Testing Normality

Hanno Förster*

Introduction

As stated in the well known Lindeberg-Levy Central Limit Theorem (CLT) any sum of standardized *iid* random variables that have finite variance converges in distribution to the standard normal, a result that is commonly used in Econometrics to approximate distributions of test statistics and estimators. Yet, there are a number of cases where neither the Lindeberg Levy Theorem nor similar CLTs can be invoked. E.g. when large samples are unavailable, one cannot rely on asymptotics in order to approximate the distribution of the least squares estimator in the linear model. In this instance the error terms are commonly assumed to be normally distributed to make the derivation of the estimator's distribution feasible. Likewise maximum likelihood estimation requires knowledge of the exact distribution of the error term and hence frequently relies on the assumption that a random sample is independently $\mathcal{N}(0, \sigma^2)$ distributed. For these and similar cases it is desirable to have a statistical test to assess if normality is a reliable assumption. Importantly since the need for testing normality sometimes arises in the first place because large samples are unavailable (like in the linear model), finite sample performance deserves special attention when assessing the viability of such tests.

* Hanno Förster received his degree in Economics (B. Sc.) from the University of Bonn in 2011. The present article refers to his bachelor thesis submitted in September 2011.

Arguably the most cited and most commonly applied examples in the existing literature on testing normality are the Kolmogorov Smirnov test and the Jarque Bera test (JB test). The Kolmogorov Smirnov test proposed in Kolmogorov (1933) and Smirnov (1948) is a nonparametric test, which evaluates the sample cumulative distribution function. The JB test proposed in Jarque and Bera (1980) and Jarque and Bera (1987) is based on third and fourth sample moments.

The objective of my work is to draw a comparison between the 'traditional' JB test and an approach that has been proposed in Bontemps and Meddahi (2005) and which is based on the Generalized Method of Moments (GMM) framework. In particular I present a theoretical derivation of each test's statistic and its distribution, a simulation based assessment of their finite sample properties and a brief conclusion on the merits and demerits of each test.

In this summary I present a complete derivation of the JB test's distribution, a sketch of the derivation of the GMM approach, a summary of the simulation results and a brief conclusion on the viability of each test.

Derivation of the Asymptotic Distribution of the Jarque Bera Statistic

In Jarque and Bera (1980) the JB test is developed by applying the lagrange multiplier principle to the pearson family of distributions. In contrast I present in the following a derivation of the test statistic's asymptotic distribution by straightforward application of the delta method.[1]

Let X_i denote the random Variable of interest. For an *iid* sample $X_1, ..., X_n$ the JB-statistic is defined

$$JB \quad = \quad n \left(\frac{1}{6} \, Skewness^2 + \frac{1}{24} \, (Kurtosis - 3)^2 \right),$$

[1] See Van der Vaart (2007) for details on the Delta Method.

where

$$Skewness = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^3}{s^3}, \quad Kurtosis = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4}{s^4},$$

and $\overline{X}$, s are sample mean and sample standard deviation respectively. In order to determine the asymptotic distribution of $Skewness$ and $Kurtosis$ (and ultimately of JB) define

$$g(a,b,c,d) := \frac{c-3ab+2a^3}{(b-a^2)^{\frac{3}{2}}}, \qquad h(a,b,c,d) := \frac{d-4ac+6ba^2-3a^4}{(b-a^2)^2},$$
$$y_n := \left(\overline{X}, \overline{X^2}, \overline{X^3}, \overline{X^4}\right)^T,$$

such that $Skewness = g(y_n)$ and $Kurtosis = h(y_n)$. The Delta Method states that for any function f that is continuously differentiable at μ

$$\sqrt{n}(y_n - \mu) \xrightarrow{d} \mathcal{N}(0,\Sigma) \quad \text{implies} \quad \sqrt{n}\left(f(y_n) - f(\mu)\right) \xrightarrow{d} \mathcal{N}\left(0, f'(\mu)\Sigma f'(\mu)^T\right).$$

Note that if the sample is indeed taken from the standard normal, we have[2]

$$\sqrt{n}\begin{pmatrix} \overline{X} \\ \overline{X^2} - 1 \\ \overline{X^3} \\ \overline{X^4} - 3 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0,\Sigma^0) \quad \text{where} \quad \Sigma^0 = \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 12 \\ 3 & 0 & 15 & 0 \\ 0 & 12 & 0 & 96 \end{pmatrix}$$

and

$$g(\mu^0) = 0\,,\ g'(\mu^0) = (-3,0,1,0)\,,\ h(\mu^0) = 3\,,\ h'(\mu^0) = (0,-6,0,1)\,.$$

[2] Population moments of the standard normal can be found in Table 1 in the Appendix.

