Tests for Skewness, Kurtosis, and Normality for Time Series Data

Jushan Bai
Department of Economics, New York University, New York, NY 10022 (*jushan.bai@nyu.edu*)

Serena Ng
Department of Economics, University of Michigan, Ann Arbor, MI 48109 (*serena.ng@umich.edu*)

We present the sampling distributions for the coefficient of skewness, kurtosis, and a joint test of normality for time series observations. We show that when the data are serially correlated, consistent estimates of three-dimensional long-run covariance matrices are needed for testing symmetry or kurtosis. These tests can be used to make inference about any conjectured coefficients of skewness and kurtosis. In the special case of normality, a joint test for the skewness coefficient of 0 and a kurtosis coefficient of 3 can be obtained on construction of a four-dimensional long-run covariance matrix. The tests are developed for demeaned data, but the statistics have the same limiting distributions when applied to regression residuals. Monte Carlo simulations show that the test statistics for symmetry and normality have good finite-sample size and power. However, size distortions render testing for kurtosis almost meaningless except for distributions with thin tails, such as the normal distribution. Combining skewness and kurtosis is still a useful test of normality provided that the limiting variance accounts for the serial correlation in the data. The tests are applied to 21 macroeconomic time series.

KEY WORDS: Jarque–Bera test; Kurtosis; Normality; Symmetry.

1. INTRODUCTION

Consider a series $\{X_t\}_{t=1}^T$ with mean μ and standard deviation σ. Let $\mu_r = E[(x-\mu)^r]$ be the rth central moment of X_t with $\mu_2 = \sigma^2$. The coefficients of skewness and kurtosis are defined as

$$\tau = \frac{\mu_3}{\sigma^3} = \frac{E[(x-\mu)^3]}{E[(x-\mu)^2]^{3/2}} \tag{1}$$

and

$$\kappa = \frac{\mu_4}{\sigma^4} = \frac{E[(x-\mu)^4]}{E[(x-\mu)^2]^2}. \tag{2}$$

If X_t is symmetrically distributed, then μ_3 and thus τ will be 0. Sample estimates of τ and κ can be obtained on replacing the population moments μ_r by the sample moments $\hat{\mu}_r = T^{-1}\sum_{t=1}^T (X_t - \bar{X})^r$. If X_t is iid and normally distributed, then $\sqrt{T}\hat{\tau} \xrightarrow{d} \mathrm{N}(0,6)$ and $\sqrt{T}(\hat{\kappa}-3) \xrightarrow{d} \mathrm{N}(0,24)$ (see, e.g., Kendall and Stuart 1969). This article presents the limiting distributions for $\hat{\tau}$ and $\hat{\kappa}$ when the data are weakly dependent. The tests can be applied to the observed data whose population mean and variance are unknown, as well as least squares regression residuals.

Whether time series data exhibit skewed behavior has been an issue of macroeconomic interest. Some authors (e.g., Neftci 1984; Hamilton 1989) have used parametric models to see whether economic variables behave similarly during expansions and recessions. Others use simple statistics to test skewness. In a well-known article, Delong and Summers (1985) studied whether business cycles are symmetrical by applying the skewness coefficient to GDP, industrial production, and the unemployment rate. However, because the sampling distribution of the skewness coefficient for serially correlated data is not known, these authors obtained critical values by simulating an AR(3) model with normal errors. These critical values are correct only if the AR(3) model is the correct data-generating process and the errors are indeed normal. The results developed in this article allow us to test for symmetry without making such assumptions.

The coefficient of kurtosis is informative about the tail behavior of a series, an issue that has drawn substantial interest in the finance literature. However, we argue that measuring the tails using the kurtosis statistic is not a sound approach. As we show later, the true value of κ will likely be substantially underestimated in practice, because a very large number of observations is required to get a reasonable estimate. This bias translates into size distortion for testing kurtosis. Exceptions are distributions with thin tails, such as the normal distribution. But concerns for heavy tails are quite rare in such cases. As such, testing for kurtosis is not a very useful exercise per se.

Normality is often a maintained assumption in estimation and finite-sample inference. The Gaussian distribution has $\tau = 0$ and $\kappa = 3$. When $\kappa > 3$, the distribution of X_t is said to have fat tails. A joint test of $\tau = 0$ and $\kappa - 3 = 0$ is often used as a test of normality. Jarque and Bera (1980) and Bera and Jarque (1981) showed that $T(\hat{\tau}^2/6 + (\hat{\kappa}-3)^2/24) \xrightarrow{d} \chi_2^2$. We extend their results developed for iid data to weakly dependent data. Although the extension is natural, such a result apparently has not yet been documented. Lomnicki (1961) considered testing for normality in linear stochastic processes using the skewness and kurtosis coefficients, but did not consider a joint test of these two or other moments. Our tests do not require that the process be linear. We also consider a regression model with dependent errors and examine finite-sample properties of the tests.

