Institut für Wirtschaftsforschung Halle
Halle Institute for Economic Research

A Simple Representation of the Bera-Jarque-Lee Test for Probit Models

Joachim Wilde

Dezember 2007 No. 13

IWH-Diskussionspapiere
IWH-Discussion Papers

A Simple Representation of the Bera-Jarque-Lee Test for Probit Models

Joachim Wilde

Dezember 2007 No. 13

Author: Dr Joachim Wilde

Chief Econometrician

Email: Joachim.Wilde@iwh-halle.de

Tel.: ++49 345 77 53 836

Herausgeber:
INSTITUT FÜR WIRTSCHAFTSFORSCHUNG HALLE – IWH
Prof. Dr. Ulrich Blum (Präsident), Dr. Hubert Gabrisch (Forschungsdirektor)
Das IWH ist Mitglied der Leibniz-Gemeinschaft

Hausanschrift: Kleine Märkerstraße 8, 06108 Halle (Saale)
Postanschrift: Postfach 11 03 61, 06017 Halle (Saale)
Telefon: (0345) 77 53-60
Telefax: (0345) 77 53-8 20
Internetadresse: http://www.iwh-halle.de

A Simple Representation of the Bera-Jarque-Lee Test for Probit Models

Abstract

The inference in probit models relies on the assumption of normality. However, tests of this assumption are not implemented in standard econometric software. Therefore, the paper presents a simple representation of the Bera-Jarque-Lee test, that does not require any matrix algebra. Furthermore, the representation is used to compare the Bera-Jarque-Lee test with the RESET-type test proposed by Papke and Wooldridge (1996).

Keywords: probit model, Lagrange multiplier test, normality assumption, artificial regression

JEL Classification: C25

Zusammenfassung

Die statistische Inferenz in Probitmodellen beruht maßgeblich auf der Gültigkeit der Normalverteilungsannahme für die Störgrößen. Dennoch sind Tests auf Normalverteilung in vielen ökonometrischen Paketen nicht implementiert und werden daher oft auch nicht durchgeführt. Das Papier präsentiert deshalb eine einfache Darstellung des Bera-Jarque-Lee Tests, der die Normalverteilung gegen die vergleichsweise allgemeine Alternative der Pearson-Verteilungen testet. Die Darstellung beruht auf einer Hilfsregression, wie sie von Davidson und MacKinnon (1984) für einen Test gegen Heteroskedastie vorgeschlagen wurde. Sie kann mit jeder Statistik- oder Ökonometriesoftware, die die Schätzung eines Probitmodells ermöglicht, problemlos berechnet werden.

Darüber hinaus zeigt die Darstellung, dass der Test auf Normalverteilung asymptotisch identisch ist mit einem von Papke und Wooldridge (1996) vorgeschlagenen RESET Test, der die Linearität des Erwartungswertes der latenten Variablen überprüft. Folglich kann ohne zusätzliche Informationen nicht zwischen der funktionalen Form des Erwartungswertes und dem Verteilungstyp der Störgrößen unterschieden werden. Sofern allerdings die Linearität des Erwartungswertes unterstellt werden kann, ist die künstliche Regression des Bera-Jarque-Lee Tests informativer, da sie Hinweise darauf gibt, ob die Ablehnung der Nullhypothese durch Schiefe oder eine von der Normalverteilung abweichende Kurtosis verursacht wurde.

Schlagwörter: Probitmodell, Lagrange Multiplier Test, Normalverteilungsannahme

JEL-Codes: C25

A Simple Representation of the Bera-Jarque-Lee Test for Probit Models*

1 Introduction

Probit models are often applied without testing the normality assumption. This is rather problematic because the standard maximum likelihood estimator of the probit model is mostly biased for nonnormal disturbances (cf. Greene 2003, p. 673). One reason might be that suitable tests are not implemented in standard econometric software. Therefore, the paper presents a simple representation of the test of Bera, Jarque, and Lee (1984). It only requires the estimation of a probit model and some standard transformations. Bera, Jarque, and Lee test the assumption of normality against a rather general alternative, whereas other authors assume more specific distributions (e.g. Silva 2001).

The representation is based on an artificial regression which was suggested by Davidson and MacKinnon (1984) for a similar test on homoscedastic errors. The dependent variable is the so-called standardized residual of the probit estimation (cf. Wooldridge 2002, p. 462). An alternative artificial regression is based on a regressand of ones. However, the small sample properties are rather poor (cf. Davidson and MacKinnon 1993, p. 477). Hence, the paper does not consider this alternative.

