<table>
<thead>
<tr>
<th>Title</th>
<th>A simple artificial regression based Lagrange multiplier test of normality in the probit model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors(s)</td>
<td>Murphy, Anthony</td>
</tr>
<tr>
<td>Publication date</td>
<td>1994-07</td>
</tr>
<tr>
<td>Series</td>
<td>UCD Centre for Economic Research Working Paper Series; WP94/22</td>
</tr>
<tr>
<td>Publisher</td>
<td>University College Dublin. School of Economics</td>
</tr>
<tr>
<td>Item record/more information</td>
<td><a href="http://hdl.handle.net/10197/1951">http://hdl.handle.net/10197/1951</a></td>
</tr>
<tr>
<td>Notes</td>
<td>A hard copy is available in UCD Library at GEN 330.08 IR/UNI</td>
</tr>
</tbody>
</table>
"A Simple Artificial Regression Based Lagrange Multiplier Test of Normality in The Probit Model"

by

Anthony Murphy
Department of Economics
University College Dublin

October 1993

Working Paper WP94/22
A SIMPLE ARTIFICIAL REGRESSION BASED
LAGRANGE MULTIPLIER TEST OF NORMALITY IN THE PROBIT MODEL*

Anthony Murphy
Department of Economics
University College Dublin
Dublin 4, Ireland

July 1994

Abstract: A convenient artificial regression based LM test of non-normality in the probit model is derived using a Gram Charlier type A alternative. The test is simply derived and may be extended to the bivariate probit case. The outer product gradient form of LM test is not used so the proposed test is likely to perform reasonably well in small samples. The test is compared with two other existing tests.

Keywords: Probit, Lagrange Multiplier Test, Non-Normality, Gram Charlier Series, Artificial Regression, Outer Product Gradient.

JEL No: C35.

Introduction

In this paper a simple artificial regression based Lagrange Multiplier (LM) test of normality in the probit model is derived. The LM test uses the Gram Charlier type A alternative. The test is simply derived and may be extended to the bivariate probit case. The proposed LM test is likely to have reasonable small sample properties since it is not based on the outer product gradient (OPG) form of the test. The information matrix is calculated as the expectation of the outer product of the contributions to the score and not just approximated by this outer product.

---

*This paper was written up whilst attending the Warwick Summer Economics Workshop. The Workshop was was supported by the Economic and Social Research Council (UK) and the Human Capital and Mobility Programme of the European Community.


2 Kiefer and Salmon (1983) use the Gram Charlier type A expansion to derive a LM test for non-normality in the linear regression model.

3 See Murphy (1994c) for details.


5 Davidson and MacKinnon (1984b) originally proposed the non-OPG artificial regression based approach to LM tests for omitted variables and neglected heteroscedasticity in logit and probit models. Engle (1984) also suggested using this approach. Murphy (1994a) shows that, using this approach, many other mis-specification tests in both binary choice models and some more general discrete choice
Bera, Jarque and Lee (1982) and Chesher and Irish (1987) also derive LM tests of non-normality for the probit model. Bera et. al. use a Pearson distribution alternative which complicates the derivation. Their LM test does not use the OPG approximation to the information matrix and they do not suggest calculating the LM test statistic using an artificial regression. Chesher and Irish use a heuristic argument to derive, what is in effect, a conditional moment based test\(^a\). The conditional moment based approach naturally leads to the use of the OPG form of LM test. The scores in Chesher and Irish test are linear combinations of those in Bera et. al. so the two LM test statistics are numerically equal if both are calculated in the same way. The Bera et. al. test is compared with the new test proposed in this paper. Since the two tests examine different scores, to see if they are significantly different from zero, their small sample properties are investigated.

**The Type A Gram Charlier Expansion**

Both Kendall and Stuart (1977) and Ord (1972) show that, under suitable regularity conditions, any probability density function \( f(u) \) has a series expansion in terms of the standard normal density \( \phi(u) \) and the Hermite polynomials \( H_j(u) \):

\[
f(u) = \phi(u) \sum_{j=0}^{\infty} c_j H_j(u) \tag{1}
\]

The Hermite polynomials are orthogonal and are defined as:

\[
H_j(u) = \frac{(-D)^j \phi(u)}{\phi(u)} \tag{2}
\]

where \( D \) is the differentiation operator. The \( c_j \) coefficients satisfy:

\[
j! c_j = \int H_j(u)f(u)du \tag{3}
\]