The literature on normality is large, and a commonly used nonparametric test is the Kolmogorov–Smirnov (KS) statistic. In the present setting, the KS test will depend on nuisance parameters relating to serial correlation in the data, and its limit will no longer be distribution-free. One can use the block bootstrap

Journal of Business & Economic Statistics
January 2005, Vol. 23, No. 1
DOI 10.1198/073500104000000271

method to obtain the critical values, or use martingale transformation methods (as in Bai 2003) to obtain distribution-free tests. Empirical implementation of the KS test in a time series setting thus can be rather computationally intensive. The tests that we consider in this article are, in contrast, very simple to construct.

Several authors have used the generalized methods of moments (GMM) to test the null hypothesis that the data are Gaussian. Richardson and Smith (1993) considered testing multivariate normality, focusing on the case when the data are cross-sectionally correlated (and also discussed how serial correlation can be accommodated). Their test is based on the overidentifying restrictions from matching the first four moments of the data with those implied by the normal distribution. More recently, Bontemps and Meddahi (2002) considered testing the validity of the so-called "Stein equations," which are moment conditions that should hold under normality. Both of these normality tests are GMM-based tests of overidentifying restrictions. As is well known, the optimal weighting matrix used in GMM is the inverse of the long-run variance of the moments under consideration, which can be consistently estimated using kernel methods. We directly test the skewness and kurtosis coefficients, properly scaled by their corresponding long-run variances, in view of the fact that the data are serially correlated. Because we also estimate long-run variances, it might appear as though we are also implementing GMM. This is not the case, however. Testing overidentifying restrictions by GMM is not equivalent to testing skewness and kurtosis, at least for nonnormal distributions. As we explain later, even for normal distribution, the theoretical weighing matrix must be used to interpret testing overidentifying restrictions as testing skewness and kurtosis.

We do not assume normal distribution in deriving the skewness and kurtosis tests. The results can be used to test any given value of skewness and kurtosis coefficients. Testing normality is no more than a joint test that can be conveniently obtained within our framework. The tests used to determine whether the data are Gaussian can be easily amended to determine whether the data are consistent with, for example, the lognormal (which has skewness and kurtosis coefficients of 6.18 and 113.9). This is in contrast with the tests just described, which are designed for testing normality. The motivation of our tests is closer to that of Dufour, Khalaf, and Beaulieu (2002), although we allow the data (or residuals) to be serially correlated. Although our focus is on asymptotic results, these authors implemented exact tests via the bootstrap. Simplicity and generality are two noteworthy aspects of our tests.

2. THE TEST STATISTICS

For any integer $r \geq 1$, we first note that

$$\frac{\hat{\mu}_r}{\hat{\sigma}^r} - \frac{\mu_r}{\sigma^r} = \frac{(\hat{\mu}_r - \mu_r)}{\hat{\sigma}^r} - \frac{\mu_r}{\sigma^r}\left[\frac{(\hat{\sigma}^2)^{r/2} - (\sigma^2)^{r/2}}{\hat{\sigma}^r}\right],$$

$$= \left[\frac{T^{-1}\sum_{t=1}^{T}(X_t - \bar{X})^r - \mu_r}{\hat{\sigma}^r}\right] - \frac{\mu_r}{\sigma^r}\left[\frac{((\hat{\sigma}^2)^{r/2}) - ((\sigma^2)^{r/2})}{\hat{\sigma}^r}\right].$$

Lemma A.1 in Appendix A provides large-sample approximations to the central moments. These are used to obtain the sampling distributions of $\hat{\tau}$ and $\hat{\kappa}$.

2.1 Testing for Skewness

We first derive the limiting distribution of the sample skewness coefficient under arbitrary τ (not necessarily 0), and then specialize the general result to $\tau = 0$. Throughout, we assume that the central limit theorem holds for the 4×1 vector series $\mathbf{W}_t' = [X_t - \mu, (X_t - \mu)^2 - \sigma^2, (X_t - \mu)^3 - \mu_3, (X_t - \mu)^4 - \mu_4]$ $(t = 1, \dots, T)$. This requires finite $(8 + \delta)$th $(\delta > 0)$ moment and some mixing conditions. When testing symmetry, finite $(6 + \delta)$th moment and some mixing conditions are sufficient.

