A Generalized Jarque-Bera Test of Conditional Normality

Yi-Ting Chen[a] and **Chung-Ming Kuan**[b]

[a]Institute for Social Sciences and Philosophy, Academia Sinica
[b]Institute of Economics, Academia Sinica

Abstract

We consider testing normality in a general class of models that admits nonlinear conditional mean and conditional variance functions. We derive the asymptotic distribution of the skewness and kurtosis coefficients of the model's standardized residuals and propose an asymptotic χ^2 test of normality. This test simplifies to the Jarque-Bera test only when: (i) the conditional mean function contains an intercept term but does not depend on past errors, and (ii) the errors are conditionally homoskedastic. Beyond this context, it is shown that the Jarque-Bera test has size distortion but the proposed test does not.

Keywords: conditional heteroskedsaticity, conditional normality, Jarque-Bera test

JEL classification: C12, C22

Address for correspondence:

Yi-Ting Chen, Institute for Social Sciences and Philosophy, Academia Sinica, 128 Academia Road, Sec. 2, Taipei 115, Taiwan. Tel:+886-2-2789-8183. Fax:+886-2-2785-4160. e-mail: ytchen@gate.sinica.edu.tw.

1   Introduction

A typical approach to testing departure from normality is to check the third and fourth moments of the random variables. The omnibus test proposed by D'Agostino and Pearson (1973) and Bowman and Shenton (1975) is an early example. Jarque and Bera (1980, 1987) and White and MacDonald (1980) showed that this test is applicable to the ordinary least squares residuals of linear regressions with an intercept term and i.i.d. errors. This test, now also known as the Jarque-Bera (JB) test in the econometrics literature, is a popular diagnostic tool in practice. Despite its original context, the JB test is usually applied to various general models, such as nonlinear regressions and conditional heteroskedasticity models. Yet its applicability to such models has not been carefully examined.

This paper is concerned with testing normality in a general class of models that admits nonlinear conditional mean and conditional variance functions. We first derive the asymptotic distribution of the sample skewness and kurtosis coefficients of the model's standardized residuals. We then construct an asymptotic χ^2 test of normality based on this result. The proposed test may be interpreted as a generalized JB test because it simplifies to the JB test when: (i) the conditional mean function contains an intercept term but does not depend on past errors, and (ii) the errors are conditionally homoskedastic. Beyond this context, our simulation shows that the JB test suffers from serious size distortion but the proposed test does not.

This paper is organized as follows. In Section 2, the main distribution result is derived and a normality test is proposed. In Section 3, we discuss the implementation of the proposed test and the limitation of the JB test. Simulation results are reported in Section 4. The paper is concluded by Section 5.

2   The Proposed Normality Test

Let y_t be the variable of interest and x_t be a vector of exogenous variables taken from the information set at time t. Consider the following general model of y_t that admits nonlinear conditional mean and conditional variance functions:

$$y_t = m(x_t, \gamma; \delta) + u_t,$$
$$u_t = \varepsilon_t h(x_t, \delta; \gamma)^{1/2},$$

(1)

where $\gamma \in \Gamma \subseteq \mathbb{R}^p$ and $\delta \in \Delta \subseteq \mathbb{R}^q$ are parameter vectors, u_t denotes the regression error, and ε_t is the standardized error. Here, $m_t := m(x_t, \gamma; \delta)$ is the conditional mean function

of y_t which explicitly depends on the parameter γ, and $h_t := h(x_t, \delta; \gamma)$ is the conditional variance function which explicitly depends on the parameter δ. Moreover, m_t (h_t) may also implicitly depend on δ (γ) through the presence of lagged conditional variances h_{t-j} and/or lagged regression errors u_{t-j} for some $j \geq 1$. For example, h_t in GARCH models depends on h_{t-j} and u_{t-j}^2, and m_t depends on functions of h_{t-j} (or u_{t-j}) when there is GARCH in mean (or bilinearity).

Throughout this paper, our maintained assumption is:

[A] The postulated model is correctly specified in the sense that ε_t are i.i.d. with mean zero and variance one and are independent of x_s for all s.

