

YARI LOGARİTMİK MODELLERDE KUKLA DEĞİŞKENLERİN KATSAYILARININ YORUMU

Yrd. Doç. Dr. Ebru ÇAĞLAYAN*
Arş. Gör. Burak GÜRİŞ**

ÖZET

Büyüme modelleri, hedonik fiyat modelleri ve ücret modelleri olarak karşımıza çıkan yarı logaritmik modellerde sürekli açıklayıcı değişkenler yanında kukla değişkenlerin de yer aldığı görülmektedir. Literatürdeki ders kitaplarında bu modellerdeki kukla değişkenlerin katsayılarının yorumlanması için kullanılan Halvorsen–Palmquist yaklaşımı yapılan çalışmalarda eleştirilmekte ve yeni yaklaşımlar önerilmektedir. Bu yaklaşımlar Kennedy ve Tam Minimum Varyans Sapmasız Tahminci yaklaşımıdır. Bu nedenle çalışmamızda hem Halvorsen-Palmquist yaklaşımı hem de diğer yeni yaklaşımlar açıklanmış ve İstanbul'da çalışanlar için tahmin edilen yarı logaritmik ücret modeli ile bu yaklaşımların sonuçları karşılaştırılmıştır.

Anahtar Kelimeler : Yarı Logaritmik Model, Kukla Değişken, Tam Minimum Varyans Sapmasız Tahminci

1.GİRİŞ

Ekonometrik modellerde katsayıların yorumu değişkenler arasındaki ilişkiyi ortaya koyabilmek için önemlidir. Modelde yer alan katsayıların yorumları modelin fonksiyonel şekline göre farklılık göstermektedir. Bilindiği gibi, yarı logaritmik modellerde katsayı yorumu yapılırken, yorumlanacak açıklayıcı değişken katsayısı 100 ile çarpılmaktadır. Açıklayıcı değişkenleri sürekli değişkenler olan yarı logaritmik regresyon modelinde,

$$\ln Y_i = \gamma + \sum_i \alpha_i X_i + \varepsilon_i$$

* Marmara Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, Öğretim Üyesi

** İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü

α_i , sürekli değişkenlerin katsayılarını ifade etmektedir. α_i 'nin 100 ile çarpılması sonucu elde edilen değer X_i 'deki çok küçük bir değişimin Y 'de meydana getireceği yüzde etkiyi ifade etmektedir. Bu değeri p_i ile ifade edersek,

$$p_i = 100 \cdot \frac{1}{Y} \cdot \frac{\partial Y}{\partial X_i} = 100 \cdot \frac{\partial \ln Y}{\partial X_i} = 100 \cdot \alpha_i$$

olacaktır. Yarı logaritmik modellerde sürekli açıklayıcı değişkenler olabileceği gibi kukla değişkenler de yer alabilmektedir. Açıklayıcı değişkenleri hem sürekli değişkenlerden hem de kukla değişkenlerden oluşan yarı logaritmik regresyon modelinde,

$$\ln Y_i = \gamma + \sum_i \alpha_i X_i + \sum_j \beta_j D_j + \varepsilon_i$$

α_i sürekli değişkenlerin katsayılarını, β_j ise kukla değişkenlerin katsayılarını ifade etmektedir. Sürekli açıklayıcı değişkenlerin bağımlı değişken üzerindeki etkilerini belirlemek için yapılan katsayı yorumu, regresyon modelinde açıklayıcı değişkenler arasında kukla değişken olması durumunda değişecektir.

Kukla değişkenin katsayısının uygun yorumu incelenen modelin dönüşümü ile daha iyi yapılabilir. Modelin dönüşümü,

$$Y = (1 + g)^D e^{(\gamma + \sum_i \alpha_i X_i)}$$

şeklinde yapılmaktadır. Burada g , modelde yer alan kukla değişkenin bağımlı değişken üzerindeki nispi etkisini ifade etmektedir ve

$$g = \frac{Y_1 - Y_0}{Y_0} = \frac{Y_1}{Y_0} - 1 = \frac{e^{\gamma + \sum_i \alpha_i X_i + \beta \cdot 0 + \varepsilon}}{e^{\gamma + \sum_i \alpha_i X_i + \beta \cdot 1 + \varepsilon}} - 1 = e^\beta - 1$$

olarak elde edilir. Burada Y_1 ve Y_0 , 0 veya 1 değerini alan kukla değişkenin değerleridir. Orjinal denklemdeki kukla değişken katsayısı,

$$\beta = \ln(1 + g)$$

ifadesine eşit olacaktır.

