

Effect of Multicollinearity on Power Rates of the Ordinary Least Squares Estimators

¹O.O. Alabi, ²Kayode Ayinde and ¹T.O. Olatayo

¹Department of Mathematical Sciences, Olabisi Onabanjo University,
P.M.B. 2002, Ago-Iwoye, Ogun State, Nigeria

²Department of Pure and Applied Mathematics, Ladoke Akintola University of Technology,
P.M.B. 4000, Ogbomoso, Oyo State, Nigeria

Abstract: Inferences on the parameter estimates of Ordinary Least Square (OLS) estimator in regression model when regressors exhibit multicollinearity is a problem in that large standard errors of the regression coefficients which cause low t-statistic value often result into the acceptance of the null hypothesis. This paper, therefore, makes efforts to investigate the effect of multicollinearity on the power rates of the OLS estimator. A regression model with constant term (β_0) and two independent variables (with β_1 and β_2 as their respective regression coefficients) that exhibit multicollinearity was considered. A Monte Carlo study of 1000 trials was conducted at eight levels of multicollinearity (0, 0.25, 0.5, 0.7, 0.75, 0.8, 0.9 and 0.99) and sample sizes (10, 20, 40, 80, 100, 150, 250 and 500). At each specification, the true regression coefficients were set at unity while 1.5, 2.0 and 2.5 were taken as their hypothesized values. Results show that at each hypothesized value of β_0 the power rate is the same at all the levels of multicollinearity at a specified sample size and that the error rate decreases asymptotically. Furthermore as the hypothesized value increases, results do not only show that the power rate increases but tends faster to one asymptotically. The pattern of effect of power rate of β_1 and β_2 is the same as that of β_0 except that at each hypothesized value the power rate decreases as level of multicollinearity increases at a specified sample size. Consequently, increasing the sample size increase the power rate of the OLS estimator in all the levels of multicollinearity.

Key words: Regression model, OLS estimator, multicollinearity, power rates

INTRODUCTION

In classical linear regression model, regressors are assumed to be independent. When this assumption fails, the problem of multicollinearity arises. Multicollinearity is found in business and economics. For instance, the independent variables such as family income and assets or store sales and number of employees or age and years of experience would tend to be highly correlated. With strongly interrelated regressors, interpretation given to the regression coefficients may no longer be valid because the assumption under which the regression model is built has been violated^[4]. Although the estimates of the regression coefficients provided by the OLS estimator is still unbiased as long as multicollinearity is not perfect^[5], the regression coefficients may have large sampling errors which affect both the inference and forecasting that is based on the model^[4]. Oduntan^[6]

pointed out that multicollinearity is less serious when attention is on prediction or forecasting of values for the dependent variable than when interest is on estimates of the parameters of the model. With high standard errors of the regression coefficients, the value of the t-statistic of each of the regression coefficients is low causing the null hypothesis to be often accepted. Consequently with generated collinear data, this paper attempts to investigate the power rate ($1-\beta$) of the OLS estimators at different levels of multicollinearity and sample size through Monte Carlo studies.

MATERIALS AND METHODS

Consider the regression model of the form:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \quad (1)$$

where,

$$e_i \sim N(0, \sigma^2) \quad i = 1, 2, \dots, n$$

Corresponding Author: O.O. Alabi, Department of Mathematical Sciences, Olabisi Onabanjo University, P.M.B. 2002, Ago-Iwoye, Ogun State, Nigeria

y is the dependent variable, x_1 and x_2 are regressors which exhibit correlation (multicollinearity) and β_0, β_1 and β_2 are the regression coefficient (parameters) of the model.

Now, suppose $X_i \sim N(\mu_i, \sigma_i^2) i=1,2$. If these variables are correlated, then X_1 and X_2 can be generated with the equations

$$\begin{aligned} X_1 &= \mu_1 + \sigma_1 z_1 \\ X_2 &= \mu_2 + \rho\sigma_2 z_1 + \sigma_2 z_2 \sqrt{1-\rho^2} \end{aligned} \quad (2)$$

where $Z_i \sim N(0,1) i=1,2$ and $|\rho| < 1$ is the value of correlation between the two variables^[2,3].

