Journal of Economics and International Finance Vol. 3(2), pp.116-120, February 2011
Available online at http://www.academicjournals.org/JEIF
ISSN 2006-9812 ©2011 Academic Journals

Full Length Research Paper

On multicollinearity in nonlinear econometric models with mis-specified error terms in large samples

Isaac D. Essi[1]*, E. C. Chukuigwe[2] and Nathaniel A. Ojekudo[3]

[1]Centre for Econometric and Allied Research (CEAR), University of Ibadan, Ibadan, Nigeria.
[2]Department of Agricultural Economics and Extension, Rivers State University of Science and Technology, Port Harcourt, Nigeria.
[3]Department of Mathematics, Rivers State College of Education Port Harcourt, Nigeria.

Accepted December 3, 2010

The Cobb-Douglas model is a common occurrence in econometrics and other areas of research. Earlier results show that the consequence is more serious when a multiplicative error plagued data set is fitted with an additive error based model than vice-versa. In this study, involving large samples, we investigate the impact of multicollinearity in nonlinear econometric models with mis-specified error terms. .As it was in small samples, we observe that in large samples the above result and trend also hold in the presence of multicollinearity. It is also observed that the effect of multicollinearity is not purged by large sample size.

Key words: Cobb-Douglass model, mis-specified error terms, multicollinearity.

INTRODUCTION

As observed by econometricians and scientists in other disciplines the p- input variable function:

$$f = \theta_0 \prod_{i=1}^{p} X_i^{\theta_i} \qquad (1.1)$$

plays a significant role in modeling certain phenomena. In economics, for instance, it is widely applied in research works on production, demand (including transport demand) and cost functions. In biometry, the function (1.1) may be used in carrying out leaf rectangularity index analysis as in Essi (2005), (2007). In economics, the powers of the X's are called elasticity's and their sum is interpreted as a measure of returns to scale. Each θ_i (i =1, 2, ... , p) under some conditions gives the factor share of the associated input variable X^i.

Econometric model demands the incorporation of an error term as well as the specification of its distribution. The specification of the error term is a major problem in applied econometrics. The functional form f in Equation (1.1) cannot be decided in isolation from the specification of the error term. Since economic theory cannot give precisely always what the functional should be, as cited in Dhyrmes et al.(1972) and Zarembka (1966) this and the related problem of error specification may be resolved empirically.

THEORETICAL FRAMEWORK AND LITERATURE REVIEW

Assumption

Let us assume a sequence of real valued responses y_t with the form

$$y_t = f_t(\theta) + u_t, \; t = 1, 2, \ldots\ldots T \qquad (2.1)$$

Where, $f_t = f(X_t, \theta)$ are known continuous functions on a compact subset Θ of the Euclidean space IR^p and the u_t are independently and identically distributed errors with zero mean and finite variance $\sigma^2 > 0$. The values of θ and σ^2 are unknown but fixed. X_t is an input vector for the period t and is fixed.

*Corresponding author. E-mail: ddmetra_utibe@yahoo.com.

Definition 2.1 (least square estimate)

Any vector $\hat{\theta}$ in Θ which minimizes

$$Q_T(\theta) = \frac{1}{T}\sum_{i=1}^{T}\left(y_t - f_t(\theta)\right)^2 \quad (2.2)$$

is called a least square estimate of θ based on the first T values of y_t.

Theorem 2.1 (minimum mean square error criterion (Essi, 2002)

(a) Let Ho: $y_t = f_t(x_t, \theta) + u_{ot}$, t = 1, 2, …T (2.3)

be the model that is to be used to estimate θ, where $f_t(X_t, \theta)$ is a known continuous function on a compact subset Θ of a p-dimensional Euclidean space IR^p and the u_{ot} are identically distributed errors with zero mean and finite various $\sigma^2 > 0$.

(b) Let H_1: $y = G(Z_t, \lambda)$ (2.4)

be an alternative structure that may be used to explain y as well, where $G(Z_t, \lambda)$ embodies both the deterministic and stochastic parts of y_t and the stochastic disturbance term may be additive or multiplicative.

(c) The function $f_t(X_1, \theta)$ and the deterministic part of the model H_1 may be linear or nonlinear.

(d) Let the estimated adjusted $\bar{R}^2$ for H_o and H_1 be respectively $\hat{\bar{R}}_0^2$ and $\hat{\bar{R}}_1^2$. If Ho is the correct model such that $E(\hat{\bar{R}}_0^2) = \bar{R}_0^2 = 1$,

implies that $\hat{\bar{R}}_0^2 \geq \hat{\bar{R}}_1^2$. (2.6)

Then, MSE $(\hat{\bar{R}}_0^2)$ < MSE $(\hat{\bar{R}}_1^2)$. (2.5)

on the overage, where MSE $(\hat{\bar{R}}_0^2)$ is the mean square error of $\hat{\bar{R}}_0^2$. MSE $(\hat{\bar{R}}_1^2)$ is similarly defined.

