t_{l0}^2}{t_{ll}}\right\}^{-1} \quad (7)$$

formülü ile hesaplanır. Burada; $t_{pq} = \sum_j {}^* \frac{\gamma_{pj}\gamma_{qj}}{\lambda_j}$ ve dolayısıyla,

$$t_{00} = \sum_j {}^* \frac{\gamma_{oj}\gamma_{0j}}{\lambda_j}, \quad (p=0,\ q=0),$$

$$t_{\ell 0} = \sum_j {}^* \frac{\gamma_{\ell j}\gamma_{0j}}{\lambda_j}, \quad (p=\ell,\ q=0)$$

$$t_{\ell\ell} = \sum_j {}^* \frac{\gamma_{\ell j}\gamma_{lj}}{\lambda_j}, \quad (p=l,\ q=\ell).$$

4. BULGULAR (FINDINGS)

Latent Root Regresyonun güncel bir veriye uygulaması düşünüldü. Yöntem Karadeniz Bölgesi'nde tütün üretim miktarını etkileyen değişkenlerin seçimi için uygulandı. Kullanacağımız veri 1990-2003 yılları arasında toplanmıştır.

Çalışmada kullanılan değişkenler aşağıdaki gibi tanımlanmıştır (Terzi, Terzi ve Şenel, 2004):

Y : Karadeniz bölgesinde tütün üretim miktarı,
X_1 : Ekim alanı (hk.),
X_2 : Verim (kg/hk.)
X_3 : Fiyat (Tl./kg.)
X_4 : Ortalama sıcaklık (C^0)
X_5 : Yüzde nem oranı
X_6 : Ortalama yağış miktarı (mm.)

Bağımlı ve bağımsız değişkenlere ait hesaplanan korelasyon matrisi,

$$R = \begin{bmatrix} 1{,}00000 & 0{,}85145 & 0{,}47002 & 0{,}14748 & 0{,}80304 & 0{,}78324 & 0{,}52890 \\ & 1{,}00000 & 0{,}31633 & 0{,}20614 & 0{,}68426 & 0{,}69449 & 0{,}49863 \\ & & 1{,}00000 & 0{,}08188 & 0{,}84482 & 0{,}69161 & 0{,}84495 \\ & & & 1{,}00000 & 0{,}25008 & 0{,}06224 & 0{,}06641 \\ & & & & 1{,}00000 & 0{,}81665 & 0{,}81468 \\ & & & & & 1{,}00000 & 0{,}75987 \\ & & & & & & 1{,}00000 \end{bmatrix}$$

Korelasyon matrisinin Latent kökleri (λ_j) ve onlara karşılık gelen uygun Latent vektörleri Tablo 1'de verilmiştir.

Tablo 1. Latent kökleri ve latent vektörleri
(Table 1. Latent roots ve latent vectors)

Latent Root	Y	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
λ_j	γ_{0j}	γ_{1j}	γ_{2j}	γ_{3j}	γ_{4j}	γ_{5j}	γ_{6j}
4,52910	-0,40153	0,247517	-0,381375	-0,449596	0,005165	0,582721	-0,300548
1,07368	-0,36568	0,376898	-0,429736	0,366453	-0,388543	-0,482594	-0,146282
0,89016	-0,38011	-0,380921	0,407620	-0,327397	-0,116856	-0,363334	-0,539819
0,22519	-0,09426	0,742094	0,639212	0,069277	0,135214	0,049557	-0,079247
0,18509	-0,45411	-0,012335	0,138691	-0,353339	-0,241921	-0,093842	0,763026
0,08124	-0,42885	-0,077024	-0,144791	0,165316	0,846913	-0,190322	0,091501
0,01554	-0,40190	-0,307886	0,234380	0,631989	-0,203183	0,498038	0,002715

Tablo 1'deki sonuçlar kullanılarak iki yöntem için parametre tahminleri ve hata kareler toplamları aşağıdaki gibi elde edilir:

Tablo 2. Parametre tahminleri
(Table 2. Parameter estimation)

	β_1	β_2	β_3	β_4	β_5	β_6	RSS
OLS (λ_0) dahil	-0,7057	46,1194	0,3527	275,7047	1499,7365	45,8580	20049000,7
MLS (λ_0) hariç	-0,0456	-4,8041	-0,0016	2601,3899	-736,9354	167,4019	75771162,8

En küçük Latent köke ($\lambda_0 = 0.01554$) karşılık gelen Latent vektör kaldırıldığı zaman hesaplanan MLS ve bütün vektörler (λ_0 dahil) üzerinden hesaplanan Ordinary Least Square (OLS) karşılaştırması yapıldı. Buna göre iki yönteme göre hesaplanan hata kareler toplamlarında önemli bir farklılık olduğu görüldü.