Application of the delta method to g and h (both differentiable at $\mu^0 = (0,1,0,3)^T$) yields that for a sample drawn from the standard normal $\sqrt{n}$ *Skewness* $\xrightarrow{d} \mathcal{N}(0,6)$ and $\sqrt{n}$ *Kurtosis* $\xrightarrow{d} \mathcal{N}(0,24)$ and hence ultimately $JB \xrightarrow{d} \chi^2_{(2)}$.

The GMM Approach, a brief Outline

The derivation of the approach proposed in Bontemps and Meddahi (2005) requires introduction of some terminology. The toehold of GMM estimation is a *population moment condition* $E\left(m(X_i)\right) = 0$ where m is a vector of functions $m_1, \dots, m_l$ and X_i is an observed random variable.[3] Since the validity of this condition is a crucial assumption in GMM estimation, it is often desirable to test if the population moment condition is indeed satisfied. This can be done using a *test for overidentifying restrictions*

$$J = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(X_i)^T \, W^* \, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(X_i) \quad \text{where} \quad W^* = \; E\left(m(X_i) m(X_i)^T\right)^{-1} .$$

If the population moment condition holds, we have $J \xrightarrow{d} \chi^2_{(l)}$. Replacing W^* by a consistent estimate leaves the asymptotic distribution of J unaffected.

Bontemps and Meddahi (2005) combine these concepts with a relationship known as the Stein equation, namely that $X \sim \mathcal{N}(0,1)$ if and only if $E(\, q'(X) - Xq(X)\,) = 0$ for all continuously differentiable q that satisfy $\int |q'(x)| e^{-x^2/2} \mathrm{d}x < \infty$.[4] With the help of the Stein equation for any suitable function q_j we can define functions $m_j(X) =: q_j'(X) - Xq(X)$ that stacked into a vector $m(X)$ satisfy a population moment condition if and only if X is standard normal. The test for overidentifying restrictions thusly turns into a test for normality. Note that interestingly by inserting Skewness and Excess Kurtosis as population moment

[3] Note m does not depend on unobserved parameters. In fact this is a special case of a moment condition that is interesting for our purpose.

[4] The Stein equation was derived in Stein (1972).

conditions, the JB test as well can be motivated as special case of the GMM approach to testing normality.

As appealing choice of functions q_j Bontemps and Meddahi (2005) suggest Hermite polynomials

$$H_j(X) = \frac{(-1)^j}{\sqrt{j!}} e^{\frac{X^2}{2}} \frac{\partial^j}{\partial X^j} e^{\frac{-X^2}{2}}.$$

These polynomials have some useful properties. For instance they are orthonormal, i.e. $\mathbf{E}\left(H_j(X)H_k(X)\right) = \mathbf{I}_{(j=k)}$ and satisfy the Stein Equation if and only if $\mathbf{E}[H_j(X)] = 0$ for all $j > 0$.[5] For one thing these properties simplify the computation of the test statistic. Inserting a selection of Hermite polynomials $\{H_j\}_{j\in I}$ yields

$$J_H = \frac{1}{n} \sum_{j\in I} \left(\sum_{i=1}^{n} H_j(X_i) \right)^2 .$$

Apart from causing computational convenience the properties of the Hermite Polynomials prove to be very beneficial in situations where the unobserved variable of interest is not directly observed. This case arises for instance in testing normality of the unobserved error term in the linear model. Although it seems natural to replace the variable of interest by a consistent estimate (e.g. regression residuals in the above mentioned example), this can distort the distribution of the test statistic, a fact that is sometimes refered to as *parameter uncertainty problem*. In order to formalize this problem consider a vector m that depends on a random vector X of observed data, but also on an unobserved parameter vector θ^0 that can be estimated consistently by $\hat{\theta}$. A first order Taylor expansion

[5] In my work I fill in the details to a proof presented in Bontemps and Meddahi (2002).

of m around θ^0 yields

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(X_i,\hat{\theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} m(X_i,\theta^0) + \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(X_i,\theta^0)}{\partial\theta^T}\right)}_{=:\mathcal{D}} \sqrt{n}\left(\hat{\theta}-\theta^0\right) + \epsilon\,.$$

It is easy to see that the solution to the problem can be boiled down to finding f such that $\lim_{n\to\infty}\mathcal{D} = 0$. Bontemps and Meddahi (2005) show that in a wide range of cases of parameter uncertainty, including the linear regression example described above, the Hermite Polynomials satisfy this condition.[6] The key to this feature lies in a combination of the orthonormality property with other properties that are specific to Hermite Polynomials. Thus by inserting Hermite polynomials into the test statistic for overidentifying restrictions we get a test for normality that is robust towards parameter uncertainty.

