

# TESTS OF NORMALITY OF REGRESSION ERRORS

Herman J. Bierens

Pennsylvania State University

In this note I will derive the Jarque-Bera and Salmon-Kiefer tests of the normality of the regression errors. I will occasionally refer to my lecture notes on Modes of Convergence, hereafter indicated by "MC", and to my lecture notes on the Multivariate Normal Distribution, hereafter indicated by "MND", but some steps in the derivations are left as exercises, where the readers have to figure out for themselves which results in my lecture notes MND and MC and my notes on the Linear Regression Model are applicable.

Consider the linear regression with an intercept

$$y_j = \theta^T x_j + u_j, \quad x_j^T = (1, x_{2,j}, \dots, x_{k,j}), \quad j = 1, \dots, n, \quad (1)$$

where

**ASSUMPTION 1:** *The vectors  $(u_j, x_{2,j}, \dots, x_{k,j})^T$  are independent random drawings from a  $k$ -variate distribution, and  $E(u_j | x_{2,j}, \dots, x_{k,j}) = 0$  with probability 1.*

We want to test the null hypothesis that (conditionally on the  $x$  variables),

$$H_0: u_j | x_j \sim N(0, \sigma^2). \quad (2)$$

Under the null hypotheses we have:

$$E(u_j^3) = 0, \quad E(u_j^4) = 3\sigma^4, \quad (3)$$

and it will be more feasible to test the latter implication of the null hypothesis than the null hypothesis (2) itself.

If the errors  $u_j$  would be observable and the variance  $\sigma^2$  known, we could test the hypothesis (3) on the basis of the random vector

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^n (u_j/\sigma)^3 \\ \sum_{j=1}^n ((u_j/\sigma)^4 - 3) \end{pmatrix}, \quad (4)$$

which under the normality hypothesis, due to the central limit theorem, converges in distribution to the bivariate normal distribution with zero mean vector and variance matrix

$$\begin{pmatrix} E(u_1/\sigma)^6 & E(u_1/\sigma)^7 - 3E(u_1/\sigma)^3 \\ E(u_1/\sigma)^7 - 3E(u_1/\sigma)^3 & E(u_1/\sigma)^8 - 6E(u_1/\sigma)^4 + 9 \end{pmatrix} = \begin{pmatrix} 15 & 0 \\ 0 & 96 \end{pmatrix}, \quad (5)$$

where the equality involved follows from the fact that under the hypothesis (2),

$$\begin{aligned} E(u_j^{2m-1}) &= 0 \quad \text{for } m = 1, 2, 3, \dots \\ E(u_j^2) &= \sigma^2 \\ E(u_j^4) &= 3\sigma^4 \\ E(u_j^6) &= 15\sigma^6 \\ E(u_j^8) &= 105\sigma^8 \end{aligned} \quad (6)$$

The results in (6) can be derived (*exercise*: Try it!) using the moment generating function of the standard normal distribution. Thus under the null hypothesis (2) we have

$$\frac{1}{\sqrt{n}} \begin{pmatrix} \sum_{j=1}^n (u_j/\sigma)^3 \\ \sum_{j=1}^n ((u_j/\sigma)^4 - 3) \end{pmatrix} \rightarrow N_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 15 & 0 \\ 0 & 96 \end{pmatrix} \right] \text{ in distr.} \quad (7)$$

and consequently, using MND Theorems 4 and 9, it follows that

$$\frac{\left(\sum_{j=1}^n (u_j/\sigma)^3\right)^2}{15n} + \frac{\left(\sum_{j=1}^n ((u_j/\sigma)^4 - 3)\right)^2}{96n} \rightarrow \chi_2^2 \text{ in distr.} \quad (8)$$

However, we do not observe the  $u_j$ 's and the variance  $\sigma^2$ , so that this test is not feasible.