Bollerslev and Wooldridge (1992) imposed the same condition to establish consistency and asymptotic normality of the Gaussian quasi-maximum-likelihood estimator (QMLE) for model (1). The condition [A] is also crucial for other conditional distribution tests; see e.g., Jarque and Bera (1980) and Bai and Ng (2001). Testing whether y_t are conditionally normally distributed amounts to testing

$$H_0: \ \varepsilon_t \sim N(0,1).$$

It is well known that under the null hypothesis,

$$\mathbb{E}(\varepsilon_t^n) = \begin{cases} 0, & n = 1, 3, 5, \ldots \\ \prod_{i=1}^{n/2}(2i-1), & n = 2, 4, 6, \ldots \end{cases} \tag{2}$$

It is thus natural to check if the sample counterparts of $\mathbb{E}(\varepsilon_t^3)$ and $[\mathbb{E}(\varepsilon_t^4) - 3]$ are sufficiently close to zero.

Let T be the sample size and $\hat{\theta}_T := (\hat{\gamma}_T', \hat{\delta}_T')'$ be the QMLE for $\theta := (\gamma', \delta')'$, obtained by maximizing the Gaussian quasi-log-likelihood function:

$$L_T(\gamma, \delta) = -\frac{1}{2} \log 2\pi - \frac{1}{2T} \sum_{t=1}^{T} \log h_t - \frac{1}{2T} \sum_{t=1}^{T} \left[\frac{y_t - m_t}{\sqrt{h_t}}\right]^2.$$

We will not state explicitly the regularity conditions required for our asymptotic result. For simplicity, we simply assume that the data (and its functions) are stationary and ergodic and obey suitable law of large numbers and central limit theorem. Moreover, $T^{1/2}(\hat{\theta}_T - \theta)$ is asymptotically normally distributed. In what follows, when m_t (h_t) is evaluated at $\theta = \hat{\theta}_T$, we write $\hat{m}_t$ $(\hat{h}_t)$. The standardized residual is thus $\hat{\varepsilon}_t = (y_t - \hat{m}_t)/\hat{h}_t^{1/2}$. The test we consider is based on

$$S_T = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^3 \qquad \text{and} \qquad K_T = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^4,$$

the sample skewness and kurtosis coefficients of the standardized residuals.

Define the $(p+q) \times 1$ vectors:

$$w_t = \begin{bmatrix} w_{\gamma t} \\ w_{\delta t} \end{bmatrix} = \begin{bmatrix} \nabla_\gamma m_t / \sqrt{h_t} \\ \nabla_\delta m_t / \sqrt{h_t} \end{bmatrix},$$

$$z_t = \begin{bmatrix} z_{\gamma t} \\ z_{\delta t} \end{bmatrix} = \begin{bmatrix} \nabla_\gamma h_t / h_t \\ \nabla_\delta h_t / h_t \end{bmatrix},$$

which are determined by the specifications of m_t and h_t. Also let

$$\Upsilon = \mathbb{E}(w_t w_t') + \frac{1}{2}\mathbb{E}(z_t z_t'),$$

$\kappa_{ww} = \mathbb{E}(w_t')\Upsilon^{-1}\mathbb{E}(w_t)$, $\kappa_{zz} = \mathbb{E}(z_t')\Upsilon^{-1}\mathbb{E}(z_t)$, and $\kappa_{wz} = \mathbb{E}(w_t')\Upsilon^{-1}\mathbb{E}(z_t)$. In Appendix it is shown that

$$T^{1/2}S_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \phi_s(\varepsilon_t, x_t) + o_p(1),$$

$$T^{1/2}(K_T - 3) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \phi_k(\varepsilon_t, x_t) + o_p(1), \tag{3}$$

where

$$\phi_s(\varepsilon_t, x_t) = \varepsilon_t^3 - 3\mathbb{E}(w_t')\Upsilon^{-1}\left[w_t\varepsilon_t + \frac{1}{2}z_t(\varepsilon_t^2 - 1)\right],$$

$$\phi_k(\varepsilon_t, x_t) = (\varepsilon_t^4 - 3) - 6\mathbb{E}(z_t')\Upsilon^{-1}\left[w_t\varepsilon_t + \frac{1}{2}z_t(\varepsilon_t^2 - 1)\right],$$

both are martingale difference sequences. From (2) we know that the odd moments of ε_t are all zero and the even moments are: $\mathbb{E}(\varepsilon_t^4) = 3$, $\mathbb{E}(\varepsilon_t^6) = 15$, $\mathbb{E}(\varepsilon_t^8) = 105$. We thus have

$$\mathrm{var}[\phi_s(\varepsilon_t, x_t)] = \mathbb{E}(\varepsilon_t^6) - 6\kappa_{ww}\mathbb{E}(\varepsilon_t^4) + 9\kappa_{ww} = 15 - 9\kappa_{ww}.$$

