A Test of the Normality Assumption
in the Ordered Probit Model[*]

Paul A. Johnson

Working Paper No. 34

March 1996

This paper presents a simple and easily implemented test of the assumption of a normally distributed error term for the ordered probit model. As this assumption is the central maintained hypothesis in all estimation and testing based on this model the test ought to serve as a key specification test in the applied literature. A small Monte Carlo experiment suggests that the test has good size and power properties.

1. Introduction

Despite the growing number of applications, the literature does not include a test of the normality assumption for the ordered probit model. As Bera, Jarque and Lee [1984] (BJL) point out, the validity of the normality assumption is more important in limited dependent variables models than in the usual regression model as, if the assumption does not hold, maximum likelihood estimation will not, in general, yield consistent parameter estimates. The assumption is also the central maintained hypothesis in any statistical inference based on the parameter estimates. In addition, normality of the error term is crucial to the interpretation of the effects of changes in the explanatory variables as these effects are usually expressed in terms of changes in the probabilities of each of the outcomes.[1]

In this paper I extend the work of BJL and derive a Lagrange multiplier (LM) test of the normality assumption for the ordered probit model. The test is easily implemented and should also serve as a general specification test of the ordered probit model. I examine the properties of the test in a small Monte Carlo experiment and find that, while the actual size of the test may exceed its nominal size somewhat, the test has good power properties, at least against the class of alternatives considered.

2. A Test for Normality

Let y_i^* be the dependent variable of interest and assume $y_i^* = \beta' x_i + \epsilon_i$ where x_i is a k–vector of exogenous variables, β is a k–vector of parameters, and ϵ_i is a zero-mean error term, distributed identically and independently across i with distribution function $F(\cdot\,;\theta)$ having parameters θ. Rather than observe y_i^*, all that is known is which of m intervals, forming a partition of the real line, contains y_i^*. Define $-\infty = \alpha_0 < \alpha_1 < \alpha_2 < \ldots < \alpha_{m-1} < \alpha_m = \infty$ and let

---

[1]For a graphical exposition of the ordered probit model see Becker and Kennedy [1992] who also discuss the pitfalls and subtleties in calculating and interpreting these probabilities.

1

$$s_{i,j} = \begin{cases} 1 & \text{if } \alpha_{j-1} < y_i^* < \alpha_j \\ 0 & \text{otherwise} \end{cases}$$

for $j = 1, \ldots, m$. It follows that $Pr\{s_{i,j} = 1 | x_i\} = F(\alpha_j - \beta' x_i; \theta)$ $- F(\alpha_{j-1} - \beta' x_i; \theta)$. The log likelihood for a sample of n observations is

$$\mathcal{L}(\alpha, \beta, \theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} s_{i,j} \, log\big[F(\alpha_j - \beta' x_i; \theta) - F(\alpha_{j-1} - \beta' x_i; \theta)\big].$$

The parameters $\alpha = (\alpha_1, \ldots, \alpha_{m-1})', \beta$ and θ may be estimated consistently by the maximizers of this function under suitable regularity conditions.[2]

When F is the standard normal distribution, Φ, the model outlined above is the ordered probit model. In this note I develop a test of the hypothesis that F is the standard normal distribution against the alternative that it is some other member of the Pearson family of distributions. The Pearson family has distribution functions which can be written as $F(u; c_0, c_1, c_2) = G(u)/G(\infty)$ where $G(u) = \int_{-\infty}^{u} e^{q(t)} dt$ and $q(t) = \int \frac{c_1 - t}{c_0 - c_1 t + c_2 t^2} dt$ (see BJL or Johnson and Kotz [1970] for details). When $c_0 = 1$ and $c_1 = c_2 = 0$, F is the standard normal distribution. Because $c_0 = 1$ is the normalization imposed to identify the parameters of the ordered probit model, the null hypothesis to be tested here is $c_1 = c_2 = 0$.

Defining $h_{i,j} = \alpha_j - \beta' x_i$, the log likelihood under the alternative hypothesis is $\mathcal{L}(\alpha, \beta, c_0, c_1, c_2) = \sum_{i=1}^{n} A_i$ where $A_i = \sum_{j=1}^{m} s_{i,j} \big\{ log\big[G(h_{i,j}) - G(h_{i,j-1})\big] - log\, G(\infty)\big\}$. Evaluated at the null values of $c_1 = c_2 = 0$, after imposing the normalization $c_0 = 1$, the derivatives of A_i with respect to the parameters are

---

[2]See, for example, Amemiya [1985].

