

## Logaritmik ve Yarı Logaritmik Ölçüm Hatalı Modeller: SIMEX Yönteminin Etkinliği

Şahika GÖKMEN\*

Rukiye DAĞALP\*\*

Geliş Tarihi (Received): 26.11.2019 – Kabul Tarihi (Accepted): 27.01.2020

### Öz

Doğrusal regresyon analizinde, açıklayıcı değişkenler hata ile ölçüldüğünde regresyon parametreleri sapmalı tahmin edilmektedir. Sapmalı tahminler ise yanlış sonuç çıkarımları yapmaya, değişkenler arası ilişki yapısını bozmaya ve kestirimlerin sapmalı olması gibi sonuçlara neden olmaktadır. Ölçüm hatasına sahip açıklayıcı değişkenin olduğu bu tip modellere ölçüm hatalı modeller denilmekte ve Simülasyon-Ekstrapolasyon (SIMEX), Regresyon Kalibrasyon gibi yöntemler ile bu modellerin parametreleri daha sapmasız olarak tahmin edilebilmektedir (Carroll v.d., 2006). Pek çok ekonomik verinin tam olarak ölçülememesi günümüzde, özellikle sosyal bilimlerde, bu konuyu daha popüler hale getirmektedir. Diğer yandan, parametrik istatistiksel yöntemlerde normallik, doğrusallık ve sabit varyanslılık varsayımları genel olarak dikkate alınmakta ve bu varsayımların sağlanmasında etkin olan logaritmik dönüşümler, özellikle istatistiksel sonuç çıkarımı için Gauss dağılımına yaklaşım amacıyla, sıklıkla kullanılmaktadır. Bu bakımdan, “logaritmik dönüşümler açıklayıcı değişkenlerde ortaya çıkan ölçüm hatasının etkisini azaltır mı?” sorusu, bu çalışmanın temel amacını oluşturmaktadır. Çalışmada ölçüm hatalı logaritmik ve yarı logaritmik modellerin parametre tahminleri Monte Carlo simülasyon çalışması ile incelenmiş ve ölçüm hatalı modellerin parametre tahmininde en başarılı yöntem olan SIMEX yönteminin logaritmik dönüşümler karşısındaki başarısı da araştırılmıştır.

**Anahtar Kelimeler:** *Logaritmik Dönüşüm, Logaritmik Model Kalıpları, SIMEX Yöntemi, Ölçüm hataları.*

---

\* Arş. Gör. Dr., Ankara Hacı Bayram Veli Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, [sahika.gokmen@hbv.edu.tr](mailto:sahika.gokmen@hbv.edu.tr).

\*\* Doç. Dr., Ankara Üniversitesi, Fen Fakültesi, İstatistik Bölümü, [rdagalp@ankara.edu.tr](mailto:rdagalp@ankara.edu.tr).

## **Logarithmic and Semi-Logarithmic Measurement Error Models: Effectiveness of SIMEX Method**

### **Abstract**

In linear regression, the estimates of regression coefficients are biased when explanatory variables are measured with error. These biased estimates produce false inferences, disrupt the relationship between variables and cause biased prediction. These types of models with the error-prone regressors are called measurement error models and the unbiased estimators of the parameters of these models are determined by Simulation-Extrapolation (SIMEX), Regression Calibration methods. In reality, the fact that the economic data can never be accurately measured; especially in the social sciences, makes this subject more popular. On the other hand, in parametric statistical methods, the assumptions of normality, linearity and constant variance are generally considered and logarithmic transformations, which are effective in providing these assumptions, are often used to approach the Gaussian distribution for statistical inference. From this perspective, the question “do logarithmic transformations reduce the effect of measurement error on the explanatory variables?” is the main purpose of this research. In the study, the parameter estimators of logarithmic and semi-logarithmic measurement error models were examined by Monte Carlo simulation study and it was also investigated that the success of SIMEX method, the most successful parameter estimation technique of measurement error models, for the logarithmic transformations.

**Keywords:** *Logarithmic Transformation, Logarithmic Models, SIMEX Method, Measurement Error.*

## Giriş

Doğrusal regresyon modelinde parametre tahmini yapmak için hata terimlerinin birbirinden bağımsız ve sabit varyanslı olduğu varsayımı ve istatistiksel sonuç çıkarımı için de normallik varsayımı yapılmaktadır. Bu varsayımlardan biri ihlal edildiğinde de dönüşümler varsayımların sağlanabilmesi rolünü üstlenebilmektedir. Bu tip dönüşümler arasında en sık kullanılan dönüşüm logaritmik dönüşüm<sup>†</sup> olup, sivri veya çarpık dağılım göstermiş verilere uygulanabilmektedir. Böylece bu dönüşüm ile verilerin normal dağılıma uygun olması sağlanarak istatistiksel analizler kolaylaştırılabilmektedir. Bunun dışında, logaritmik dönüşümler, ekonomideki marjinal etkiler, esneklikler, oransal değişiklikler gibi farklı davranış yapılarının bulunduğu modellerin parametre tahminleri için de özel olarak tercih edilebilmektedir.

Logaritmik dönüşüm yapılmış değişkenler için kurulan doğrusal modellere genel olarak “logaritmik veya yarı logaritmik doğrusal modeller” adı verilmektedir. Basit doğrusal regresyon modeli göz önüne alındığında, hiç dönüşüm yapılmamış ve logaritmaları içeren olası dönüşümler: dönüşümü olmayan doğrusal durum, doğrusal-log modeli, log-doğrusal modeli ve log-log modeli olup Tablo 1’de gösterilmektedir. Bu tabloda yer alan modeller, logaritmik dönüşümün yapıldığı değişkene göre tek-logaritmik (dog-log, log-dog) ve çift-logaritmik (log-log) modeller isimlerini de almaktadır (Gujarati ve Porter, 2009: 159-162).

**Tablo 1.** Doğrusal, logaritmik ve yarı logaritmik model türlerinin gösterimi.

		X	
		X	log X
Y	Y	<b>Doğrusal (dog-dog)</b> $Y = \alpha + \beta x + \varepsilon$	<b>Doğrusal-log (dog-log)</b> $Y = \alpha + \beta(\log x) + \varepsilon$
	log Y	<b>Log-doğrusal (log-dog)</b> $\log Y = \alpha + \beta x + \varepsilon$	<b>Çift logaritmik (log-log)</b> $\log Y = \alpha + \beta(\log x) + \varepsilon$

Modellerde  $\alpha$  sabit terimi,  $\beta$  eğim katsayısı ve  $\varepsilon$  ise hata terimini ifade etmektedir.