It can also be shown that

$$\mathrm{var}[\phi_k(\varepsilon_t, x_t)] = \mathbb{E}(\varepsilon_t^8) - 6\mathbb{E}(\varepsilon_t^4) + 9 - 6\kappa_{zz}[\mathbb{E}(\varepsilon_t^6) - \mathbb{E}(\varepsilon_t^4) - 3\mathbb{E}(\varepsilon_t^2) + 3)] + 36\kappa_{zz}$$

$$= 96 - 36\kappa_{zz},$$

and that $\mathrm{cov}[\phi_s(\varepsilon_t, x_t), \phi_k(\varepsilon_t, x_t)] = -18\kappa_{wz}$. By invoking a central limit theorem for martingale difference sequences (e.g., Billingsley, 1961), we obtain the main distribution result below.

3

Theorem 2.1 *When [A] holds for model (1),*

$$\begin{bmatrix} \sqrt{T} S_T \\ \sqrt{T}(K_T - 3) \end{bmatrix} \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 15 - 9\kappa_{ww} & -18\kappa_{wz} \\ -18\kappa_{wz} & 96 - 36\kappa_{zz} \end{bmatrix} \right),$$

under the null hypothesis.

Theorem 2.1 shows that for the general model (1), $T^{1/2}S_T$ and $T^{1/2}(K_T - 3)$ are not asymptotically independent unless $\kappa_{wz} = 0$. Even when they are independent, their asymptotic variances depend on the parameters κ_{ww} and κ_{zz}. Basing on this result, a generally applicable test of normality can be easily constructed. Let V_o denote the asymptotic variance-covariance matrix in Theorem 2.1 and $\widetilde{V}_T$ its consistent estimator; the computation of $\widetilde{V}_T$ will be discussed in next section. The proposed test of conditional normality for model (1) is

$$N_T := T\, [S_T,\ K_T - 3]\, \widetilde{V}_T^{-1} \begin{bmatrix} S_T \\ K_T - 3 \end{bmatrix}. \tag{4}$$

The result below is immediate from Theorem 2.1 and the continuous mapping theorem.

Corollary 2.2 *When [A] holds for model (1), $N_T \xrightarrow{D} \chi^2(2)$ under the null hypothesis.*

3 Implementing the Proposed Test

To compute $\widetilde{V}_T$ for the proposed N test, we may replace the parameters κ_{ww}, κ_{wz}, and κ_{zz} in V_o with their sample counterparts:

$$\tilde{\kappa}_{ww} = \left[\frac{1}{T}\sum_{t=1}^{T} \hat{w}_t'\right] \left[\frac{1}{T}\sum_{t=1}^{T} \hat{w}_t\hat{w}_t' + \frac{1}{2T}\sum_{t=1}^{T} \hat{z}_t\hat{z}_t'\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} \hat{w}_t\right],$$

$$\tilde{\kappa}_{zz} = \left[\frac{1}{T}\sum_{t=1}^{T} \hat{z}_t'\right] \left[\frac{1}{T}\sum_{t=1}^{T} \hat{w}_t\hat{w}_t' + \frac{1}{2T}\sum_{t=1}^{T} \hat{z}_t\hat{z}_t'\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} \hat{z}_t\right],$$

$$\tilde{\kappa}_{wz} = \left[\frac{1}{T}\sum_{t=1}^{T} \hat{w}_t'\right] \left[\frac{1}{T}\sum_{t=1}^{T} \hat{w}_t\hat{w}_t' + \frac{1}{2T}\sum_{t=1}^{T} \hat{z}_t\hat{z}_t'\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} \hat{z}_t\right],$$

where $\hat{w}_t$ and $\hat{z}_t$ are, respectively, w_t and z_t evaluated at the Gaussian QMLE $\hat{\theta}_T$. It is not too difficult to see that, when the QMLE is consistent and the functions above all satisfy a uniform law of large numbers, these sample counterparts are consistent for the true parameters; we omit the details.