Thus the first few Hermite polynomials \( H_j(u) \) are:

\[
\begin{align*}
H_0(u) &= 1 \\
H_1(u) &= u \\
H_2(u) &= u^2 - 1 \\
H_3(u) &= u^3 - 3u \\
H_4(u) &= u^4 - 6u^2 + 3
\end{align*}
\]

and the corresponding \( c_j \) coefficients are as follows:

---

\(^a\) See Newey (1985) and Tauchen (1985) for details of conditional moment tests.
\[ \begin{align*}
    c_0 &= 1 \\
    c_1 &= \mu = 0 \\
    2! \cdot c_2 &= \mu^2 - 1 = \sigma^2 - 1 = 0 \\
    3! \cdot c_3 &= \mu^3 = \kappa_3 \\
    4! \cdot c_4 &= \mu^4 - 6\sigma^2 + 3 = \mu_4 - 3 = \kappa_4
\end{align*} \]

where \( \kappa_3 \) and \( \kappa_4 \) are cumulants and \( u \) is standardised by assumption i.e. \( u \) has a mean of zero and a unit variance. \( \kappa_3 \) and \( \kappa_4 \) are measures of asymmetry and excess kurtosis which are zero when \( u \) is normally distributed.

The type A Gram Charlier approximation for \( f(u) \) is based on the first four terms in the expansion (1):

\[ f(u) \approx \phi(u) \left[ 1 + \frac{\kappa_3}{3!} H_2(u) + \frac{\kappa_4}{4!} H_4(u) \right] \]  

Since:

\[ \int_{-\infty}^{\infty} H_i(u) \phi(u) \, du = H_{i-1}(-z) \phi(-z) \]

the probability that \( u > -z \) is approximately:

\[ 1 - F(-z) \approx 1 - \Phi(-z) + \frac{\kappa_3}{3!} H_2(-z) \phi(-z) + \frac{\kappa_4}{4!} H_4(-z) \phi(-z) \]

\[ = \Phi(z) + \frac{\kappa_3}{6} H_2(z) \phi(z) - \frac{\kappa_4}{24} H_4(z) \phi(z) \]  

(5)

where \( \Phi \) is the cumulative standard normal distribution.

The Probit Model and the Gram Charlier Alternative

In the probit model the unobserved variable \( y^* \) is generated by the latent regression:

\[ y^* = x' \beta + u \]  

(6)

where \( u \) has a standard normal distribution. The sign of \( y^* \) is indicated by the dummy variable \( y \) i.e. \( y \) equals 1 if \( y^* > 0 \) and \( y \) equals 0 otherwise. Then:
\[
\text{prob}(y = 1) = \text{prob}(y^* > 0) = \text{prob}(u > x'\beta) = 1 - \Phi(-x'\beta) = \Phi(x'\beta)
\] (7)

The Gram Charlier alternative to the probit model has:

\[
\text{prob}(y = 1) = \Phi(x'\beta) + \frac{\kappa_3}{6} H_2(x'\beta) \Phi(x'\beta) - \frac{\kappa_4}{24} H_3(x'\beta) \Phi(x'\beta)
\] (8)

which collapses to the probit model under the null \(\kappa_3 = \kappa_4 = 0\) ie. when \(u\) is normal or, strictly speaking, when \(u\) is symmetric and has no excess kurtosis.

The Likelihood, Score and Information Matrix

With a random sample of size \(N\), the log likelihood is:

\[
I = \sum_{i=1}^{N} [y_i \ln p_i + (1 - y_i) \ln (1 - p_i)]
\] (9)

where the subscript \(i\) refers to individuals and \(p_i = \Phi(x'_i\beta)\). Under the null hypothesis \(\kappa_3 = \kappa_4 = 0\), the scores are:

\[
\frac{\delta I}{\delta \beta} = \sum_{i} \frac{y_i - p_i}{p_i(1 - p_i)} \phi(x'_i\beta) x_i
\]

\[
\frac{\delta I}{\delta \kappa_3} = \sum_{i} \frac{y_i - p_i}{p_i(1 - p_i)} \phi(x'_i\beta) H_2(x'_i\beta)
\] (10)

\[
= \sum_{i} \frac{y_i - p_i}{p_i(1 - p_i)} \phi(x'_i\beta) \frac{1}{6} [(x'_i\beta)^2 - 1]
\]

\[
\frac{\delta I}{\delta \kappa_4} = -\sum_{i} \frac{y_i - p_i}{p_i(1 - p_i)} \phi(x'_i\beta) \frac{1}{24} H_3(x'_i\beta)
\]

\[
= -\sum_{i} \frac{y_i - p_i}{p_i(1 - p_i)} \phi(x'_i\beta) \frac{1}{24} [(x'_i\beta)^3 - 3x'_i\beta]
\]

