Ekonometri ve İstatistik Sayı:15 2011 1-14

İSTANBUL ÜNİVERSİTESİ
İKTİSAT FAKÜLTESİ
EKONOMETRİ VE İSTATİSTİK
DERGİSİ

KUKLA DEĞİŞKENLERİN *T* İSTATİSTİĞİ İLE AYKIRI GÖZLEMLER TESPİT EDİLEMEZ

Arzdar KİRACI*

Özet

Güncel yazında, bir gözlemi aykırı gözlem (outlier) olarak tespit edilebilmek için bu gözlem kukla değişken ile temsil edilmekte ve kukla değişkenin istatistiksel olarak anlamlı olup olmadığına bakılmaktadır. Bir gözlemin aykırı gözlem olması için kukla değişkenin t-istatistiğinin istatistiksel olarak anlamlı olma tezi kullanılmaktadır. Bu çalışma, basit bağlaşım modelini kullanarak, bu tezin doğru olmadığını kuramsal olarak ispatlamakta ve bir karşı örnekle bu tezi çürütmektedir. Bu çalışma için türetilmiş örnekte, dirençli (robust) bağlaşım yöntemi ile aykırı gözlem olarak tespit edilen bir gözlemin, kukla değişkenin t-istatistiği ile tespit edilemediği gösterilmektedir. Buna ilaveten, kukla değişken eklemenin önemli bağlaşım istatistiklerini nasıl etkilediğini de irdelemektedir.

Anahtar Kelimeler: Dirençli bağlaşım, t-istatistiği, kukla değişken, aykırı gözlem, tez çürütme, basit bağlaşım modeli, tespit sorunu, örnek

Jel Sınıflaması: C2, C3, C51, C52

Abstract

In the current literature, in order to be able to detect a single observation as an outlier observation, this observation is represented by a dummy variable and the dummy variable is checked for statistical significance. For an observation to be an outlier observation, the thesis of significant t-statistics of dummy variable is used. This paper proves using a theoretic proof for simple regression model that this thesis is wrong and refutes this thesis using a counterexample. The example derived for this paper illustrates that an outlier observation detected by robust regression methods cannot be detected by the t-statistics of dummy variable. In addition, the effect of adding a dummy variable to regression on important regression statistics is investigated.

Keywords: Robust Regression, *t*-statistics, dummy variable, outlier, refute the thesis, simple regression model, detection problem, example

***Jel Classification**:* C2, C3, C51, C52

* Yrd. Doç. Dr., Başkent Üniversitesi, İİBF Ekonomi Bölümü B209, 06810 Bağlıca Ankara, Tel: 312 234 10 10 / 1714, E-Mail: arzdar@baskent.edu.tr

1. GİRİŞ

Aykırı gözlemler, modelde yer almayan etkili değişkenlerin devreye girmesi sonucu yüksek hata içeren ve bu sebeple veri kümesindeki diğer gözlemlerin sahip olduğu (doğrusal) davranışa aykırı bir davranış sergileyen gözlem olarak tanımlanabilir. En küçük kareler (EKK) tahmin edicisi ya da verinin tamamını dikkate alan başka bir tahmin edici böyle bir durumda çökmekte (Rousseeuw ve Leroy, 1987) ve aykırı gözlem miktarına göre EKKler hem yansızlık hem de tutarlılık özelliğini kaybetmektedir. Veride tedbir almadan aykırı gözlem bırakılması durumunda, diğer gözlemler ile ulaşılacak bilimsel çıkarsama sonuçlarının tam tersi sonuçlara ulaşılmasına sebep olabilmektedir (Rousseeuw ve Van Aelst, 1999). Bu durum büyük bir tehlike yaratmaktadır, çünkü EKK tahminleri çok yaygın olarak kullanılmakta ve uzun zamandan beridir en kaliteli verilerin bile aykırı gözlem içerdiği, buna ilaveten kaliteli iktisadi veri olmaya çok az sayıda aday veri olduğu bilinmektedir (Zaman *vd.*, 2001).

Gözlemler içinde aykırı gözlem bulunması şüphesinde, şüpheli gözlemlere kukla değişkenler atayarak aykırı gözlem olup olmadığını sınamak, en kolay aykırı tespit yöntemidir. Bunun sebebi, makalenin takip eden bölümlerinde gösterilebileceği gibi, bir gözleme kukla değişken atamak, o gözlemin etkisini bağlaşım tahminlerinden (Greene, 2002) ve değişirlikten (varyanstan), o verinin silinmesi (Studenmund, 2002) ile aynı etkiyi yaratmaktadır. Kukla değişken atanmış gözlemlerin sonuçlar üzerindeki etkisi bu şekilde elenirken kukla değişken tarafından etkisi elenmeyen bir tane bile aykırı gözlem, EKKlerin yine çökmesini sağlayacaktır. Böyle bir şüphe durumunda kukla değişken kullanılmadan dirençli (robust) bağlaşım yöntemleri ile bütün aykırı gözlemlerin eş-anlı tespiti zorunludur.