4

The proposed test may be simplified in some special cases of model (1). Consider first the case that m_t does not depend on δ and h_t does not depend on γ:

$$y_t = m(x_t, \gamma) + u_t,$$
$$u_t = \varepsilon_t h(x_t, \delta)^{1/2}. \tag{5}$$

This model excludes bilinear models, GARCH models, and models with conditional variance in mean. In this case, $w_t = (w'_{\gamma t}, 0')'$ and $z_t = (0', z'_{\delta t})'$. It follows that

$$\tilde{\kappa}_{ww} = \left[\frac{1}{T}\sum_{t=1}^{T} \hat{w}'_{\gamma t}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\hat{w}_{\gamma t}\hat{w}'_{\gamma t}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{w}_{\gamma t}\right],$$

$$\tilde{\kappa}_{zz} = 2\left[\frac{1}{T}\sum_{t=1}^{T}\hat{z}'_{\delta t}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\hat{z}_{\delta t}\hat{z}'_{\delta t}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{z}_{\delta t}\right], \tag{6}$$

and $\tilde{\kappa}_{wz} = 0$. The N test (4) now becomes

$$\frac{TS_T^2}{15 - 9\tilde{\kappa}_{ww}} + \frac{T(K_T - 3)^2}{96 - 36\tilde{\kappa}_{zz}}, \tag{7}$$

and Corollary 2.2 applies. In fact, as $\kappa_{wz} = 0$ for model (5), there is no need to estimate κ_{wz} anyway. The test (7) is therefore a natural simplification of the N test.

In view of (7), the N test would further reduce to the JB test

$$\mathrm{JB}_T = \frac{TS_T^2}{6} + \frac{T(K_T - 3)^2}{24}, \tag{8}$$

provided that $\tilde{\kappa}_{ww} = 1$ and $\tilde{\kappa}_{zz} = 2$. This is the case when the conditional mean function of (5) contains an intercept term and the errors are conditionally homoskedastic:

$$y_t = \gamma_o + m_1(x_t, \gamma_1) + u_t, \qquad u_t = \varepsilon_t \delta_o^{1/2}. \tag{9}$$

To see this, note that $\tilde{\kappa}_{ww}$ in (6) is the sample average of the fitted values for the regression of the constant one on $\hat{w}_{\gamma t} = \nabla_\gamma \hat{m}_t / \hat{h}_t^{1/2}$, where $\nabla_\gamma \hat{m}_t$ is $\nabla_\gamma m_t$ evaluated at $\hat{\theta}_T$. When $\hat{w}_{\gamma t}$ contains a constant term, this average equals the sample average of the dependent variable which is one. Clearly, model (9) is a case that $\hat{w}_{\gamma t}$ does contain a constant term. Similarly, $\tilde{\kappa}_{zz}$ in (6) is 2 times the sample average of the fitted values for the regression of the constant one on $\hat{z}_{\delta t} = \nabla_\delta \hat{h}_t / \hat{h}_t$, where $\nabla_\delta \hat{h}_t$ is $\nabla_\delta h_t$ evaluated at $\hat{\theta}_T$. It can then be seen that for model (9), $\hat{z}_{\delta t}$ contains a constant term due to homoskedasticity, so that $\tilde{\kappa}_{zz} = 2$.

5

Straightforward calculation shows that for model (9), the true parameters $\kappa_{ww} = 1$ and $\kappa_{zz} = 2$, so that

$$\begin{bmatrix} \sqrt{T}S_T \\ \sqrt{T}(K_T - 3) \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 16 & 0 \\ 0 & 24 \end{bmatrix} \right).$$

This result also justifies that model (9) is the context in which the JB test is valid. Comparing with Jarque and Bera (1980, 1987), model (9) is more general and allows nonlinear conditional mean functions, as long as it does not depend on δ and includes an intercept term. On the other hand, conditional heteroskedasticity, even correctly modeled, would invalidate the JB test. The examples below illustrate.

Example 3.1: Consider a conditionally homoskedastic AR(1) model without the intercept term:

$$y_t = \gamma y_{t-1} + u_t, \qquad u_t = \varepsilon_t \delta_o^{1/2}.$$

Estimating this model yields $w_{\gamma t} = y_{t-1}/\delta_o^{1/2}$ which has mean zero. Thus, $\kappa_{ww} = 0$, and the asymptotic variance of $T^{1/2}S_T$ is 15 by Theorem 2.1, instead of 6. As $z_{\delta t} = 1/\delta_o$ and $\kappa_{zz} = 2$, the asymptotic variance of $T^{1/2}(K_T - 3)$ remains 24. Consequently, the statistic

$$\frac{TS_T^2}{15} + \frac{T(K_T - 3)^2}{24}$$

has a limiting $\chi^2(2)$ distribution, but the JB test does not. Note also that, as $\tilde{\kappa}_{ww}$ converges to $\kappa_{ww} = 0$ and $\tilde{\kappa}_{zz} = 2$, the N test (7) for this model is asymptotically equivalent to the statistic above.