Monte Carlo experiments were performed 1000 times for eight sample sizes ($n = 10, 20, 40, 80, 100, 150, 250$ and 500) and eight levels of multicollinearity. ($\rho = 0, 0.25, 0.5, 0.7, 0.75, 0.8, 0.9$ and 0.99) with stochastic regressors that are normally distributed. At a particular specification of n and ρ (a scenario), the first replication was obtained by generating $e_i \sim N(0,1), i=1,2,\dots,n$. Next, $X_{1i} \sim N(0,1)$ and $X_{2i} \sim N(0,1)$ were generated using equation (2) such that they exhibit correlation ρ . The values y_i in equation (1) were obtained by taking the true regression coefficient as unity. This process continued until all the 1000 replications had been done. Another scenario then started until all the scenarios were completed. For each replication in the scenario, the OLS method of parameter estimation was used to obtain estimates of the regression coefficients and the hypothesis about the hypothesized values of the regression coefficient was tested at 0.05 level of significance using the t-statistic to examine the type 11 error (β) of each of the regression coefficients. These values were taken as 1.5, 2.0 and 2.5. All these were accomplished by a computer programme using Time Series Processor (TSP version 5.0) software. Then, the power rates was obtained by taken the values of type 11 error rates away from one. These was done for every multicollinearity level at every selected sample sizes.

To facilitate a good understanding of the behaviours of the estimates as the true value is being changed to the hypothesized values, the following subdivisions are considered.

- Case 1:** True values of β_1 and β_2 are maintained while that of β_0 is allowed to change.
- Case 2:** True values of β_0 and β_2 are maintained while that of β_1 is allowed to change.
- Case 3:** True values of β_0 and β_1 are maintained while that of β_2 is allowed to change.
- Case 4:** The true value of β_0 is maintained while those of β_1 and β_2 are allowed to change.
- Case 5:** All the values of β_0, β_1 and β_2 are allowed to change.

RESULTS AND DISCUSSION

- Case 1:** True value of β_1 and β_2 are maintained while that of β_0 is allowed to change.

The summary of the power rate of β_0 when the true value β_1 and β_2 is maintained and that of β_0 is allowed to change at different levels of multicollinearity and sample size are respectively shown in Tables 1, 2 and 3.

From the tables, it can be observed that at each changed value of β_0 the power rate is the same for all levels of multicollinearity at a specified sample size; and that the power rate increases asymptotically. Furthermore, as the value of β_0 increases, the power rate increases and tends faster to one asymptotically. Alabi, (2007) show that type 1 error rate of β_1 and β_2 are still the same when β_0 changes Thus, the type 11 error rate of β_0 does not affect the type 1 error rate of β_1 and β_2 . Consequently, the type 11 error rate of β_0 depends on the departure of the hypothesized value from the true value and sample size.

Table 1: Power rates of β_0 when $\beta_0 = 1.5$ at different levels of multicollinearity and sample sizes

Sample size ρ	10	20	40	80	100	150	250	500
0	0.268	0.502	0.846	0.985	0.998	0.998	1.000	1.000
0.25	0.266	0.501	0.846	0.985	0.995	0.998	1.000	1.000
0.5	0.266	0.502	0.844	0.983	0.996	0.998	1.000	1.000
0.7	0.266	0.502	0.843	0.982	0.996	0.997	1.000	1.000
0.75	0.266	0.502	0.843	0.982	0.996	0.996	1.000	1.000
0.8	0.266	0.502	0.843	0.982	0.996	0.996	1.000	1.000
0.9	0.266	0.502	0.843	0.982	0.995	0.996	1.000	1.000
0.99	0.267	0.503	0.844	0.982	0.994	0.997	1.000	1.000

Table 2: Power rates of β_0 when $\beta_0 = 20$. at different levels of multicollinearity and sample sizes

