A NOTE ON TESTING THE ERROR SPECIFICATION IN NONLINEAR REGRESSION

Dennis Leech

NUMBER 48

WARWICK ECONOMIC RESEARCH PAPERS

DEPARTMENT OF ECONOMICS

UNIVERSITY OF WARWICK
COVENTRY

A NOTE ON TESTING THE ERROR SPECIFICATION IN NONLINEAR REGRESSION

Dennis Leech

NUMBER 48

April, 1974

A NOTE ON TESTING THE ERROR SPECIFICATION IN NONLINEAR REGRESSION [1]

In typical economic applications of nonlinear regression methods the "systematic part" of the equation is specified with reference to theoretical considerations. The error term however, is usually introduced at the estimation stage in a manner most appropriate to the use of linear regression methods. A common example is the case of a Cobb-Douglas type function which is typically estimated in a form which is linear in the logarithms. The convenient assumption implicit in this procedure is that the error term is multiplicative with a lognormal distribution. While this assumption confers the practical advantage of allowing linear methods to be used it may not be the most realistic. It may well be more appropriate to introduce the error term additively and normally distributed. Ideally, of course, the form of the distribution of the error term would emerge in the derivation of the estimating equation from an underlying stochastic hypothesis. The most appropriate estimation technique could then be applied and the properties of the estimator derived. In many cases, however, the underlying theory is nonstochastic and nothing a priori is known about the form of the error distribution. In these circumstances if the disturbance term is incorrectly introduced into the model, then a misspecification arises. This misspecification has been referred to by Kmenta {5}, Malinvaud {6}, Goldfeld and Quandt {4} and Bodkin and Klein {1}. It is not the purpose of this note to analyse this misspecification further, but to suggest a procedure which can be followed to choose the form of the error distribution on statistical grounds in circumstances where the form of the nonstochastic part of the model can be taken as given.

Consider a nonlinear relation which we assume can be specified in either of two ways depending on whether the error term is additive or multiplicative. [2]

Either

(1) $y_i = g(X_i, \beta) + \varepsilon_i, \; E(\varepsilon_i) = 0, \; E(\varepsilon_i^2) = \sigma_\varepsilon^2, \; E(\varepsilon_i\varepsilon_j) = 0, i \neq j$

or

(2) $y_i = g(X_i, \beta)e^{\eta_i}, \; E(\eta_i) = 0, \; E(\eta_i^2) = \sigma_\eta^2, \; E(\eta_i\eta_j) = 0, i \neq j$

In both cases y_i is the dependent variable, X_i is a vector of independent variables, β is a vector of parameters, $g(\cdot)$ is a given function, ε_i and η_i are error terms and n is the sample size.

In both cases parameter estimation is straightforward using either least squares or maximum likelihood.

If model (1) is assumed to hold, then least squares estimates of β are obtained by minimising the residual sum of squares

(3) $S_1(\hat{\beta}) = \sum_{i=1}^{n} \left[y_i - g(X_i, \hat{\beta}) \right]^2$

with respect to $\hat{\beta}$. If it can be further assumed that the error term ε_i is normally distributed then $\hat{\beta}$ is also a maximum likelihood estimator and therefore has the properties of consistency, asymptotic efficiency and has an asymptotic normal distribution. Inferences concerning β can be made using standard maximum likelihood theory. [3]

If specification (2) is assumed, the standard approach is to find a least squares estimator of β in the model.

$$(4) \qquad \log y_i = \log g(X_i, \beta) + \eta_i$$

by minimising

$$(5) \qquad S_2(\hat{\hat{\beta}}) = \sum_{i=1}^{n} \left[\log y_i - \log g(X_i, \hat{\hat{\beta}})\right]^2$$

with respect to $\hat{\hat{\beta}}$. In this case $\hat{\hat{\beta}}$ is a maximum likelihood estimator if it can be assumed that η_i is normally distributed.

In both cases the minimisation can be achieved either by solving the normal equations or by applying a numerical minimisation algorithm directly to the sum of squared residuals.[4] In the general case the normal equations are nonlinear and estimation can only be achieved using iterative methods.

In this situation we are faced with two related questions. First, how do we choose on statistical grounds alone which assumption about the form of the error distribution is the more appropriate ? Second, if one model is found to be superior, is the difference between the two significant ?

Goldfeld and Quandt {4} summarise several approaches which treat the problem as essentially one of discriminating between discrete families of hypotheses.[5] They also put forward an approach in the specific context of Cobb-Douglas type functions by which (1) and (2) are special cases of a more general, mixed model involving both types of error term in the same equation.