The elements of the information matrix are:
\[ I_{p\theta} = \lim_{N \to \infty} E \frac{1}{N} \frac{\delta l}{\delta \theta} \frac{\delta l}{\delta \theta'} \]

\[ = \lim_{N \to \infty} \frac{1}{N} \sum_{i} \frac{1}{p_i(1-p_i)} \phi(x_i^2) x_i x_i' \]

(11)

\[ I_{p\kappa_3} = \lim_{N \to \infty} E \frac{1}{N} \frac{\delta l}{\delta \kappa_3} \frac{\delta l}{\delta \kappa_3} \]

\[ = \lim_{N \to \infty} \frac{1}{N} \sum_{i} \frac{1}{p_i(1-p_i)} \phi(x_i^2) x_i x_i' \frac{1}{6} [(x_i^2)^2 - 1] x_i \]

etc. since \( E(y_i - p_i)^2 - p_i(1-p_i). \) The information matrix is assumed to be non-singular. The other standard regularity conditions are also assumed to hold.

**LM Test Statistic**

Let \( \theta' = (\beta', \kappa') \) denote the vector of parameters where \( \kappa' = (\kappa_3, \kappa_4) \) and let a tilde denote the restricted vector of parameter estimates under the null \( \kappa_3 = \kappa_4 = 0. \) Then the LM test statistic is:

\[ \text{LM} = \frac{1}{N} \frac{\delta l}{\delta \hat{\theta}} \frac{\delta l}{\delta \hat{\theta}'} \]

\[ = \frac{1}{N} \frac{\delta l}{\delta \hat{\kappa}'} (I_{\hat{\theta} \theta} \hat{I}_{\theta \theta} \hat{I}_{\theta \theta}^{-1} \hat{I}_{\theta \kappa})^{-1} \frac{\delta l}{\delta \hat{\kappa}} \]

(12)

using the score and observed information matrix evaluated at the restricted parameter estimates. The observed information matrix is:

\[ I_{\hat{\theta} \theta} = E \frac{1}{N} \frac{\delta l}{\delta \hat{\theta}} \frac{\delta l}{\delta \hat{\theta}'} \]

\[ = \frac{1}{N} \frac{1}{\hat{p}_{ij}(1-\hat{p}_{ij})} \frac{\delta p_{ij}}{\delta \hat{\theta}} \frac{\delta p_{ij}}{\delta \hat{\theta}'} \]

(13)

Under weak regularity conditions, the LM test statistic is distributed as \( \chi^2(2) \) under the null.
Artificial Regression Based LM Test

The observed score and information matrix may be rewritten as:

\[ \frac{\delta l}{\delta \tilde{\beta}} = \sum \tilde{r}_i \tilde{s}_i \]
\[ \frac{\delta l}{\delta \tilde{k}_3} = \sum \tilde{r}_i \tilde{k}_{3i} \]
\[ \frac{\delta l}{\delta \tilde{k}_4} = \sum \tilde{r}_i \tilde{k}_{4i} \]

\[ l_0^{\tilde{\beta} \tilde{\beta}} = \frac{1}{N} \sum \tilde{s}_i \tilde{s}_i' \]
\[ l_0^{\tilde{\beta} \tilde{s}_3} = \frac{1}{N} \sum \tilde{k}_{3i} \tilde{s}_i \]
\[ l_0^{\tilde{\beta} \tilde{k}_3} = \frac{1}{N} \sum \tilde{k}_{3i}^2 \]
\[ l_0^{\tilde{\beta} \tilde{k}_4} = \frac{1}{N} \sum \tilde{k}_{3i} \tilde{k}_{4i} \]

etc where:

\[ \tilde{p}_i = \Phi(x_i' \tilde{\beta}) \]
\[ \tilde{r}_i = \frac{y_i - \tilde{p}_i}{\sqrt{\tilde{p}_i(1 - \tilde{p}_i)}} \]
\[ \tilde{s}_i = \frac{\Phi(x_i' \tilde{\beta}) x_i}{\sqrt{\tilde{p}_i(1 - \tilde{p}_i)}} \]
\[ \tilde{k}_{3i} = \frac{\Phi(x_i' \tilde{\beta})}{\sqrt{\tilde{p}_i(1 - \tilde{p}_i)}} \frac{1}{6} [(x_i' \tilde{\beta})^2 - 1] \]
\[ \tilde{k}_{4i} = \frac{\Phi(x_i' \tilde{\beta})}{\sqrt{\tilde{p}_i(1 - \tilde{p}_i)}} \frac{1}{24} [-3(x_i' \tilde{\beta})^3 + 3x_i' \tilde{\beta}] \]