Buna karşılık, model parametrelerinin tahmin sürecinde bu tip dönüşümler yapılsa dahi, doğrusal modellerdeki açıklayıcı değişkenlerin tam olarak ölçülebildiği ya da verilerin elde edilmesinde kayda değer bir hata yapılmadığı varsayılmaktadır (Edwards ve Hamilton, 1995; Belsley vd., 1980). Ancak uygulamada, özellikle iktisadi verilerin ölçülmesinde, değişkenleri hatasız/tam/kesin olarak gözlemek mümkün olmamaktadır. Bu tip değişkenlerin kullanılmasıyla tahmin edilen modellere ölçüm hatalı modeller denmektedir (Buonaccorsi, 2010). Burada belirtilmek istenen ölçüm hatası kavramı, yalnızca ölçümden değil, araştırmanın yapısından veya değişkenlerin tanımından da kaynaklanabilmekle birlikte değişkenlerin tam/gerçek değerlerinin gözlenememesi anlamına gelmektedir. Özellikle iktisadi verilerin derlenmesinde ortaya çıkan yastık altı paralar, kaçakçılık gibi sorunların yanı sıra verilerin

<sup>†</sup> Burada logaritmik dönüşüm ile sözü edilen “doğal logaritma” yani “ln” dönüşümüdür.

toplulaştırılması veya yalnızca tam olarak gözlenemiyor olması (yıllık gelir, işsizlik oranı... vb.) bu tip verilerin ölçümlerinde hataya yol açmaktadır (Bound vd., 2000; TÜİK, 2012). Yapılan analizlerde, açıklayıcı değişkende ölçüm hatalarının olması, parametre tahminlerinin sapmalı olmasına ve dolayısıyla gelecek kestirimlerinin maskelenmesine veya yanlış politikaların belirlenmesine neden olabilmektedir.

Açıklayıcı değişkende ölçüm hatası olmasıyla ortaya çıkan bu gibi etkiler literatürde “ölçüm hatası problemleri” olarak adlandırılmaktadır (Carroll v.d., 2006: 1). Ölçüm hatası problemlerinde yalnızca açıklayıcı değişkende ölçüm hataları ele alınmaktadır. Bunun nedeni, açıklanan değişkende ölçüm hatasının parametre tahminine bir etkisinin olmaması, yalnızca parametre tahminlerinin etkinlik özelliğini bozmasıdır. Parametrelerin sapsızlık özelliğini bozan, açıklayıcı değişkende ölçüm hatasıdır. Ölçüm hataları konusu Schneeweiss (1976) ve Fuller (1980) aracılığıyla dikkati üzerine çekmeye başlamakta, günümüzde de bu alanda yapılan çalışmaların artışı bu konunun önemini gittikçe artırdığını göstermektedir.

Günümüzde, yapılan çalışmalar ölçüm hatalı modellerde sapsız parametre tahminleri elde edebilmek üzerinde yoğunlaşmakta ve ölçüm hatalarının neden olduğu sorunlar daha çok dikkate alınmaktadır. Edwards ve Hamilton (1995), açıklayıcı değişkeninin tam olarak ölçülemediği durumlarda ölçüm hatalı modellerin parametre tahmin yöntemlerinden faydalanılması gerektiğini belirtmektedir. Ölçüm hatalı modellerdeki parametre tahmin yöntemleri arasında özellikle Cook ve Stefanski (1992) tarafından geliştirilen SIMEX (Simulation-Extrapolation) yöntemi en sık kullanılan yöntemlerden biridir.

Diğer yandan, literatürde, logaritmik dönüşüm yapılmış ölçüm hatalı değişkenlerin model kestirimine etkilerinin incelenmesi sürecinde Edwards ve Hamilton’ın (1995) çalışmaları dışında bir çalışmaya rastlanılamamıştır. Edwards ve Hamilton’ın (1995) çalışmasında ise ölçüm hatasının etkisini genel olarak Box-Cox dönüşüm yöntemleri üzerinden incelenmiş ve çalışmayı simülasyon bulguları ile tamamlamışlardır. Ardından Buonaccorsi (2010), ölçüm hatalarını ele aldığı kapsamlı çalışmasında logaritmik dönüşümlere özel olarak değinmiştir. Ancak burada da ölçüm hatalı gözlemlenen açıklayıcı değişkenin logaritmasının ortalama üzerindeki etkisinin tam olarak gösterilemeyeceğini, logaritmik dönüşümle birlikte ölçüm hatalarının etkilerinin de küçüleceğini belirtmiştir. Ayrıca, bu koşul altında kuramsal olarak parametre tahminlerinin nasıl etkilendiğinin gösterilememesi sorununu da dile getirmiştir. Buonaccorsi (2010) çalışmasında, bir örnek üzerinden logaritmik dönüşüm yapılsa dahi ölçüm hatasının etkisinin sürdüğünü göstermiştir. Daha sonraları ise çalışmasıyla bu alanda dikkati çeken Richardson vd. (2018) logaritmik regresyon modelleri üzerinde ölçüm hatasının etkilerini, hızlandırılmış yaşam testleri üzerindeki araştırmış, ancak bunu yalnızca Ağırlıklandırılmış En Küçük Kareler yöntemi üzerinden göstermiştir. Bu çalışmada ise, logaritmik ve yarı logaritmik ölçüm hatalı modellerin parametre tahminlerinin elde edilmesi aşamasında simülasyona dayanan SIMEX yönteminden yararlanılmıştır.