Bilimsel endeksler tarandığında, 2000 yılından sonra yapılmış çok sayıda çalışmada, halen bağlaşım analizinde bir gözlemin aykırı gözlem olduğundan şüphelenildiğinde o gözlem için bir kukla değişken atanmaktadır. Bir gözlemin aykırı gözlem olduğuna ise t-istatistiğine bakarak karar verilmektedir. Bu çalışma, eğer bir ekonomik şok, teknolojik buluş, doğal afet veya yanlış bir kayıt sonucu oluşmuş bir aykırı gözlem varsa eksik olan değişkenin kukla değişkenler tarafından temsil edilebilirliği tezini irdelemektedir. Buna ilaveten basit bağlaşım modeli için EKK yöntemi kullanıldığında istatistikler kukla değişken kullanımından nasıl etkilenmekte olduğu kuramsal olarak incelenmektedir. Bu çalışmada t-istatistiğinin her zaman bu konuda başarılı olmadığı kuramsal olarak ispatlanmakta ve bu çalışma için hazırlanmış örneklerde gösterilmektedir. Bu sayede bağlaşımda şüpheli gözlemlerin aykırı gözlem olup olmayacağı daha güvenilir bir biçimde tespit edilebilecektir.

Bu makalenin ikinci bölümü, kukla değişkenlerle t-istatistiğinin birlikte kullanılması sonucunda istatistiklerin nasıl etkilendiğini kuramsal olarak göstermektedir. Bu teorik bulgularla, bir aykırı gözlemi veri kümesinden çıkarmak ya da kukla değişkenle bırakmak konusunda da yorum yapılmaktadır. Takip eden üçüncü bölüm kuramsal bulgulardan çıkarsama yapmakta ve dirençli bağlaşım yazınında aykırı gözlem tespit etmek için kullanılan yöntemlerin bağlaşım sonuçlarından nasıl elde edileceği gösterilmektedir. Yine bu bölümde bir örnekle t-istatistiğinin yetersiz kaldığı gösterilmektedir. Bir tezi çürütmek için bir karşı bir tez ya da örnek yeterli olmaktadır, bu çalışma bu çürütme bilimsel katkısını sağlamaktadır. Son bölüm sonuçları özetlemektedir.

2. KUKLA DEĞİŞKEN MODELİ

Bir tezi çürütmek için basit bir örnek bile yeterli olabileceğinden, sonuçların kolay türetilebilmesi için $y_i= \beta_1+ \beta_2 x_i+ \beta_k k_i+ u_i$ (i=1,...,n) basit bağlaşım modeli kullanılmaktadır. Bu modelde β anakütle bağlaşım parametrelerini, y_i bağımlı değişken gözlemlerini, x_i bağımsız değişken gözlemlerini, u Normal dağılımdan gelen hata terimini, k_i kukla değişken değerlerini, b EKK tahmin edicileri parametrelerini, $\hat{u}_i = y_i - b_1 - b_2 x_i - b_k k_i$ kalıntı değerlerini ve n gözlem sayısını temsil etmektedir. Bu çalışmada aykırı gözlemin son gözlem olduğu varsayılmakta, bu sebeple son aykırı gözlem için k_n=1 değerini almakta, diğer gözlemler için k_i=0 (i=1,...,n-1) değerini almaktadır. Bu model için EKK tahmin edicileri aşağıdaki gibi çıkmaktadır.