II. KUKLA DEĞİŞKENLERİN KATSAYILARININ YORUMU

Yarı logaritmik modellerde kukla değişkenlerin katsayılarının yorumu için literatürdeki ders kitaplarında da yaygın olarak önerilen yaklaşım Halvorsen-Palmquist yaklaşımıdır. Bu yaklaşımın bazı sorunları içermesi nedeni ile Kennedy ve Tam Minimum Sapmasız Varyans yaklaşımları geliştirilmiştir. Bu yöntemler burada yer alacaktır.

Halvorsen ve Palmquist (1980) yarı logaritmik modelde yer alan kukla değişkenin katsayısının, sürekli açıklayıcı değişken gibi 100 ile çarpılarak yorumlandığı çalışmalarını eleştirmişlerdir. Kukla değişkenin katsayısının sürekli değişken katsayısı gibi karşılaştırılması durumunda $c=g$ olduğunun varsayıldığını ve bunun yanlış bir varsayım olduğunu belirtmişlerdir. Bunun nedeni; g 'nin değeri pozitif olduğunda β 'nin g 'den küçük; g 'nin değeri negatif olduğunda ise matematiksel olarak β 'nin g 'den daha küçük ve mutlak değer olarak daha büyük değer almasıdır. g 'nin küçük değerleri için β yaklaşık olarak g 'ye eşit olmaktadır. g 'nin mutlak değeri 1'den daha küçük olduğunda c ile g arasında,

$$c = \ln(1 + g) = g - \frac{1}{2}g^2 + \frac{1}{3}g^3 - \dots |g| < 1$$

ilişkisi bulunmaktadır. Bu nedenle $c=g$ ilişkisinin doğru bir varsayım olmadığını belirtmişler ve bu yaklaşım yerine Halvorsen ve Palmquist, kukla değişkenin bağımlı değişken üzerindeki nispi etkisini,

$$g = e^\beta - 1$$

ve yüzde etkisini,

$$p_H = 100.g = 100.(e^\beta - 1)$$

olarak tanımlanan yaklaşımı ileri sürmüşlerdir. Kukla değişken katsayılarının bu şekilde yorumlanması ancak β 'nin değerinin bilinmesi durumunda geçerlidir; fakat β bilinmemekte ve tahmin edilmesi gerekmektedir. Halvorsen ve Palmquist'in yaklaşımlarında $\hat{\beta}$, $\beta = \ln(1 + g)$ 'nin tahmini olarak ele alınmaktadır. Yaygın olarak kullanılan bu yaklaşım, β ve g arasındaki ilişki belirtilmesine rağmen, g 'yi tahmin etmek için $\hat{\beta}$ 'in nasıl kullanılacağına açıklanmaması nedeni ile Kennedy (1981) tarafından eleştirilmiştir.

Kennedy, Halvorsen- Palmquist yaklaşımına göre g 'nin,

$$\hat{g} = e^{\hat{\beta}} - 1.$$

şeklinde tahmin edildiğinde sonuçların sapmalı olacağını belirtmişlerdir. Bu şekilde tahminin geçerli olabilmesi ancak $\hat{\beta}$ 'in β 'nin ve \hat{g} 'in ise g 'nin tutarlı tahminçisi olması durumunda sözkonusudur; fakat bu durumun gerçekleşmesi mümkün olmamaktadır (Kennedy, 1981,s:801 ve Garderen,2001,s:150).

İncelenen kukla ve sürekli değişkenli yarı logaritmik modellerde hata terimi log-normal dağılımlıdır ve en küçük kareler yöntemi uygulandığında, $\hat{\beta}$ etkin ve sapmasız olacaktır. Bu nedenle Kennedy (1981), kukla değişkenleri yorumlamak için $\hat{\beta}$ 'nin varyansını kullanmayı önermiş ve \hat{g} yerine g 'nin tahminçisi olarak,