Literatürdeki çalışmalar incelendiğinde, özel olarak logaritmik dönüşümlerin etkisinin farklı durumlar altında incelenmediği dikkati çekmektedir. Bu durum, özellikle ekonometri alanındaki analizlerde ölçüm hatasının etkisinin kaybolup kaybolmadığının incelenmesi bakımından önem taşımaktadır. Aynı zamanda, literatürde, ölçüm hatalı modellerin kestirilmesi üzerine kullanılan yöntemlerin logaritmik ve yarı logaritmik modellerdeki parametre tahmin

etme noktasındaki başarısının incelendiğine de rastlanılmamıştır. Dolayısıyla bu çalışmanın amacı, ekonomik değişkenler için oldukça önemli bir yeri olan logaritmik dönüşümlerin ölçüm hatasının etkisini arındırıp arındırmadığı ve ölçüm hatasının etkisi kaybolmuyorsa, SIMEX yönteminin buradaki başarısının araştırılması çalışmanın ana fikrini oluşturmaktadır. Bu bağlamda çalışmanın iki aşaması bulunmaktadır. Çalışmada, öncelikle, açıklayıcı değişkendeki ölçüm hatasının logaritmik ve yarı logaritmik regresyon modelinde parametre tahminleri üzerinde nasıl bir etkiye sahip olduğunu araştırmak amacıyla simülasyon tekniği kullanılmıştır. Bu bağlamda, farklı ölçüm hatası düzeyine sahip açıklayıcı değişkenin yer aldığı basit doğrusal regresyon modeli üzerinden farklı örneklem büyüklükleriyle simülasyon çalışmaları yapılmış ve sonuçları irdelenmiştir. Genel olarak logaritmik dönüşüm yapılmış olsa dahi ölçüm hatasının etkisinin baskın çıktığı sonucuna varılmıştır. Çalışmanın ikinci aşamasında, yine farklı ölçüm hatası seviyeleri altında ölçüm hatalı açıklayıcı değişkene sahip logaritmik ve yarı logaritmik model parametreleri SIMEX yöntemi ile tahmin edilmiştir. Bu aşamada, doğrusal modellerde etkinliğini kanıtlamış olan SIMEX yönteminin logaritmik ve yarı logaritmik modellerdeki etkinliği tartışılmıştır.

Bu çalışma dört bölümden oluşacak şekilde düzenlenmiştir. Çalışmanın ikinci bölümünde ölçüm hatasının basit doğrusal regresyon modelindeki parametre tahminine etkisi ve logaritmik dönüşüm üzerine bir tartışma yer almaktadır. Bu bölümde aynı zamanda SIMEX yönteminin algoritmasından da söz edilmiştir. Üçüncü bölümde simülasyon çalışmasının çerçevesi anlatılmış ve elde edilen bulgular incelenmiştir. Son bölümde ise çalışmada elde edilen çıkarımlar sunulmuş ve çeşitli değerlendirmeler yapılmıştır.

## 2. Ölçüm Hatalı Basit Doğrusal Regresyon Modeli ve SIMEX Yöntemi

### 2.1. Basit Doğrusal Regresyon Modelinde Ölçüm Hatası ve Logaritmik Dönüşüm

Ölçüm hatasız gözlemlenebilen  $(Y, X)$  değişkenleri arasındaki basit doğrusal regresyon modeli

$$Y = \alpha + \beta X + \varepsilon \quad (1)$$

ile gösterilmek üzere, burada  $(Y, X)$  ölçüm-hatasız (error-free) değişkenler olarak adlandırılmaktadır. Ayrıca modelde yer alan  $\alpha$  sabit terimi,  $\beta$  eğim katsayısını ve  $\varepsilon$  ise modelin hata terimini ifade etmektedir. Regresyon varsayımları altında hata terimi olan  $\varepsilon$ , sıfır beklenen değer ve sabit varyansa sahip olmakla beraber, açıklayıcı değişkenden istatistiksel olarak bağımsızdır.

Eşitlik (1)'de gerçeği ölçülemeyen  $X$  değişkeninin yerine ölçüm hatalı (error-prone)  $W$  değişkeninin gözlenebildiği varsayılmaktadır. Çalışmada açıklanan değişkendeki ölçüm hatasının etkisi dikkate alındığı ve bu etkinin daha açık ortaya konması bakımından, açıklanan değişkende ölçüm hatası olmadığı varsayılmaktadır. Bu durumda açıklayıcı değişken ile açıklayıcı değişken yerine gözlenen değişken arasındaki ilişki

$$W = X + U \quad (2)$$

modeli ile tanımlanmaktadır. Eşitlik (2)'de verilen modelde,  $U$  değişkeni  $N(0, \sigma_U^2)$  dağılımına sahip olmak üzere ölçüm hatası olarak adlandırılmakta,  $X$  ve  $U$  değişkenleri birbirinden bağımsız olduğu varsayılmaktadır. Dolayısıyla  $W$  değişkeni de  $N(\mu_X, \sigma_W^2)$  dağılımına sahip ve

varyansı  $\sigma_W^2 = \sigma_X^2 + \sigma_U^2$  olarak tanımlanmaktadır. Burada ölçüm hatasının klasik (toplamsal) ölçüm hatası olduğu varsayılmakta ve buna ilişkin varsayımlara ait detaylı bilgi Schneeweiss (1976) ve Fuller (1980) de verilmektedir. Başka ölçüm hatası modelleri ise Carroll vd. (2006) ve Buonnaccorsi (2010) de bahsedilmiştir. Ölçüm hatalı  $W$  değişkeninin açıklayıcı değişken olarak yer aldığı Eşitlik (1)'deki regresyon parametrelerinin EKK ile tahmin edilmesiyle elde edilen parametre tahminleri naif tahminler (naive estimators) olarak adlandırılmaktadır. Bu durumda Eşitlik (1) modelinin kestirim denklemi,

$$\hat{Y} = \hat{\alpha}_{naif} + \hat{\beta}_{naif} W \quad (1')$$

olarak gösterilmektedir. Eşitlik (1) ve (1')'de verilen doğrusal regresyon modelindeki eğim katsayıları

$$\hat{\beta}_{gerçek} = \frac{S_{XY}}{S_{XX}} \quad ve \quad \hat{\beta}_{naif} = \frac{S_{WY}}{S_{WW}} \quad (3)$$

ile tahmin edilebilmektedir. Burada  $S_{XY}$  ve  $S_{WY}$  sırasıyla  $(X, Y)$  ve  $(W, Y)$  arasındaki kovaryanslara,  $S_{XX}$  ve  $S_{WW}$  ise  $X$  ve  $W$  değişkenlerinin varyanslarına karşılık gelmektedir. Ölçüm hatası  $U$ 'nun açıklayıcı ve açıklanan değişkenler ile istatistiksel olarak bağımsız olduğunun varsayılmasından dolayı  $S_{WY} \cong S_{XY}$  olarak ifade edilebilmektedir. Bu durumda Eşitlik (3)'de yer alan  $\hat{\beta}_{naif}$ ,