Sample size p	10	20	40	80	100	150	250	500
0	0.697	0.973	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.696	0.973	1.000	0.999	1.000	1.000	1.000	1.000
0.5	0.695	0.973	1.000	0.999	1.000	1.000	1.000	1.000
0.7	0.694	0.973	1.000	0.999	1.000	1.000	1.000	1.000
0.75	0.694	0.973	1.000	0.999	1.000	1.000	1.000	1.000
0.8	0.693	0.027	1.000	1.000	1.000	1.000	1.000	1.000
0.9	0.694	0.973	1.000	1.000	1.000	1.000	1.000	1.000
0.99	0.693	0.973	1.000	1.000	1.000	1.000	1.000	1.000

Table 3: Power rates of β_0 when $\beta_0 = 2.5$ at different levels of multicollinearity and sample sizes

Sample size p	10	20	40	80	100	150	250	500
0	0.945	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.945	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.944	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.7	0.942	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.75	0.940	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.8	0.940	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.9	0.940	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.99	0.940	0.999	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: The Power rates of β_1 when $\beta_1 = 1.5$ at different levels of multicollinearity and sample sizes

Sample Size p	10	20	40	80	100	150	250	500
0	0.249	0.530	0.834	0.987	0.995	1.000	1.000	1.000
0.25	0.245	0.498	0.805	0.982	0.995	1.000	1.000	1.000
0.5	0.200	0.410	0.744	0.954	0.984	1.000	1.000	1.000
0.7	0.167	0.291	0.569	0.866	0.936	0.984	1.000	1.000
0.75	0.146	0.256	0.502	0.814	0.912	0.971	1.000	1.000
0.8	0.131	0.224	0.436	0.753	0.843	0.950	1.000	1.000
0.9	0.097	0.140	0.266	0.476	0.594	0.737	0.919	1.000
0.99	0.069	0.062	0.088	0.094	0.122	0.144	0.217	0.354

Table 5: The Power rates of β_1 when $\beta_1 = 2.0$ at different levels of multicollinearity and sample sizes

Sample Size p	10	20	40	80	100	150	250	500
0	0.671	0.954	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.651	0.945	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.568	0.821	0.999	1.000	1.000	1.000	1.000	1.000
0.7	0.454	0.790	0.985	1.000	1.000	1.000	1.000	1.000
0.75	0.399	0.790	0.967	1.000	1.000	1.000	1.000	1.000
0.8	0.344	0.642	0.934	0.997	1.000	1.000	1.000	1.000
0.9	0.213	0.403	0.739	0.958	0.985	0.998	1.000	1.000
0.99	0.176	0.088	0.16	0.225	0.313	0.411	0.578	0.891

Table 6: The Power rates of β_1 when $\beta_1 = 2.5$ at different levels of multicollinearity and sample sizes

Sample Size p	10	20	40	80	100	150	250	500
0	0.999	0.995	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.886	0.997	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.827	0.992	1.000	1.000	1.000	1.000	1.000	1.000
0.7	0.719	0.963	1.000	1.000	1.000	1.000	1.000	1.000
0.75	0.666	0.947	1.000	1.000	1.000	1.000	1.000	1.000
0.8	0.594	0.917	0.999	1.000	1.000	1.000	1.000	1.000
0.9	0.389	0.708	1.000	1.000	1.000	1.000	1.000	1.000
0.99	0.104	0.132	0.274	0.458	0.570	0.711	0.911	0.994

Case 2: True values of β_0 and β_2 are maintained while that of β_1 is allowed to change.

