# Score Tests of Normality in Bivariate Probit Models

**Anthony Murphy**

**Nuffield College, Oxford OX1 1NF, UK**

## Abstract:

A relatively simple and convenient score test of normality in the bivariate probit model is derived. Monte Carlo simulations show that the small sample performance of the bootstrapped test is quite good. The test may be readily extended to testing normality in related models.

**Keywords**: Score test, bivariate probit, normality, Gram-Charlier series.

**JEL Codes:** C25.

## Introduction

In this paper I show how to construct a simple score (LM) test of the normality assumption in bivariate probit and related models. To date, the normality assumption in these models has seldom been tested, even though the parameter estimates are inconsistent when the distribution of the random error terms is mis-specified.

A score test of normality has an obvious advantage over likelihood ratio or Wald tests. I follow convention and focus on skewness and excess kurtosis when deriving the test statistic. The alternative hypothesis used is based on a truncated or type AA bivariate Gram Charlier series used by Lee (1984) and Smith (1985) for example. The score test involves conditional / truncated expectations of terms such as $u_1^j u_2^k$, where u' = $(u_1,u_2)$' is a bivariate normal random vector. Unfortunately, the cited papers do not contain explicit expressions for these expectations. These may be simulated but explicit expressions are more accurate and convenient.

## The Bivariate Probit Model

The bivariate probit model may be derived from a pair of regression or reduced form equations $y_1^* = x_1'\beta_1 + u_1$ and $y_2^* = x_2'\beta_2 + u_2$ with latent dependent variables $y_1^*$ and $y_2^*$. The random errors $u_1$ and $u_2$ are distributed as standard bivariate normal variables with correlation coefficient $\rho$. Only the signs of the latent variables are observed. It is useful to define the indicator and sign variables $y_j$ and $s_j$, with $y_j = 1(y_j^* > 0)$ and $s_j = 2y_j - 1$ for $j = 1,2$.

In order to derive the log likelihood for this model, expressions for the four probabilities $P_{11}$, $P_{10}$, $P_{01}$ and $P_{00}$ are required, where $P_{10} = \text{prob}(y_1 = 1, y_2 = 0) = \text{prob}(u_1 > -x_1'\beta_1, u_2 \leq -x_2'\beta_2)$ for example. In the bivariate probit model, the probabilities are given by:

$$P_{y_1y_2} = \Phi(s_1x_1'\beta_1, s_2x_2'\beta_2, s_1s_2\rho) = \Phi_{y_1y_2} \tag{1}$$

where $\Phi$ is the standard normal bivariate c.d.f., and $P_{y_1y_2}$ and $\Phi_{y_1y_2}$ are used for short. Thus, for example, $\Phi_{10} = \Phi(x_1'\beta_1, -x_2'\beta_2, -\rho)$.

## The Type AA Gram Charlier Alternative

The results in Lee (1984) and Smith (1985), inter alia, suggest that a truncated or type AA bivariate Gram Charlier series may be a suitable alternative to the standard bivariate normal density. The Gram Charlier expansion for a regular standardized bivariate density $f(u_1,u_2)$ with correlation coefficient $\rho$ is:

$$\begin{aligned} f(u_1,u_2) &= \phi(u_1,u_2,\rho) + \sum_{j+k}\sum_{\geq 3} (-1)^{j+k} \frac{c_{jk}}{j!k!} D_1^j D_2^k \phi(u_1,u_2,\rho) \\ &= \phi(u_1,u_2,\rho)\left\{1 + \sum_{j+k}\sum_{\geq 3} \frac{\kappa_{jk}}{j!k!} H_{jk}(u_1,u_2,\rho)\right\} \end{aligned} \tag{2}$$

where the $H_{jk}(u_1,u_2,\rho) = \left((-1)^{j+k} D_1^j D_2^k \phi(u_1,u_2,\rho)\right) / \phi(u_1,u_2,\rho)$ are bivariate Hermite polynomials, the D's are differentiation operators, the $\kappa_{jk}$'s are cumulants and $\phi(u_1,u_2,\rho)$ is the bivariate standard normal density.

Truncating (2) by omitting all terms with $j + k > 4$ yields the type AA Gram Charlier series, which may not be a proper p.d.f. However, it is only being used to generate a test statistic with, it is hoped, some power against local departures from normality. Pagan and Vella

(1989) suggest using the density in Gallant and Nychka (1987) to test for normality. However, this approach is no simpler than the one used in this paper.