Simulation Evidence

As emphasized earlier, the necessity for testing normality often arises in the first place, because large samples are unavailable. Thus in testing normality finite sample performance is of special importance. In my work I assess the finite sample properties of the considered approaches by the means of simulation. In particular I revisit simulations presented in Bontemps and Meddahi (2005) and consider additional settings that seemed to hold promise to reveal advantages and disadvantages of the considered tests. A selection of results is presented in Tables 2 through 7 in the Appendix.

Like Bontemps and Meddahi (2005) I find that all considered tests exhibit good size properties, although most tests based on the GMM approach tend to overreject a little bit, in general they still it perform better than the JB test that underrejects a bit, especially when applied to samples of 100 observations

[6] For a proof see Bontemps and Meddahi (2005).

or less. Both tests are very powerful against a strongly skewed exponential distribution ($\lambda = 1$). Against t-distributions in contrast all considered tests exhibit unsatisfyingly low power.

Moreover I apply both approaches to Bimodal distributions. For the considered Bimodal distributions the GMM approach is outperformed by the JB test by a bit, if even Hermite polynomials are included and by far, if used exclusively with odd Hermite Polynomials.

Note that throughout odd Hermite Polynomials seem to be good at detecting skewed distributions whereas even Hermite Polynomials seem to be best at detecting deviations in curvature.

Conclusion

All things considered, both approaches to testing normality seem to do a good though not excellent job. The JB test works well when applied to large samples, but does not exhibit entirely satisfying size properties when applied to small samples. Moreover it has poor power against symmetric alternatives like t- or bimodal distributions.

Bontemps and Meddahi (2005) contribute an insightful framework for testing normality that shows a new perspective on preceding approaches and holds promise to be a useful basis for future research in this and related subfields. The specifically proposed tests for normality exhibit good size properties, but have poor power against several alternatives. Ultimately the GMM approach can certainly not be said to perform generally better than the JB test. It might be of interest however, that if a specific alternative is suspected, a test based on carefully selected Hermite polynomials is likely to perform better than the JB test.

I conclude by remarking that although it is true that the shortcomings that both tests exhibit when applied to samples of 100 observations or less make them unreliable for testing normality in small samples, in instances where large samples are available (and yet the usual CLTs do not spare us the test), both discussed approaches make well suited candidates for testing normality.

References

Bontemps, C., and N. Meddahi (2002): "Testing Normality: a GMM Approach," *Working Paper, University of Montreal*, 10, 12–200.

——— (2005): "Testing Normality: a GMM Approach," *Journal of Econometrics*, 124, 149–186.

Casella, G., and R. L. Berger (2002): *Statistical Inference*. Duxbury.

Davidson, R., and J. G. MacKinnon (2004): *Econometric Theory and Methods*. Oxford University Press.

Eichelsbacher, P. (2003): "Die Steinsche Methode," Manuscriript. Ruhr Universitä Bochum.

Greene, W. H. (2008): *Econometric Analysis*. Pearson Education.

Hall, A. (2005): *Generalized Method of Moments*. Oxford University Press.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Jarque, C., and A. K. Bera (1980): "Efficient Tests for Normality, Homoscedasticity and serial Independence of Regression Residuals," *Economics Letters*, 6, 225–259.

——— (1987): "A Test for Normality of Observations and Regression Residuals," *International Statistical Review*, 55, 163–172.

Kolmogorov, A. (1933): "Sulla determinazione empirica di una legge di distribuzione," *G. Inst. Ital. Attuari*, 4, 83–91.

Smirnov, N. (1948): "Tables for estimating the goodness of fit of empirical distributions," *Annals of Mathematical Statistics*, 19, 279.

Stein, C. (1972): "A Bound for the Error in the normal approximation to the distribution of a Sum of Dependent Random Variables," in *Proc. Sixth Berkeley Symposium on Mathematical Statistics and Probability.*

Van der Vaart, A. W. (2007): *Asymptotic Statistics*. Cambridge University Press.

Appendix

k	$E(Y^k)$, $Y \sim \mathcal{N}(\mu, \sigma^2)$	$E(Z^k)$, $Z \sim \mathcal{N}(0,1)$
1	μ	0
2	$\mu^2 + \sigma^2$	1
3	$\mu^3 + 3\mu\sigma^2$	0
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	3
5	$\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	15
7	$\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6$	0
8	$\mu^8 + 28\mu^6\sigma^2 + 210\mu^4\sigma^4 + 420\mu^2\sigma^6 + 105\sigma^8$	105