The artificial regression of the Bera-Jarque-Lee test can be transformed into the artificial regression of the RESET-type test proposed by Papke and Wooldridge (1996). Although two regressors are numerically changed by the transformation, the explained sums of squares of both regressions and thus the test statistics are identical. Therefore, without additional information the distributional form of the disturbances is not distinguishable from the functional form of the mean of the latent variable.

Section 2 summarizes the ideas of Bera, Jarque and Lee in a slightly different notation which makes it easy to present the results in Section 3. In Section 3 the simple representation of the test will be derived and interpreted. Section 4 concludes with a comparison of the two tests.

* I thank an anonyomous referee for rather useful comments. The usual disclaimer applies.

The Bera-Jarque-Lee (BJL) test

BJL consider the usual probit model:

$$y_i^* = \beta' x_i + u_i, \quad u_i \text{ iid } N\left(0, \sigma^2\right), \quad i = 1, ..., N$$
$$y_i = \begin{cases} 1, & y_i^* > 0 \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

x_i being a (K×1)-vector of exogenous variables, β a vector of unknown parameters, y_i^* a latent tendency, y_i an observed indicator, and u_i a stochastic error term, which is stochastically independent of the exogenous variables. They derive a Lagrange-multiplier test of the null hypothesis of normally distributed error terms. Under the alternative hypothesis BJL assume the Pearson family of distributions. Implying a mean of zero, the density function f(u) of this family is characterized by the following differential equation (BJL 1984, p. 565):

$$\frac{\partial \ln f\left(u\right)}{\partial u} = \frac{c_1 - u}{c_0 - c_1 u + c_2 u^2}. \tag{2}$$

The normal distribution is a special case of (2) with $c_0=\sigma^2$ and $c_1=c_2=0$. Hence, the following hypotheses are tested:

H_0: $c_1=c_2=0$ against H_1: $c_1 \neq 0$ and/ or $c_2 \neq 0$.

"$c_1 \neq 0$" implies a skew distribution, "$c_2 \neq 0$" a kurtosis different from 3 (given the distribution is symmetric) (cf. Johnson, Kotz and Balakrishnan 1994, p. 22). The variance σ^2 is normalized to 1 as usual (BJL 1984, p. 566). Therefore, the vector θ of the remaining unknown parameters is equal to (β, c_1, c_2).

The Pearson family of distributions contains many common distributions as special cases, for instance the t-distribution or the gamma distribution (Johnson, Kotz and Balakrishnan 1994, pp. 18 and 21). Nevertheless, the alternative hypothesis is not as general as that of the well established Jarque-Bera test in the linear regression model (BJL 1984, p. 564). However, it is the most general alternative hypothesis in the literature.

Using the information matrix equality the Lagrange multiplier test statistic can be denoted as

$$LM = \left(\frac{\partial \ln L\left(y, \hat{\theta}_r\right)}{\partial \theta} \right)' \left[I_{H_0}\left(\hat{\theta}_r\right) \right]^{-1} \left(\frac{\partial \ln L\left(y, \hat{\theta}_r\right)}{\partial \theta} \right),$$

$$I_{H_0}\left(\theta_r\right) = E_y \left(\frac{\partial \ln L\left(y, \theta_r\right)}{\partial \theta} \right) \left(\frac{\partial \ln L\left(y, \theta_r\right)}{\partial \theta} \right)', \tag{3}$$

lnL being the log-likelihood function under H_1, $y = (y_1 \; ... \; y_N)'$, $\theta_r = (\beta \;\, 0 \;\, 0)'$, $\hat{\theta}_r = \left(\hat{\beta}_r \;\, 0 \;\, 0\right)'$, and $\hat{\beta}_r$ the probit maximum likelihood estimator of β.

Define $X_i = \left(x_i' \quad (-1/3)\left[\left(\beta' x_i\right)^2 - 1\right] \quad (1/4)\left[\left(\beta' x_i\right)\left(3 + \left(\beta' x_i\right)^2\right)\right]\right)'$, $\varphi_i = \varphi\left(\beta' x_i\right)$,