Expressions for the four probabilities $P_{00}$, $P_{10}$, $P_{01}$ and $P_{11}$ are required. Under the type AA Gram Charlier alternative, the probabilities are:

$$\begin{aligned} P_{y_1 y_2} &= \int\int_R f(u_1, u_2, \rho)\, du_1 du_2 \\ &= \Phi_{y_1 y_2} + \sum_{j+k=3,4}\sum \frac{\kappa_{jk}}{j!k!} \int\int_R H_{jk}(u_1, u_2, \rho)\, \phi(u_1, u_2, \rho)\, du_1 du_2 \qquad (3) \\ &= \Phi_{y_1 y_2}\left( 1 + \sum_{j+k=3,4}\sum \frac{\kappa_{jk}}{j!k!} E_{y_1 y_2} H_{jk}(u_1, u_2, \rho) \right) \end{aligned}$$

where R is the relevant region of integration and $E_{y_1 y_2}$ denotes conditional expectations (given $y_1$, $y_2$ and the x's) taken with respect to the p.d.f. $\phi(u_1, u_2, \rho)$. The third and fourth order bivariate Hermite polynomials in (3) involve products of $u_1$ and $u_2$ raised up to the combined power of four.

— Table 1 About Here ---

Closed form expressions for the conditional expectations in (3) are set out in Table 1. These were derived using the moment generating function in Tallis (1961) and the mathematical package Maple. The calculations, which are very tedious, were checked in a number of ways. For example, for a range of parameter values, truncated data were generated using the Gibbs sampler and sample moments calculated. When these were regressed on the corresponding theoretical moments, the joint restrictions that the intercepts and slopes were zero and one respectively were accepted.

## The Score Test

The score test is entirely standard. The log likelihood for a random sample of N individuals is:

$$\ell(\theta) = \sum_{i=1}^{N} \{(y_1 y_2 \ln P_{11}(\theta) + (1-y_1) y_2 \ln P_{01}(\theta) + y_1 (1-y_2) \ln P_{10}(\theta) + (1-y_1)(1-y_2) \ln P_{00}(\theta)\}$$

where individual i subscripts have been omitted and $\theta$ is a vector of parameters. Under the null hypothesis that the bivariate probit model is the DGP, the probabilities are given by (1) and $\theta$ consists of $\beta_1, \beta_2$ and $\rho$ only, since all 9 $\kappa_{jk}$ parameters are zero. Under the alternative hypothesis, the probabilities are given by (3) and $\theta$ now includes the $\kappa_{jk}$'s as well as $\beta_1, \beta_2$ and $\rho$.

The derivatives of (3) under the null, which are needed to calculate the score test statistic, are as follows:

$$\frac{\partial P_{y_1 y_2}}{\partial \beta_1} = T_1 x_1 \qquad \frac{\partial P_{y_1 y_2}}{\partial \beta_2} = T_2 x_2 \qquad \frac{\partial P_{y_1 y_2}}{\partial \rho} = T_3 \qquad \frac{\partial P_{y_1 y_2}}{\partial \kappa_{jk}} = \frac{1}{j!\,k!} \Phi_{y_1 y_2} E_{y_1 y_2} H_{jk}$$

These expressions are easily evaluated using the results in Table 1. Let $\tilde{\theta}$ denote the restricted parameter estimates. Under the null, the observed score is:

$$\frac{\partial \ell(\tilde{\theta})}{\partial \theta} = \sum_i \{ \frac{y_1 y_2}{\Phi_{11}(\tilde{\theta})} \frac{\partial P_{11}(\tilde{\theta})}{\partial \theta} + \frac{y_1 (1-y_2)}{\Phi_{10}(\tilde{\theta})} \frac{\partial P_{10}(\tilde{\theta})}{\partial \theta} + \frac{(1-y_1) y_2}{\Phi_{01}(\tilde{\theta})} \frac{\partial P_{01}(\tilde{\theta})}{\partial \theta} + \frac{(1-y_1)(1-y_2)}{\Phi_{00}(\tilde{\theta})} \frac{\partial P_{00}(\tilde{\theta})}{\partial \theta} \}$$

and the observed information matrix is:

$$\mathbb{I}(\tilde{\theta}) = \sum_i \left( \sum_{y_1} \sum_{y_2} \frac{1}{\Phi_{y_1 y_2}(\tilde{\theta})} \frac{\partial P_{y_1 y_2}(\tilde{\theta})}{\partial \theta} \frac{\partial P_{y_1 y_2}(\tilde{\theta})}{\partial \theta'} \right)$$

The score test statistic that all nine $\kappa_{ij}$ terms are zero is just $\frac{\partial \ell(\tilde{\theta})}{\partial \theta'} \mathbb{I}(\tilde{\theta})^{-1} \frac{\partial \ell(\tilde{\theta})}{\partial \theta}$, which, under the null, has a limiting chi-squared distribution with 9 degrees of freedom given standard regularity conditions. It may be calculated directly or using an artificial regression, whichever is most convenient.