Writing

$$(6) \qquad y_i = \beta_0 X_{1i}^{\beta_1} X_{2i}^{\beta_2} \ldots\ldots X_{ki}^{\beta_k} e^{\eta_i} + \varepsilon_i$$

the Goldfeld and Quandt approach is to maximise the likelihood of the sample of observations on the dependent variable under the assumption that η_i and ε_i are normally distributed with standard errors respectively of σ_η and σ_ε. When $\sigma_\eta = 0$ the model reduces to the form of (1) and when $\sigma_\varepsilon = 0$ it assumes the form of (2) as a special case. Goldfeld and Quandt therefore suggest that hypotheses concerning the form of the error distribution might be tested by likelihood ratio tests on σ_η and σ_ε respectively. This approach is promising and certainly capable of producing satisfactory results. One drawback, however, is that the likelihood function involves the evaluation of a definite integral for which no closed expression can be found in general. Maximisation of the likelihood function therefore requires the use of a numerical integration algorithm, in conjunction with one for maximisation, and consequently significant rounding errors are likely to appear.

An approach similar in spirit to that of Goldfeld and Quandt, although easier to follow in practice, is based on the method of transformations of Box and Cox {2}. Using this approach we define a family of transformations of both the dependent variable and the "systematic part" of the relationship by some parameter λ. For particular values of λ this family must give (1) and either (2) or (4) as special cases. Testing the form of the error specification could then be approached by testing hypotheses about λ.

Consider the simple power transformation for $a > 0$,

$$a^{(\lambda)} = \begin{cases} \frac{a^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log a & \lambda = 0 \end{cases}$$

which is continuous in λ.

Using this transformation we can define a family of models

$$y_i^{(\lambda)} = [g(X_i, \beta)]^{(\lambda)} + \delta_i \tag{7}$$

where δ_i is an error term with $E(\delta_i) = 0$, $E(\delta_i^2) = \sigma^2$, $E(\delta_i\delta_j) = 0$, $i \neq j$. Equations (1) and (4) are clearly both special cases of (7) corresponding to $\lambda = 1$ and 0 respectively.

If it is assumed that for some value of λ the error term δ_i has an independent normal distribution, then λ and β can be estimated by maximum likelihood.

Under the assumption that, for some λ,

$$y_i^{(\lambda)} \sim N([g(X_i, \beta)]^{(\lambda)}, \sigma^2)$$

its probability density is

$$\frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{1}{2\sigma^2} (y_i^{(\lambda)} - [g(X_i, \beta)]^{(\lambda)})^2 \right\}$$

The density of the untransformed dependent variable is therefore

$$\frac{y_i^{\lambda-1}}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left(y_i^{(\lambda)} - \left[g(X_i, \beta\right]^{(\lambda)}\right)^2\right\}$$

and hence the log likelihood in relation to the sample observations is

$$L\left(\beta,\lambda\right) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 + (\lambda-1) \sum_{i=1}^{n} \log y_i - \frac{1}{2\sigma^2} S(\beta,\lambda),$$

where

$$S(\beta,\lambda) = \sum_{i=1}^{n} \left(y^{(\lambda)} - \left[g(X_i, \beta)\right]^{(\lambda)}\right)^2 .$$

Replacing σ^2 by its maximum likelihood estimator,

$$\tilde{\sigma}^2 = \frac{1}{n} S(\tilde{\beta},\tilde{\lambda})$$

where $\tilde{\beta}\ \tilde{\lambda}$ are maximum likelihood estimators, gives the condensed log likelihood function

$$(8) \quad L(\tilde{\beta},\tilde{\lambda}) = -\frac{n}{2}\left(1 + \log \frac{2\pi}{n}\right) - \frac{n}{2} \log S(\tilde{\beta},\tilde{\lambda}) + (\tilde{\lambda}-1) \sum_{i=1}^{n} \log y_i \;.$$

Maximising (8) with respect to $\tilde{\beta}$ and $\tilde{\lambda}$ produces an estimate of λ on which a test might be based.

A confidence interval for λ can be found using the distribution of the likelihood ratio. In large samples the logarithm of the likelihood ratio is proportional to a χ^2 variable. In this case, therefore, we can construct an asymptotic confidence interval for λ using the quantity

$$(9) \quad 2\{L(\tilde{\beta},\tilde{\lambda}) - L(\tilde{\tilde{\beta}}, \lambda)$$

$$= n \log S(\tilde{\tilde{\beta}},\lambda) - n \log S(\tilde{\beta},\tilde{\lambda}) + 2(\tilde{\lambda} - \lambda) \sum_{i=1}^{n} \log y_i \ ,$$

from (8). This statistic is distributed asymptotically as χ^2 with one degree of freedom if $\tilde{\tilde{\beta}}$ is the maximum likelihood estimator of β conditional upon λ.[6]

This method was used to test the form of the error specification in an aggregate CES production function for Soviet industry. The data used was annual data for the period 1950-69 given by Weitzman {8} who estimated the function directly assuming constant returns to scale, Hicks neutral technical change and a multiplicative, serially independent error term. The "systematic part" of the equation is written

$$g(\cdot) = \gamma e^{\mu t} \left[\delta K_t^{-\rho} + (1 - \delta) L_t^{-\rho} \right]^{-\frac{1}{\rho}} ,$$

where K_t is aggregate capital and L_t aggregate labour at time t. Three sets of estimates were obtained by maximising (8) with, respectively, the restrictions $\lambda = 1$, $\lambda = 0$ and also with λ unrestricted but nonzero.