Theorem 1. Suppose that X_t is stationary up to sixth order. Then

$$\sqrt{T}(\hat{\tau} - \tau) = \frac{\boldsymbol{\alpha}}{\hat{\sigma}^3}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_t + o_p(1),$$

where

$$\boldsymbol{\alpha} = [\,1 \quad -3\sigma^2 \quad -\tfrac{3\sigma\tau}{2}\,],$$

$$\mathbf{Z}_t = \begin{bmatrix} (X_t - \mu)^3 - \mu_3 \\ (X_t - \mu) \\ (X_t - \mu)^2 - \sigma^2 \end{bmatrix},$$

and $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_t \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma} = \lim_{T\to\infty} TE(\bar{\mathbf{Z}}\bar{\mathbf{Z}}')$, with $\bar{\mathbf{Z}}$ being the sample mean of $\mathbf{Z}_t$. Moreover,

$$\sqrt{T}(\hat{\tau} - \tau) \xrightarrow{d} \mathrm{N}\left(0, \frac{\boldsymbol{\alpha}\boldsymbol{\Gamma}\boldsymbol{\alpha}'}{\sigma^6}\right).$$

Serial dependence in X_t is explicitly taken into account through $\boldsymbol{\Gamma}$, the spectral density matrix at frequency 0 of $\mathbf{Z}_t$. The foregoing result thus permits testing the skewness coefficient at any arbitrary value of τ, even when the data are serially correlated. Notice that $\mathbf{Z}_t$ is three-dimensional, not one-dimensional. This is because the population mean and variance are unknown and must be estimated, as is typically the case in practice.

In the special case when $\tau = 0$ (or, equivalently, $\mu_3 = 0$), the last element of α is 0. Thus one need only consider the sampling properties of

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathbf{Z}_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{bmatrix} (X_t - \mu)^3 \\ (X_t - \mu) \end{bmatrix}. \tag{3}$$

This leads to the following result.

Corollary 1. Under the null hypothesis that $\tau = 0$,

$$\sqrt{T}\hat{\tau} \xrightarrow{d} \mathrm{N}\left(0, \frac{\boldsymbol{\alpha}_2\boldsymbol{\Gamma}_{22}\boldsymbol{\alpha}_2'}{\sigma^6}\right), \tag{4}$$

where $\boldsymbol{\alpha}_2 = [1, -3\sigma^2]$ and $\boldsymbol{\Gamma}_{22}$ is the first 2×2 block matrix of $\boldsymbol{\Gamma}$.

Similarly, one can easily show that under $\tau = \mu_3 = 0$,

$$\sqrt{T}\hat{\mu}_3 \xrightarrow{d} \mathrm{N}(0, \boldsymbol{\alpha}_2\boldsymbol{\Gamma}_{22}\boldsymbol{\alpha}_2').$$

The only difference between the limiting distributions of $\hat{\mu}_3$ and $\hat{\tau}$ is the normalizing constant σ^6. If the asymptotic standard deviations are estimated by $s(\hat{\mu}_3) = (\hat{\boldsymbol{\alpha}}_2\hat{\boldsymbol{\Gamma}}_{22}\hat{\boldsymbol{\alpha}}_2')^{1/2}$ and

$s(\hat{\tau}) = (\hat{\boldsymbol{\alpha}}_2\hat{\boldsymbol{\Gamma}}_{22}\hat{\boldsymbol{\alpha}}_2'/\hat{\sigma}^6)^{1/2}$, then we have the numerical identity, $\hat{\mu}_3/s(\hat{\mu}_3) = \hat{\tau}/s(\hat{\tau})$. We summarize the foregoing results in the following theorem.

Theorem 2. Suppose that X_t is stationary up to sixth order and let $\hat{\boldsymbol{\alpha}}_2 = [1, -3\hat{\sigma}^2]$. Let $\hat{\sigma}^2$ and $\hat{\boldsymbol{\Gamma}}_{22}$ be consistent estimates of σ^2 and $\boldsymbol{\Gamma}_{22}$. Let $s(\hat{\mu}_3) = (\hat{\boldsymbol{\alpha}}_2\hat{\boldsymbol{\Gamma}}_{22}\hat{\boldsymbol{\alpha}}_2')^{1/2}$ and $s(\hat{\tau}) = (\hat{\boldsymbol{\alpha}}_2\hat{\boldsymbol{\Gamma}}_{22}\hat{\boldsymbol{\alpha}}_2'/\hat{\sigma}^6)^{1/2}$. Then, under the null hypothesis of $\tau = \mu_3 = 0$, we have

$$\hat{\pi}_3 = \frac{\sqrt{T}\hat{\mu}_3}{s(\hat{\mu}_3)} = \frac{\sqrt{T}\hat{\tau}}{s(\hat{\tau})} \xrightarrow{d} \mathrm{N}(0, 1).$$

That is, $\hat{\mu}_3$ and $\hat{\tau}$ are the same. To construct $\hat{\pi}_3$, one need only a consistent estimate of this long-run variance, which can be obtained nonparametrically by kernel estimation. The test is valid even if the null distribution is not normally distributed, albeit symmetric. Simple calculations show that if X_t is iid normal, then the variance of $\hat{\tau}$ is 6.