## Some Monte Carlo Results

The small sample performance of the score test was examined using Monte Carlo experiments. The results are based on 1000 replications and sample sizes of 500 or 1000 observations. Details of the various DGP's used are given in the notes to Tables 2 and 3. The correlation between the random error terms in the DGP's ranges from -¼ to +¾. The empirical sizes and powers of the tests at the 5% and 10% levels are calculated using bootstrapped critical values (Horowitz, 1994).

— Table 2 About Here —

The results in Table 2 suggest that bootstrapping the critical values controls the size of the test. This result is not surprising since the test statistic is asymptotically pivotal and, as Horowitz (1994) and others have shown, bootstrapped critical values should lead to a better (higher order) asymptotic approximation in this case.

— Table 3 About Here —

In Table 3 the power of the bivariate test is examined using two non-normal DGP's. In the first DGP, the random error terms are distributed as bivariate Student's t variables with 3 degrees of freedom (d.f.). On average, the absolute values of the estimated coefficients are biased downwards by about 10% in the first and 20% in the second equation. The score test is reasonably powerful. At the 5% significance level, the power of the test ranges from 18.7% to 25.3% when the sample size is 500, and, from 26.7% to 29.8% when the sample size is 1000. The test is mainly picking up the kurtosis in the random error terms.

The second DGP in Table 3 involves a lot of skewness and kurtosis since the random error terms are distributed as a correlated mixture of standardized chi-squared variables with 3 d.f. Surprisingly, the bias of the estimated slope coefficients in two equations is relatively small. The bivariate test is fairly powerful and picks up both the skewness and kurtosis in the random error terms. At the 5% significance level, the power of the bivariate test ranges from 35.5% to 46.2% when the sample size is 500, and, from 71.0% to 85.1% with 1000 observations.

Butler and Chatterjee (1997) suggest using the generalized method of moments (GMM) over-identifying restrictions test to test the normality (and exogeneity) assumption in an ordered bivariate probit model. I looked at this test and found that it has very low power. This is not surprising since the over-identifying restrictions tested have little to do with skewness or kurtosis.

## Summary

Although the score test of normality in the bivariate probit model is tedious to derive, it is relatively simple and convenient to apply. Monte Carlo simulations show that the small sample performance of the bootstrapped test is quite good. The test may be readily extended to testing normality in grouped and ordered bivariate probit models, multinomial probit models with three outcomes as well as more general bivariate models which, following Lee (1983), have been transformed to some form of the bivariate probit model. Using simulation based estimation and inference methods, the score test may be extended to multivariate and multinomial probit models.

## References

Butler, J.S. and Chatterjee, P. (1997), “Tests of Specification of Univariate and Bivariate Ordered Probit”, *Review of Economic Statistics*, 79, 343-47.

Gallant, A. R. and Nychka, D. W. (1987), “Semi-parametric Maximum Likelihood Estimation”, *Econometrica*, 55, 363-93.

Horowitz, J. L. (1994), “Bootstrap-Based Critical Values for the Information Matrix Test”, *Journal of Econometrics*, 61, 395-411.

Lee, L. F. (1983). "Generalized Econometric Models with Selectivity”, *Econometrica*, 51, 507-12.

Lee, L. F. (1984), "Tests for the Bivariate Normal Distribution in Econometric Models with

Selectivity", *Econometrica*, 52, 843-63.

Smith, R. J. (1985), "Some Tests of Mis-Specification in Bivariate Limited Dependent Variable Models", *Annales de l'INSEE*, 59/60, 97-123.

Tallis, G. M. (1961), "The Moment Generating Function of the Truncated Multi-Normal Distribution", *Journal of the Royal Statistical Society*, Series B, 23, 223-9.