Maximising (8) with $\lambda \neq 0$ is equivalent to minimising

$$(10) \quad \frac{n}{2} \log \sum_{t=1}^{n} \frac{1}{\lambda^2} \left(y_t^{\lambda} - \gamma^{\lambda} e^{\mu\lambda t} \left[\delta K_t^{-\rho} + (1 - \delta) L_t^{-\rho} \right]^{-\frac{\lambda}{\rho}} \right)^2$$

$$- (\lambda - 1) \sum_{t=1}^{n} \log y_t$$

In the case $\lambda = 1$, estimates were found by minimising (10) subject to this restriction and in the case $\lambda = 0$, by minimising

$$(11) \quad \sum_{t=1}^{n} \left(\log y_t - \log \gamma - \mu t + \frac{1}{\rho} \log\left[\delta K_t^{-\rho} + (1 - \delta) L_t^{-\rho}\right]\right)^2 .$$

The minimisation algorithm used was a variant of the conjugate gradient method of Powell.[7] The results are shown in theTable. These estimates are, of course, subject to the caveat that the likelihood function may have more than one maximum and that the algorithm used is only capable of finding a local maximum. In general, therefore, we are unsure of whether we have attained the required global maximum. In this case we have re-estimated the parameters a number of times using each time a different vector of initial approximations and no improvement in the likelihood function has been observed. It therefore seems reasonable to accept these figures as maximum likelihood estimates.

Aggregate CES production function for the USSR:
Maximum likelihood estimates

	(a)	(b)	(c)
ρ	1.7757 (0.3558)	1.4812 (0.1665)	1.5314
δ	0.4617 (0.0785)	0.6390 (0.0628)	0.6775
γ	0.6803 (0.0414)	0.7876 (0.0420)	0.8145
μ	0.0340 (0.0057)	0.0205 (0.0048)	0.0176
λ	1	0	-0.4773
log likelihood (apart from constant)	-38.689	-32.394	-30.9194

Estimated asymptotic standard errors in brackets.[8]

In the Table, column (a) contains estimates obtained under the assumption of an additive error term (i.e. model (1)) by minimising (10) subject to $\lambda = 1$. The estimates in column (b) were found under the hypothesis of a multiplicative error (i.e. model (2)) by minimising (11) and thosein column (c) by minimising (10) without any restriction on λ except $\lambda \neq 0$. The log likelihood is given in each case as the value of (8) apart from the constant.

It will be seen that the coefficient estimates are quite sensitive to the way in which the error term is introduced. The choice between form (1) and form (4) must be made in favour of (4) since the likelihood is higher in this case. This conclusion is borne out by computing the statistic (9) for $\lambda = 1$ and $\lambda = 0$. For $\lambda = 1$ its value is 15.54 and for $\lambda = 0$ its value is 2.95. Since $\chi_1^2(0.95) = 3.84$, we conclude that the 95% confidence interval includes $\lambda=0$ but not $\lambda = 1$. We can therefore argue that, in this example, the data tends to support the hypothesis that the error term is multiplicative rather than additive.

Footnotes

1. The author acknowledges Graham Pyatt for comments and Diane Ellwood for computing assistance.

2. It is assumed that the probability of negative values of y_i is negligible.

3. See, for example, Mood and Graybill {7}.

4. See Goldfeld and Quandt {4}, Ch. 1.

5. See Chapter 5.

6. This is the case provided certain regularity conditions are satisfied. In particular the likelihood function must be twice continuously differentiable which it clearly is in this case if $g(\cdot)$ is. Also the range of variation of the dependent variable must be independent of the parameter values. This is guaranteed by footnote 2.

7. See Brent {3}.

8. These were obtained by evaluating the inverse of the relevant information matrix at the estimated maximum of the likelihood function. They are consistent estimates of the population standard errors provided a set of jointly sufficient estimators exist for the parameters.

References

{1} Bodkin, R.C. and L. R. Klein, "Nonlinear Estimation of Aggregate Production Functions," Review of Economics and Statistics 49, pp.28-44.

{2} Box, G. E. P. and D. R. Cox, "An Analysis of Transformations", Journal of the Royal Statistical Society, Series B, Vol. 26, 1964, pp.211-243.

{3} Brent, R. P., Algorithms for Minimisation without Derivatives, Prentice-Hall, Englewood Cliffs, N.J., 1973.

{4} Goldfeld, S. M. and R. E. Quandt, Nonlinear Methods in Econometrics, North-Holland, London, 1972.

{5} Kmenta, J., Elements of Econometrics, MacMillan, New York, 1971.

{6} Malinvaud, E., Statistical Methods of Econometrics, North-Holland, 1966.

{7} Mood, A. M. and F. A. Graybill, Introduction to the Theory of Statistics, McGraw-Hill, New York, 1973.

{8} Weitzman, M. K. "Soviet Postwar Economic Growth and Capital-Labor Substitution", American Economic Review, 1970, Vol. LX, No.4, pp.676-692.