The idea behind the Jarque-Bera and Salmon-Kiefer tests is to replace in (7) the  $u_j$ 's by the least squares residuals

$$\hat{u}_j = y_j - \hat{\theta}^T x_j = u_j - (\hat{\theta} - \theta)^T x_j, \quad (9)$$

where  $\hat{\theta}$  is the OLS estimator of  $\theta$ , and  $\sigma^2$  by the estimator

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n \hat{u}_j^2, \quad (10)$$

and to prove that the asymptotic normality result (7) goes through, although with a different variance matrix.

Note that, due to the presence of an intercept,

$$\sum_{j=1}^n \hat{u}_j = 0. \quad (11)$$

(*exercise: why?*). Therefore, we can also write

$$\hat{u}_j = u_j - \bar{u} - (\hat{\theta} - \theta)^T (x_j - \bar{x}), \quad (12)$$

where  $\bar{u} = (1/n) \sum_{j=1}^n u_j$  and  $\bar{x} = (1/n) \sum_{j=1}^n x_j$ . This will be convenient in proving Lemmas 2 and 3

below.

Next assume:

**ASSUMPTION 2:**  $Q = E(x_j x_j^T)$  is nonsingular, and  $E(\|x_j\|^4) \leq \infty$ ,

where  $\|x\| = \sqrt{x^T x}$  is the Euclidean norm. Then:

**LEMMA 1:** *Under Assumptions 1-2 and the null hypothesis (2) we have*

$$\text{plim}_{n \rightarrow \infty} [(1/\sqrt{n})\sum_{j=1}^n \hat{u}_j^2 - (1/\sqrt{n})\sum_{j=1}^n u_j^2] = 0.$$

*Proof:* It follows from (9) that

$$\begin{aligned} \sum_{j=1}^n \hat{u}_j^2 &= \sum_{j=1}^n \left( u_j - (\hat{\theta} - \theta)^T x_j \right)^2 \\ &= \sum_{j=1}^n u_j^2 - 2(\hat{\theta} - \theta)^T \sum_{j=1}^n u_j x_j + (\hat{\theta} - \theta)^T \sum_{j=1}^n x_j x_j^T (\hat{\theta} - \theta) \\ &= \sum_{j=1}^n u_j^2 - \sqrt{n}(\hat{\theta} - \theta)^T \left( \frac{1}{n} \sum_{j=1}^n x_j x_j^T \right) \sqrt{n}(\hat{\theta} - \theta). \end{aligned} \quad (13)$$

(*exercise:* prove the last step). Moreover, under the conditions of the lemma we have

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N_k(0, \sigma^2 Q^{-1}) \text{ in distr.} \quad (14)$$

and by the law of large numbers (MC, Theorem 1),

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n x_j x_j^T = Q. \quad (15)$$

(*exercise:* Verify the conditions of the law of large numbers). These two results, together with MC

Theorem 4, imply that

$$\sum_{j=1}^n u_j^2 - \sum_{j=1}^n \hat{u}_j^2 = \sqrt{n}(\hat{\theta} - \theta)^T \left( \frac{1}{n} \sum_{j=1}^n x_j x_j^T \right) \sqrt{n}(\hat{\theta} - \theta) \rightarrow \chi_k^2 \text{ in distr.} \quad (16)$$

(*exercise:* which result in MND has been applied?). The lemma follows now from (16) and MC,

Theorem 5. (*exercise:* why?).

**LEMMA 2:** *Under Assumptions 1-2 and the null hypothesis (2) we have*

$$\text{plim}_{n \rightarrow \infty} [(1/\sqrt{n})\sum_{j=1}^n \hat{u}_j^3 - (1/\sqrt{n})\sum_{j=1}^n (u_j^3 - 3\sigma^2 u_j)] = 0.$$