$$\left. \frac{\partial A_i}{\partial \alpha_p} \right|_{H_0} = \left[\frac{s_{i,j}}{\Phi(h_{i,j}) - \Phi(h_{i,j-1})} - \frac{s_{i,j+1}}{\Phi(h_{i,j+1}) - \Phi(h_{i,j})} \right] \phi(h_{i,j}) \text{ for } j = 1, \dots, m-1;$$

$$\left. \frac{\partial A_i}{\partial \beta} \right|_{H_0} = \sum_{j=1}^{m} s_{i,j} \frac{\phi(h_{i,j-1}) - \phi(h_{i,j})}{\Phi(h_{i,j}) - \Phi(h_{i,j-1})} x_i;$$

$$\left. \frac{\partial A_i}{\partial c_1} \right|_{H_0} = \frac{1}{3} \sum_{j=1}^{m} s_{i,j} E(3z - z^3 | h_{i,j-1} < z < h_{i,j}); \text{ and,}$$

$$\left. \frac{\partial A_i}{\partial c_2} \right|_{H_0} = -\frac{1}{4} \left[3 - \sum_{j=1}^{m} s_{i,j} E(z^4 | h_{i,j-1} < z < h_{i,j}) \right],$$

where $\phi(\,\cdot\,)$ is the standard normal density function and z is a standard normal random variable.

It can be shown that[3]

$$E(z | h_{i,j-1} < z < h_{i,j}) = \frac{\phi(h_{i,j-1}) - \phi(h_{i,j})}{\Phi(h_{i,j}) - \Phi(h_{i,j-1})};$$

$$E(z^3 | h_{i,j-1} < z < h_{i,j}) = \frac{(2 + h_{i,j-1}^2)\phi(h_{i,j-1}) - (2 + h_{i,j}^2)\phi(h_{i,j})}{\Phi(h_{i,j}) - \Phi(h_{i,j-1})}; \text{ and,}$$

$$E(z^4 | h_{i,j-1} < z < h_{i,j}) = 3 + \frac{(3 + h_{i,j-1}^2)h_{i,j-1}\phi(h_{i,j-1}) - (3 + h_{i,j}^2)h_{i,j}\phi(h_{i,j})}{\Phi(h_{i,j}) - \Phi(h_{i,j-1})}.$$

These results may be used to write

$$\left. \frac{\partial A_i}{\partial c_1} \right|_{H_0} = \frac{1}{3} \sum_{j=1}^{m} s_{i,j} \frac{(1 - h_{i,j-1}^2)\phi(h_{i,j-1}) - (1 - h_{i,j}^2)\phi(h_{i,j})}{\Phi(h_{i,j}) - \Phi(h_{i,j-1})}; \text{ and,}$$

$$\left. \frac{\partial A_i}{\partial c_2} \right|_{H_0} = \frac{1}{4} \sum_{j=1}^{m} s_{i,j} \frac{(3 + h_{i,j-1}^2)h_{i,j-1}\phi(h_{i,j-1}) - (3 + h_{i,j}^2)h_{i,j}\phi(h_{i,j})}{\Phi(h_{i,j}) - \Phi(h_{i,j-1})}.$$

In order to compute the test statistic define

---

[3]The results in Johnson and Kotz [1970] p81-83 and the recursion $\int_a^b z^{p+1}\phi(z)dz = p\int_a^b z^{p-1}\phi(z)dz + a^p\phi(a) - b^p\phi(b)$ for $a < b$ and $p \geq 1$ are used here.

$$
\Xi_i = \begin{pmatrix}
\left[\frac{s_{i,1}}{\Phi(h_{i,1})} - \frac{s_{i,2}}{\Phi(h_{i,2})-\Phi(h_{i,1})}\right]\phi(h_{i,1}) \\[2ex]
\vdots \\[2ex]
\left[\frac{s_{i,m-1}}{\Phi(h_{i,m-1})-\Phi(h_{i,m-2})} - \frac{s_{i,m}}{1-\Phi(h_{i,m-1})}\right]\phi(h_{i,m-1}) \\[2ex]
\sum\limits_{j=1}^{m} s_{i,j}\frac{\phi(h_{i,j-1})-\phi(h_{i,j})}{\Phi(h_{i,j})-\Phi(h_{i,j-1})}x_i \\[2ex]
\xi_i
\end{pmatrix}
$$

where

$$
\xi_i = \begin{pmatrix}
\sum\limits_{j=1}^{m} s_{i,j}\frac{(1-h_{i,j-1}^2)\phi(h_{i,j-1})-(1-h_{i,j}^2)\phi(h_{i,j})}{\Phi(h_{i,j})-\Phi(h_{i,j-1})} \\[3ex]
\sum\limits_{j=1}^{m} s_{i,j}\frac{(3+h_{i,j-1}^2)h_{i,j-1}\phi(h_{i,j-1})-(3+h_{i,j}^2)h_{i,j}\phi(h_{i,j})}{\Phi(h_{i,j})-\Phi(h_{i,j-1})}
\end{pmatrix}.
$$