Suppose that we estimate the AR(1) model with the intercept term even when the true intercept is zero. This model is still correctly specified in the sense of [A], so that the N and JB tests are algebraically equivalent and have asymptotic $\chi^2(2)$ distribution. In contrast with the discussion above, $w_{\gamma t}$ now reads $(1/\delta_o^{1/2}, y_{t-1}/\delta_o^{1/2})'$. It can be verified that $\kappa_{ww} = 1$, so that the asymptotic variance of $T^{1/2}S_T$ is still 6. This illustrates how different specifications may affect the asymptotic variance.

Example 3.2: Consider now the model that contains an intercept term and conditionally heteroskedastic errors:

$$y_t = \gamma + u_t, \qquad u_t = \varepsilon_t h_t(x_t, \delta)^{1/2}.$$

It can be verified that $\mathbb{E}(w_t) = (\mathbb{E}[h_t^{-1/2}], 0)'$, $\mathbb{E}(z_t) = (0, \mathbb{E}[z_{\delta t}])'$, and

$$\Upsilon = \begin{bmatrix} \mathbb{E}[h_t^{-1}] & 0 \\ 0 & \frac{1}{2}\mathbb{E}[z_{\delta t}^2] \end{bmatrix}.$$

Then as long as h_t and $z_{\delta t}$ are nondegenerate random variables (with positive variances), we have

$$\kappa_{ww} = \frac{[\mathbb{E}(h_t^{-1/2})]^2}{\mathbb{E}(h_t^{-1})} < 1, \qquad \kappa_{zz} = 2\frac{[\mathbb{E}(z_{\delta t})]^2}{\mathbb{E}(z_{\delta t}^2)} < 2.$$

The asymptotic variance of $T^{1/2}S_T$ is thus greater than 6, and that of $T^{1/2}(K_T - 3)$ is greater than 24. Consequently, the JB test is not valid and would reject the null more often than it should.

4   Monte Carlo Simulation

In this section, we conduct a Monte Carlo simulation to compare the finite sample performance of the JB test and the N test. We consider the following models:

AR_c	$y_t = \gamma_o + \gamma_1 y_{t-1} + u_t,$	$u_t = \varepsilon_t \delta_o^{1/2};$
AR_n	$y_t = \gamma_1 y_{t-1} + u_t,$	$u_t = \varepsilon_t \delta_o^{1/2};$
$Bilinear$	$y_t = \gamma_o + \gamma_1 y_{t-2} u_{t-1} + u_t,$	$u_t = \varepsilon_t \delta_o^{1/2};$
$ARCH$	$y_t = \gamma_o + \gamma_1 y_{t-1} + u_t,$	$u_t = \varepsilon_t h_t^{1/2}, \quad h_t = \delta_o + \delta_1 u_{t-1}^2;$
$IGARCH$	$y_t = \gamma_o + \gamma_1 y_{t-1} + u_t,$	$u_t = \varepsilon_t h_t^{1/2}, \quad h_t = \delta_o + \delta_1 h_{t-1} + (1 - \delta_1) u_{t-1}^2;$
MD	$y_t = u_t,$	$u_t = \varepsilon_t h_t^{1/2}, \quad h_t = \delta_o + \delta_1 u_{t-1}^2.$

These models are quite common in practice. In particular, AR_c and AR_n are familiar dynamic models but differ by the intercept term; *Bilinear* is a nonlinear model that depends on u_{t-1}; the ARCH-type models involve conditionally heteroskedastic errors. Note that *Bilinear* and ARCH-type models are usually employed to characterize volatility clustering in empirical studies; see e.g., Bera and Higgins (1997). As discussed earlier, the JB and N tests are algebraically equivalent for AR_c, a special case of model (9), but not for the other models.