Proof

The proof can be seen in Essi (2002) and Essi et al. (2007).

We shall use this criterion later in comparing two competing models. The minimum variance criterion of Heben (1983: 90-98) is replaced by minimum mean square criterion of Essi in Essi (2002). The Essi criterion, which is in terms of the mean squared error (MSE) of the adjusted coefficient of determination is more preferred, for some obvious reasons, to that of Theil, as a basis for comparing estimates of the true and mis-specified models, especially where we use simulated models with replications. In using Theil minimum variance criterion, the response variables in both H_o and H_1 must be in the same units and the models should be linear. The use of MSE $(\hat{\bar{R}}^2)$ overcomes these drawbacks. We also benefit as the ratio of MSE $(\hat{\bar{R}}^2)$ from two competing models, can be used to assess the overall relative efficiency of a set of one model estimates to that another.

The consequences of an incorrect form for the disturbance term, according to Greene (2003) are bias and inconsistency in the least square estimate of the parameters. Heben (1983: 25) observes that there is trouble with the multiplicative error model (MEM) in that one may "encounter severe multicollinearity between K and L, especially with cross-section (rather than time – series) data and especially if our observations are for firms in a fairly homogenous industry". This, Heben says, is " because for such an industry, the capital-labour mix is fairly uniform across firms, since all use more or less the same technique, hence if the K/L ratio is, say, 3, then for all observations we would have approximately K = 3L, and hence very strong collinearity. "

Fabrycy in (Essi, 2000) observes that "using linear least squares regressions induce us to adopt functions which are linear in parameters. Often this imposes unrealistic-cally rigid constraints which may create multicollinearity. Using more realistic nonlinear forms and nonlinear least squares regressions is likely to overcome this problem." The papers (Essi and Iyaniwura, 2007) and (Essi et al., 2007; 2(1): 41- 48) consider the consequences of mis-specifying the error term for the Cobb-Douglas production model. The articles (Essi and Iyaniwura, 2007) and (Essi et al., 2007; 2(1): 41- 48) observe that the consequence is more serious when a multiplicative error plagued data set is fitted with an additive error based model than vice-versa .This trend, it is observed, persists in the presence of multicollinearity.

METHODOLOGY AND DATA

Two competing models are considered and they are

$$AEM: y = \theta_1 K^{\theta_2} L^{\theta_2} + U_0 \quad (3.1)$$

and

$$MEM: y = \theta_1 K^{\theta_2} L^{\theta_3} e^{U_1} \quad (3.2)$$

Table 1. Values of MSE ($\hat{\bar{R}}^2$) for all the models ($\sigma_0^2 = 0.16,$ T = 20, N = 20).

Model	Cor(K, L) = 0.03	Cor(K, L) = 0.24	Cor(K, L) = 0.45
AED/AEM	3.6E-10	4.36E-10	2.81E-10
AED/MEM (H_1)	43.61E-10	69.64E-10	67.93E-10
MED/AEM (H_2)	0.115188	0.123676	0.109820
MED/MEM	0.046419	0.050798	0.035785
Ratio of MSE($\hat{\bar{R}}^2$) in H_2 and H_1	2.64E07	1.78E07	1.62E07

Table 2. Values of MSE ($\hat{\bar{R}}^2$) for all the models ($\sigma_0^2 = 0.16,$ T = 40, N = 20).

Model	Cor(K, L) = 0.03	Cor(K, L) = 0.24	Cor(K, L) = 0.45
AED/AEM	3.40E-10	4.25E-10	2.41E-10
AED/MEM (H_1)	47.53E-10	177.28E-10	64.24E-10
MED/AEM (H_2)	0.1290918	0.146557167	0.1033563
MED/MEM	0.0546240	0.031281871	0.03304124
Ratio of MSE($\hat{\bar{R}}^2$) in H_2 and H_1	2.72E07	0.83E07	1.61E07

when one is held to be true, the other becomes its mis-specification and vice-versa. The model (3.2) is intrinsically linear and OLS estimation of the Log-transformed version provides estimate for $(\theta = \theta_1, \theta_2, \theta_3)'$. The modified Gauss-Newton algorithm is used in estimating the intrinsically non-linear model (3.1). The choice of model parameters $(\theta_1, \theta_2, \theta_3)$ is such that $\theta_2 + \theta_3 < 1, \theta_2 + \theta_3 = 1$ and $\theta_2 + \theta_3 > 1$ but $\theta_2 + \theta_3 < 2$ while the value of θ_1 is arbitrary and kept at $\theta = 10.0$. We use as our set of parameters $(\theta_1, \theta_2, \theta_3)$ = (10.0, 0.45 and 0.50).