$$\begin{aligned} \hat{\beta}_{naif} &\cong \frac{S_{XY}}{S_{WW}} \\ &\cong \frac{S_{XY}}{S_{XX} + S_{UU}} \\ &\cong \frac{S_{XX}}{S_{XX} + S_{UU}} \cdot \frac{S_{XY}}{S_{XX}} \\ &\cong \hat{\lambda} \cdot \beta_{gerçek} \end{aligned} \quad (4)$$

olarak düzenlenebilmektedir. Burada,  $\hat{\lambda}$  *güvenirlilik oranı* (reliability ratio) olarak adlandırılan ve  $[0,1]$  aralığında değer alan  $\lambda$  parametresinin tahmin edicisidir. Basit doğrusal regresyon modelinde bu katsayıya aynı zamanda *azaltıcı faktör* (attenuation factor) de denmektedir. Bunun nedeni,  $\sigma_U^2 > 0$  olduğunda  $\lambda$  güvenirlilik oranı nedeniyle gerçek (ölçüm-hatasız) eğim katsayısının x-eksenine yakınsaması yani eğimin sifıra yakınsamasıdır.

Daha önce söz edildiği gibi, literatürde açıklayıcı değişkende logaritmik dönüşüm yapılmasının etkisine pek fazla değinilmediği dikkati çekmektedir. Edwards ve Hamilton (1995) çalışmalarında, ölçüm hatasının etkisini Box-Cox dönüşümleri üzerinden göstermiş ancak özel olarak logaritmik dönüşüme değinmemişlerdir. Buonnaccorsi (2010: 159-160) ise kitabında, logaritmik dönüşümlere ayrıca yer vermiş ve söz konusu kitapta, açıklayıcı değişkendeki logaritmik dönüşümün, değişkenin gözlenmesinde ortaya çıkan ölçüm hatasının etkisini ne ölçüde yok ettiğinin önemli olmakla birlikte, bunun kuramsal olarak ortaya konmasında soru işaretleri olduğunu belirtmişlerdir.

## 2.2. SIMEX Yöntemi

Cook ve Stefanski (1994) tarafından geliştirilmiş SIMEX yöntemi simülasyona dayalı bir parametre tahmin tekniğidir. Bu yöntem ile, Eşitlik (1)'deki açıklayıcı değişkenin Eşitlik (2)'deki gibi bir ölçüm hatasına sahip olması varsayımı altında; sapmalı olan Eşitlik (1)'deki eğim katsayısı sapmasız olarak tahmin edilmeye çalışılmaktadır. Carroll vd. (2006)'ne göre bu yöntemin temel fikri, ölçüm hatasının etkisinin simülasyon yardımıyla belirlenmesi ve bu bilgiden yola çıkarak ölçüm hatasının etkisinin parametre tahmini üzerinden arındırılmasıdır. SIMEX yöntemi, ölçüm hatalı modellerin parametrelerinin tahmin edilmesinde kullanılan en etkin yöntemlerden birisi olarak literatürde yerini almaktadır. Bunun nedeni, SIMEX yöntemi dışındaki yöntemlerin uygulanabilmesi için ölçüm hatası varyansının bilinmesinin veya tahmin edilmesinin gerekmesidir. SIMEX yönteminde ise böyle bir ön koşul bulunmamaktadır. Bu yöntem öncelikle klasik ölçüm hatası için geliştirilmiştir. Ancak zaman içerisinde Kuchenoff vd. (2006) gibi bazı araştırmacılar ile farklı alanlara da uyarlanmıştır.

SIMEX yöntemi ile parametre tahmini yapabilmek için simülasyon ve ekstrapolasyon olmak üzere iki adım bulunmaktadır. Simülasyon adımında, gerçek değerinin gözlenemediği veri setine ilaveten,  $0 = \zeta_1 < \zeta_2 < \dots < \zeta_M$  bilinen sıralı değerlerine bağlı olarak oluşturulan artan ölçüm hatası varyansına sahip

$$W_{b,i}(\zeta) = W_i + \zeta^{1/2} \sigma Z_{b,i} \quad (5)$$

veri setleri oluşturulmaktadır. Burada  $\{Z_{b,i}\}_{i=1}^n$  normal dağılımdan üretilmiş tesadüfi değişken ve  $b = 1, 2, \dots, B$  olmak üzere iterasyon sayısını temsil etmektedir. Eşitlik (5)'den elde edilen her bir  $W_{b,i}(\zeta_m)$  değişkeni ile  $\hat{\beta}_b(\zeta_m)$  katsayıları EKK ile tahmin edilmektedir. Ardından,

$$\bar{\beta}(\zeta_m) = \sum_b \hat{\beta}_b(\zeta_m) / B \quad (6)$$

ortalama değerleri kullanılarak, her bir  $\zeta_m$ 'e karşılık  $\bar{\beta}(\zeta_m)$  değerlerinin grafiği çizilir.

Bu aşamada ekstrapolasyon adımına geçilmekte ve değer aralığı dışında tahmin yapılmaktadır. Bunun için öncelikle oluşturulan saçılım grafiğinden uygun bir fonksiyonla model belirlenmektedir. Rasyonel ekstrapolant, kuadratik ekstrapolant gibi farklı fonksiyonel yaklaşımlar bulunmaktadır. Ancak literatürde en başarılı bulunan model rasyonel ekstrapolant model olduğundan, bu çalışmada o model dikkate alınmıştır. Eşitlik (6)'da dikkate alınan ortalama fonksiyonu, asimptotik olarak

$$E(\hat{\beta}_m | \zeta) = f(\zeta) = \frac{\beta \sigma_x^2}{\sigma_x^2 + (1 + \zeta) \sigma_u^2}, \quad \zeta \geq 0 \quad (7)$$

modeli formunda gösterildiğinden; oluşturulan rasyonel model yardımıyla  $f(\zeta = -1)$  noktasına ekstrapolasyon yapılmaktadır. Bu şekilde tahmin edilen katsayı,

$$\hat{\beta}_{SIMEX} = f(-1) \quad (6)$$

ile gösterilmekte ve SIMEX tahmin edicisi olarak adlandırılmaktadır (Cook ve Stefanski, 1994; Carroll vd., 2006: 97-112).