The summary of the power rates of β_1 at different levels of multicollinearity and sample size are shown in Tables 4, 5 and 6. From these tables, it can be observed that at each value of β_1 the power rate decreases as the

level of multicollinearity increases at a specified sample size and that the power rate increases asymptotically at each level of multicollinearity. This is more rapid as the hypothesized value increases for all levels of multicollinearity. Furthermore, as hypothesized value of β_1 increases, the type 11 error rate decreases and

Table 7: The Power rates of β_2 when $\beta_2 = 1.5$ at different levels of multicollinearity and Sample sizes

Sample Size ρ	10	20	40	80	100	150	250	500
0	0.253	0.548	0.836	0.989	0.997	1.000	1.000	1.000
0.25	0.242	0.521	0.810	0.984	0.996	1.000	1.000	1.000
0.5	0.200	0.398	0.721	0.961	0.985	0.002	1.000	1.000
0.7	0.148	0.301	0.543	0.888	0.929	0.982	1.000	1.000
0.75	0.063	0.372	0.534	0.827	0.892	0.974	1.000	1.000
0.8	0.116	0.236	0.460	0.744	0.823	0.949	1.000	1.000
0.9	0.079	0.140	0.279	0.503	0.567	0.750	0.934	1.000
0.99	0.058	0.062	0.073	0.102	0.123	0.146	0.208	0.374

Table 8: The Power rates of β_2 when $\beta_2 = 2.0$ at different levels of multicollinearity and Sample sizes

Sample Size ρ	10	20	40	80	100	150	250	500
0	0.650	0.949	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.626	0.942	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.560	0.904	0.996	1.000	1.000	1.000	1.000	1.000
0.7	0.423	0.795	0.985	1.000	1.000	1.000	1.000	1.000
0.75	0.376	0.731	0.962	1.000	1.000	1.000	1.000	1.000
0.8	0.329	0.676	0.930	0.997	0.001	1.000	1.000	1.000
0.9	0.204	0.448	0.725	0.962	0.985	0.998	1.000	1.000
0.99	0.063	0.098	0.151	0.260	0.288	0.399	0.595	0.878

Table 9: The Power rates of β_2 when $\beta_2 = 2.5$ at different levels of multicollinearity and Sample sizes

Sample Size ρ	10	20	40	80	100	150	250	500
0	0.894	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.880	0.999	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.824	0.994	1.000	1.000	1.000	1.000	1.000	1.000
0.7	0.704	0.968	1.000	1.000	1.000	1.000	1.000	1.000
0.75	0.646	0.948	1.000	1.000	1.000	1.000	1.000	1.000
0.8	0.582	0.917	0.997	1.000	1.000	1.000	1.000	1.000
0.9	0.367	0.718	0.957	1.000	1.000	0.998	1.000	1.000
0.99	0.082	0.149	0.263	0.468	0.543	0.712	0.913	0.991

power rate increases and tends faster to one asymptotically.

Case 3: True values of β_0 and β_1 are maintained while that of β_2 is allowed to change.

When the true value β_0 and β_1 are maintained and that of β_2 is allowed to change, the power rate of β_2 at different levels of multicollinearity and sample size are shown in Tables 7, 8 and 9.

A comparative examination of these tables reveals that the pattern of the results of the power rate of β_2 are essentially the same with that of β_1 already discussed in case 1.

Consequently, the power rates of the OLS estimator on the basis of β_1 and β_2 are affected by the levels of multicollinearity, sample size and the hypothesized value. Moreover with increasing sample size, the power rate of the OLS estimator increases at all levels of multicollinearity.

Case 4: The true value of β_0 is maintained while those of β_1 and β_2 are allowed to change.

When the true value of β_0 is maintained while that of β_1 and β_2 are allowed to change, the following combinations are possible:

- $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 1.5$
- $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 2.0$
- $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 2.5$
- $\beta_0 = 1, \beta_1 = 2.0, \beta_2 = 1.5$
- $\beta_0 = 1, \beta_1 = 2.0, \beta_2 = 2.0$
- $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 2.5$
- $\beta_0 = 1, \beta_1 = 2.5, \beta_2 = 1.5$
- $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 2.0$
- $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 2.5$

However, the results of these two combinations are considered:

Case 4.1: $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 2.5$

Case 4.2: $\beta_0 = 1, \beta_1 = 2.0, \beta_2 = 2.0$

Case 4.1: $\beta_0 = 1, \beta_1 = 1.5, \beta_2 = 2.5a$. The type 11 error rates of β_1 and β_2 are given in Tables 10 and 11.