\( \tilde{p}_i \) is the predicted probability of success evaluated using the restricted parameter estimates etc. The \( \tilde{r}_i \) and \( \tilde{s}_i \) are just the scaled residuals and regressors. For example, \( \tilde{r}_i \) has a mean of zero and a unit variance.
Thus the LM test statistic may be rewritten as:

\[
LM = \sum \hat{r}_i \hat{\bar{w}}_i \left( \sum \hat{w}_i \hat{\bar{w}}_i \right)^{-1} \sum \hat{r}_i \hat{\bar{w}}_i \\
= \sum \hat{r}_i \hat{k}_i \left( \sum \hat{k}_i \hat{k}_i - \sum \hat{k}_i \hat{s}_i \left( \sum \hat{s}_i \hat{s}_i \right)^{-1} \sum \hat{s}_i \hat{k}_i \right)^{-1} \sum \hat{k}_i \hat{r}_i
\]

(16)

where \( \hat{w}_i = (\hat{s}_i, \hat{k}_i) \) and \( \hat{k}_i = (\hat{k}_{3i}, \hat{k}_{4i}) \). From (16) it is clear that the LM test statistic for \( \kappa_3 = \kappa_4 = 0 \) is just the explained sum of squares from the uncentred regression of \( \hat{r}_i \) on \( \hat{s}_i, \hat{k}_{3i} \) and \( \hat{k}_{4i} \). This proposition holds since:

\[
\frac{\delta l}{\delta \beta} = \sum \hat{r}_i \hat{s}_i = 0
\]

It is easy to show that \( N \) times the \( R^2 \) from this regression is asymptotically equal to the LM test statistic since \( NR^2 = LM / \left( \frac{1}{N} \sum_i \hat{r}_i^2 \right) \) and \( \lim \frac{1}{N} \sum_i \hat{r}_i^2 - 1 \).

**Comparison With Other Tests of Normality in the Probit Model**

It is useful to compare this test with the tests proposed by Bera, Jarque and Lee (1982) and Chesher and Irish (1987). Using the recursions:

\[
\int_{-\infty}^{\infty} u^{j} \phi (u) du = (-z)^{j-1} \phi (z) + (j-1) \int_{-\infty}^{\infty} u^{j-2} \phi (u) du
\]

\[
\int_{-\infty}^{z} u^{j} \phi (u) du = (-z)^{j-1} \phi (z) + (j-1) \int_{-\infty}^{z} u^{j-2} \phi (u) du
\]

and substituting the restricted parameter estimates for the true but unknown parameters, one may show that the "moment residuals" for the probit model are7:

---

7 The terminology comes from Chesher and Irish (1987). The first moment residual is called the generalised residual by Gourieroux et. al. (1987).
\( \hat{\epsilon}_i^{(1)} = \sum_i \hat{\epsilon}_i^{(1)} = \sum_i \{ \hat{E}(u_i | y_i, x_i) - E(u_i) \} = \sum_i \hat{E}(u_i | y_i, x_i) = \sum_i \frac{y_i - \hat{\beta}_{i1} x_i}{\hat{\beta}_i(1 - \hat{\beta}_i)} \phi(x_i/\hat{\beta}) \)

\( \hat{\epsilon}_i^{(2)} = \sum_i \{ \hat{E}(u_i^2 | y_i, x_i) - E(u_i^2) \} = \sum_i \{ \hat{E}(u_i^2 | y_i, x_i) - 1 \} = -\sum_i x_i/\hat{\beta} \hat{\epsilon}_i^{(1)} \)

(17)

\( \hat{\epsilon}_i^{(3)} = \sum_i \{ \hat{E}(u_i^3 | y_i, x_i) - E(u_i^3) \} = \sum_i \{ \hat{E}(u_i^3 | y_i, x_i) \} = \sum_i (x_i/\hat{\beta})^2 + 2) \hat{\epsilon}_i^{(1)} \)