Depending on the distribution under investigation, a large number of observations might be required to detect symmetry. The possibility of low power can be remedied in two ways. The first is to exploit the fact that most economic time series are bounded below by 0. Hence one can test symmetry against positive skewness. Second, the odd moments of symmetric distributions are 0, if they exist. Therefore, in theory one can construct a joint test of several odd moments to increase power. To illustrate, consider a joint test of two odd moments, r_1 and r_2. Let

$$\mathbf{Y}_T = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T (X_t - \bar{X})^{r_1} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T (X_t - \bar{X})^{r_2} \end{pmatrix}.$$

By Lemma A.1 in Appendix A, we can show that

$$\mathbf{Y}_T = \boldsymbol{\alpha}\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{Z}_t + o_p(1)$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} 1 & 0 & -r_1\mu_{r_1-1} \\ 0 & 1 & -r_2\mu_{r_2-1} \end{bmatrix} \quad \text{and} \quad \mathbf{Z}_t = \begin{pmatrix} (X_t - \mu)^{r_1} \\ (X_t - \mu)^{r_2} \\ (X_t - \mu) \end{pmatrix}.$$

Assuming that a central limit theorem holds for $\mathbf{Z}_t$ such that $\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{Z}_t \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma} = \lim_{T\to\infty} TE(\bar{\mathbf{Z}}\bar{\mathbf{Z}}')$, we have $\mathbf{Y}_T \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\alpha}\boldsymbol{\Gamma}\boldsymbol{\alpha}')$ under the null hypothesis of symmetry. Let $\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\alpha}}'$ be a consistent estimate for $\boldsymbol{\alpha}\boldsymbol{\Gamma}\boldsymbol{\alpha}'$ (which is easy to obtain); we then have

$$\hat{\mu}_{r_1,r_2} = \mathbf{Y}_T'(\hat{\boldsymbol{\alpha}}\hat{\boldsymbol{\Gamma}}\hat{\boldsymbol{\alpha}}')^{-1}Y_T \xrightarrow{d} \chi_2^2.$$

In principle, this is a more powerful test than the test based on the third moment alone. The cost is that this test requires the finiteness of $(2r_2)$th moment ($r_1 < r_2$). Furthermore, even if the population moments exist, precise estimates may be difficult to obtain unless we have an enormous number of observations. In the simulations, we consider only the test for $r_1 = 3$ and $r_2 = 5$, that is, $\hat{\mu}_{35}$. Interestingly, $\hat{\mu}_{35}$ performs well even for distributions that do not have finite fifth moment.

2.2 Testing for Kurtosis

Again we derive the limiting distribution of the estimated kurtosis under arbitrary true κ and then specialize it to $\kappa = 3$ under normality. By Lemma A.1 in Appendix A, we have the following result.

Theorem 3. Suppose that X_t is stationary up to eighth order. Then

$$\sqrt{T}(\hat{\kappa} - \kappa) = \frac{\boldsymbol{\beta}}{\hat{\sigma}^4}\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{W}_t + o_p(1),$$

where

$$\boldsymbol{\beta} = [\,1 \quad -4\mu_3 \quad -2\sigma^2\kappa\,], \qquad \mathbf{W}_t = \begin{bmatrix} (X_t - \mu)^4 - \mu_4 \\ (X_t - \mu) \\ (X_t - \mu)^2 - \sigma^2 \end{bmatrix}, \tag{5}$$

and $\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{W}_t \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega} = \lim_{T\to\infty} TE(\bar{\mathbf{W}}\bar{\mathbf{W}}')$. Let $\hat{\sigma}^2$ and $\hat{\boldsymbol{\Omega}}$ be consistent estimates of σ^2 and $\boldsymbol{\Omega}$. Then

$$\hat{\pi}_4(\kappa) = \frac{\sqrt{T}(\hat{\kappa} - \kappa)}{s(\hat{\kappa})} \xrightarrow{d} \mathrm{N}(0, 1),$$

where $s(\hat{\kappa}) = (\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Omega}}\hat{\boldsymbol{\beta}}'/\hat{\sigma}^8)^{1/2}$.

To test kurtosis, estimation of a three-dimensional long-run covariance matrix is necessary, as the population mean and variance are both unknown and must be estimated. Note that the first component of $\mathbf{W}_t$ depends on the fourth moment of $(X_t - \mu)^4$, which itself is a highly skewed random variable even if X_t is not skewed. The convergence to normality could be extremely slow, and the sample estimate of $\hat{\kappa}$ can deviate substantially from its true value even with a large number of observations. Thus for moderate sample sizes, the kurtosis test cannot be expected to be accurate. This is confirmed by simulations in the next section.