Table 10: The Power rates of β_1 when $\beta_1 = 2.0$ at different levels of multicollinearity and Sample sizes

Sample Size ρ	10	20	40	80	100	150	250	500
0	0.671	0.954	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.651	0.945	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.568	0.821	0.999	1.000	1.000	1.000	1.000	1.000
0.7	0.464	0.790	0.985	1.000	1.000	1.000	1.000	1.000
0.75	0.399	0.726	0.977	1.000	1.000	1.000	1.000	1.000
0.8	0.344	0.642	0.944	0.997	1.000	1.000	1.000	1.000
0.9	0.213	0.403	0.739	0.958	1.000	1.000	1.000	1.000
0.99	0.176	0.088	0.160	0.225	0.313	0.411	0.578	0.891

Table 11: The Power rates of β_2 when $\beta_2 = 2.0$ at different levels of multicollinearity and Sample sizes

Sample Size ρ	10	20	40	80	100	150	250	500
0	0.650	0.949	1.000	1.000	1.000	1.000	1.000	1.000
0.25	0.626	0.942	1.000	1.000	1.000	1.000	1.000	1.000
0.5	0.560	0.904	0.996	1.000	1.000	1.000	1.000	1.000
0.7	0.423	0.795	0.985	1.000	1.000	1.000	1.000	1.000
0.75	0.376	0.731	0.962	1.000	1.000	1.000	1.000	1.000
0.8	0.329	0.676	0.930	0.997	0.999	1.000	1.000	1.000
0.9	0.204	0.448	0.725	0.962	0.985	0.998	1.000	1.000
0.99	0.063	0.098	0.151	0.260	0.288	0.389	0.595	0.878

It can be observed that the results from Tables 10 and 11 are identical to those of Tables 4 and 9. Thus, the explanations given earlier about the power rates of the parameters still hold.

Alabi (2007) also show that type1 error rate of β_1 is the same at different hypothesized values. Thus, the effect of the presence of multicollinearity on the type 1 and type 11 error rate of a particular regression coefficient does not affect the value of the other parameters.likewise for the power rates of the parameters.

Case 4.2: $\beta_0 = 1, \beta_1 = 2.0, \beta_2 = 2.0$. The power rates of β_1 and β_2 obtained are identical to the one given in Tables 5 and 8. Inferences based on Tables 5 and 8 also apply. It was also observed that for all other different combinations similar results are obtained though the results are not shown in this paper.

Case 5: All the values of β_0, β_1 and β_2 are allowed to change.

When all the values of β_0, β_1 and β_2 are allowed to change, the following combinations are possible:

- $\beta_0 = 1.5, \beta_1 = 1.5, \beta_2 = 1.5$
- $\beta_0 = 1.5, \beta_1 = 1.5, \beta_2 = 2.0$
- $\beta_0 = 1.5, \beta_1 = 1.5, \beta_2 = 2.5$
- $\beta_0 = 1.5, \beta_1 = 2.0, \beta_2 = 1.5$
- $\beta_0 = 1.5, \beta_1 = 2.0, \beta_2 = 2.0$
- $\beta_0 = 1.5, \beta_1 = 2.0, \beta_2 = 2.5$
- $\beta_0 = 1.5, \beta_1 = 2.5, \beta_2 = 1.5$