$\Phi_i = \Phi\left(\beta' x_i\right)$, $\hat{X}_i$, $\hat{\varphi}_i$, $\hat{\Phi}_i$ the analogous functions after substituting β by $\hat{\beta}_r$, φ and Φ being the density function and the distribution function of a standard normal distribution. Given the independence of the sample units the test statistic (3) is obtained as (cf. ibid., pp. 570, 571)

$$LM = \left(\sum_{i=1}^{N} \frac{\left(y_i - \hat{\Phi}_i\right)\hat{\varphi}_i}{\hat{\Phi}_i\left(1 - \hat{\Phi}_i\right)} \hat{X}_i \right)' \left(\sum_{i=1}^{N} \frac{\hat{\varphi}_i^{\,2}}{\hat{\Phi}_i\left(1 - \hat{\Phi}_i\right)} \hat{X}_i \hat{X}_i' \right)^{-1} \left(\sum_{i=1}^{N} \frac{\left(y_i - \hat{\Phi}_i\right)\hat{\varphi}_i}{\hat{\Phi}_i\left(1 - \hat{\Phi}_i\right)} \hat{X}_i \right). \tag{4}$$

It is asymptotically $\chi^2(2)$ distributed.

3 A simple representation of the test

The statistic (4) is equal to the explained sum of squares of an artificial regression. Define r as vector of the standardized residuals of a probit estimation, i.e.

$$r = \left(r_1, \quad ..., \quad r_N\right)', \; r_i = \frac{y_i - \hat{\Phi}_i}{\sqrt{\hat{\Phi}_i\left(1-\hat{\Phi}_i\right)}}, \; i=1, ..., N, \; \text{and} \tag{5}$$

$$R = \begin{pmatrix} R_1 \\ \vdots \\ R_N \end{pmatrix}, \; R_i = \frac{\hat{\varphi}_i}{\sqrt{\hat{\Phi}_i\left(1-\hat{\Phi}_i\right)}} \hat{X}_i', \quad i = 1, ..., N. \tag{6}$$

Using (5) and (6) the statistic (4) can be written as follows:

$$LM = r' \, R\left(R' \, R\right)^{-1} R' \, r\,. \tag{7}$$

Defining the artificial regression

$$r = R\delta + \varepsilon, \tag{8}$$

and letting $\hat{\delta}$ be the ordinary least squares estimator of δ (7) is equivalent to

$$LM = \left(R\hat{\delta}\right)' R\hat{\delta}\,, \tag{9}$$

i.e. the LM statistic is the explained sum of squares of (8).

Based on (7)-(9) LM statistic and artificial regression can be interpreted as follows: The test rejects the hypothesis of normally distributed errors if the variables on the right hand side of (8) "explain" the standardized residuals of the probit model. The products of those variables with the dependent variable are already equal to the gradient of the log-likelihood function under H_1 (cf. (3) and (4)). The first K components of this vector are equal again to the gradient of the log-likelihood function of a standard probit model (cf. Greene 2003, p. 671). Hence, the first K regressors of the artificial regression are orthogonal to the dependent variable.

Furthermore, in case the null hypothesis is valid $\hat{\theta}_r$ is close to the unrestricted maximum likelihood estimate. Therefore, the derivates with respect to c_1 and c_2 are close to zero, i.e. the last two regressors of the artificial regression are almost orthogonal to the dependent variable. Thus, no regressor explains the dependent variable, and the LM statistic becomes small. However, in case of skewness or a kurtosis unequal to 3 $\hat{\theta}_r$ might be rather different from the unrestricted maximum likelihood estimate. Then, the last two regressors are not orthogonal to the vector of the standardized residuals, and the LM statistic becomes large.

Concerning a sample unit regression and defining $\hat{f}_i = \frac{\hat{\varphi}_i}{\sqrt{\hat{\Phi}_i\left(1-\hat{\Phi}_i\right)}}$ (8) is

$$\frac{y_i - \hat{\Phi}_i}{\sqrt{\hat{\Phi}_i\left(1-\hat{\Phi}_i\right)}} = \delta_1 \hat{f}_i x_{i1} + \ldots + \delta_K \hat{f}_i x_{iK}$$

$$+\delta_{K+1}\hat{f}_i\left(-\frac{1}{3}\right)\left[\left(\hat{\beta}_r{}'x_i\right)^2 - 1\right] + \delta_{K+2}\hat{f}_i\,\frac{1}{4}\left[\left(\hat{\beta}_r{}'x_i\right)\left(3+\left(\hat{\beta}_r{}'x_i\right)^2\right)\right] + \varepsilon_i\,. \tag{10}$$

Using (10) all variables of the artificial regression can be calculated easily without matrix algebra. Merely standard forecasts of a probit estimation are needed. Afterwards, the LM statistic is computed via (9) and compared with the critical value of a $\chi^2(2)$ distribution. The LIMDEP code of the complete procedure can be found in the appendix.

