The Economic Journal of Nepal, Vol.17, No. 4, Issue 68, Oct.-Dec.,1994 © DOE-TU

A Note on Significance Tests in Quadratic Regression

Bishwa Nath Tiwari*

SETTING

This note provides a description of significance tests for quadratic regression. Indeed,it does not attempt to add anything significant to the theory of quadratic function. Instead, it aims to reveal some inconsistencies concerning with the relationships of t and F tests that come across when one runs a polynomial regression using OLS method. With this aim, the author heavily draws upon earlier works, Maddala (1977), and Geary and Leser (1968),which focused only on linear functions. Hence, this exercise primarily helps statisticians to get the feel of statistics. Moreover, it may prove helpful to those involved in the field of neoclassical production/cost economics.

INTRODUCTION

The quality of t^2 and F ratio (t^2=F) is a case of simple regression but not of multiple regression. This feature therefore, brings in different relationship between b_1 and R^2 in multiple regression. Writing b_1 for the partial regression coefficients and R^2 for the coefficients of determination, Geary and Leser(1968) enumerate the following six situations.

(1) R^2 and all b1 significant

(2) R^2 and some but not all b_i significant

* Dr. Tiwari is lecturer at the Central Department of Economics, Tribhuvan University, Kathmandu, Nepal.

(3) R^2 but none of the b_i significant

(4) All b_i significant but not R^2

(5) Some b_i significant but not all nor R^2

(6) Neither R^2 nor any b_i significant

There are theoretical *a priori* criteria, statistical criteria and econometric criteria to evaluate the results of estimated regression. According to these tests, case(1) passes the statistical and econometric criteria, and so does the case (6). But they may not pass *a priori* criteria on the ground of incorrect specification of the model. Similar is the case(2) which suggests a respecification of the model, based on the theory, for obtaining all or most of the coefficients as significant. Also respecification of the model seems to be the solution for the case (5). However, there is an important difference between the cases (2) and (5), i.e., whereas the case (2) passes the statistical and econometric criteria, the case (5) does not. Moreover, the latter is more serious concern. However, in both cases some manipulation in the inclusion of variables in the regression model could result in the harmony of t and F test. Of particular note in this respect is therefore, the procedure of inclusion of significant and omission of insignificant variable(s), which Maddala(1977), and Geary and Leser(1968) do not seem to suggest as a usual procedure.

> If we search over a large number of variables, we are often bound to hit on a few that produce significant coefficients. But if we use this"search" procedure, we cannot use the usual t and F tables. What the appropriate t and F values should be is not known–all we can say is that they are higher than the tabulated values and perhaps much higher if we found the "significant" variables after searching over a large number of variables (Maddala 1977,pp. 122- 3) .

Cases (3) and (4) are the most severe ones. The former is the famous case of multicollinearity which occupies an important place in the econometric literature, but the latter is, in true sense, a new one and has never captured a wide attention due to the econometricians' facile excuse that it is not commonly observed in econometric work.

QUADRATIC FUNCTION: SOME PRELIMINARIES

A quadratic function is a second degree polynomial function. The word polynomial means *multiterm* and a polynomial function has the general form:

$$y = b_0 + b_1 x + b_2 x^2 + \ldots\ldots + b_n x^n \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

in which each term contains a coefficient as well as a non negative interior power to the variable x. Depending on the value of integer n (which specifies the highest power of x, called degree), there are different subclasses of polynomial function. Thus, the second-degree polynomial or quadratic function is written as:

$$y = b_0 + b_1 x + b_2 x^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

If we take positive square root of the independent variable x, we have:

$$y = b_0 + b_1 x^{1/2} + b_2 x \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

The function (3) may be called as a square root quadratic function. To distinguish it from function (2), the function (2) can be referred to as a simple quadratic function. The square root quadratic function provides a simple compromise between the exponential function and the simple quadratic function. Similar modifications in the simple quadratic function my be used to characterize different situations (see Heady and Dillon 1961, p.79)

The simple quadratic function plots as a parabola– roughly, a curve with a single build-in bump or wiggle. Similarly, the square root quadratic function plots roughly as a hockey-stick shape ( ) or inverted hockey-stick shape curve in which the change in the ordinate value is successively smaller after the extreme point. Thus, there is sharp turn on the left and a gradual turn on the right of the extremum point of the curve depicting the square root quadratic function. In particular, when $b_2 < o$ the quadratic function displays a hill, but when $b_2 > o$ it displays a valley.

Sometimes a quadratic function is devised on *a priori* considerations. One particular instance is its specification according to the principle of the neo-classical theory of production to depict production and/or cost relationships. More often than not, it is fitted simply because a straight line appears not to give a good fit to the data. In the latter case, it is very often employed to test for deviation from linearity in data which have already ben used to estimate a linear function.

The quadratic function and other higher degree ploynomials are generally estimated by the method of OLS, by making suitable transformations of data before estimation of the parameters. In such a transformation the x's with different exponents are treated as separate variables. For example, in a second-degree polynomial, besides a linear term there is a quadratic term which decides the curvature of the line introducted by the liner term. In general, it can be said that the sign of the quadratic coefficient determines the nature of wiggle or turn over of the curve, and the constant and linear coefficient determine the location of the wiggle. More specifically, the sign of the constant term determines whether the curve turns over to above or below the origin, whereas the sign of the linear coefficient determines whether it turns over to the right or left of the origin. If the constant term is positive, and the linear and quadratic coefficients are of alternate signs, the wiggle takes place in the first quadrant of the cartesian coordinate plane. Of particular importance is the instance where the linear coefficient is positive and the quadratic one is negative. In this case the curve will appear in conformity with the behaviour of average product or marginal product of the neo-classical theory of production.