## 3. Simülasyon Bulguları

Ölçüm hatalı logaritmik ve yarı logaritmik modellerin parametre tahminlerindeki sapma ve bu tip modeller üzerinde SIMEX yönteminin etkinliğinin incelenmesi bakımından simülasyon

çalışması yapılmıştır. Bu tip simülasyon çalışmalarında, seçilen tek bir regresyon modeli üzerinden hareket edilmesi bir kısıt oluştursa da farklı ölçüm hatası varyansları, örnek çapları gibi diğer etkenlerin gözlemlenmesine de olanak sağlamaktadır. Dolayısıyla simülasyon çalışmalarının yapımında Fuller (1987)'in önerdiği güvenilirlik oranını yaklaşımı dikkate alınmış ve belirlenen güvenilirlik oranı seviyelerine göre ölçüm hatasının seviyeleri belirlenmiştir. Bu yaklaşıma göre,  $\sigma_u = 0,5; 1,5; \dots; 6,5$  değerlerini almakta ve güvenilirlik oranı da buna bağlı olarak  $(0,17; 0,97)$  aralığında değer almaktadır.

Tablo 1'de söz edilen her bir logaritmik ve yarı logaritmik model için bu ölçüm hatası seviyelerinin yanı sıra 30, 50, 100 ve 200 olacak şekilde farklı örnek büyüklükleri 10000 döngü ile çalıştırılmıştır. Çalışmanın bu aşamasında, söz konusu tüm farklı koşulları kapsayan R-project programında yazılan kodlar kullanılmıştır. Simülasyonda çalışılan veriler, Tablo 1'de yazılı modeller dikkate alınarak, aşağıda verilen regresyon katsayılarından üretilmiştir.

$$Y = 1,5 + 2,4X + \varepsilon$$

$$Y = 1,5 + 2,4(\log X) + \varepsilon$$

$$\log Y = 1,5 + 2,4X + \varepsilon$$

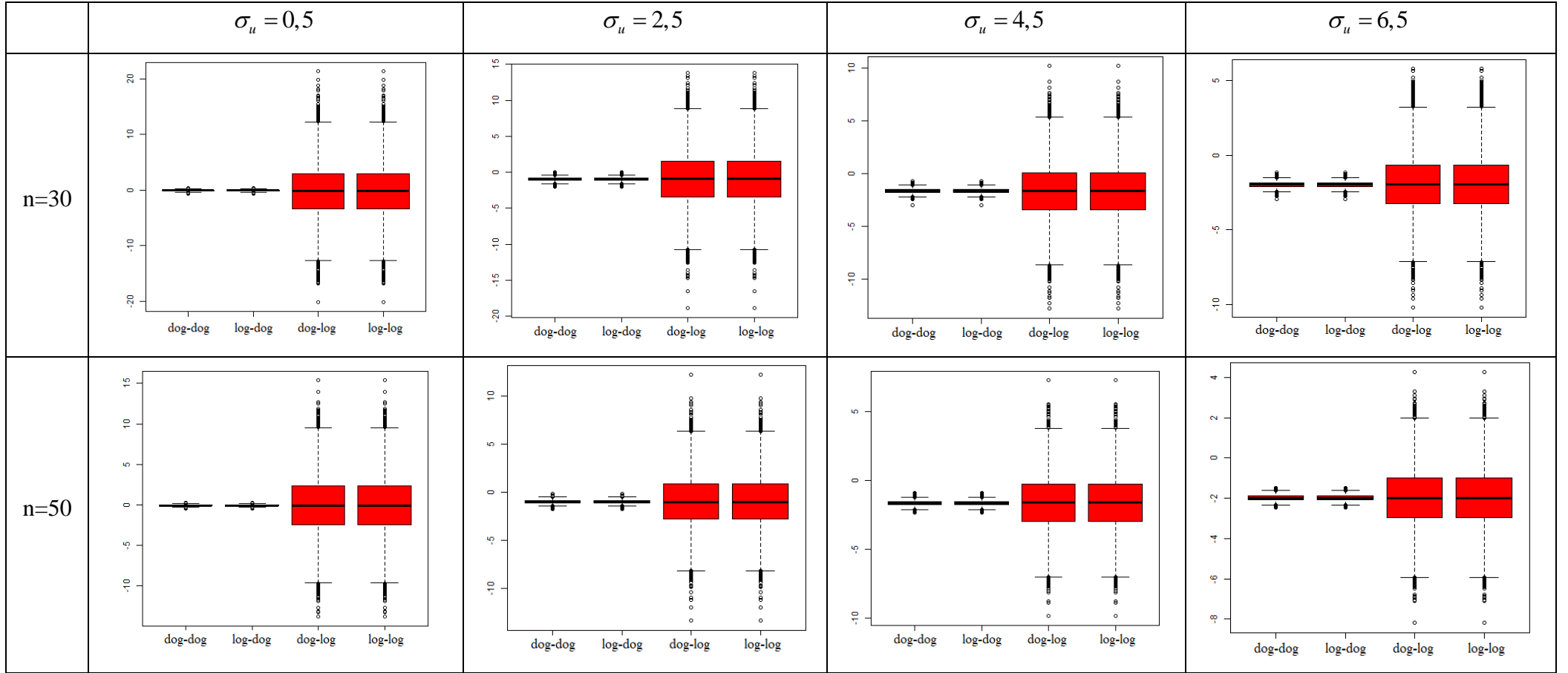
$$\log Y = 1,5 + 2,4(\log X) + e$$

olacak şekilde üretilmiştir. Bu aşamada  $\varepsilon \sim N(0; 1,5)$  ve  $X \sim N(50; 9)$  tesadüfi değerlerinden yararlanılmıştır.