- $\beta_0 = 1.5, \beta_1 = 2.5, \beta_2 = 2.0$
- $\beta_0 = 1.5, \beta_1 = 2.5, \beta_2 = 2.5$
- $\beta_0 = 2.0, \beta_1 = 1.5, \beta_2 = 1.5$
- $\beta_0 = 2.0, \beta_1 = 1.5, \beta_2 = 2.0$
- $\beta_0 = 2.0, \beta_1 = 1.5, \beta_2 = 2.5$
- $\beta_0 = 2.0, \beta_1 = 2.0, \beta_2 = 1.5$
- $\beta_0 = 2.0, \beta_1 = 2.0, \beta_2 = 2.0$
- $\beta_0 = 2.0, \beta_1 = 2.0, \beta_2 = 2.5$
- $\beta_0 = 2.0, \beta_1 = 2.5, \beta_2 = 1.5$
- $\beta_0 = 2.0, \beta_1 = 2.5, \beta_2 = 2.0$
- $\beta_0 = 2.0, \beta_1 = 2.5, \beta_2 = 2.5$
- $\beta_0 = 2.5, \beta_1 = 1.5, \beta_2 = 1.5$
- $\beta_0 = 2.5, \beta_1 = 1.5, \beta_2 = 2.0$
- $\beta_0 = 2.5, \beta_1 = 1.5, \beta_2 = 2.5$
- $\beta_0 = 2.5, \beta_1 = 2.0, \beta_2 = 1.5$
- $\beta_0 = 2.5, \beta_1 = 2.0, \beta_2 = 2.0$
- $\beta_0 = 2.5, \beta_1 = 2.0, \beta_2 = 2.5$
- $\beta_0 = 2.5, \beta_1 = 2.5, \beta_2 = 1.5$
- $\beta_0 = 2.5, \beta_1 = 2.5, \beta_2 = 2.0$
- $\beta_0 = 2.5, \beta_1 = 2.5, \beta_2 = 2.5$

However, Alabi (2007) considered the following among all the possible combinations.

Case 5.1: $\beta_0 = 2.0, \beta_1 = 1.5, \beta_2 = 2.5$

Case 5.2: $\beta_0 = 1.5, \beta_1 = 2.5, \beta_2 = 2.0$

CASE 5.3: $\beta_0 = 2.5, \beta_1 = 2.5, \beta_2 = 2.5$.

The results for case 5 show that the same trend as for other cases. That is:

- The power rates increases asymptotically.
- Power rates decreases with increase in multicollinearity and
- Power rate is unaffected by change in the values of other parameters.

CONCLUSION

This study has revealed when multicollinearity is present in a data set to which regression analysis is to be applied, the power rate of the OLS estimator of β_0 at a specified sample size is the same in all the levels of multicollinearity; and that the power rate increases asymptotically. Moreover as the hypothesized value increases, the power rate tends faster to one asymptotically. It further revealed that the pattern of effect of the power rate of the OLS estimator of β_1 and β_2 is the same with that of β_0 except that at each hypothesized value the the power rate decreases as the levels of multicollinearity increases at a specified sample size. Consequently, increasing the sample size increases the power rate of the OLS estimator in all the levels of multicollinearity. Hence, the sample size with

the power rate value of one at each level of multicollinearity is the tolerable sample size .

REFERENCES

1. Alabi, O.O., 2007. Effects of Presence of Multicollinearity on Type 1 and Type II errors of the Ordinary Least Squares Estimators. Unpublished M.Sc. Thesis. University of Ilorin, Nigeria.
2. Ayinde, K., 2006. A comparative study of the performances of the OLS and some GLS estimators when regressors are both stochastic and collinear. *West Afr. J. Biophys. Biomath.*, 2: 54-67.
3. Ayinde, K. and B.A. Oyejola, 2007. A comparative study of the performances of the OLS and some GLS estimators when regressors are correlated with error terms. *Res. J. Applied Sci.*, 2: 215-220.
4. Chatterjee, S., A.S. Hadi and B. Price, 2000. *Regression Analysis by Example*. 3rd Edn. A Wiley-Interscience Publication. John Wiley and Sons.
5. Johnston, J., 1984. *Econometric Methods*. 3rd Edn., McGraw Hill, New York.
6. Odutan, 2004. A Monte Carlo Study of the Problem of Multicollinearity in a Simultaneous Equation system. Unpublished Ph.D Thesis. University of Ibadan, Nigeria.