$$b_1 = \frac{\sum_{i=1}^{n-1} x_i^2 \sum_{i=1}^{n-1} y_i - \sum_{i=1}^{n-1} x_i y_i \sum_{i=1}^{n-1} x_i}{(n-1)\sum_{i=1}^{n-1} x_i^2 - (\sum_{i=1}^{n-1} x_i)^2} \quad b_2 = \frac{(n-1)\sum_{i=1}^{n-1} x_i y_i - \sum_{i=1}^{n-1} x_i \sum_{i=1}^{n-1} y_i}{(n-1)\sum_{i=1}^{n-1} x_i^2 - (\sum_{i=1}^{n-1} x_i)^2}$$

$$s_k^2 = \frac{\sum_{i=1}^{n-1}(y_i - b_1 - b_2 x_i)^2}{n-3} = \frac{\sum_{i=1}^{n-1}\hat{u}_i^2}{n-3}$$

$$b_k = y_n - b_1 - b_2 x_n \tag{1}$$

Aykırı gözlem kukla değişkenle temsil edildiğinde ve tahminler incelendiğinde, aykırı gözlemin etkisinin sadece denklem (1)de yer alan kukla değişkenin tahmin edilen parametre katsayısında olduğu görülmektedir, çünkü *n*. gözlemle ilgili bilgi sadece bu denklemde yer almaktadır. Tahmin ediciler, daha önceki çalışmalarda belirtildiği gibi, bir gözlem için kukla değişken kullanıldığında Greene (2002:117) EKK parametrelerinin, bu gözlemi silip bağlaşım yapıldığında elde edilen parametreler ile aynı olduğu, buna ilaveten Studenmund (2002:224) değişirliğin (s^2) de aynı özelliği gösterdiği sonucunu desteklemektedir.

$$\hat{u}_n = y_n - b_1 - b_2 x_n - b_k = 0 \tag{2}$$

Önemli bir özellik, kukla değişken ile temsil edilen gözlemin kalıntı değerinin denklem (2)de gösterildiği gibi sıfır çıkmasıdır. Bu özellik çoklu bağlaşım modeli için de geçerli olmaktadır, çünkü yapısı gereği kukla değişken sadece atandığı gözlem tarafından etkilenmektedir. EKK amaç fonksiyonunda indirgeme yapıldığında, EKKler yöntemi kukla değişkeninin kalıntısını sıfır yaparak hem ilgili kalıntıyı hem de genel toplamı indirgemektedir. Peki burada akla gelecek bir soru aykırı gözlemin sahip olduğu yüksek hata miktarının nereye gittiği olacaktır. Denklem (1) incelendiğinde kukla değişkenin katsayısının aslında hatanın hepsini temsil ettiği görülecektir.

2.1 Kukla Değişkenin Katsayısı Bir Kalıntıdır

Denklem (1) incelendiğinde *n*. aykırı gözlemin etkisin içermeyen ve sadece diğer gözlemlerin yaptığı b_1 ve b_2 tahminleri kullanılarak bir kalıntı hesaplanmaktadır. Kısaca, kukla değişkenin katsayısı aykırı gözlemin sahip olduğu yüksek hata miktarı temsil eden bir kalıntıdır.

Dirençli bağlaşım analizinde aykırı gözlemleri tespit etmek için değişik yöntemler mevcuttur. Genel olarak şüpheli gözlemler veri kümesinden çıkarılıp EKK veya başka bir yöntemle parametreler hesaplanmakta ve bu parametreler sonucunda şüpheli gözlemlerin kalıntıları incelenmektedir. Eğer şüpheli gözlemler bu tahminden çok sapma gösterirse o zaman aykırı gözlem olarak tespit edilmektedir. Bu sebeple, kukla değişkenin katsayısı şüpheli gözlemin diğer gözlemlerin göstermiş olduğu doğrusallıktan uzaklığı göstereceği için, bu yöntem aykırı değerleri tespit etmek için kullanılabilir. Fakat biraz sonraki bölümlerde gösterileceği gibi *t*-istatistiği bu iş için kullanılamaz.

2.2 Kukla Değişkenin Bağlaşım Sonuçlarına Etkisi

Tahmin ediciler ve değişirlik açısından aykırı gözlem olduğundan şüphelenilen gözlemi veriden çıkarmakla onu kukla değişkenle temsil etmek açısından bir fark yoktur, fakat diğer önemli istatistikler olan belirleme katsayısı (R^2) bu durumdan etkilenmektedir. Eğer kukla değişkenli bağlaşımın belirleme katsayısı R_k^2 ve aykırı gözlem olmadan bağlaşımın belirleme katsayısı, R^2 ile gösterilecek olursa o zaman alttaki ilişki oluşmaktadır:

$$1-\frac{\sum_{i=1}^{n-1}\hat{u}_i^2}{\sum_{i=1}^{n}(y_i-\frac{1}{n}\sum_{i=1}^{n}y_i)^2}=R_k^2\geq R^2=1-\frac{\sum_{i=1}^{n-1}\hat{u}_i^2}{\sum_{i=1}^{n-1}(y_i-\frac{1}{(n-1)}\sum_{i=1}^{n-1}y_i)^2} \tag{3}$$