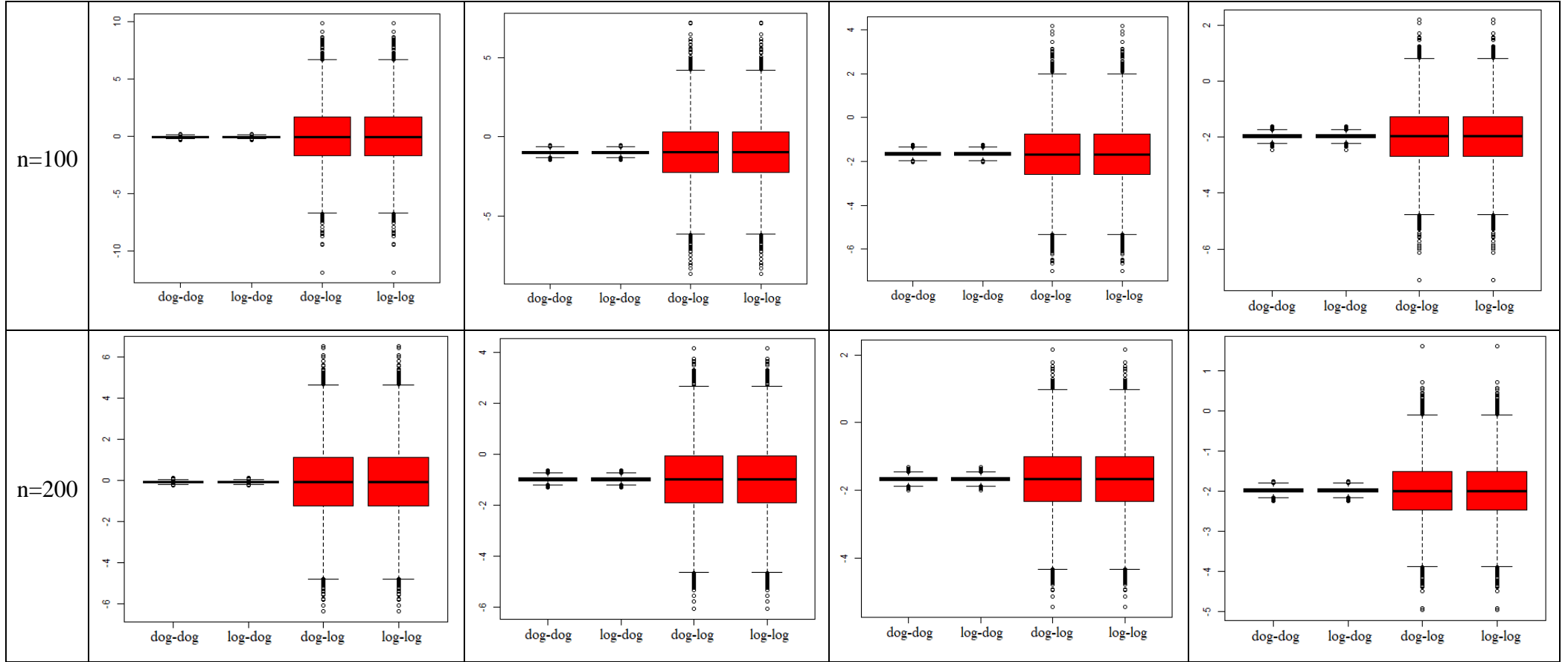
Simülasyon çalışması sonuçları ile öncelikli olarak model parametrelerindeki sapmalar incelenmiştir. Bu amaçla parametre tahminlerinde ölçüm hatası sonucu ortaya çıkan sapma miktarları için kutu grafikleri Şekil 1'de gösterilmiştir. Bu grafiklerde ilk dikkati çeken nokta, ölçüm hatasının dog-dog modellerde ortaya çıkardığı azaltıcı etkinin her modelde de gözlenmiş olmasıdır. Bu azaltıcı etkiyle, eğim parametreleri olduğundan daha düşük tahmin edilmiş ve tahmin edilen regresyon doğruları orijine yaklaşmıştır. Açıklayıcı değişkendeki ölçüm hatasının etkileri dikkate alındığında dog-dog ve log-dog modelin benzer hareket etmesi beklentisinin gerçekleşmiş olduğu da grafiklerde açıkça görülmektedir. Bununla birlikte, çalışmanın esas sorusunu oluşturan, açıklayıcı değişkendeki logaritmik dönüşümün ölçüm hatasının etkisini de gizleyebileceği düşüncesinin yanlış olduğu gözlenmektedir. Dog-log ve log-log modellerin aynı azaltıcı etkiye maruz kalmış olması ve logaritmik dönüşüm ile ölçüm hatasının etkisinin giderilmediğinin gözlenmesi, özellikle ekonometrik çalışmalarda sıklıkla kullanılan bu dönüşümler aracılığıyla ölçüm hatalarının etkisinin de yok edilebileceği inancıyla çelişmektedir. Üstelik logaritmik dönüşüm yapılmış açıklayıcı değişkende ölçüm hatasının etkisinin tahminlerde güvensizliğe yani standart hatalarında büyümeye yol açtığı da kutu grafiklerinin genişliğinden anlaşılmaktadır. Simülasyon çalışmalarının sonuçlarına göre, açıklayıcı değişkendeki ölçüm hatası, logaritmik dönüşüm de yapıldığı varsayımı altında, parametre tahminlerinde yalnızca sapmasızlık özelliğinin değil tutarsızlık özelliğinin de bozulmasına neden olmuştur. Şekil 1'e göre bir diğer öne çıkan özellik de örnek büyüklüğünün etkisidir. Örnek büyüklüğünün artması ile ölçüm hatasının etkisinin hafiflediği, buna karşılık yüksek ölçüm hatası büyüklüklerinde yine de parametre tahminlerinde sapma meydana geldiği görülmüştür.



Model parametre tahminlerinde, ölçüm hatası sonucu ortaya çıkan sapmaları daha detaylı inceleyebilmek için Tablo 2’de farklı ölçüm hatası seviyeleri için elde edilen logaritmik ve yarı logaritmik modellerin parametre tahminleri ve standart hataları yer almaktadır. Tablo 2’de elde edilen bulgular, Şekil 1’de gösterilen kutu grafikleri ile tutarlıdır. Burada da öncelikli olarak vurgulanması gereken nokta, logaritmik dönüşümün küçük ölçüm hatalarının etkisini de tam olarak yok edememiş olmasıdır. Hatta ölçüm hatasının artmasıyla, özellikle dog-log ve log-log modellerinin parametre tahminleri diğer modellerin parametre tahminlerine göre çok daha küçük elde edilmiştir. Dog-dog ve dog-log modellerin parametre tahminleri ile log-log- ve log-dog modellerin parametre tahminlerinin birbirlerine yakın olması ise bir diğer dikkati çeken noktadır. Buna göre ölçüm hatalı açıklayıcı değişken olması durumunda, açıklayıcı değişkene logaritmik dönüşüm uygulanmasından çok, açıklanan değişkene logaritmik dönüşüm uygulanması, modellerin parametre tahminleri üzerinde daha etkili olmaktadır. Ayrıca Tablo 2’ye göre örneklem büyüklüğünün artmasının, ölçüm hatasının etkisini kısmen azalttığı ancak parametrenin sapmasız tahmin edilmesi yönünde pek bir etkisinin olmadığı da görülmektedir.