In our simulation, the data are generated from the models above with the parameters $(\gamma_o, \gamma_1, \delta_o, \delta_1) = (0, 0.5, 1, 0.9)$ and the sample sizes $T = 200, 1000$. The innovations ε_t are i.i.d. $N(0, 1)$ random variables in the size experiments, but they are i.i.d. standardized t random variables with the degrees of freedom seven in the power experiments. The number of replications is 1000. The empirical sizes (at the 5% and 10% nominal levels) and empirical powers (at the 5% level) of the JB and N tests are summarized in Table 1.

From Table 1 we can see that for AR_c, the JB test (and hence the N test) is slightly under-sized for $T = 200$ and has good power properties. For all other models, the JB

Table 1: The empirical sizes and powers of the JB test and the N test.

model	$T=$	JB$_{\text{size}(5\%)}$		$N_{\text{size}(5\%)}$		JB$_{\text{size}(10\%)}$		$N_{\text{size}(10\%)}$		JB$_{\text{power}*}$		N_{power}	
		200	1000	200	1000	200	1000	200	1000	200	1000	200	1000
AR_c		3.8	4.3	3.8	4.3	7.7	9.1	7.7	9.1	–	–	65.5	99.5
AR_n		16.4	17.5	4.7	5.2	22.9	24.0	8.3	8.8	58.2	98.6	64.1	99.8
$Bilinear$		9.6	11.6	5.9	5.1	15.5	17.3	9.6	9.9	62.9	99.4	64.2	99.8
$ARCH$		9.2	10.2	6.0	5.4	14.4	17.2	8.9	10.1	57.8	99.5	60.5	99.7
$IGARCH$		13.8	19.3	6.1	6.5	21.3	28.5	10.6	11.3	47.5	95.6	47.4	95.5
MD		17.1	17.9	4.6	5.7	24.6	25.2	8.4	10.2	65.5	98.1	61.6	99.8

Note: The entries are rejection frequencies in percentages. The column "JB$_{\text{power}*}$" reports the size-corrected powers of the JB test.

test tends to over-reject under the null; the rejection percentages are even two times higher than the nominal sizes for AR_n, $IGARCH$ and MD. Moreover, the size distortions deteriorate when the sample size increases. For example, at the 5% level, the JB test has the empirical size 13.8% (19.3%) for $IGRACH$ when $T = 200$ (1000). This suggests that existing applications of the JB test to conditionally heteroskedastic models, e.g., Baillie and DeGennaro (1990) and Vlaar and Palm (1993), may be vulnerable. On the other hand, the N test maintains roughly correct empirical sizes in all cases considered. The powers of the N test are quite high and similar to the size-corrected powers of the JB test. These results support the analysis in the preceding sections.

5 Conclusion

In this paper, we derive the asymptotic distribution of the sample skewness and kurtosis coefficients of a general model's standardized residuals and construct a test of conditional normality that generalizes the well known JB test. It is shown that the JB test is valid only when the conditional mean function contains an intercept but does not involve past errors and when the errors are conditionally homoskedastic. For other models, the JB test suffers from size distortion in finite samples. In contrast, the proposed test is not subject to this limitation and, therefore, can serve as a substitute for the JB test in various empirical applications.

Appendix

Proof of Equation (3): As $\nabla_\theta \varepsilon_t = -(w_t + 2^{-1} z_t \varepsilon_t)$, the first-order expansion of $\sqrt{T} S_T$ about θ is

$$\sqrt{T} S_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t^3 - 3 \left[\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \left(w_t + \frac{1}{2} z_t \varepsilon_t \right)' \right] \sqrt{T}(\hat{\theta}_T - \theta) + o_p(1).$$

By [A] and the facts that $\mathbb{E}(\varepsilon_t^2) = 1$ and $\mathbb{E}(\varepsilon_t^3) = 0$ under the null, we have, with a suitable law of large number effect,

$$\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 w_t' \xrightarrow{\mathbb{P}} \mathbb{E}(\varepsilon_t^2) \mathbb{E}(w_t') = \mathbb{E}(w_t'),$$

$$\frac{1}{2T} \sum_{t=1}^{T} \varepsilon_t^3 z_t' \xrightarrow{\mathbb{P}} \mathbb{E}(\varepsilon_t^3) \mathbb{E}(z_t') = 0,$$

where $\xrightarrow{\mathbb{P}}$ stands for convergence in probability. It is also easy to calculate the score function of the Gaussian log-likelihood function as

$$g_T(\theta) := \frac{1}{T} \sum_{t=1}^{T} \left[w_t \varepsilon_t + \frac{1}{2} z_t (\varepsilon_t^2 - 1) \right].$$