2.3 Connection With Generalized Method of Moments

We directly test the coefficients of skewness and kurtosis, and it is of interest to compare this with an overidentifying test on moments of the data. Consider testing symmetry by testing $\mu_3 = 0$ using GMM. Let $\bar{\mathbf{g}}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^T \mathbf{g}_t$, where $\boldsymbol{\theta} = (\mu, \sigma^2)$ and $\mathbf{g}_t$ is $\mathbf{Z}_t$ in Theorem 1 with $\mu_3 = 0$; that is,

$$\mathbf{g}_t(\boldsymbol{\theta}) = \begin{bmatrix} (X_t - \mu)^3 \\ (X_t - \mu) \\ (X_t - \mu)^2 - \sigma^2 \end{bmatrix}.$$

With three moments and two parameters, there is one overidentifying restriction. Let $\hat{\boldsymbol{\Gamma}}$ be a consistent estimator of $\boldsymbol{\Gamma} = \mathrm{var}(\sqrt{T}\bar{\mathbf{g}}_T(\boldsymbol{\theta}_0))$, the long-run variance of $\mathbf{g}_t$. The GMM estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}}_{\mathrm{GMM}} = \arg\min_{\boldsymbol{\theta}} \bar{\mathbf{g}}_T(\boldsymbol{\theta})'\hat{\boldsymbol{\Gamma}}\bar{\mathbf{g}}_T(\boldsymbol{\theta}).$$

Note that in general, $\hat{\boldsymbol{\theta}}_{\mathrm{GMM}}$ is not equal to $\hat{\boldsymbol{\theta}} = (\bar{X}, \hat{\sigma}^2)$, which is what we use in our tests. Indeed, in the present case, estimating μ by GMM entails solving a nonlinear equation. Define

$$\mathbf{J}_T = T\bar{\mathbf{g}}_T(\hat{\boldsymbol{\theta}}_{\mathrm{GMM}})'\hat{\boldsymbol{\Gamma}}\bar{\mathbf{g}}_T(\hat{\boldsymbol{\theta}}_{\mathrm{GMM}}),$$

where $\bar{\mathbf{g}}_T(\hat{\boldsymbol\theta}_{\text{GMM}})$ is $\bar{\mathbf{g}}_T(\boldsymbol\theta)$ evaluated at $\boldsymbol\theta = \hat{\boldsymbol\theta}_{\text{GMM}}$. Then $\mathbf{J}_T$ converges in distribution to χ_1^2. But it follows from $\hat{\boldsymbol\theta}_{\text{GMM}} \neq (\bar{X}, \hat\sigma^2)$ that $\bar{\mathbf{g}}_T(\hat{\boldsymbol\theta}_{\text{GMM}}) \neq (\frac{1}{T}\sum_{t=1}^T (X_t - \bar{X})^3, 0, 0)'$. Thus the $\mathbf{J}_T$ is not simply testing $\frac{1}{T}\sum_{t=1}^T (X_t - \bar{X})^3$, the third sample moment. GMM testing of overidentifying restrictions is thus different from testing the skewness coefficient.

However, under normality assumption and with the theoretical $\boldsymbol\Gamma$ as the weighting matrix, Richardson and Smith (1993) showed that the GMM estimator of $\boldsymbol\theta$ coincides with the sample mean and sample variance so that $\bar{\mathbf{g}}_T(\hat{\boldsymbol\theta}_{\text{GMM}}) = (\frac{1}{T} \times \sum_{t=1}^T (X_t - \bar{X})^3, 0, 0)'$. In this special case, the $\mathbf{J}_T$ test is equivalent to testing the skewness coefficient. This result is due to the fact that the odd moments of the normal distribution are 0, and thus the theoretical weighting matrix $\boldsymbol\Gamma$ has a block structure under normality. Nevertheless, even for normal distributions, when a consistent estimator for $\boldsymbol\Gamma$ (not $\boldsymbol\Gamma$ itself) is used, the $\mathbf{J}_T$ test will not be identical to testing the skewness coefficient. Similarly, a GMM test of the overidentifying restriction on the fourth moment (e.g., choosing $\mathbf{g}_t$ to be $\mathbf{W}_t$ in Thm. 3 with $\mu_4 = 3\sigma^4$) is not equivalent to testing the kurtosis coefficient.

Whereas the GMM test requires reestimation of the model parameters (μ, σ^2) every time that a different null hypothesis is tested (i.e., different null values of μ_3 and μ_4), our statistics do not require reestimation of these parameters. Our statistics directly test skewness and kurtosis coefficients. We simply construct the statistics once, and compare them with the null values of interest. Because our statistics are t-tests, we can also test one-sided hypotheses. The GMM test is a two-sided test.