The input matrixes is made of two variables K (capital) and L (Labour) and are randomly generated and normally distributed independent variables such that they are typical of data set on capital and labour as that of Theil (1957).We adopt the data generating process outlined in Essi and Iyaniwura (2007) and (Essi et al., 2007; 2(1),41-48).

The noisy Y's are obtained according to the relations (3.1) and (3.2). The Monte Carlo study uses sample size of 20, 40 and 80 with each experiment replicated 20 times under three levels of collinearity between K and L. The levels of collinearity are furnished by simple Pearson correlation coefficient between K and L. We denote this coefficient by Cor (K, L) and the values for this work are 0.03 , 0.24 and 0.45.

Two correct specifications are used in the work. They are AED/AEM and MED/MEM. The specification AED/AEM is the one where an additive error-plagued data (AED) is fitted with an additive error-based model (AEM). The model MED/MEM is one where a multiplicative error-plagued data (MED) is fitted with a multiplicative error-based model (MEM). A situation in which the model building process is based on a different structure from that used to generate the date gives rise to two in-appropriate specifications in the study. They are designated as AED/MEM and MED/AEM. AED/MEM is the case in which additive error-based data (AED) is fitted with multiplicative error model (MEM). MED/AEM is the case in which multiplicative error-based data (MED) is fitted with an additive error-based model (AEM). We shall refer to AED/MEM and MED/AEM respectively as first and second mis-specifications or simply as H_1 and H_2.

EMPIRICAL RESULTS

Altogether we estimated 720 equations. Some of the numerical results obtained are summarized and presented in Tables 1, 2 and 3.
The competing models to be considered first are:

$$\text{AED / AEM}: Y = \theta_1 K^{\theta_2} L^{\theta_3} + U_0 \quad (4.1)$$

and

$$\text{AED / MEM}: Y = \theta_1 K^{\theta_2} L^{\theta_3} e^{U_1} \quad (4.2)$$

Where, U_0 and U_1 respectively follow $N(0, \sigma_0^2)$ and $N(0, \sigma_1^2)$. Monte Carlo results showing estimates of mean square error of $\hat{\bar{R}}^2$, MSE($\hat{\bar{R}}^2$) are presented in Table 1. In this table $\sigma_0^2 = 0.16,$ θ = (10, 0.45, 0.50) with sample size T = 20 and replication N = 20. $\sigma_0^2 = 0.16,$ On

Table 3. Values of MSE ($\hat{\bar{R}}^2$) for all the models ($\sigma_0^2 = 0.16,$ T = 80, N = 20).

Model	Cor(K, L) = 0.03	Cor(K, L) = 0.24	Cor(K, L) = 0.45
AED/AEM	23.29E-10	4.09E-10	2.60E-10
AED/MEM (H_1)	49.30E-10	237.324E-10	60.25E-10
MED/AEM (H_2)	0.126702	0.144621	0.117384
MED/MEM	0.044565	0.029809	0.032017
Ratio of MSE($\hat{\bar{R}}^2$) in H_2 and H_1	2.57E07	0.61E07	1.95E07

Table 4. Ratio of MSE ($\hat{\bar{R}}^2$) in H_2 and H_1 for various sample sizes and levels of multicollinearity.

Sample size (n)	Cor(K, L) = 0.03	Cor(K, L) = 0.24	Cor(K, L) = 0.45
20	2.64E07	1.78E07	1.62E07
40	2.72 E07	0.83E07	1.61E07
80	2.57E07	0.61E07	1.95E07

reversing the roles of the models (4.1) and (4.2) we obtain respectively.

$$\text{MED / MEM}: Y = \theta_1 K^{\theta_2} L^{\theta_3} e^{U_0} \tag{4.3}$$

and

$$\text{MED / AEM}: Y = \theta_1 K^{\theta_2} L^{\theta_3} + U_1 \tag{4.4}$$

as competing models. We still use $\sigma_0^2 = 0.16,$ θ = (10, 0.45, 0.50), T = 20, N =20 to obtain required estimates.

One wonders why many decimal places are allowed in the computations. The outstanding reason is that while the MSE (values for two different models are small, the size of their ratios could be very high. For instance, when the level of collinearity between the inputs K and L is 0.03, MSE ($\hat{\bar{R}}^2$) values for MED/AEM and AED/MEM are respectively 0.115188 and 43.61E-10. (Recall that we refer to AED/MEM and MED/AEM respectively as first and second mis-specifications or simply as H_1 and H_2). The ratio of these values is 2.64E07. This is high. (Table1).

The results of Monte Carlo experiments for samples of sizes 40 and 80 are seen in Tables 2 and 3.