Denklem (3) incelendiğinde iki belirlilik katsayısı arasındaki fark $(y_n-\frac{1}{n}\sum_{i=1}^{n}y_i)^2\geq 0$ teriminden kaynaklanmaktadır. Veri kümesinde yer alan bir aykırı gözleminin y_n değerinin istisnai durumlar dışında, diğer gözlemlerin ortalamasına eşit olma olasılığının sıfıra yakın olduğu düşünülecek olursa o zaman denklem (3) eşitlik sağlanmayacaktır. Buradan çıkan önemli bir sonuç bağlaşıma eklenmiş kukla değişkenlerin belirlilik katsayılarının arttırıcı yönde etki yaratabilecekleridir. Ne kadar çok kukla değişken kullanılacak olursa o kadar yüksek belirlilik katsayıları elde edilecektir. Bu durum çoklu bağlaşım modeli için de geçerli olacaktır, çünkü R^2 kalıntıları ve bağımlı değişkeni dikkate alan bir istatistiktir. Kalıntılar çok sayıda bağımsız değişkenin etkisini içerdiği için R^2 kolaylıkla hesaplanabilmektedir.

Basit bağlaşım modeli için *F*-istatistiği kullanılamamaktadır, fakat kukla değişkenlerin kalıntıları kullanılarak çoklu bağlaşım modeli için *F*-istatistiğinin davranışı belirlenebilir.

Kukla değişkenli bağlaşımda aykırı gözlem olan gözlemin kalıntısı denklem (2) kullanılarak $u_n = 0$ bulunmaktadır. Bunun anlamı kukla değişkenleri kalıntılarının değişirlik çözümleme (ANOVA) analizinde etkisinin olmamasıdır. Bu sebeple bağlaşımda ne kadar değişken olursa olsun ANOVA yönteminde etkili olacak istatistikler kukla değişkenler tarafından temsil edilmeyen gözlemler tarafından belirlenecektir. Yapılacak F-testi ise kukla değişken katsayısı dışındaki katsayıların eşanlı olarak sıfır olma önsavı olacaktır. Denklem (2)'de gösterildiği gibi her kukla değişkenle R^2 artacaktır, eğer l bağlaşımdaki kukla değişken sayısı ise o zaman $dR^2/dl > 0$ olacaktır. Buna ilaveten, d terimi bağlaşımdaki kukla değişken dışındaki değişken sayısını verecek olursa, çoklu bağlaşım analizinde aşağıdaki F-istatistiği formülü ve R^2 kullanılarak alttaki yorum yapılabilir.

$$\frac{dF}{dl} = \frac{(1-n)R^2(1-R^2) + (n-d-l)(d+l-1)\dfrac{dR^2}{dl}}{(d+l-1)^2(1-R^2)^2} \tag{4}$$

Denklem (4)'te çıkan terim için kesin bir işaret belirlemek mümkün değildir, dolayısıyla bağlaşıma eklenen kukla değişken sayısı arttıkça F-istatistiği artabilir ya da azalabilir. Fakat gözlem sayısı fazla ($n >> d, l$) ise bu denklem (4) alttaki denklem biçiminde yazılabilir:

$$\frac{dF}{dl} = \frac{n[(d+l-1)\dfrac{dR^2}{dl} - R^2(1-R^2)]}{(d+l-1)^2(1-R^2)^2}$$

Bu denklem incelendiğinde $\max R^2(1-R^2) = 0.25$ olmakta (çünkü $0<R^2<1$) ve büyük ihtimalle $(d+l-1)dR^2/dl$ terimi (eğer kukla değişken aykırı terimi temsil ediyorsa dR^2/dl büyük bir değer olacağı için) daha büyük olacaktır. Buna ilaveten çoklu bağlaşım modelinde $d>1$ olacağı için $dF/dl > 0$ çıkma olasılığı çok kuvvetlidir. Bu durum takip eden örnekte de gösterilmiştir.

$$\frac{d\overline{R}^2}{dl} = (n-1)\frac{\dfrac{dR^2}{dl}(n-d-l) - (1-R^2)}{(n-d-l)^2} \tag{5}$$