*Proof:* Using (12) and the well-known equality  $(a-b)^3 = a^3 - 3a^2b + 3ab^2 - b^3$ , we have

$$\begin{aligned}
\sum_{j=1}^n \hat{u}_j^3 &= \sum_{j=1}^n \left( u_j - \bar{u} - (\hat{\theta} - \theta)^T (x_j - \bar{x}) \right)^3 \\
&= \sum_{j=1}^n (u_j - \bar{u})^3 - 3(\hat{\theta} - \theta)^T \sum_{j=1}^n (u_j - \bar{u})^2 (x_j - \bar{x}) + 3(\hat{\theta} - \theta)^T \sum_{j=1}^n (u_j - \bar{u}) (x_j - \bar{x}) (x_j - \bar{x})^T (\hat{\theta} - \theta) \\
&\quad - \sum_{j=1}^n \left( (\hat{\theta} - \theta)^T (x_j - \bar{x}) \right)^3
\end{aligned} \tag{17}$$

Since under the conditions of the lemma,

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (u_j - \bar{u})^2 (x_j - \bar{x}) = 0 \tag{18}$$

and

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (u_j - \bar{u}) (x_j - \bar{x}) (x_j - \bar{x})^T = O \tag{19}$$

(*exercise: why?*), it follow that

$$\text{plim}_{n \rightarrow \infty} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n \hat{u}_j^3 - \frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j - \bar{u})^3 \right) = 0. \tag{20}$$

Next, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j - \bar{u})^3 &= \frac{1}{\sqrt{n}} \sum_{j=1}^n u_j^3 - 3\sqrt{n}\bar{u} \frac{1}{n} \sum_{j=1}^n u_j^2 + 2\sqrt{n}\bar{u}^3 \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j^3 - 3\sigma^2 u_j) - 3\sqrt{n}\bar{u} \frac{1}{n} \sum_{j=1}^n (u_j^2 - \sigma^2) + 2\sqrt{n}\bar{u}^3,
\end{aligned} \tag{21}$$

hence

$$\text{plim}_{n \rightarrow \infty} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j - \bar{u})^3 - \frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j^3 - 3\sigma^2 u_j) \right) = 0. \tag{22}$$

(*exercise: why?*). Combining (20) and (22), the lemma follows.

**LEMMA 3:** Under Assumptions 1-2 and the null hypothesis (2) we have

$$\text{plim}_{n \rightarrow \infty} \left( (1/\sqrt{n}) \sum_{j=1}^n \hat{u}_j^4 - (1/\sqrt{n}) \sum_{j=1}^n u_j^4 \right) = 0.$$

*Proof:* Again using (12), and the well-known equality  $(a-b)^4 = a^4 - 4a^3b + 6a^2b^2 - 4ab^3 + b^4$ , we have

$$\begin{aligned} \sum_{j=1}^n \hat{u}_j^4 &= \sum_{j=1}^n \left( u_j - \bar{u} - (\hat{\theta} - \theta)^T (x_j - \bar{x}) \right)^4 \\ &= \sum_{j=1}^n (u_j - \bar{u})^4 - 4(\hat{\theta} - \theta)^T \sum_{j=1}^n (u_j - \bar{u})^3 (x_j - \bar{x}) + 6(\hat{\theta} - \theta)^T \sum_{j=1}^n (u_j - \bar{u})^2 (x_j - \bar{x}) (x_j - \bar{x})^T (\hat{\theta} - \theta) \\ &\quad - 4 \sum_{j=1}^n (u_j - \bar{u}) \left( (\hat{\theta} - \theta)^T (x_j - \bar{x}) \right)^3 + \sum_{j=1}^n \left( (\hat{\theta} - \theta)^T (x_j - \bar{x}) \right)^4 \end{aligned} \quad (23)$$

Therefore, similar to the proof of Lemmas 2 we have:

$$\text{plim}_{n \rightarrow \infty} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n \hat{u}_j^4 - \frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j - \bar{u})^4 \right) = 0. \quad (24)$$

(Exercise: Prove this). Moreover,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j - \bar{u})^4 = \frac{1}{\sqrt{n}} \sum_{j=1}^n u_j^4 - 4\sqrt{n\bar{u}} \frac{1}{n} \sum_{j=1}^n u_j^3 + 6 \frac{1}{\sqrt{n}} (\sqrt{n\bar{u}})^2 \frac{1}{n} \sum_{j=1}^n u_j^2 - 3 \frac{1}{n\sqrt{n}} (\sqrt{n\bar{u}})^4. \quad (25)$$