The first-order expansion of $T^{1/2} g_T(\hat{\theta}_T) = 0$ about θ is

$$0 = \sqrt{T} g_T(\theta) + \nabla_\theta g_T(\theta) \sqrt{T}(\hat{\theta}_T - \theta) + o_p(1).$$

where

$$\nabla_\theta g_T(\theta) = \frac{1}{T} \sum_{t=1}^{T} \left[(\nabla_\theta w_t) \varepsilon_t + (\nabla_\theta \varepsilon_t) w_t' + \frac{1}{2} (\nabla_\theta z_t)(\varepsilon_t^2 - 1) + (\nabla_\theta \varepsilon_t) z_t' \varepsilon_t \right].$$

It can be verified that, upon taking expectation,

$$\mathbb{E}[\nabla_\theta g_T(\theta)] = - \left[\mathbb{E}(w_t w_t') + \frac{1}{2} \mathbb{E}(z_t z_t') \right],$$

which is just $-\Upsilon$. The normalized QMLE now can be expressed as

$$\sqrt{T}(\hat{\theta}_T - \theta) = \Upsilon^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[w_t \varepsilon_t + \frac{1}{2} z_t (\varepsilon_t^2 - 1) \right] + o_p(1).$$

Putting these results together, we obtain

$$\sqrt{T} S_T = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon_t^3 - 3 \mathbb{E}(w_t') \Upsilon^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[w_t \varepsilon_t + \frac{1}{2} z_t (\varepsilon_t^2 - 1) \right] + o_p(1),$$

9

as asserted in (3). Similarly, the first-order expansion of $\sqrt{T}(K_T - 3)$ is

$$\sqrt{T}(K_T - 3) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_t^4 - 3) - 4 \left[\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^3 \left(w_t + \frac{1}{2} z_t \varepsilon_t \right)' \right] \sqrt{T}(\hat{\theta}_T - \theta) + o_p(1).$$

Again by [A] and the facts that $\mathbb{E}(\varepsilon_t^3) = 0$ and $\mathbb{E}(\varepsilon_t^4) = 3$ under the null, the terms in square bracket converges in probability to $3\,\mathbb{E}(z_t')/2$. We obtain

$$\sqrt{T}(K_T - 3) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon_t^4 - 3) - 6\mathbb{E}(z_t')\Upsilon^{-1} \left[w_t \varepsilon_t + \frac{1}{2} z_t (\varepsilon_t^2 - 1) \right] + o_p(1),$$

as the second assertion of (3). □

References

[1] Bai, J. and S. Ng (2001). A test for conditional symmetry in time series models, *Journal of Econometrics*, **103**, 225–258.

[2] Baillie, R. T. and R. P. DeGennaro (1990). Stock returns and volatility, *Journal of Financial and Quantitative Analysis*, **25**, 203–214.

[3] Bera, A. K. and M. L. Higgins (1997). ARCH and bilinearity as competing models for nonlinear dependence, *Journal of Business and Economic Statistics*, **15**, 43–50.

[4] Billingsley, P. (1961). The Lindeberg-Levy theorem for martingales. *The Proceeding of American Mathematical Scoiety*, **12**, 788–792.

[5] Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, **11**, 143–172.

[6] Bowman, K. O. and L. R. Shenton (1975). On omnibus contours for departures from normality based on $\sqrt{b_1}$ and b_2, *Biometrika*, **62**, 243–250.

[7] D'Agostino, R. B. and E. S. Pearson (1973). Tests for departure from normality: Empirical results for the distributions of b_2 and $\sqrt{b_1}$, *Biometrika*, **60**, 613–622.

[8] Jarque, C. M. and A. K. Bera (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Economics Letters*, **6**, 255–259.

[9] Jarque, C. M. and A. K. Bera (1987). A test for normality of observations and regression residuals, *International Statistical Review*, **55**, 163–172.

[10] Vlaar, P. J. G. and F. C. Palm (1993). The message in weekly exchange rates in the European monetary system: Mean reversion, conditional heteroskedasticity, and jumps, *Journal of Business and Economic Statistics*, **11**, 351–360.

[11] White, H. and G. M. MacDonald (1980). Some large-sample tests for nonnormality in the linear regression model, *Journal of the American Statistical Association*, **75**, 16–28.