SIGNIFICANCE TESTS IN QUADRATIC REGRESSION

The purpose of the preceding observations is to tailor the background for approaching the problem in the contest of quadratic functions. In the present context, we waive the cases (1) and (6), for the discordance lies in other cases. Let us proceed from case (2) to (5). Case (2) exhibits a condition when R^2 and some but not all b_i are significant. With regard to the quadratic function, it states a situation when either linear or quadratic coefficient turns up significant with significant R^2. If R^2 is significant with significant quadratic coefficient, the quadratic function depicts a parabola with turns at the ordinate, Thus, from the point of view of the theory of production this combination does not yield a meaningful result as it does not incorporate both the decreasing and increasing ordinate value in the (positive) domain of the input. On the other hand, the other combination, i.e. significant R^2 and linear coefficient, precludes the possibility of any curvature in the data. Thus, it suggests a linear relation.

Now coming to case (3) this is a situation of multi-collinearity which comes about when both linear and quadratic coefficients are found insignificant despite a significant R^2. In spite of the positive correlation existing between the linear and quadratic terms of the quadratic function, this situation does not occur much in econometric work. However, if it occurs the usual remedies of multi-collinearity problem break down. Thus, to contrive from the situation one is bound to change the functional form. Those who use quadratic functions to test the deviation from linearity assume their linear counterparts (Raj Krishna 1964). Finally, the remaining two cases, (4) and (5) are:

(4) All b_i significant but not R^2

(5) Some b_i significant, but not all, or R^2

These situations occur both in simple and square root quadratic function (see Tiwari 1988). As regards case (5), greater possibility exists for a significant linear coefficient but insignificant R^2. If this is so, one could run a linear regression but it is not certain that he will come up with significant R^2. However, this situation certainly calibrates our eye estimates of curvature. Thus, one may solve by respecifying the model.

The case (4) is more problematical, however, Geary and Leser (1968) point out that given $b_i > 0$ the situation of multi-collinearity, i.e. case (3), occurs when simple correlations between independent variables are positive and very high, whereas case (4) occurs when such a correlation is predominnatly (moderately) negative. Maddala (1977) gives a lucid illustration of the latter situation by presenting a zero order correlation matrix for the dependent variable (y) and two independent variables, x_1 and x_2 (Table1).

Table 1

Zero-Order Correlation Matrix

	Y	X_1	X_2
Y	1.0	0.1	0.1
X1	0.1	1.0	-0.5
X_2	0.1	-0.5	1.0

In this case the partial correlations will be given by:

$$r^2_{y1.2} = r^2_{y2.1} = 0.023$$

and the multiple correlation coefficient is given by $R^2 = 0.0237$. He further mentions that if we have 175 observations, noting that $t^2/(t^2+df)$ = partial r^2, we see that each of the t ratios is 2.01, which for 172 degrees of freedom is significant at the five percent level. The F test for R^2 is:

$$F = \frac{0.0327}{1 - 0.0327} X \frac{172}{2} = 2.91$$

which for degrees of freedom 2 and 172 is not significant at the 5 percent level. Thus, Maddala says that one can argue in this case hat the regression equation is useless anyway. However, he fruther notes that the tabulated 5 percent value of F for the corresponding degrees of freedom is 3.0, which is only slightly above the computed F value. The above example therefore, illustrates how one can obtain each individual coefficient significant but the regression equation insignificant.

However, the above illustration is a case of linear multiple regression. Therefore, it may not directly fit in quadratic polynomials. It is evident from the fact that Maddala's illustration is based on two assumptions: (i) the simple correlations between the dependent variable and independent variable (x_{y1}, r_{y2}) are positive and small; and (ii) the simple correlation between independent variables is moderately negative. But these two assumptions generally does not hold in case of a quadratic function. Given the positive domain of the independent variable in the quadratic polynomial, the simple correlation between the linear term (x) and the quadratic term (x^2) appears positive. Also, their coefficients should assume alternate signs for a meaningful economic interpretation. Therefore, it indeed becomes difficult to explain how the significant b_i with insignificant R^2 comes about in a quadratic polynomial regression.

CONCLUSION

Thus, of all the cases, the case (4) seems the most intricate one. To contrive from this bizarre relationship requires a brilliant feat; otherwise the antidote becomes a remote possibility. By then one had better bring forward such problems and take caution in drawing conclusions.

SELECTED REFERENCES

Geary, R.C. and Leser, C.E.V. (1968) "Significance Tests in Multiple Regression" *The American Statistician* Vol. 22, No.1, pp.20-1.

Heady, E.O. and Dillon, J.L. (1961) *Agricultural Production Function,* Ludhiana, Kalyani Publishers.

Maddala, G.S. (1977), *Econometrics,* Auckland, Mc Graw-Hill International Book company.

Krishna, Raj (1964), "Some Production Functions for the Punjab", *Indian Journal of Agricultural Economics,* Vol. 19, Nos. 3 and 4.

Tiwari, Bishwa N. (1988) , *Farm Size and Productivity in Nepalese Agriculture: A Case Study of Nawal Parasi District,* An unpublished Ph.D. Thesis submitted to the University of Rajasthan, Jaipur, India.