DISCUSSION

We want to focus on the mis-specified models and the mean square error (MSE) of the estimated gives the magnitude of the impact of mis-specification in the presence of multicollinearity. The correlation coefficient Cor(K, L), between the inputs K and L gives the level of multicollinearity between K and L. Consider Table 1 (Sample size T= 20) and multicollinearity level Cor(K, L) = 0.03. The value of MSE in model H_2 is higher than MSE in H_1. The trend is the same for Cor(K,L)=0.24 and Cor(K,L)=0.45. When we go to higher sample sizes in Tables 2 and 3, H_2 still has higher MSE than H_1. In Table 4, the ratio of MSE in H_2 to that in H_1 is far greater than unity irrespective of sample size and level of multi-collinearity. The models are brought together in Table 5. In each of the entries, H_2 has higher MSE than H_1.

Conclusion

We, from the beginning do not focus on the detection of multicoliinearity as attempted by Fabrycy (1975) but rather investigate the consequences and the seriousness of the consequences of mis-specifying the error term in the presence of multicollinearity. Earlier results, stated in Essi (2002), Essi and Iyaniwura (2007), Essi et al. (2007; 2(1), 41- 48) and Essi (2000) that the consequence is more serious when a multiplicative error plagued data set is fitted with an additive error based model than vice-versa. This result and trend also hold in the presence of multicollinearity in this work. That is, the adverse effect of misspecification in H_2 is worse than the adverse effect of mis-specification H_1. The higher values of mean square error in H_2 attest to this, that however, as the level of

Table 5. Values of MSE ($\hat{\bar{R}}^2$) in H_2 and H_1 for various sample sizes and levels of multicollinearity.

Sample size (T)	Cor(K, L) = 0.03	Cor(K, L) = 0.24	Cor(K, L) = 0.45
20	43.61E-10 (H_1) 0.115188 (H_2)	0.69.64E-10 (H_1) 0.123676(H_2)	67.93E-10 (H_1) 0.109820(H_2)
40	47.53E-10 (H_1) 0.1290918(H_2)	177.28E-10 (H_1) 0.14655716(H_2)	64.24E-10 (H_1) 0.1033563(H_2)
80	49.30E-10 (H_1) 0.126702(H_2)	37.324E-10 (H_1) 0.144621(H_2)	60.25E-10 (H_1) 0.117384(H_2)

multicollinearity between the inputs rises, the relative efficiency of the estimates of H_2 to that of H_1 follows ambiguous trend and at best said to be oscillating for large samples (Table 4). We therefore submit that the robustness of the mis-specification H_2 relative to that of H_1 depends not only on returns to scale, as was earlier advanced in Essi and Iyaniwura (2007) but on multicollinearity of the inputs also. We also observe that effect of multicollinearity is not purged by large sample size in mis-pecified models. These results should be taken into consideration when we encounter studies involving production functions, and leaf rectangularity index analysis, among others.

REFERENCES

Essi ID (2005). On Constructing Leaf Rectangularity Index by Bootstrap Regression, AMSE J., 66(6): 13 – 22.

Essi ID (2007). Computing Leaf Regularity Index under Alternative Error Specifications, AMSE J. Modeling. C., 70(1): 67 -79.

Essi ID (2002). Econometric Models with Mis-specified Error Terms, Abacus: J. Math. Assoc. Nig., 29(2): 152- 160.

Essi ID, Iyaniwura JO (2007). On Robustness and Choosing Between Two Nonlinearities, Adv. Appl. Stat., 7(3): 451-462.

Essi ID, Iyaniwura JO, Amadi SN (2007). Further on Monte Carlo Simulation of Inputs in Production Theory, AMSE, J. Modeling., 28(2): 23-35

Essi ID, Iyaniwura JO, Ojekudo NA (2007). On Multicollinearity in Nonlinear Econometric Models with Mis-specified Error Terms in Small Samples Forthcoming in Int. J. Stat. Syst.,(IJSS), 2(1): 41-48

Essi N (2000). Robustness of Estimators of Nonlinear Econometric Models with Mis-specified Error Terms - A Ph.D Thesis, University of Ibadan, Ibadan.

Fabrycy MZ (1975). Multicollinearity Caused by Specification Errors, Appl. Stat., 24(1): 250 -260.

Greene WH (2003). Econometric Analysis,5th ed. Prentice- Hall Englewood Cliffs, N.J.

Heben J (1983). Application of Econometrics, Oxford, Philip Allen, p. 25

Heben J (1983). Application of Econometrics, Oxford Philip Allen, pp. 90-98.

Theil H (1957). Specification Errors and Estimation of Economic Relationships Rev. Int. Stat. Inst., 25: 41 – 51.

Zarembka P (1966) Manufacturing and Agricultural Production Functions in International Trade. J. Econ., 47: 952-959.