2.4 Testing for Normality

The skewness and kurtosis tests developed herein can be used to test whether the data conform to any distribution of interest, provided that the theoretical coefficients of skewness and kurtosis are known. A case of general interest is the Gaussian distribution. Under normality, $\tau = 0$ and $\kappa = 3$. Let $\hat\pi_3$ be the test defined earlier for testing $\tau = 0$, and let $\hat\pi_4$ be the test statistic for kurtosis evaluated at $\kappa = 3$; that is, $\hat\pi_4 = \hat\pi_4(3)$. It can be shown that $\hat\pi_3$ and $\hat\pi_4$ are asymptotically independent under normality even for time series data (see the proof of Thm. 4 and discussion in Lomnicki 1961). Thus a direct generalization of the Jarque–Bera test to dependent data is

$$\hat\pi_{34} = \hat\pi_3^2 + \hat\pi_4^2 \xrightarrow{d} \chi_2^2.$$

An asymptotically equivalent test, based directly on the third and fourth central moments, can be constructed as follows. Let

$$\mathbf{Y}_T = \begin{pmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T (X_t - \bar{X})^3 \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T [(X_t - \bar{X})^4 - 3(\hat\sigma^2)^2] \end{pmatrix}.$$

Under normality, it can be shown that

$$\mathbf{Y}_T = \boldsymbol\gamma \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathbf{Z}_t + o_p(1),$$

$$\boldsymbol\gamma = \begin{bmatrix} -3\sigma^2 & 0 & 1 & 0 \\ 0 & -6\sigma^2 & 0 & 1 \end{bmatrix}, \quad \text{and}$$

$$\mathbf{Z}_t = \begin{bmatrix} (X_t - \mu) \\ (X_t - \mu)^2 - \sigma^2 \\ (X_t - \mu)^3 \\ (X_t - \mu)^4 - 3\sigma^4 \end{bmatrix}, \tag{6}$$

with $\frac{1}{\sqrt{T}}\sum_{t=1}^T \mathbf{Z}_t \xrightarrow{d} \text{N}(\mathbf{0}, \boldsymbol\Phi)$ and $\boldsymbol\Phi = \lim_{T\to\infty} TE(\bar{\mathbf{Z}}\bar{\mathbf{Z}}')$. Thus $\mathbf{Y}_T \xrightarrow{d} \text{N}(\mathbf{0}, \boldsymbol\gamma\boldsymbol\Phi\boldsymbol\gamma')$. Let $\hat{\boldsymbol\gamma}$ and $\hat{\boldsymbol\Phi}$ be consistent estimators of $\boldsymbol\gamma$ and $\boldsymbol\Phi$; then we have

$$\hat\mu_{34} = \mathbf{Y}_T'(\hat{\boldsymbol\gamma}\hat{\boldsymbol\Phi}\hat{\boldsymbol\gamma}')^{-1}\mathbf{Y}_T \xrightarrow{d} \chi_2^2.$$

Summarizing the foregoing results, we have the following.

Theorem 4. Suppose that X_t is stationary. Then, under the null hypothesis of normality,

$$\hat\pi_{34} \xrightarrow{d} \chi_2^2$$

and

$$\hat\mu_{34} \xrightarrow{d} \chi_2^2.$$

It is notable from Theorem 4 that both two-dimensional tests necessitate estimation of a four-dimensional long-run covariance matrix. In contrast, the one-dimensional tests in Theorems 2 and 3 necessitate estimation of three-dimensional spectral density matrices at the frequency 0. These formulas define the vector series that should be used to construct a serial-correlation consistent covariance matrix. For testing normality, the covariance for the vector series $\mathbf{Z}_t$ can be estimated by the covariance of $\hat{\mathbf{Z}}_t = ((X_t - \bar{X}), (X_t - \bar{X})^2 - \hat\sigma^2, (X_t - \bar{X})^3, (X_t - \bar{X})^4 - 3\hat\sigma^4)'$. Providing these formulas is one contribution of this article.

Under the theoretical weighting matrix, the GMM test of Richardson and Smith (1993) can be written in our notation as $\mathbf{Y}_T'\boldsymbol\Phi^{22}\mathbf{Y}_T$, where $\boldsymbol\Phi^{22}$ is the second diagonal block of the inverse of $\boldsymbol\Phi$. In actual implementation, the unknown $\boldsymbol\Phi^{22}$ must be replaced by a consistent estimator. When the weighting matrix is not the theoretical one, then, as explained in Section 2.3, the optimal GMM estimator for (μ, σ^2) is not the sample mean and sample variance. The GMM test is thus not a function of $\mathbf{Y}_T$ (or, equivalently, skewness and kurtosis coefficients). In contrast, our test $\hat\mu_{34}$ is of the form $\mathbf{Y}_T'(\hat{\boldsymbol\gamma}'\hat{\boldsymbol\Phi}\hat{\boldsymbol\gamma})^{-1}\mathbf{Y}_T$, which is always a function of skewness and kurtosis. It is thus a direct generalization of the Jarque–Bera test to dependent data. The GMM test and $\hat\mu_{34}$ are asymptotically equivalent under the null of normality, but will have different power properties under the alternative.