The LM test statistic for the hypothesis that ϵ_i is normally distributed may then be computed as

$$
\nu = \left(\sum_{i=1}^{n}\xi_i\right)' J'\left(\sum_{i=1}^{n}\Xi_i\Xi_i'\right)^{-1} J\left(\sum_{i=1}^{n}\xi_i\right)
$$

where J is a matrix consisting of the last two columns of an $k+m+1$ dimensional identity matrix and all of the elements of Ξ_i are evaluated using the maximum likelihood (ML) estimates for the ordered probit model. The proofs in the appendix to BJL may be modified to show that, under the null hypothesis, ν is asymptotically distributed as a χ^2 random variable with two degrees of freedom.

3. A Monte Carlo Experiment

To examine the properties of the test I conduct a small Monte Carlo experiment. The model is specified as $y_i^* = \beta x_i + \epsilon_i$ with x_i distributed uniformly on $[-1,1]$, $\beta = 1$,

$\alpha_1 = -\frac{1}{3}$ and $\alpha_2 = \frac{1}{3}$. I perform 10000 replications for each of several distributions for ϵ_i having zero mean and unit variance and each of the sample sizes 250, 500, 750, and 1000. To examine the size of the test ϵ_i is drawn from an $N(0,1)$ distribution. To examine the power of the test ϵ_i is drawn from gamma distributions.

The gamma distribution is the member of the Pearson family generated by setting $c_2 = 0$ with $c_1 \neq 0$. Here $c_1 = \frac{1}{\sqrt{\pi}}$ so that the density of ϵ_i under the alternatives is $f_\epsilon(u; \pi) = \frac{\sqrt{\pi}^\pi (u + \sqrt{\pi})^{\pi-1} e^{-\sqrt{\pi}(u+\sqrt{\pi})}}{\Gamma(\pi)}$ for $u \geq -\sqrt{\pi}$. Note that, as $\pi \to \infty$, $c_1 = \frac{1}{\sqrt{\pi}} \to 0$ so $f(u; \pi) \to \phi(u)$. I use $\pi = 2$, 5 and 10.[4] Figure 1 plots these three density functions and $\phi(u)$. As the figure indicates, $f_\epsilon(u; \pi)$ becomes less skewed to the right as π increases and for $\pi = 10$ it is quite close to $\phi(u)$.

Figure 1: Density Functions for ε_i
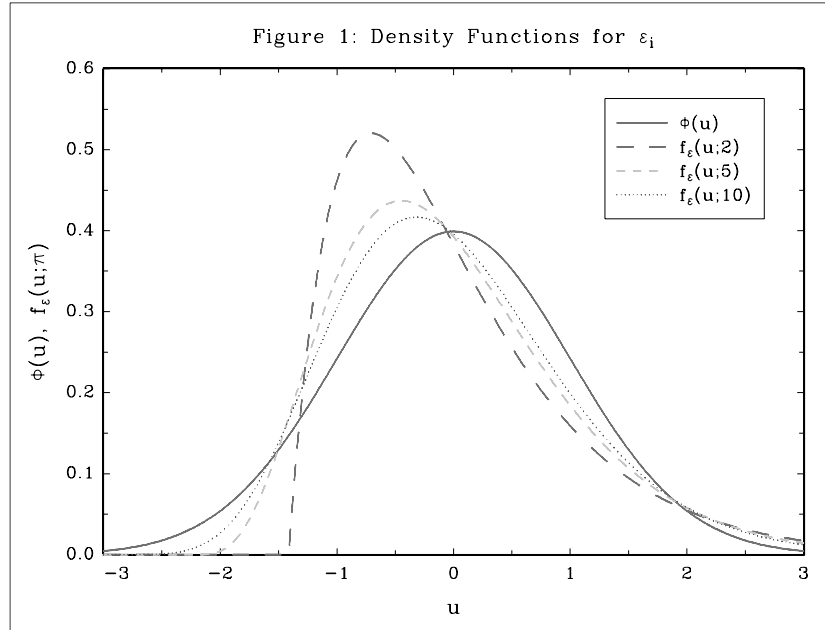
Tables 1 and 2 give the results of the Monte Carlo experiment. Table 1 shows the estimated mean and mean squared error (MSE) of the estimates of each of the parameters

---

[4]For $\pi = 2$, 5 and 10, ϵ_i was constructed as $\epsilon_i = -\frac{1}{\sqrt{\pi}} \sum_{h=1}^{\pi} \log z_h - \sqrt{\pi}$ with the z_h drawn from a $U(0,1)$ distribution.

when the sample size is 1000. The results for the other sample sizes are similar and not reported to save space. The first row shows that when the assumption of a normally distributed error is true the distributions of the ML estimates are centered on the respective true values of the parameters. The other three rows show that, as the deviation of the distribution of the error term from normality increases, as measured by the decline in π (or, equivalently, the skewness of the distribution) so does both the bias and MSE of the ML estimates. This is particularly so for $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$ with the result that the gap between them increases with the skewness of the distribution of the error term. Estimated outcome probabilities and their derivatives with respect to x will be accordingly biased.