**Table 1**

**Conditional Expectations of Bivariate Hermite Polynomials**

$$\Phi_{y_1y_2} E_{y_1y_2} H_{10} = T_1$$

$$\Phi_{y_1y_2} E_{y_1y_2} H_{20} = a_1 T_1 - \rho T_3$$

$$\Phi_{y_1y_2} E_{y_1y_2} H_{11} = T_3$$

$$\Phi_{y_1y_2} E_{y_1y_2} H_{30} = (a_1^2 - 1) T_1 + \rho [(\rho^2 - 2) a_1 + \rho a_2] T_3 / (1 - \rho^2)$$

$$\Phi_{y_1y_2} E_{y_1y_2} H_{21} = (a_1 - \rho a_2) T_3 / (1 - \rho^2)$$

$$\Phi_{y_1y_2} E_{y_1y_2} H_{40} = a_1 (a_1^3 - 3) T_1 - \rho [(\rho^4 - 3\rho^2 + 3) a_1^2 + (\rho^3 - 3\rho) a_1 a_2 + \rho^2 a_2^2 + (1 - \rho^2)(2\rho^2 - 3)] T_3 / (1 - \rho^2)^2$$

$$\Phi_{y_1y_2} E_{y_1y_2} H_{31} = [a_1^2 - 2\rho a_1 a_2 + \rho^2 a_2^2 - (1 - \rho^2)] T_3 / (1 - \rho^2)^2$$

$$\Phi_{y_1y_2} E_{y_1y_2} H_{22} = -[\rho a_1^2 - (1 + \rho^2) a_1 a_2 + \rho a_2^2 - \rho (1 - \rho^2)] T_3 / (1 - \rho^2)^2$$

Notes: The following notation is used: $a_1 = -x_1'\beta_1$, $a_2 = -x_2'\beta_2$ and:

$$T_1 = \frac{\partial \Phi_{y_1y_2}}{\partial x_1'\beta_1} = s_1 \phi(x_1'\beta_1) \Phi\left( \frac{s_2 (x_2'\beta_2 - \rho x_1'\beta_1)}{\sqrt{1 - \rho^2}} \right)$$

$$T_2 = \frac{\partial \Phi_{y_1y_2}}{\partial x_2'\beta_2} = s_2 \Phi\left( \frac{s_1 (x_1'\beta_1 - \rho x_2'\beta_2)}{\sqrt{1 - \rho^2}} \right) \phi(x_2'\beta_2)$$

$$T_3 = \frac{\partial \Phi_{y_1y_2}}{\partial \rho} = s_1 s_2 \phi(x_1'\beta_1, x_2'\beta_2, \rho)$$

The conditional expectations of $H_{02}$, $H_{12}$, $H_{03}$, $H_{13}$ and $H_{04}$ are obtained by symmetry.

**Table 2: The Empirical Size of Score Tests of Normality in the Bivariate Probit Model**

| Sample Size N | Correlation Coefficient | Univariate Score Tests of Skewness & Kurtosis | | | | Bivariate Score Tests | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Equation 1 | | Equation 2 | | Skewness Only | | Kurtosis Only | | Skewness and Kurtosis | |
| | | 5% Level | 1% Level | 5% Level | 1% Level | 5% Level | 1% Level | 5% Level | 1% Level | 5% Level | 1% Level |
| N = 500 | $\rho$ = -0.25 | 5.6% | 1.3% | 4.9% | 1.3% | 4.4% | 0.8% | 5.7% | 1.0% | 5.2% | 0.9% |
| | $\rho$ = 0.00 | 5.5% | 1.2% | 4.8% | 1.2% | 5.4% | 1.1% | 5.3% | 1.2% | 4.6% | 1.2% |
| | $\rho$ = 0.25 | 4.8% | 0.9% | 4.8% | 1.3% | 4.5% | 1.4% | 5.1% | 1.4% | 5.4% | 0.9% |
| | $\rho$ = 0.50 | 5.1% | 1.0% | 4.6% | 0.8% | 4.7% | 0.9% | 5.3% | 1.7% | 5.6% | 0.7% |
| | $\rho$ = 0.75 | 5.0% | 1.5% | 5.5% | 1.2% | 5.8% | 1.8% | 5.1% | 1.5% | 4.8% | 0.8% |
| N = 1000 | $\rho$ = -0.25 | 4.8% | 0.7% | 4.9% | 0.8% | 5.0% | 0.7% | 5.2% | 1.3% | 5.1% | 1.2% |
| | $\rho$ = 0.00 | 4.9% | 0.7% | 4.7% | 1.1% | 4.9% | 0.8% | 5.0% | 4.7% | 4.6% | 0.9% |
| | $\rho$ = 0.25 | 4.8% | 1.2% | 5.1% | 1.0% | 5.2% | 1.1% | 4.6% | 1.1% | 5.2% | 1.1% |
| | $\rho$ = 0.50 | 5.2% | 0.7% | 5.4% | 1.2% | 4.8% | 1.0% | 5.6% | 1.3% | 4.8% | 1.2% |
| | $\rho$ = 0.75 | 4.5% | 0.9% | 4.5% | 0.7% | 5.5% | 1.3% | 4.9% | 0.9% | 5.3% | 1.0% |