2.5 Testing Regression Residuals

We have thus far referred to our tests as unconditional tests, in the sense that apart from the sample mean, no other information is being removed before testing. In earlier work (Bai and Ng 2001), we discussed how to test whether the regression residuals are symmetrically distributed. The test is based on martingale transformations. It has very good finite-sample properties, but its computation is more demanding. In view of the simplicity of the unconditional tests for skewness, kurtosis, and normality just proposed, one might wonder whether the same tests can be applied to regression residuals, and, if so, to what extent does estimation of the parameters of the regression function affect the sampling distribution of the tests. Consider the regression model

$$y_t = \alpha + \mathbf{x}_t'\boldsymbol{\beta} + e_t,$$

where e_t is a sequence of mean-0, serially correlated disturbances. Lutkepohl (1993) considered autoregressive models ($\mathbf{x}_t$ is the lag of y_t) with e_t being iid, and Kilian and Demiroglu (2000) considered autoregressions with integrated variables. To ensure consistency of least squares estimators for general serially correlated disturbances, we assume that e_t is independent of $\mathbf{x}_s$ for all t and s. We also assume that a central limit theorem is valid for the vector $(e_t, e_t^2 - \sigma_e^2, e_t^3 - Ee_t^3, e_t^4 - Ee_t^4)$. The objective is to test hypotheses on the skewness coefficient, the kurtosis, and the normality of e_t. Because e_t is unobservable, we use $\hat{e}_t$ instead, where $\hat{e}_t = y_t - \hat{\alpha} - \mathbf{x}_t'\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are the least squares estimators of α and $\boldsymbol{\beta}$.

Theorem 5. Assume that $E\|\mathbf{x}_t\|^4 \le M$ for all t, and $\frac{1}{T} \times \sum_{t=1}^{T} \mathbf{x}_t\mathbf{x}_t' \xrightarrow{p} Q > 0$. Let $\hat{\tau}$, $\hat{\pi}_3$, $\hat{\pi}_4(\kappa)$, $\hat{\pi}_{34}$, and $\hat{\mu}_{34}$ be constructed as in previous sections with $\mathbf{X}_t$ replaced by $\hat{e}_t$ $(t = 1, \dots, T)$. Assume that e_t is stationary up to the eighth order; then Theorems 1–3 hold. In addition, under normality of e_t, Theorem 4 also holds.

Theorem 5 is a consequence of the following:

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{e}_t^j = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (e_t - \bar{e})^j + o_p(1), \qquad j = 2, 3, 4. \quad (7)$$

Result (7) implies the asymptotic equivalence of tests based on $\hat{e}_t$ and those on e_t. Note that the sum is divided by $\sqrt{T}$ rather than by T; thus (7) is much stronger than saying that the jth moment of the residuals is asymptotically the same as the jth moment of e_t. White and Macdonald (1980) proved (7) when $\sqrt{T}$ is replaced by T, and for iid disturbances.

The main implication of Theorem 5 is that inference on disturbances can be performed using the estimated residuals. Estimation of slope coefficients $\boldsymbol{\beta}$ does not affect the limiting distributions of the skewness coefficient, kurtosis, and normality tests. Indeed, Theorems 1–4 may be considered special cases of Theorem 5 in which the nonconstant regressors $\mathbf{x}_t$ are absent. Theorem 5 says that even if such regressors had been present, the limiting distributions of these test statistics would not have changed. Note, however, that when an intercept is not included in the regression, $\hat{e}_t$ in the foregoing equation should be replaced by $\hat{e}_t - \bar{\hat{e}}$.

3. SIMULATIONS

To assess the size and power of the tests, we consider well-known distributions, such as the normal, t, and chi-squared, as well as distributions from the generalized lambda family. This family encompasses a range of symmetric and asymmetric distributions that can be easily generated because they are defined in terms of the inverse of the cumulative distribution $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $0 < u < 1$. The λ parameters are taken from table 1 of Randles, Fligner, Policello, and Wolfe (1980). Specifically, data are generated from seven symmetric and eight skewed distributions, listed in Appendix B.

To evaluate the size of the test for skewness, we draw e_t from seven symmetric distributions. The power of the tests is assessed by considering eight asymmetric distributions. Because the kurtosis of all 15 distributions are known (and are given in Table 2), the size and power of $\hat{\pi}_4$ as well as the normality tests can be easily obtained.

The data used in the simulations are generated as $X_t = \rho X_{t-1} + e_t$. Many values of ρ were considered, and results are presented for $\rho = 0$, .5, and .8. The long-run covariance matrix is estimated by the kernel method with the truncation lag selected using the automatic procedure of Andrews (1991). We report results for the Bartlett kernel as discussed by Newey and West (1987). We also considered the Parzen window, where the weights are defined as $w(x) = 1 - 6x^2 + 6|x|^3$ if $0 \le |x| \le 1/2$, and $w(x) = 2(1 - |x|)^3$ if $1/2 \le |x| \le 1$. These additional results are available in the working version of this article, which are available from the authors on request.