Table 1: Mean Parameter Estimates and MSEs

Distribution of ϵ_i	$\widehat{\beta}$		$\widehat{\alpha}_1$		$\widehat{\alpha}_2$	
	Mean	MSE	Mean	MSE	Mean	MSE
$\phi(u)$	1.004	.005	-.333	.002	.335	.002
$f_\epsilon(u; 10)$	1.017	.005	-.267	.006	.412	.008
$f_\epsilon(u; 5)$	1.032	.006	-.244	.010	.449	.015
$f_\epsilon(u; 2)$	1.091	.013	-.215	.016	.530	.040

This table shows, for a sample size of 1000, the estimated means and MSEs of the ML estimates of the parameters in the Monte Carlo experiment described in the text. The true values are $\beta = 1$, $\alpha_1 = -\frac{1}{3}$ and $\alpha_2 = \frac{1}{3}$. The means and MSEs are estimated using 10000 replications for ϵ_i having each of the densities $\phi(u)$ and $f_\epsilon(u; \pi)$ for $\pi = 10, 5,$ and 2. The $\phi(u)$ case is equivalent to $\pi = \infty$.

Table 2 shows the percentage of rejections for a 5% test of the hypothesis that the error term is normally distributed for each of the sample sizes and each of the densities of

the error term. When ϵ_i has a normal distribution this percentage is an estimate of the size of the test. In the other cases the percentage of rejections is an estimate of the power against that particular alternative.

Table 2: Percentage of Rejections of Normality Hypothesis

Sample	Density of ϵ_i			
Size	$\phi(u)$	$f_\epsilon(u; 10)$	$f_\epsilon(u; 5)$	$f_\epsilon(u; 2)$
250	6.6	16.2	28.1	79.4
500	5.3	24.4	48.8	97.7
750	5.1	34.4	66.5	99.8
1000	5.0	43.7	78.5	100.0

This table shows, for each of the indicated sample sizes, the percentage of rejections of the hypothesis that ϵ_i has a normal distribution using the test described in the text with an asymptotic size of 5% for ϵ_i having the densities $\phi(u)$ and $f_\epsilon(u; \pi)$ for $\pi = 10$, 5, and 2. In the first case, which is equivalent to $\pi = \infty$, this percentage estimates the size of the test, while in the others, it indicates the power against that particular deviation from the null. In each case 10000 replications were performed.

Overall, conditional on the setup of the Monte Carlo experiment, the size and power properties of the test appear to be quite good. The size of the test may slightly exceed the nominal value in small samples but rapidly approaches 5% as the sample size rises. The power of the test increases with both the magnitude of the deviation from the null, as measured by the skewness of the distribution of the error term, and with the sample size. In the case of ϵ_i having the density $f_\epsilon(u; 2)$, which is the furthest from the null of normality, the power is never less than 79% and reaches 100% for the 1000 observation sample. As the ordered probit model reduces to the binary probit model

when there are only two outcomes, the conclusions suggest that the LM normality test for the bivariate probit model proposed by BJL also has desirable size and power properties. In addition, the conclusion that the test has good power properties when the error term has a gamma distribution most likely also holds for other skewed deviations from the normal distribution.

4. Conclusions

This paper presents a simple and easily implemented test of the assumption of a normally distributed error term for the ordered probit model. As this assumption is the central maintained hypothesis in all estimation and testing based on this model the test ought to serve as a key specification test in the applied literature. The Monte Carlo experiment suggests that the test has good size and power properties.

References

Amemiya, Takeshi, [1985], *Advanced Econometrics*, Harvard University Press, Cambridge.

Becker, William E., and Peter E. Kennedy, [1992], "A Graphical Exposition of the Ordered Probit," *Econometric Theory*, 8:127–31.

Bera, Anil K., Carlos M. Jarque and Lung-Fei Lee, [1984], "Testing the Normality Assumption in Limited Dependent Variable Models," *International Economic Review*, 25:563–578.

Johnson, Norman L., and Samuel Kotz, [1970], *Continuous Univariate Distributions*–1, Houghtin-Mifflin, Boston.