\( \hat{\epsilon}_i^{(4)} = \sum_i \{ \hat{E}(u_i^4 | y_i, x_i) - E(u_i^4) \} = \sum_i \{ \hat{E}(u_i^4 | y_i, x_i) - 3 \} = -\sum_i (x_i/\hat{\beta})^2 + 3x_i/\hat{\beta}) \hat{\epsilon}_i^{(1)} \)

Then the three LM tests may be interpreted as testing whether the following scores are significantly different from zero:

Bera, Jarque and Lee (1982)\(^8\):

\[ \sum_i \hat{\epsilon}_i^{(1)} x_i = 0 \]

\[ -\frac{1}{3} \sum_i (\hat{\epsilon}_i^{(3)} - 3\hat{\epsilon}_i^{(1)}) = -\frac{1}{3} \sum_i ((x_i/\hat{\beta})^2 - 1) \hat{\epsilon}_i^{(1)} \]

\[ -\frac{1}{4} \sum_i \hat{\epsilon}_i^{(4)} = -\frac{1}{4} \sum_i ((x_i/\hat{\beta})^3 + 3x_i/\hat{\beta}) \hat{\epsilon}_i^{(1)} \]

\(^8\) Note that Bera et. al. (1982) appear to have omitted the minus sign from the third score in their test. Compare their equations (4.10) and (4.10)'.
Chesher and Irish (1987):

\[
\sum \hat{e}_{i}^{(1)} x_i = 0 \\
\sum \hat{e}_{i}^{(3)} = \sum ((x_i/\hat{\beta})^2 + 2) \hat{e}_{i}^{(1)} \\
\sum \hat{e}_{i}^{(4)} = -\sum ((x_i/\hat{\beta})^3 + 3x_i/\hat{\beta}) \hat{e}_{i}^{(1)}
\]

Murphy (1994):

\[
\sum \hat{e}_{i}^{(1)} x_i = 0 \\
\frac{1}{3!} \sum H_2(x_i/\hat{\beta}) \hat{e}_{i}^{(1)} = \frac{1}{6} \sum ((x_i/\hat{\beta})^2 - 1) \hat{e}_{i}^{(1)} \\
-\frac{1}{4!} \sum H_3(x_i/\hat{\beta}) \hat{e}_{i}^{(1)} = -\frac{1}{24} \sum ((x_i/\hat{\beta})^3 - 3x_i/\hat{\beta}) \hat{e}_{i}^{(1)}
\]

The first score, which is common to the three tests, is just the first order or likelihood equation for \( \beta \) and is identically zero. However since the information matrix is not block diagonal one cannot omit it when calculating the LM test statistics. The test statistics may be calculated using artificial regression of either the OPG form or the alternative form used in this paper. The OPG form approximates the information matrix by the outer product of the matrix of contributions to the score. The alternative form uses the expectation of this outer product.

Since all three LM test statistics may be expressed as the ESS’s from artificial regressions, one may ignore the signs and scaling and only consider linear combinations of the scores. Since \( x_i \) includes a constant the second score in Chesher and Irish (1987) is a linear combination of the first and second scores in Bera et al. (1982). Thus the two LM statistics are numerically equal if calculated in the same way.

The second and third scores in Chesher and Irish (1987) are just the third and fourth order moment residuals and so are easy to interpret. Apart from scaling, the second and third scores in Bera et. al. are also easy to interpret. They are proportional to the difference between the third order moment residual and three times the first order moment residual and to the fourth order moment residual respectively.

The second and third scores for the test proposed in this paper are just the scores...
for the omission of the second and third order Hermite polynomials. Tests for the omission of higher order polynomial terms are very easy to implement. Apart from scaling the second score is the same as in Bera et al. However the third score is not a linear combination of the scores in Bera et al. which suggests that the small sample performance of the two tests is worth investigating using Monte Carlo methods.

Preliminary Monte Carlo results suggest that the third scores in the two tests are very similar. Thus the non-OPG forms of the two tests have similar size and power properties. Nevertheless the approach used in this paper has two important advantages. The derivation is simpler than in Bera et al. and the approach may be simply extended to the bivariate probit model.

Conclusions

In this paper a simple artificial regression based Lagrange Multiplier (LM) test of non-normality in the probit model is derived using a Gram Charlier type A alternative. The test is simply derived and may be extended to the bivariate probit case. The proposed LM test is likely to have reasonable small sample properties since it is not based on the outer product gradient (OPG) form of the test. The test is compared with two other existing tests for non-normality in the probit model.

References