**Şekil 1.** Farklı ölçüm hatası seviyeleri ve örnek büyüklükleri için doğrusal, logaritmik ve yarı logaritmik modellerde meydana gelen sapmalar için kutu grafikleri.



**Şekil 1. (devam)** Farklı ölçüm hatası seviyeleri ve örnek büyüklükleri için doğrusal, logaritmik ve yarı logaritmik modellerde meydana gelen sapmalar için kutu grafikleri.

Simülasyon çalışması sırasında ortaya çıkarılan her bir ölçüm hatalı modelin parametreleri SIMEX yöntemi ile de tahmin edilmiştir. Buradan elde edilen tahminlere ve standart hatalarına ise Tablo 3'te yer verilmiştir. Bu sonuçların incelenmesinde ilk dikkati çeken nokta, düşük ölçüm hatasının bulunduğu durumda SIMEX yönteminin oldukça başarılı sonuçlar vermesidir. Üstelik bu başarı, logaritmik dönüşüm uygulanan açıklayıcı değişkenin bulunduğu modellerde de geçerlidir. Buna karşılık özellikle yüksek ölçüm hatası seviyeleri için SIMEX yönteminin başarısı her model türünde düşmüş, hatta büyük örnek çapları için SIMEX tahmin edicileri hesaplanamamıştır.

**Tablo 2.** Farklı ölçüm hatası seviyeleri ve örnek büyüklükleri için, ölçüm hatalı doğrusal, logaritmik ve yarı logaritmik modellerin parametre tahminleri ( $\beta_{gerçek} = 2,4$ ).

		$\sigma_u = 0,5$	$\sigma_u = 2,5$	$\sigma_u = 4,5$	$\sigma_u = 6,5$
dog-dog	n=30	2,421 (0,126)	1,997 (0,369)	1,217 (0,370)	0,706 (0,309)
	n=50	2,393 (0,089)	2,017 (0,300)	1,194 (0,313)	0,757 (0,235)
	n=100	2,335 (0,004)	1,418 (0,016)	0,738 (0,013)	0,424 (0,009)
	n=200	2,334 (0,002)	1,418 (0,007)	0,740 (0,005)	0,422 (0,003)
log-dog	n=30	2,424 (0,134)	2,000 (0,370)	1,220 (0,375)	0,704 (0,307)
	n=50	2,391 (0,095)	2,021 (0,297)	1,188 (0,311)	0,759 (0,237)
	n=100	2,335 (0,004)	1,417 (0,016)	0,736 (0,013)	0,422 (0,008)
	n=200	2,334 (0,021)	1,418 (0,071)	0,737 (0,007)	0,421 (0,004)
dog-log	n=30	2,844 (4,229)	1,372 (3,407)	0,686 (2,610)	0,417 (1,943)
	n=50	2,148 (3,479)	1,234 (2,795)	0,377 (2,528)	0,566 (1,559)
	n=100	2,363 (6,204)	1,427 (3,774)	0,724 (1,930)	0,415 (1,080)
	n=200	2,319 (3,065)	1,417 (1,871)	0,736 (0,942)	0,408 (0,522)
log-log	n=30	2,844 (4,229)	1,372 (3,407)	0,686 (2,609)	0,416 (1,943)
	n=50	2,148 (3,478)	1,233 (2,795)	0,378 (2,529)	0,568 (1,559)
	n=100	2,362 (6,205)	1,421 (3,775)	0,723 (1,938)	0,411 (1,082)
	n=200	2,310 (3,069)	1,414 (1,879)	0,731 (0,949)	0,406 (0,526)

Parantez içindeki değerler tahmin edicilerin standart hatalarını göstermektedir.

Tablo 3'e göre dikkati çeken bir diğer nokta ise açıklayıcı değişkendeki logaritmik dönüşümün, küçük örnek çapı ve düşük ölçüm hatası seviyesinde, SIMEX tahminlerinin standart hatalarının yüksekliğidir. SIMEX yöntemi işleyişi gereği, veri kümesi içerisinde tekrar örneklem çekerek simülasyon sürecini sürdürmektedir. Bu durumda tahmin edicinin de tahmininin dikkate alınıyor olmasının standart hatanın yüksek çıkmasına neden olan etkenlerden biri olduğu

düşünülmektedir (Boos ve Stefanski, 2013: 385). Ancak artan örnek çapı ile bu durum tolere edilebilmektedir.

**Tablo 3.** Ölçüm hatalı doğrusal, logaritmik ve yarı logaritmik modellerinin SIMEX yöntemiyle elde edilen parametre tahminleri ( $\beta_{gerçek} = 2,4$ ).

		$\sigma_u = 0,5$	$\sigma_u = 2,5$	$\sigma_u = 4,5$	$\sigma_u = 6,5$
dog-dog	n=30	2,421 (0,016)	1,997 (0,136)	1,217 (0,137)	0,706 (0,963)
	n=50	2,393 (0,008)	2,017 (0,090)	1,194 (0,098)	0,757 (0,055)
	n=100	2,402 (0,068)	2,002 (0,195)	1,213 (0,204)	0,729 (0,163)
	n=200	2,403 (0,046)	2,005 (0,139)	-	-
log-dog	n=30	2,424 (0,018)	2,000 (0,137)	1,220 (0,141)	0,704 (0,094)
	n=50	2,391 (0,008)	2,020 (0,088)	1,182 (0,097)	0,759 (0,051)
	n=100	2,402 (0,068)	2,001 (0,196)	1,213 (0,204)	0,731 (0,169)
	n=200	2,402 (0,043)	2,004 (0,138)	-	-
dog-log	n=30	2,844 (17,883)	1,372 (11,611)	0,686 (6,810)	0,417 (3,777)
	n=50	2,148 (12,101)	1,233 (7,813)	0,376 (6,391)	0,568 (2,434)
	n=100	2,407 (2,421)	1,470 (1,915)	0,733 (1,348)	0,431 (1,010)
	n=200	2,416 (1,675)	1,389 (1,437)	-	-
log-log	n=30	2,844 (17,888)	1,372 (11,612)	0,686 (6,812)	0,416 (3,779)
	n=50	2,148 (12,112)	1,233 (7,815)	0,376 (6,393)	0,567 (2,431)
	n=100	2,407 (2,420)	1,471 (1,912)	0,732 (1,349)	0,432 (1,011)
	n=200	2,416 (1,673)	1,388 (1,437)	-	-

Parantez içindeki değerler tahmin edicilerin standart hatalarını göstermektedir.