Aynı durum ayarlanmış belirlilik katsayısı $\bar{R}^2 = 1 - ({}^{n-1}/_{n-d-l})(1-R^2)$ için de geçerlidir. Denklem (5) için kesin hüküm vermek imkânsızdır, fakat gözlem sayısı fazla ise ($n >> d$, l) denklem (5) alttaki denklem biçiminde yazılabilir:

$$\frac{d\bar{R}^2}{dl} = \frac{dR^2}{dl} - \frac{(1-R^2)}{n}$$

Bu denklemde $(1 - R^2)$ teriminin etkisi gözlem arttıkça azalmakta ve alttaki denklemde gösterildiği gibi ayarlanmış belirlilik katsayısı $\bar{R}^2$, belirlilik katsayısı ile aynı davranışı göstermektedir. Bu sebeple gözlem sayısı fazla ise $d\bar{R}^2/dl > 0$ olacaktır.

$$\lim_{n\to\infty} \frac{d\bar{R}^2}{dl} = \frac{dR^2}{dl} > 0$$

Modele kukla değişken eklemenin önemli istatistikler üzerinde etkili olduğu görülmektedir. Özellikle model kurma sürecinde etkili olan R^2 değerinin artması yanıltıcı sonuçlar doğuracaktır. Buna ilaveten, aykırı değeri bağlaşımda kukla değişkenle bırakmalı mı yoksa aykırı gözlemleri veri kümesinde çıkarmalı mı sorusuna bu şekilde yanıt bulunabilmektedir. Bu yanıt, kukla değişkenlerin modeli iyileştirmeden önemli istatistikleri iyileştirmesi sebebi ile aykırı gözlemlerle kukla değişkenleri çıkarmanın daha bilimsel olacağıdır.

3. KUKLA DEĞİŞKEN İLE AYKIRI GÖZLEM TESPİTİ

Bölüm 2.1'de belirtildiği gibi kukla değişkenlerin katsayısı diğer gözlemlerin oluşturduğu doğrusallığa uzaklığı ölçen dirençli bir kalıntıdır. Buna ilaveten, daha önce belirtildiği gibi kukla değişkenli bağlaşımda bağlaşım değişirliği aykırı gözlemlerden etkilenmemektedir. Bu bilgiler biraraya getirildiğinde altta yer alan istatistik dirençli bağlaşımda aykırı gözlem tespit etmek için kullanılan standardize edilmiş dirençli kalıntı (*SDK*) değeri olmaktadır. Bu değer 2 veya 2.5'ten büyük olması durumunda bir gözlem dirençli bağlaşım yöntemleri tarafından aykırı değer olarak tespit edilmektedir (Rousseeuw ve Leroy, 1987).

$$SDK = \frac{b_k}{\sqrt{s_k^2}} = \frac{b_k}{s_k} \tag{6}$$

Basit bağlaşım modeli için denklem (6) ve kukla değişkenin t-değeri karşılaştırılacak olursa ikisi arasındaki fark ortaya çıkmaktadır.

$$t_k = \frac{SDK}{\sqrt{1 + \dfrac{\sum_{i=1}^{n-1} x_i^2 - 2x_n \sum_{i=1}^{n-1} x_i + (n-1)x_n^2}{(n-1)\sum_{i=1}^{n-1} x_i^2 - (\sum_{i=1}^{n-1} x_i)^2}}} \tag{7}$$

Öneri: Kukla değişkenin t-istatistiği x yönünde aşırı değer alan aykırı değerleri tespit edemez.

İspat:

Denklem (7) incelenecek olursa EKKde kukla değişkenin t-değeri hem *SDK* değerinden farklı olmakta, hem de aykırı gözlemden (x_n) etkilenmektedir. Karekök içindeki terim, aykırı gözlemin değerine göre çok küçük değerler alabilmekte, böylece t-değerini *SDK* değerinin altına indirebilmekte ve bu durumda bu istatistik güvenilmez olmaktadır.

Kritik t_c değeri için t tablosu incelendiğinde %5 anlamlılık düzeyinde çok sayıda gözlem için kritik değer en düşük değeri olan 1.96 olmaktadır. Böylece $t < 1.96 = t_c$ olan fakat SK > 2.5 olan aykırı değer gözlemi mümkün olmaktadır. Denklem (7) kullanılarak takip eden eşitsizliği sağlayan x_n aykırı değeri mevcuttur.