$$g^* = e^{\left(\hat{\beta} - \frac{V(\hat{\beta})}{2}\right)} - 1$$

tahmincisini kullanmıştır. Burada $V(\hat{\beta})$, $\hat{\beta}$ 'nin varyansıdır. Kennedy'nin önerdiği g^* değeri, \hat{g} 'den daha az sapmalıdır. Kennedy'nin önerdiği bu tahminci, g 'nin beklenen değeri için en çok benzerlik tahminçisine eşittir (Lye ve Hirschberg, 2002,s:3). Buna göre yarı logaritmik modelde yer alan kukla değişkenin bağımlı değişken üzerindeki yüzde etkisi,

$$\hat{p}_K = 100 \left(e^{\left(\hat{\beta} - \frac{V(\hat{\beta})}{2}\right)} - 1 \right)$$

olacaktır. Kennedy'nin yaklaşımı görüldüğü gibi Halvorsen-Palmquist'in yaklaşımına benzemektedir. Her iki yaklaşımda da sadece bilinmeyen değerler ile bu değerlerin tahminleri yer değiştirilmiştir. Sonuçta elde edilen tahmincinin hala sapmalı olması nedeni ile Kennedy'in bu yaklaşımı yerine Tam minimum varyans sapmasız tahminci yaklaşımı önerilmiştir. Bu yaklaşımda tam sapmasız tahminci, tahmincinin varyansı ve varyansın tam sapmasız tahminçisi kullanılmaktadır.

Tam minimum varyans sapmasız tahminci yaklaşımında, kukla değişkenin bağımlı değişken üzerindeki yüzde etkisi,

$$\hat{p}_T = 100 \left\{ e^{\left(\hat{\beta}\right)} {}_0F_1 \left(m; -\frac{1}{2} m V(\hat{\beta}) \right) - 1 \right\}$$

olacaktır. Burada $V(\hat{\beta})$, $\hat{\beta}$ 'nin varyansı ve ${}_0F_1$ ise hipergeometrik fonksiyondur. $m=(n-k)/2$ olarak elde edilir. n örnek birim sayısı, k parametre sayısıdır (Garderen ve Shah, 2002,

s:151). Hipergeometrik fonksiyon terimi parametrelerin belirsizliği için sapma düzelticisidir. Örnek birim sayısı büyüdükçe ve tahminin standart hatası sifıra yaklaştıkça bu terimin değeri bire yaklaşır. Örnek birim sayısı çok küçük olmadığı takdirde

$$\left\{ {}_0F_1\left(m; -\frac{1}{2}mV(\hat{\beta})\right) \right\}$$

ifadesi, $e^{\left(-\frac{V(\hat{\beta})}{2}\right)}$ değerine yakın olacaktır. Elde edilen değerler m büyüdükçe Kennedy'nin tahminleri ile asimtotik olarak eşit olacaklardır. Bu nedenle Kennedy'nin tahminleri ile Tam minimum varyans sapmasız tahminci benzer sonuçlar verecektir*.

Daha sapmalı sonuçlar vermesi nedeni ile uygulamalarda Halvorsen-Palmquist yaklaşımı yerine benzer sonuçlar veren Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımlarının birlikte kullanılması daha uygun olacaktır.

III.UYGULAMA

Literatürde yapılan çalışmalar incelendiğinde en sık rastlanan yarı logaritmik modellerinin; Hedonik fiyat modelleri, büyüme modelleri ve ücret modelleri olduğu görülmektedir. Çalışmamızda 2005 yılında İstanbul'da tesadüfi olarak seçilmiş 406 çalışana yapılan anketlerden elde edilen bilgiler kullanılarak ücret modeli tahmin edilmiştir. Yarı logaritmik tahmin ettiğimiz model için, eleştirilen Halvorsen-Palmquist yaklaşımı ile yeni önerilen yaklaşımlara göre kukla değişkenlerin bağımlı değişken üzerindeki yüzde etkileri hesaplanacak ve bu yaklaşımlara göre elde edilen sonuçlar karşılaştırılacaktır.