Tablo 3'e göre, SIMEX yönteminin düşük ölçüm hatası için tüm logaritmik ve yarı logaritmik modellerde başarılı sonuç verdiği görülmektedir. Buna karşılık ölçüm hatası arttıkça dog-dog ve log-dog modellerde daha iyi sonuç verdiği de önemli bir noktadır. Dolayısıyla açıklayıcı değişkene logaritmik dönüşüm yapılması SIMEX yönteminin işleyişi ile uyuşmamaktadır.

#### 4. Sonuç

İktisadi değişkenlerin tam olarak ölçülememesi yapılan uygulamalarda modellerin sapmalı tahmin edilmesine neden olmaktadır. Yapılan uygulamalarda genellikle değişkenlere logaritmik dönüşüm yapılması ve bu dönüşüm ile ölçüm hatalarının etkisinin de yok olduğu görüşü yaygındır. Bu çalışma ile açıklayıcı değişkende ölçüm hatası olmasının logaritmik dönüşümle dahi yok olmadığı öncelikli olarak vurgulanmıştır. Yalnızca küçük ölçüm hatasının

parametre tahminine etkisi tüm model kalıplarında tolere edilebilmiştir. Ancak çoğu zaman parasal büyüklüklerle ilgilenen iktisadi değişkenlerde bu kadar küçük bir hata seviyesinin gözlenmeyeceği de dikkate alınmalıdır. Buna karşılık logaritmik ve yarı logaritmik model kalıpları içerisinde ölçüm hatasından en az etkilenen model kalıbının log-log modeller olduğu görülmüş; sonuç olarak, logaritmik dönüşümler ölçüm hatasının etkisini azaltmış olsa da verilerde ölçüm hatası olmasının yine de göz ardı edilemeyecek bir sorun olduğu kanaatine varılmıştır. Bu sonuç, özellikle ekonomik verilerle yapılan analizler bakımından önem taşımaktadır. Sapmalı parametre tahminleriyle politika belirlemek ve kestirim yapmanın önüne geçebilmek bakımından çalışmalarda tahmin yöntemlerinin doğru belirlenmesi, gerek duyulduğunda ölçüm hatalı model tahmin yöntemlerinin kullanılması büyük önem taşımaktadır.

SIMEX yöntemi, ölçüm hatalı modellerin parametre tahminlerini sapmasız verebilen ve oldukça yaygın bir yöntemdir. Buna karşılık literatürde bu yöntemin logaritmik dönüşümler üzerindeki performansının incelendiğine rastlanılmamıştır. Burada yapılan çalışma ile SIMEX yönteminin de düşük ölçüm hatalarında ve tüm model kalıplarında başarılı sonuç verdiği görülmektedir. Ancak, ölçüm hatasının artması durumunda açıklayıcı değişkendeki logaritmik dönüşüm, dönüşümün uygulanmadığı modellere göre daha zayıf sonuçlar vermiştir. Yüksek ölçüm hatası seviyeleri için ise SIMEX yönteminin başarısı her model türünde düşmüş, hatta büyük örnek çapları için SIMEX tahmin edicileri hesaplanamamıştır. SIMEX yönteminin açıklayıcı değişkendeki dönüşümden bu kadar etkilenmesinin diğer bir açıklaması da logaritmik dönüşümün rasyonel ekstrapolant modeli ile iyi açıklanamıyor olmasıdır ve bu, SIMEX yönteminin incelenmesi gereken bir özelliği olarak dikkati çekmektedir. Gelecek çalışmalarda özellikle bu yöntemin, logaritmik kalıplara daha uygun ekstrapolant modellerin geliştirilmesinin, yöntemin yaygın uygulanabilirliği bakımından önemli olacağı düşünülmektedir. Özellikle iktisadi değişkenlerle yapılan uygulamalar bakımından bu yöntemin uyarlanmasının oldukça önemli bir katkı sağlaması beklenmektedir.

## Kaynakça

- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley&Sons.
- Boos, D. D. & Stefanski, L. A. (2013). *Essential Statistical Inference: Theory and Methods*. USA: Springer.
- Bound, J., Brown, C., & Mathiowetz, N. (2000). *Measurement Error in Survey Data*. United States of America: PSC Publications.
- Buonaccorsi, J. P. (2010). *Measurement Error Models, Methods and Applications*. USA: CRC Press.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. USA: Chapman&Hall/CRC.
- Cook, J. R., & Stefanski, L. (1992). *Simulation-Extrapolation Estimation in Parametric Measurement Error Models*. MIMEO Series #2224R, North Caroline: North Caroline State University.
- Cook, J. R., & Stefanski, L. A. (1994). Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association*, 89(428), 1314-1328.
- Edwards, L. J., & Hamilton, S. A. (1995). Errors-in-variables and the Box-Cox Transformation. *Computational Statistics&Data Analysis*, 20, 131-140.
- Fuller, W.A. (1980). Properties of Some Estimators of the Errors-in-variables Model. *Annual Statistics*, 8, 407-422.
- Fuller, W. A. (1987). *Measurement Error Models*. New York: Wiley.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. America: The McGraw-Hill Companies.
- Kuchenoff, H., Mwalili, S.M. & Lesaffre, E. (2006). A General Method for Dealing with Misclassification in Regression: The Misclassification SIMEX. *Biometrics*, 62, 85-96.
- Richardson, R., Tolley, H. D., Evenson, W. E. & Lunt, B. M. (2018). Accounting for Measurement Error in Log Regression Models with Applications to Accelerated Testing. *PLoS ONE*, 13(5), 1-13.
- Schneeweiss, H. (1976). Consistent Estimation of a Regression with Errors in the Variables. *Metrika*, 23(1), 101-115.
- TÜİK, T. İ. (2012). *Üretim ve Harcama Yöntemi ile Gayri Safı Yurtiçi Hasıla Tahminleri - Kavram, Yöntem ve Kaynaklar*. Ankara: Türkiye İstatistik Kurumu Matbaası.