Table 1: Population moments of the $\mathcal{N}(0,1)$ distribution

n	25	100	500	1000	5000
JB	0.0242	0.0398	0.0460	0.0477	0.0499
$H_{\{3\}}$	0.0554	0.0546	0.0512	0.0511	0.0521
$H_{\{4\}}$	0.0385	0.0428	0.0463	0.0474	0.0508
$H_{\{5\}}$	0.0220	0.0348	0.0457	0.0467	0.0513
$H_{\{6\}}$	0.0132	0.0196	0.0312	0.0358	0.0438
$H_{\{3,4\}}$	0.0541	0.0570	0.0536	0.0526	0.0513
$H_{\{5,6\}}$	0.0189	0.0287	0.0424	0.0462	0.0505
$H_{\{3,4,5,6\}}$	0.0546	0.0561	0.0592	0.0601	0.0583

Table 2: Rejection of true H_0

n	25	100	500	1000	5000
JB	0.5927	1.0000	1.0000	1.0000	1.0000
$H_{\{3\}}$	0.6276	0.9270	1.0000	1.0000	1.0000
$H_{\{4\}}$	0.6834	0.9842	1.0000	1.0000	1.0000
$H_{\{5\}}$	0.6232	0.9754	1.0000	1.0000	1.0000
$H_{\{3,4\}}$	0.6851	0.9810	1.0000	1.0000	1.0000
$H_{\{5,6\}}$	0.6467	0.9877	1.0000	1.0000	1.0000
$H_{\{2,3,4,5,6\}}$	0.7263	0.9944	1.0000	1.0000	1.0000

Table 3: Rejection of false H_0 (exp)

n	25	100	500	1000	5000
JB	0.0421	0.1033	0.2345	0.3594	0.8955
$H_{\{3\}}$	0.1122	0.1376	0.1549	0.1627	0.1684
$H_{\{4\}}$	0.0942	0.1650	0.3563	0.5223	0.9708
$H_{\{5\}}$	0.0631	0.1261	0.2372	0.2920	0.4016
$H_{\{6\}}$	0.0437	0.0892	0.1956	0.2687	0.5231
$H_{\{3,4\}}$	0.1189	0.1892	0.3550	0.5001	0.9567
$H_{\{5,6\}}$	0.0592	0.1193	0.2470	0.3293	0.5842
$H_{\{3,4,5,6\}}$	0.11045	0.1857	0.3694	0.5146	0.9475

Table 4: Rejection of false H_0 (t, df=25)

n	25	100	500	1000	5000
JB	0.0338	0.0751	0.1359	0.1902	0.5344
$H_{\{3\}}$	0.0883	0.1026	0.1077	0.1091	0.1124
$H_{\{4\}}$	0.0694	0.1089	0.1925	0.2710	0.6957
$H_{\{5\}}$	0.0459	0.0854	0.1460	0.1733	0.2356
$H_{\{6\}}$	0.0288	0.0561	0.1124	0.1470	0.2561
$H_{\{3,4\}}$	0.0925	0.1303	0.1989	0.2648	0.6457
$H_{\{5,6\}}$	0.0413	0.0770	0.1464	0.1861	0.3080
$H_{\{3,4,5,6\}}$	0.0837	0.1265	0.2156	0.2855	0.6307

Table 5: Rejection of false H_0 (t, df=40)

n	25	100	500	1000	5000
JB	0.0029	0.0112	0.9168	1.0000	1.0000
$H_{\{3\}}$	0.0097	0.0068	0.0065	0.0070	0.0069
$H_{\{4\}}$	0.0234	0.1743	0.9508	0.9998	1.0000
$H_{\{5\}}$	0.0026	0.0026	0.0021	0.0022	0.0021
$H_{\{6\}}$	0.00595	0.0796	0.8518	0.9977	1.0000
$H_{\{3,4\}}$	0.0103	0.0702	0.8594	0.9988	1.0000
$H_{\{5,6\}}$	0.0012	0.0205	0.6375	0.9832	1.0000
$H_{\{3,4,5,6\}}$	0.0101	0.0897	0.9017	0.9995	1.0000

Table 6: Rejection of false H_0 (Bimodal)

n	25	100	500	1000	5000
JB	0.9566	0.9991	0.9801	0.8156	0.2347
$H_{\{3\}}$	0.9999	0.9975	0.9914	0.5186	0.2971
$H_{\{4\}}$	1.0000	1.0000	0.9984	0.8808	0.2687
$H_{\{5\}}$	1.0000	0.9999	0.9996	0.9981	0.7230
$H_{\{6\}}$	1.0000	1.0000	0.9984	0.8808	0.2687
$H_{\{3,4\}}$	1.0000	1.0000	0.9987	0.8720	0.2474
$H_{\{5,6\}}$	1.0000	1.0000	1.0000	1.0000	0.9973
$H_{\{3,4,5,6\}}$	1.0000	1.0000	1.0000	1.0000	0.9980

Table 7: Rejection of true H_0, outlier ($x = 5$) added