$$t_k = 2.5 \cdot \frac{1}{\sqrt{1 + \dfrac{\sum_{i=1}^{n-1} x_i^2 - 2x_n \sum_{i=1}^{n-1} x_i + (n-1)x_n^2}{(n-1)\sum_{i=1}^{n-1} x_i^2 - (\sum_{i=1}^{n-1} x_i)^2}}} < 1.96$$

$$1.2755 < \sqrt{1 + \frac{\sum_{i=1}^{n-1} x_i^2 - 2x_n \sum_{i=1}^{n-1} x_i + (n-1)x_n^2}{(n-1)\sum_{i=1}^{n-1} x_i^2 - (\sum_{i=1}^{n-1} x_i)^2}}$$

$$1.6269 < 1 + \frac{\sum_{i=1}^{n-1} x_i^2 - 2x_n \sum_{i=1}^{n-1} x_i + (n-1)x_n^2}{(n-1)\sum_{i=1}^{n-1} x_i^2 - (\sum_{i=1}^{n-1} x_i)^2}$$

$$0.6269[(n-1)\sum_{i=1}^{n-1} x_i^2 - (\sum_{i=1}^{n-1} x_i)^2] < [\sum_{i=1}^{n-1} x_i^2 - 2x_n \sum_{i=1}^{n-1} x_i + (n-1)x_n^2]$$

$$0 < [\{(1.6269 - 0.6269n)\sum_{i=1}^{n-1} x_i^2 + 0.6269(\sum_{i=1}^{n-1} x_i)^2\} - 2x_n \sum_{i=1}^{n-1} x_i + (n-1)x_n^2]$$

Bu ikinci dereceden, x_n, aykırı gözlem değerine bağlı bir polinomdur. Bu polinomun $(n-1)x_n^2$ katsayısı pozitif olduğu için ($n > 1$), $x_n \rightarrow \pm\infty$ x yönünde büyük değerler aldığında, o zaman bu polinom hep pozitif olabilmektedir. Kısaca *t*-istatitiği 1.96'dan küçük olurken *SKD* 2.5tan büyük olabilmektedir.

Fakat son eşitsizlik x_n x yönünde büyük değerler almadan da sağlanabilmektedir. Eğer bu eşitsizlikte veri sebebi ile discriminant hep negatif çıkarsa o zaman herhangi bir aykırı değerde bu eşitsizlik sağlanacak ve *t*-istatistiği ile tespit edilemeyecektir.

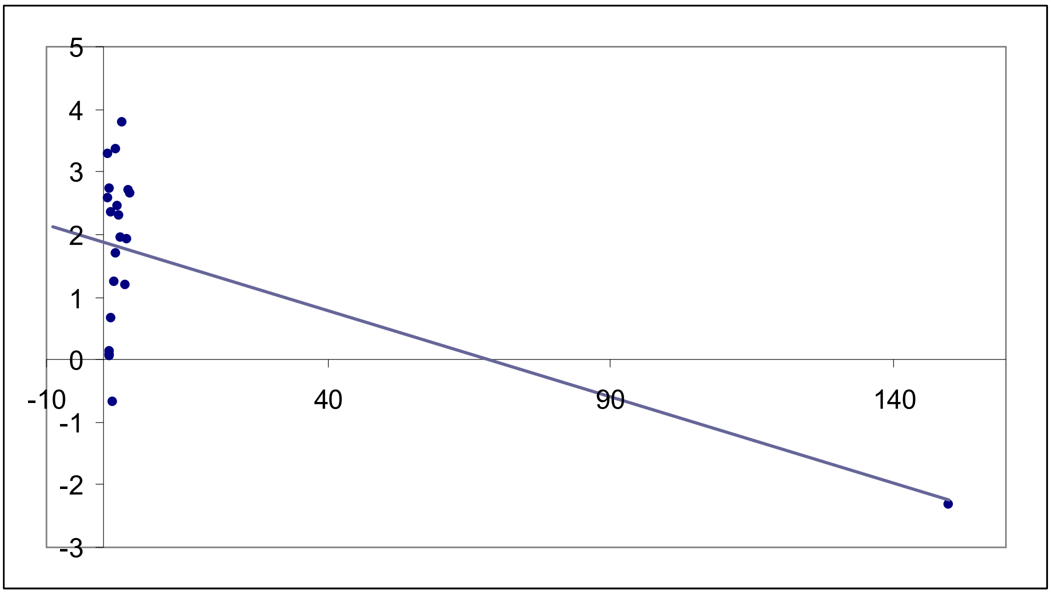
Buna ilaveten bu ispat, aykırı gözlem tespiti için kritik bir değerin tespitinin ancak aykırı gözlemin dağılımı bilinirse mümkün olacağını söylemekte ki bu durum her veri için ayrı kritik değer hesaplamasını gerektirmektedir. Takip eden bölümde bir örnekle bu durumlar gösterilmektedir.