Şimdi uygulamada elde edilen modelden $i = 1,2,\ldots,6$ için açıklayıcı (X_i) değişken çıkardıktan sonra geriye kalan değişkenler kullanılarak bulunan MLS ve OLS hata kareler toplamlarını inceleyelim.

Tablo 3. Modelden açıklayıcı değişken çıkarma
(Table 3. Explanatory variable elimination from model)

	X_1	X_2	X_3	X_4	X_5	X_6
OLS	41776120,1	33609501,2	2454108,9	38316077,3	43940770,9	21637721,4
MLS(λ_0 hariç)	75909000,6	75980436,1	75777046,2	126536591,7	200327651,3	107925458,6

Her iki tahmin edilen eşitlik için modelden ilk çıkarılması düşünülen değişken X_3=fiyat olmaktadır. Bu durum OLS'den daha açık bir şekilde görülmektedir. X_3 modelden çıkarıldıktan sonra modellerin hata kareler toplamları önemli derecede azalmıştır.

5. SONUÇ VE ÖNERİLER (CONCLUSION AND SUGGESTIONS)

Karadeniz Bölgesi tütün üretim miktarını etkileyen faktörler incelendi. Buna göre ekim alanı, verim, sıcaklık, nem oranı ve yağış miktarı üretimi etkileyen önemli açıklayıcı değişkenler olarak bulundu. Fiyatın ise modele önemli bir katkısı yoktur. Böylece söz konusu değişkenlerin modeli açıklama oranı %89.3 olarak bulunmuştur. Bu sonuç, bölgenin tarımsal üretimini önemli derecede etkileyen tütün üretimi için önemli bir bilgi vermektedir. Elde ettiğimiz sonuçlara göre tüm Latent köklerinin (λ_j) modele katılımıyla elde edilen

Ordinary Least Squary (OLS) kullanılarak bulunan tahmin eşitliği, en küçük Latent kökü (λ_0) dışındaki λ_j'ler için hesaplanan Modified Least Square (MLS) eşitliğinden daha faydalı olacaktır. Tahmin edilen model için anlamlılık testleri yapılarak ileriye doğru tahminler yapılabilir. Bu ise Karadeniz Bölgesi'nde tütün üretimi için kolaylık ve maddi kazanç sağlar.

Latent değişkenli modeller, çok değişkenli çalışmalarda açıklayıcıların seçimi ve en uygun model seçimi için oldukça uygundur. Bu modeller; sağlık, sanayi ve tarım sektörüyle ilgili çalışmalarda elde edilebilecek uygun veri türlerine de uygulanabilir.

KAYNAKLAR (REFERENCES)

1. Box, G.E.P. and Tidwell, P.W., (1962). Transformation of The Independent Variables, Technometrics, Vol.4, pp: 531-550.
2. Delbance, J.D. and Jobson, (1993). Applied Multivariate Data Analysis, New York.
3. Draper, N.R. and Smith, H., (1988). Applied Regression Analysis, Second Edition, Wiley, New York.
4. Elizabeth, A.P. and Dauglas, C.M., (1993). Introduction to Linear Regression Analysis, New York.
5. Everitt, B.S., (1984). An Introduction to Latent Variables Models, London.
6. Freund, R.J. and Minton, P.D., (1979). Regression Methods, Marcel Dekker, New York.
7. Terzi, E., Terzi, Y. ve Şenel, T., (2004). Latent Regresyon Modellerinin Gerçek Veri Kümesine Uygulaması, IV. İstatistik Günleri Sempozyumu, İzmir.