ON A CERTAIN TRANSFORMATION OF THE DEPENDENT VARIABLE IN LINEAR REGRESSION

Halûk Erlat*

In a recent paper[1] Mullet and Murray [1973] have proved that if a certain linear transformation is applied to the dependent variable in the multiple linear regression model and the parameters of the transformed and untransformed models are estimated by least squares (LS) , then (i) there will be a one-to-one correspondence between the estimates of the transformed and untransformed model parameters,(ii) the residual vector of the two models will be the same, (iii) the estimated covariance matrix of the LS estimators will also be the same; but (iv) the R^2 values will be different; i.e., their R^2's will not be comparable.[2]

A similar situation is encountered if one considers the aggregate consumption function as specified below :

---

* Assistant Professor of Economics, Middle East Technical University, Ankara, Turkey. Thanks are due to Dr. Hasan Ersel and Dr.Oktar Türel for helpful comments. All remaining errors are, of course, the authors' responsibility.

1 In fact, the proof of the theorem first appeared in Mullet and Murray [1971].

2 Since R^2's (or $\bar{R}^2$'s) are routinely utilised to choose between functional forms, this last conclusion is of importance. It implies that since the information content of the original model is invariant to this transformation, a choice based on R^2's would be erronous.

(1) $C_t = a_o + a_1Y_t + a_2C_{t-1} + u_t$, $t = 1,\ldots, T$

where C is aggregate consumption expenditures, Y is disposable income and T is the sample size. Since Y = C + S , where S denotes aggregate saving, the same information may be obtained by using the aggregate saving function instead, i.e.,

(2) $S_t = Y_t - C_t = -a_o + (1-a_1)Y_t - a_2 C_{t-1} - u_t$

$$= a_o^* + a_1^*Y_t + a_2^*C_{t-1} + u_t^*$$

where the definitions of the starred coefficients are evident from (2). Thus, even though there is no difference in estimating (1) or (2), they lead to different R^2 values (see, Rao and Miller [1971:14-16]).

Let us now formalize the transformation considered by Mullet and Murray and that implied by equation (2). Thus, if y is the T x 1 vector of observations on the dependent variable, X is the Tx(k+1) matrix of explanatory variables (including the intercept term), β the (k+1)x1 vector of associated coefficients and u, the Tx1 vector of disturbances, then our model may be written as

(3) $y = X\beta + u$

and will obey the following assumptions :

A.1 $E(u) = 0$

A.2 $E(uu') = \sigma^2 I$

A.3 X is a matrix of nonstochastic elements

A.4 $E(X'u) = 0$

A.5 $r(X) = k + 1 < T$

where "r" stands for "rank".

Suppose we transform y as

(4) $z = y + \alpha x_i$

where α is a known scalar and x_i is the i^{th} column of X. Then the model to be estimated becomes

(5) $z = X\beta^* + u$

where

(6) $\beta^* = \beta + \alpha e_i$

e_i being a (k+1)x1 vector with 1 in the i^{th} position and zeros elsewhere.

Equation (4) is the formalization of the transformation discussed in Mullet and Murray [1973]. The following results were proved therein :

(a) $\hat{\beta} = \hat{\beta}^* - \alpha e_i$

(b) $\hat{u}_z = \hat{u}_y$, hence $\hat{u}_z'\hat{u}_z = \hat{u}_y'\hat{u}_y$

(c) $\text{Cov}(\hat{\beta}) = \text{Cov}(\hat{\beta}^*)$

where "^" denotes LS estimators, and the "z" and "y" subscripts indicate the transformed and untransformed models, respectively.

On the other hand, the transformation implied by (2) may be formalized as

(7) $z^* = \alpha x_i - y = X\beta^{**} + u^*$

where

(8) $\beta^{**} = \alpha e_i - \beta$ and $u^* = -u$

It may be shown, by steps similar to the ones followed in the Mullet and Murray [1973] paper, that

(a') $\hat{\beta} = \alpha e_i - \hat{\beta}^{**}$

(b') $\hat{u}_y = \hat{u}_z{}^*$

(c') $\hat{Cov}\ (\hat{\beta}) = \hat{Cov}\ (\hat{\beta}^{**})$.

We shall now propose a generalization of the transformations given in (4) and (7), and show that both sets of results may be obtained as special cases of the implications of this generalization. We, thus, have

(9) $w = \lambda y + Xs$

where λ is a known nonzero real number and s is a (k+1)x1 vector of known real numbers. It then follows, by substituting (3) into (9), that

(10) $w = X(\lambda\beta + s) + \lambda u = X\delta + v$

where $\delta = \lambda\beta + s$ and $v = \lambda u$. Thus, we see that

(11) $E(v) = \lambda E(u) = 0,$

$E(vv') = \lambda^2 E(uu') = \lambda^2\sigma^2 I = \theta^2 I$

where $\theta = \lambda\sigma$, and

$$E(X'v) = \lambda E(X'u) = 0$$

i.e., v satisfies assumptions A.1, A.2 and A.4. We may now prove the following theorem.