Since  $\sqrt{n\bar{u}} \sim N(0, \sigma^2)$ ,  $\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n u_j^3 = 0$  and  $\text{plim}_{n \rightarrow \infty} (1/n) \sum_{j=1}^n u_j^2 = \sigma^2$ , it follows now that

$$\text{plim}_{n \rightarrow \infty} \left( \frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j - \bar{u})^4 - \frac{1}{\sqrt{n}} \sum_{j=1}^n u_j^4 \right) = 0. \quad (26)$$

(Exercise: Why?). Combining (24) and (26), Lemma 3 follows.

**LEMMA 4:** Under Assumptions 1-2 and the null hypothesis (2) we have

$$\text{plim}_{n \rightarrow \infty} \left[ (1/\sqrt{n}) \sum_{j=1}^n (\hat{u}_j^4 - 3\hat{\sigma}^4) - (1/\sqrt{n}) \sum_{j=1}^n (u_j^4 - 6\sigma^2 u_j^2 + 3\sigma^4) \right] = 0.$$

*Proof:* It follows from (10), Lemma 1, and the law of large numbers and the central limit theorem that

$$\text{plim}_{n \rightarrow \infty} \hat{\sigma}^2 = \sigma^2, \quad \sqrt{n}(\hat{\sigma}^2 - \sigma^2) \rightarrow N(0, 2\sigma^4) \text{ in distr,} \quad (27)$$

(exercise: why?) and thus

$$\text{plim}_{n \rightarrow \infty} \left( \sqrt{n}(\hat{\sigma}^4 - \sigma^4) - 2\sigma^2 \frac{1}{\sqrt{n}} \sum_{j=1}^n (u_j^2 - \sigma^2) \right) = 0 \quad (28)$$

(exercise: why?). Combining the latter result with Lemma 3, Lemma 4 follows.

Finally, it follows from (6) that

$$E \begin{pmatrix} u_j^3 - 3\sigma^2 u_j \\ u_j^4 - 6\sigma^2 u_j^2 + 3\sigma^4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{Var} \begin{pmatrix} u_j^3 - 3\sigma^2 u_j \\ u_j^4 - 6\sigma^2 u_j^2 + 3\sigma^4 \end{pmatrix} = \begin{pmatrix} 6\sigma^6 & 0 \\ 0 & 24\sigma^8 \end{pmatrix} \quad (29)$$

hence, using the central limit theorem, it follows that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \begin{pmatrix} u_j^3 - 3\sigma^2 u_j \\ u_j^4 - 6\sigma^2 u_j^2 + 3\sigma^4 \end{pmatrix} \rightarrow N_2 \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 6\sigma^6 & 0 \\ 0 & 24\sigma^8 \end{pmatrix} \right] \text{ in distr.} \quad (30)$$

Using this result and those of Lemmas 1-4, it follows that

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \begin{pmatrix} (\hat{u}_j/\hat{\sigma})^3 \\ (\hat{u}_j/\hat{\sigma})^4 - 3 \end{pmatrix} \rightarrow N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 6 & 0 \\ 0 & 24 \end{pmatrix} \right) \text{ in distr.} \quad (31)$$

hence

**THEOREM 1:** Under Assumptions 1-2 and the null hypothesis (2) we have

$$\hat{T}_N = n \left( \frac{[(1/n) \sum_{j=1}^n (\hat{u}_j/\hat{\sigma})^3]^2}{6} + \frac{[(1/n) \sum_{j=1}^n (\hat{u}_j/\hat{\sigma})^4 - 3]^2}{24} \right) \rightarrow \chi_2^2 \text{ in distr.} \quad (32)$$

*Proof: Exercise.*

The test statistic  $\hat{T}_N$  is the test statistic of the Salmon-Kiefer test. The Jarque-Bera test is derived in a slightly different way, but is essentially the same.