Notes: The DGP consists of the two latent equations $y_1^* = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + u_1$ and $y_2^* = \gamma_1 + \gamma_2 x_2 + \gamma_3 x_3 + u_2$ where the random errors are standard bivariate normal with correlation coefficient $\rho$. The explanatory variables $x_1$ and $x_2$ are generated as independent standard normal pseudo random variables. The parameter values in the DGP are $\beta_1 = -0.1$, $\beta_2 = 0.3$, $\beta_3 = 0.1$, $\gamma_1 = -0.5$, $\gamma_2 = 1.0$ and $\gamma_3 = -0.1$. The Monte Carlo results are based on 1000 replications and 99 bootstrap replications for the 95% and 99% critical values.

**Table 3: The Empirical Power of LM Tests of Bivariate Normality When the Random Error Terms in the DGP are Distributed as Bivariate Student's t and Correlated Mixtures of Standardised Chi-Squared Random Variables.**

| Sample Size N | Correlation Coefficient ρ | Bivariate Student's t (3 d.f.) | | | | | | Correlated Mixtures of Standardised Chi-Squared (3 d.f.) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Skewness | | Kurtosis | | Skewness & Kurtosis | | Skewness | | Kurtosis | | Skewness & Kurtosis | |
| | | 5% Level | 1% Level | 5% Level | 1% Level | 5% Level | 1% Level | 5% Level | 1% Level | 5% Level | 1% Level | 5% Level | 1% Level |
| N = 500 | ρ = -0.25 | 13.3% | 5.4% | 24.5% | 15.8% | 25.3% | 14.0% | 62.7% | 39.3% | 39.8% | 27.2% | 45.8% | 30.1% |
| | ρ = 0.00 | 12.5% | 5.9% | 23.4% | 14.1% | 21.3% | 12.4% | 60.4% | 38.5% | 36.6% | 24.4% | 46.2% | 28.7% |
| | ρ = 0.25 | 12.2% | 5.1% | 19.9% | 13.1% | 18.9% | 11.5% | 56.1% | 34.7% | 34.7% | 24.6% | 41.9% | 26.4% |
| | ρ = 0.50 | 11.9% | 5.8% | 20.1% | 11.8% | 18.7% | 10.2% | 47.4% | 26.7% | 31.7% | 20.4% | 35.5% | 22.9% |
| | ρ = 0.75 | 14.4% | 7.1% | 23.9% | 16.4% | 22.4% | 14.5% | 45.4% | 27.5% | 31.7% | 21.3% | 37.2% | 24.2% |
| N = 1000 | ρ = -0.25 | 16.9% | 6.8% | 34.2% | 20.9% | 29.8% | 18.8% | 93.8% | 81.6% | 63.4% | 49.1% | 84.5% | 66.9% |
| | ρ = 0.00 | 15.6% | 6.2% | 31.5% | 20.8% | 29.2% | 17.8% | 94.6% | 81.3% | 58.7% | 43.1% | 85.1% | 69.0% |
| | ρ = 0.25 | 15.2% | 6.9% | 31.2% | 19.4% | 28.2% | 16.8% | 93.0% | 80.2% | 57.2% | 41.2% | 82.1% | 63.8% |
| | ρ = 0.50 | 15.1% | 6.5% | 30.4% | 19.9% | 27.3% | 16.0% | 86.2% | 67.1% | 52.4% | 38.2% | 71.1% | 51.8% |
| | ρ = 0.75 | 16.1% | 7.3% | 30.6% | 19.5% | 26.7% | 17.5% | 86.2% | 67.0% | 52.3% | 38.2% | 71.0% | 51.7% |

Notes: The DGPs use the same two latent equations, explanatory variables and parameter values set out in the notes to Table 2. When the random errors are distributed as bivariate Student's t variables with 3 degrees of freedom, $u_1 = v_1 / v_3$ and $u_2 = v_2 / v_3$. where $v_1$ and $v_2$ are standard bivariate normal variables with correlation coefficient ρ and $v_3$ is the square root of a chi-squared variable with 3 degrees of freedom divided by it's degrees of freedom. When the random errors are generated as correlated mixtures of standardised (mean zero, unit variance) chi-squared variables with three degrees of freedom, $u_1 = v_1$ and $u_2 = \rho v_1 + \sqrt{(1 - \rho^2)}\, v_2$ where $v_1$ and $v_2$ are standardised chi-squared random variables.