3.1 Örnek

Tablo 1: Basit bağlaşım model verisi

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
x	1.010	1.040	1.090	1.160	1.250	1.360	1.490	1.640	1.810	2.000	2.210	2.440	2.690	2.960	3.250	3.560	3.890	4.240	4.610	5.000	150.000
y	2.569	3.278	0.071	2.725	0.047	0.130	0.658	2.359	-0.673	1.246	1.687	3.350	2.459	2.312	1.936	3.795	1.198	1.921	2.696	2.643	-2.310

Şekil 1: Tablo 1'de yer alan gözlemlerin gösterimi



Tablo 1, $y_i = 1 + 0.3369 \cdot x_i + u_i$ ($u_i \sim N(0,0.1)$) modeli için rastgele üretilmiş ve kuramı destekleyecek şekilde aykırı gözlem ilave edilmiş basit bağlaşım modelinin verisini içermektedir. Sonuncu (21.) gözlem, bağlaşımın kukla değişkeninin *t*-değerinin anlamsız olacağı bir aykırı gözlemdir. Şekil 1'de gözlemlerin çoğunun sahip olduğu doğrusallıktan EKK tahmin edicisinin sapması gösterilmiştir. Tablo 2 aykırı gözlemle birlikte, tablo 3 aykırı gözlemin kukla değişkenle temsil edildiğinde ve tablo 4 aykırı gözlemsiz bağlaşım sonuçlarını sunmaktadır.

Tablo 2 incelendiğinde, bağımsız değişkenin aykırı gözlemle birlikte istatistiksel olarak anlamlı çıktığı ve model kurma hatası olup olmadığına bakmak için Durbin-Watson istatistiği hesaplandığında 1.901 değeri ile böyle bir durumun söz konusu olmadığı görülmektedir. Aykırı gözlem, EKK istatistiklerini kendi lehine çevirmesi sebebi ile tahminler gerçek anakütle parametresi için yapılacak önsavları red etmektedir.

Tablo 2'de yer alan bağlaşım sonucuna tablo 3'te yer alan kukla değişkenin *t*-değeri sebep olabilmektedir. Bu tabloda kukla değişkenin *t*-değeri (-1.70754) istatistiksel olarak anlamsız çıkmakta, fakat *SDK* (-45.2299) değeri istatistiksel olarak çok anlamlı çıkmaktadır. Dirençli bağlaşımda *SDK* değeri 2 ya da 2.5ten büyük gözlemler aykırı gözlem olarak tanımlanmaktadır.

Tablo 2: Aykırı gözlemle birlikte bağlaşım sonuçları.

R^2	0.341756587			
Düzeltilmiş R^2	0.307112197			
Standart Hata	1.259099367		DW *d*	1.901389214
Gözlem	21			
F	9.864702052			
	Katsayılar	Standart Hata	*t*-ist.	*p*-değeri
Kesişim	1.883398092	0.286924309	6.564093	2.76749E-06
x	-0.02744025	0.008736673	-3.14081	0.005381921

Tablo 3: Aykırı gözlemin kukla değişkenle temsil edildiği sonuçlar.

R^2	0.433517669			
Düzeltilmiş R^2	0.370575188			
Standart Hata	1.200053138		*SDK*	-45.2299
Gözlem	21			
F	6.887521129			
	Katsayılar	Standart Hata	*t*-ist.	*p*-değeri
Kesişim	0.992949318	0.588833824	1.686298	0.10899334
x	0.339835643	0.215251061	1.578787	0.131797011
k	-54.27829576	31.78731684	-1.70754	0.104910376

Tablo 4: Aykırı gözlem olmadan bağlaşım sonuçları.

R^2	0.121632838			
Düzeltilmiş R^2	0.072834662			
Standart Hata	1.200053138			
Gözlem	20			
F	2.49256937			
	Katsayılar	Standart Hata	t-ist.	p-değeri
Kesişim	0.992949318	0.588833824	1.686298	0.10899334
x	0.339835643	0.215251061	1.578787	0.131797011

Tablo 3 ve tablo 4 karşılaştırıldığında, beklendiği gibi katsayılar ve bağlaşımın değişirliği değişmemektedir. Kukla değişkenli bağlaşımda beklendiği gibi her iki R^2 değeri ve F-değeri de artmaktadır. Katsayılar anlamlı olmamalarına rağmen anakütle parametrelerine çok yakın çıkmakta ve anakütle parametresi için yapılacak önsav testleri kabul görmektedir.