Çalışanların ücretlerinin demografik özelliklerinden ve tecrübeden nasıl etkilendiğinin incelendiği modelde bağımlı değişken, çalışanların ücret/maaşları (LGL) olarak belirlenmiş ve modelde bu değişkenin logaritması kullanılmıştır. Açıklayıcı değişkenler sürekli ve kukla değişkenlerden oluşmuştur. Sürekli açıklayıcı değişkenler, çalışanların yaşları (YAS) ve yaşlarının karesi (YAS²) ve tecrübeleri (TC)'dir. Tecrübe, çalışanların aynı iş kolunda çalıştıkları yıl olarak ele alınmıştır. Kukla değişkenler cinsiyet (CN) ve medeni hal durumları (MD) için kullanılmış,

$$D_{1i} = \begin{cases} 0 & ERKEK \\ 1 & KADIN \end{cases} \quad \text{ve} \quad D_{2i} = \begin{cases} 0 & BEKAR \\ 1 & EVLİ \end{cases}$$

* Tam minimum varyans sapmasız tahminci ile ilgili ispatlar ve ayrıntılı bilgiler için Garderen(2001) ve Garderen ve Shah(2002) çalışmalarından yararlanılabilir.

olarak oluşturulmuştur. Ankete katılanlardan dul ve boşanmış olanların sayısının çok az olması nedeni ile bunlar bekar seçeneğinde değerlendirilmiştir.

Tablo I.'de yarı logaritmik model tahminleri ve model için uygulanan test sonuçları yer almaktadır. Tanımlama hatasının incelenmesi amacıyla uygulanan Ramsey Reset Testi'nde bağımlı değişkenin tahmini değerlerinin 2,3 ve 4. kuvvetleri kullanılmıştır. Sabit varyans varsayımının geçerliliğinin incelenmesi için White Testi uygulanmıştır. Normallik varsayımı Jarque-Bera Testi ile incelenmiştir. Model için uygulanan tüm testlerin olumlu sonuçlar verdiği görülmektedir.

Ücret modelimizde tecrübe, yaş değişkeninin katsayısının pozitif, yaşın karesinin konveks olması nedeni ile negatif olması beklediğimiz bir sonuçtur. Modelde yer alan diğer değişkenlerin etkisi sabit olduğu varsayımı altında yaşın karesinin katsayısının negatif işaretli olması, yaşın çalışanların ücretlerini gittikçe azalan hadlerde arttırdığını ifade etmektedir.

Tablo II.'de tahmin edilen ücret modelinde yer alan kukla ve sürekli değişkenlerin çalışanların ücretleri üzerindeki etkilerini belirlemek için katsayılarının yorumlanmasında önerilen yaklaşımlardan yararlanılmıştır. İkinci sütunda modelde yer alan sürekli değişkenlerin bağımlı değişken üzerindeki yüzde etkisi değişkenlerin katsayı değerlerinin 100 ile çarpılması ile elde edilmiştir. Diğer değişkenler sabitken tecrübe ortalama ücretleri yüzde 3.78, yaş ortalama ücretleri yüzde 8.61 arttırmakta ve yaşın karesi ise ortalama ücretleri yüzde 0.10 azaltmaktadır.

Tablonun 3.,4. ve 5. sütunlarında sırası ile Halvorsen-Palmquist, Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımları ile kukla değişkenin bağımlı değişken üzerindeki yüzde etkisi hesaplanmıştır. Tablodan da açıkça görüldüğü gibi tüm kukla değişkenlerin bağımlı değişken üzerindeki yüzde etkisi Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımları için aynı bulunmuştur. Diğer değişkenlerin değeri sabitken kadınların ortalama ücretleri (erkeklerden) Halvorsen-Palmquist yaklaşımına göre yüzde 18.54 daha düşük, Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımlarına göre yüzde 18.79 daha düşüktür. Medeni durum değişkenini incelediğimizde ise diğer değişkenlerin değeri sabitken evlilerin ortalama ücretleri (bekarlar da) Halvorsen-Palmquist yaklaşımına göre yüzde 16.64 daha yüksek, Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımlarına göre yüzde 16.27 daha yüksek olduğu görülmektedir. Cinsiyet için Halvorsen-Palmquist yaklaşımı ile Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımı arasında yüzde 0.25; medeni hal için yüzde 0.37 fark bulunmuştur.