4 Comparison with a RESET-type test

Papke and Wooldridge (1996, p. 624) test the hypothesis that $E\left(y_i^* \mid x_i\right)$ is a linear function of x_i against the alternative that $E\left(y_i^* \mid x_i\right)$ is a nonlinear function of x_i. Using the notation above, their artificial regression is

$$\frac{y_i - \hat{\Phi}_i}{\sqrt{\hat{\Phi}_i\left(1-\hat{\Phi}_i\right)}} = \tilde{\delta}_1 \hat{f}_i x_{i1} + \ldots + \tilde{\delta}_K \hat{f}_i x_{iK} + \tilde{\delta}_{K+1} \hat{f}_i \left(\hat{\beta}_r' x_i\right)^2 + \tilde{\delta}_{K+2} \hat{f}_i \left(\hat{\beta}_r' x_i\right)^3 + \tilde{\varepsilon}_i \,. \tag{11}$$

Their test statistic is $N \cdot R_u^2$, R_u^2 being the constant-unadjusted coefficient of determination of (11). Since (1/N)r'r converges to 1 (Engle 1984, p. 818), $N \cdot R_u^2$ is asymptotically equivalent to the explained sum of squares of (11).

It is easily seen that (10) can be transformed into (11) given that x_i includes a constant. Therefore, the JBL test against a Pearson distribution of the residuals and the RESET-type test of Papke and Wooldridge against a nonlinear function of the conditional mean (approximated by a quadratic and cubic term) are asymptotically identical. Additional information is needed to clarify the reason for a rejection of the null hypothesis. However, given that the mean is a linear function of x_i, regression (10) contains more information than regression (11). For it indicates whether the null is rejected because of skewness or because of a non-standard kurtosis.

Appendix:

LIMDEP code

?= Replace *var2*, …, *varK* by your explanatory variables, and insert the dependent dummy variable

```
namelist   ;x_prob=one, var2, …, varK $

create  ;y=… $

probit  ;lhs=y; rhs=x_prob $

matrix ;bxi=x_prob*b $

create  ;bexi=bxi $

create     ;bexiph=phi(bexi) ;ya=y-bexiph

           ;yar=bexiph*(1-bexiph) ;yart=1/(yar^0.5)

           ;y_art=ya*yart ;xa=N01(bexi)

           ;xar=xa*yart ;x_art1=xar $
```

?= Complement the lines, and replace *var2*, …, *varK* by your explanatory variables

```
create     ;x_art2=xar*var2

           ;…

           ;x_artK=xar*varK $

create     ;x_arts=-1/3*xar*(bexi^2-1)

           ;x_artk=0.25*xar*bexi*(3+bexi^2)$
```

?= Complete the list of regressors for the artificial regression

```
namelist   ;x_artreg=x_art1,x_art2,...,x_artK,x_arts,x_artk$

regress    ;lhs=y_art; rhs=x_artreg $

matrix     ;y_hat=x_artreg*b$

calculate  ;list; LM=y_hat'y_hat; p_value=1-Chi(LM,2)$
```

References

Bera, A. K.; Jarque, C. M.; Lee, L. F. (1984): Testing the normality assumption in limited dependent variable models. International Economic Review 25, pp. 563-578.

Davidson, R.; MacKinnon, J. G. (1984): Convenient specification tests for logit and probit models. Journal of Econometrics 25, pp. 241-262.

Davidson, R; MacKinnon, J. G. (1993): Estimation and inference in econometrics. Oxford University Press: New York.

Engle, R. F. (1984): Wald, Likelihood Ratio, and Lagrange Multiplier tests in econometrics, in: Z. Griliches, M. Intriligator (eds), Handbook of econometrics, Vol. 2. North Holland, Amsterdam, pp. 775-826.

Greene, W. (2003): Econometric analysis 5th ed. Prentice Hall: Upper Saddle River.

*Johnson, N. L.; Kotz, S.; Balakrishnan, N. (*1994): Continuous univariate distributions Vol. 1 2nd ed. Wiley: New York.

Papke, L. E.; Wooldridge, J. M. (1996): Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. Journal of Applied Econometrics 11, pp. 619-632.

Silva, J. M. C. (2001): A score test for non-nested hypotheses with applications to discrete data models. Journal of Applied Econometrics 16, pp. 577-597.

Wooldridge, J. M. (2002): Econometric analysis of cross section and panel data. MIT Press: Cambridge MA.