4. SONUÇ

Bu çalışma, kurulan bağlaşım modeli ile kukla değişken ve aykırı gözlem mevcudiyetinde, bağlaşım tahminleri, değişirlik, t-istatistiği, belirlilik katsayısı (R^2), F-istatistiği hesaplanmakta ve teorik olarak aykırı gözlem tespiti için t-istatistiğinin kullanılması durumunda bu yöntemin bazı durumlarda yetersiz kaldığı ispatlanmaktadır. Bu çalışma bulgularına göre, kukla değişkenler ile aykırı gözlemleri tespit ederken dikkat etmek gerektiği ortaya çıkmaktadır. Örnekle de gösterildiği gibi halen sıklıkla yazında kullanılan t-istatistiğinin bu konuda yetersiz kalabilmektedir.

Çıkan sonuçlardan dikkate alınacak noktalardan birisi kukla değişken kullanmakla, kukla değişkenin temsil ettiği gözlemi veri kümesinden atmanın, bağlaşım katsayı tahmini ve değişirlik açısından bir farkının olmadığıdır. Fakat, doğru model arayışında R^2 istatistikleri incelenerek bu işlem yapılacak olursa kukla değişkenli modellerin R^2 değerlerinin daha fazla olacağı dikkate alınmalıdır. Kısaca aykırı gözlemleri kukla değişkenle modelde tutmak modele katkı yapmadığı halde bazı istatistikleri iyileştirmektedir. Bu sebeple aykırı gözlemleri ve kukla değişkenleri modelden çıkarmak daha bilimsel olacaktır.

Kukla değişkenlerle aykırı değer tespiti kolay olmakta ve aynı mantıkla şüpheli gözlemler için kullanılan Cook mesafesi (Cook, 1985), DFFITS (Billingsley *vd*, 1980), DFBETAS (Billingsley *vd*, 1980) istatistiklerinden daha kolay hesaplanabilmektedir. Fakat bu tespit yapılırken bu çalışmadan çıkan sonuçlar dikkate alınmalıdır. Buna ilaveten, her verinin aykırı gözlem tespiti için ayrı bir kritik *t*-değeri tablosu hazırlanması gerektiği çalışmadan çıkan başka sonuçlardan biridir.

Bir uyarı olarak, kukla değişken ile aykırı gözlem tespiti en kolay dirençli bağlaşım yöntemi olmakla birlikte, bütün aykırı gözlemler kukla değişkenler ile temsil edilmemesi durumunda sonuçların güvenilmez olacağı vurgulanmalıdır. Sonuçlardan şüphe duyulması durumunda dirençli bağlaşım yöntemleri ile veriler aykırı gözlem için taranmalıdır. Aykırı gözlemler bilimsel sonuç çıkarmaya darbe vurmaktadır, çünkü aykırı gözlem mevcudiyetinde sonuçlar bilimsel teoremleri desteklemeyebilmektedir. Buna karşı tedbir alınmaması durumunda önsav aşamasında bulunan bir çok teorem veya kanun, kabul edilmeden elenebilir. Bütün bu sebeplerden ötürü, bilimsel önsavların ya da teoremlerin savunduğu fikirler "dirençli" (bağlaşım sınamasından geçmiş) veriler ile sınanmadıkça kabul görmeleri imkansız olacaktır, çünkü sonuçlar yanıltıcı olmaktadır.

AÇIKLAMA NOTLARI

Bu çalışmada kullanılan ekonometrik terimler Türk Dil Kurumu Elektronik Ekonometri Terimleri Sözlüğü'nden yararlanılarak yazılmıştır.

KAYNAKÇA

Billingsley, P., Belsley, D. A., Kuh, E., Welsch, R. E. 1980 *Regression diagnostics: identifying influential data and sources of collinearity*. Wiley Series in Probability and Mathematical Statistics; Wiley, New York-Chichester-Brisbane. MR0576408

Cook, R. D. 1985. Detection of influential observation in linear regression. *Technometrics* 19:15–18. MR0436478

Greene, W. H. 2002. *Econometric Analysis 5*; Prentice Hall, NJ.

Rousseeuw, P. J., ve Leroy, A. M. 1987. *Robust regression and outlier detection.* John Wiley & Sons, Inc., New York. MR0914792

Rousseeuw, P. J., ve Van Aelst, S. 1999. Positive-breakdown robust methods in computer vision. *Computing Science and Statistics* 31:451-460.

Studenmund, A. H. 2002. *Using Econometrics: A Practical Guide*. Addison Wesley, London.

Zaman, A., Rousseeuw, P. J., ve Orhan, M. 2001. Econometric Applications of High-Breakdown Robust Regression Techniques. *Economics Letters*, 71:1–8.