IV. SONUÇ

Yarı logaritmik modellerde kukla değişkenlerin katsayılarının yorumu için önerilen Halvorsen-Palmquist yaklaşımının sapmalı sonuçlar verdiği belirlenerek yerine Kennedy ve Tam Minimum Varyans Sapmasız Tahminci yaklaşımları önerilmiştir. Çalışmamızda bu yaklaşımlar ve farkları teorik olarak incelenmiş ve yaklaşımlar arasındaki farkların ve işleyişlerin ortaya konulması amaçlanmıştır. Bu nedenle uygulamada yaygın olarak kullanılan yarı logaritmik modellerden biri olan ücret modeli, İstanbul'da çalışanlar için tahmin edilmiş ve çalışanların ücretlerinin demografik özelliklerden ve tecrübeden nasıl etkilendiği incelenmiştir. Model için klasik model belirleme testlerinden bazıları uygulanarak sonuçları verilmiştir. Tahmin edilen katsayıların işaretleri beklentilere uygun çıkmıştır. Modelde yer alan sürekli değişkenlerin bağımlı değişken üzerindeki yüzde etkisi klasik yaklaşım ile hesaplanırken; kukla değişkenlerin bağımlı değişken üzerindeki yüzde etkisi üç farklı yaklaşım ile hesaplanmıştır.

Halvorsen-Palmquist yaklaşımı parametre değeri bilindiğinde doğru yorum sağlasa da, parametreler tahmin edildiğinde sapmalı sonuç vermektedir. Bu yaklaşımı eleştiren Kennedy daha az sapmalı bir yaklaşım önermiştir. Çalışmamızda ele aldığımız son yaklaşım Tam minimum varyans sapmasız tahminci yaklaşımıdır. Teoride bahsettiğimiz gibi, uygulamada Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımlarının sonuçları aynı çıkmıştır. Halvorsen-Palmquist'in yaklaşımı ile diğer iki yaklaşımın sonuçları arasında fark olduğu görülmüştür.

Yapılacak uygulamalarda yarı logaritmik modellerde kukla değişken katsayılarının yorumlanmasında Halvorsen-Palmquist yaklaşımı yerine daha az sapmalı sonuçlar verecek Kennedy ve Tam minimum varyans sapmasız tahminci yaklaşımlarının birlikte kullanılması daha uygun olacaktır.

TABLO I. YARI LOGARİTMİK ÜCRET MODELİ

Deđiřken	Katsayı	Standart hata	t-testi	p-deđeri
SABİT	5.121134	0.399345	12.82384	0.0000
YAŐ	0.086109	0.022403	3.843675	0.0001
YAŐ ²	-0.001030	0.000270	-3.807316	0.0002
MD	0.153953	0.079330	1.940662	0.0530
TC	0.037790	0.021551	1.753513	0.0803
CN	-0.205107	0.077875	-2.633788	0.0088
Bađımlı Deđiřken : LGL				
n=406				
F=12.17273 p-deđeri : 0.0000				
TEST	Test İstatistiđi Deđeri		p-deđeri	
Ramsey Reset	0.303706		0.822719	
White	24.29847		0.111550	
Jarque-Bera	3.878170		0.143836	

TABLO II. YARI LOGARİTMİK ÜCRET MODELİNDE KATSAYILAR

Değişken	Tahmin Değerleri	Klasik Yaklaşım	Halvorsen-Palmquist Yaklaşımı	Kennedy Yaklaşımı	Tam Minimum Varyans Sapmasız Tahminci Yaklaşımı
YAŞ	0.086109 (0.022403)	8.6109			
YAŞ ²	-0.001030 (0.00027)	-0.1030			
MD	0.153953 (0.07933)		16.643606	16.277149	16.277149
TC	0.037790 (0.021551)	3.7790			
CN	-0.205107 (0.077875)		-18.543984	-18.790606	-18.790606

Parantez içindeki değerler standart hatalardır.

YARARLANILAN KAYNAKLAR

GARDEREN K.J. (2001), Optimal Prediction in Loglinear Models, Journal of Econometrics, 104, 119-140.

GARDEREN K.J., SHAH C.(2002), Exact Interpretation of Dummy Variables in Semilogarithmic Equations, Econometrics Journal, 5, 149-159.

HALVORSEN, R., PALMQUIST R.(1980), The Interpretation of Dummy Variables in Semilogarithmic Equations, American Economic Review, 70, 474-475.

KENNEDY, P.E.(1981), Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations, American Economic Review, 71, 801.

LYE J.N., HIRSCHBERG J.G. (2002), Test of Inference for Dummy Variables in Regressions with Logarithmic Transformed Dependent Variables, [www. Business.qut.edu.au/ esam02/ program/ papers/ Hirschberg_Joseph.pdf](http://www.Business.qut.edu.au/esam02/program/papers/Hirschberg_Joseph.pdf)