Theorem : Given the models described by (3) and (10), which both satisfy assumptions A.1 to A.5,

(A) $\hat{\beta} = \lambda^{-1} (\hat{\delta} - s)$

(B) $\hat{u} = \lambda^{-1} \hat{v}$

(C) $Cov(\hat{\beta}) = \lambda^{-2} Cov (\hat{\delta})$

Proof : (A) Applying LS to (10) we obtain

(12) $\hat{\delta} = (X'X)^{-1}X'w = \lambda(X'X)^{-1}X'y + (X'X)^{-1}X'Xs = \lambda\hat{\beta} + s$

or

(13) $\hat{\beta} = \lambda^{-1}(\hat{\delta} - s)$

(B) By definition,

(14) $\hat{v} = w - X(X'X)^{-1}X'w = Nw$

where $N = I - X(X'X)^{-1}X'$ is symmetric, idempotent and $NX = 0$. (See Theil [1971:40]).

Hence, substituting for w from (9) we have

(15) $\hat{v} = N(\lambda y + Xs) = \lambda Ny + NXs = \lambda Ny = \lambda\hat{u}$

Thus, $\hat{u} = \lambda^{-1}\hat{v}$.

(C) The estimated covariance matrix for $\hat{\delta}$ may be written as (Theil [1971:112-114]),

$$(16) \quad \widehat{Cov}(\hat{\delta}) = \frac{\hat{v}'\hat{v}}{T-k-1}(X'X)^{-1}$$

$$= \lambda^2 \frac{\hat{u}'\hat{u}}{T-k-1}(X'X)^{-1} = \lambda^2 \widehat{Cov}(\hat{\beta})$$

Note that if we choose $s = \alpha e_i$ and first let $\lambda=1$ we obtain the transformation in (4) and results (a), (b) and (c) would follow as corrolaries to the theorem, while alternatively letting $\lambda = -1$ would yield transformation (7) and results (a'), (b') and (c') would, in turn, be obtained as corrollaries.

If, as an additional step, we assume

A.6 $\quad u \sim N(0,\sigma^2 I)$

this would imply that $v \sim N(0,\theta^2 I)$. Letting d denote a (k+1)x1 vector of known scalars we may test any linear hypothesis on the β and δ vectors by using the t-distribution (Theil [1971:131]). Now, suppose null hypothesis for the β vector is formulated as $H_o : d'\beta = d'\beta^o$. By equation (10) $\delta^o = \lambda\beta^o + s$ and this implies the following null hypothesis for the δ vector; $H_o : d'\delta = d'\delta^o$. What we want to demonstrate is that, at a given significance level, a decision reached by one of the tests implies the decision to be reached for the other test, To see this, recall that the t-statistic for the test on $d'\beta$ is obtained as

$$(17) \quad t_y = \frac{d'\hat{\beta} - d'\beta^o}{[d' \hat{Cov}(\hat{\beta})d]^{1/2}}$$

Comparing t_y with the critical t value obtained at the preassigned significance level we reach a certain decision. Now, making use of (10), (12) and (16) we note that the t-statistic for the test on $d'\delta$ is obtained as

$$(18) \quad t_z = \frac{d'\hat{\delta} - d'\delta^o}{[d' \hat{Cov}(\hat{\delta})d]^{1/2}} = \frac{d'[\lambda\beta+s-\lambda\beta^o - s]}{\lambda[d' \hat{Cov}(\hat{\beta})d]^{1/2}} = t_y$$

Hence the decision to be reached in either case would imply the other.

This result further implies that if the tests involved are significance tests on individual coefficients, in which case $d = e_i$ and $\beta_i^o = 0$, then $\delta_i^o = s_i$. Thus, the decision about the significance of β_i does not imply anything about the significance of δ_i unless $s_i = 0$.

References :

Mullet, G.M. and T.W. Murray : "A New Method of Examining Rounding Errors in Least Squares Regression Computer Programs", Journal of the American Statistical Association, vol. 66, 1971, pp.496-498.

Mullet, G.M. and T.W. Murray : "More on the Game of Maximising $\bar{R}^2$", Australian Economic Papers, vol. 12, 1973, pp. 263-266.

Rao, P. and R.L. Miller : Applied Econometrics, Belmont, Calif. : Wasdworth, 1971.

Theil, H. : Principles of Econometrics, New York: John Wiley and Sons, 1971.

ÖZET

DOĞRUSAL REGRESYONDA BAĞIMLI DEĞİŞKENE UYGULANAN BİR DÖNÜŞÜM ÜSTÜNE

Doğrusal regresyon modelinde bağımlı değişken sıfırdan farklı bir katsayıyla çarpılır ve buna açıklayıcı değişkenlerin herhangi bir doğrusal bileşimi eklenirse, elde edilen dönüştürülmüş modelin En Küçük Kareler yöntemiyle tahmin edilmiş parametreleriyle, orijinal modelin tahmin edilmiş parametreleri arasında şu ilişkiler elde edilir : (i) dönüştürülmüş katsayı tahminleri dönüştürülmemişlerin doğrusal fonksiyonudur; (ii) dönüştürülmüş modelin artıklar vektörü, dönüştürülmemişlerin doğrusal fonksiyonudur, (iii) İki modelin tahmin edilmiş katsayı vektörlerinin tahmin edilmiş kovaryans matriksleri de doğrusal ilişki içindedir.