3.1 Skewness

We consider both one-tailed and two-tailed tests for symmetry (denoted by $\hat{\pi}_3^*$ and $\hat{\pi}_3^{**}$). Results are reported at the 5% level without loss of generality. The critical values are 1.64 (one-tailed) and 1.96 (two-tailed). We also consider $\hat{\mu}_{35}$, a joint test of the third and fifth central moments, and the 5% critical value is 5.99. Three sample sizes are considered: $T = 100$, 200, and 500. Table 1 indicates that $\hat{\pi}_3$ has accurate size even for small T, but the $\hat{\mu}_{35}$ statistic rejects less often than π_3 under the null. The size of the tests are, however, quite robust to the degree of persistence in the data. It is useful to remark that in theory, the μ_{35} test requires the data to have tenth moments. Even though the t_5 distribution does not have sixth moment, the test still has good size. When a large number of observations are available for testing symmetry, the μ_{35} test can be a good complementary test to $\hat{\pi}_3$ because it is easy to implement.

The two-sided test has low power, but imposing a priori information on the direction of skewness leads to substantial power gains. All of the tests considered have low power for A4, A5, and A6 unless the sample size is large (say, more than 200 observations). The statistics developed by Randles et al. (1980) for testing symmetry in iid data also have low power for the same distributions, all of which have large kurtosis. In general, $\hat{\mu}_{35}$ has very good power even for T as small as 50. However, whereas the size of the test is quite robust to serial correlation in the data, the power function is quite sensitive to persistence. Results for $\rho = .8$ reveal that the power of the tests drops significantly when the degree of persistence increases. The reason for this is that for AR(1) models, $y_t = \sum_{j=1}^{t} \alpha^j u_{t-j}$ (assuming that $y_0 = 0$). In the limit when $\alpha = 1$, y_t (scaled by $\sqrt{T}$) is asymptotically normal. The data thus become more symmetric as persistence increases. This is confirmed on comparison with results for iid data, given in Table 1.

Table 1. Size and Power of the Test Symmetry for $\tau = 0$ (Newey–West kernel, no prewhitening)

			T = 100			T = 200			T = 500		
	τ	κ	$\hat{\pi}_3^{**}$	$\hat{\pi}_3^{*}$	$\hat{\mu}_{35}$	$\hat{\pi}_3^{**}$	$\hat{\pi}_3^{*}$	$\hat{\mu}_{35}$	$\hat{\pi}_3^{**}$	$\hat{\pi}_3^{*}$	$\hat{\mu}_{35}$
$\rho = 0$											
S1	0	3.0	.04	.05	.02	.05	.06	.03	.06	.06	.04
S2	0	9.0	.03	.05	.02	.04	.05	.03	.03	.04	.02
S3	0	2.5	.04	.05	.03	.05	.06	.04	.05	.06	.04
S4	0	3.0	.03	.04	.01	.04	.05	.03	.05	.05	.04
S5	0	6.0	.03	.04	.02	.04	.05	.02	.04	.06	.03
S6	0	11.6	.02	.04	.02	.04	.05	.02	.03	.04	.03
S7	0	126.0	.03	.04	.01	.02	.04	.02	.03	.04	.02
A1	6.18	113.9	.43	.63	.62	.52	.69	.81	.68	.81	.95
A2	2.0	9.0	.74	.88	.94	.90	.96	1.00	.98	.99	1.00
A3	2.0	9.0	.75	.89	.96	.91	.97	1.00	.98	.99	1.00
A4	.5	2.2	.85	.93	.64	1.00	1.00	.99	1.00	1.00	1.00
A5	1.5	7.5	.67	.85	.79	.85	.94	.99	.98	1.00	1.00
A6	2.0	21.2	.23	.43	.22	.43	.64	.55	.72	.85	.93
A7	3.16	23.8	.56	.76	.79	.73	.87	.96	.87	.94	1.00
A8	3.8	40.7	.53	.71	.76	.67	.82	.95	.83	.92	.99
$\rho = .5$											
S1	0	3.0	.03	.05	.03	.04	.05	.02	.05	.05	.04
S2	0	9.0	.03	.05	.02	.04	.05	.03	.04	.04	.03
S3	0	2.5	.02	.05	.01	.04	.06	.02	.04	.05	.03
S4	0	3.0	.03	.04	.02	.04	.05	.03	.05	.06	.04
S5	0	6.0	.03	.04	.01	.04	.06	.02	.05	.05	.03
S6	0	11.6	.03	.05	.02	.04	.05	.03	.04	.05	.03
S7	0	126.0	.04	.05	.01	.02	.04	.02	.03	.04	.03
A1	6.18	113.9	.43	.64	.62	.52	.71	.81	.68	.82	.94
A2	2.0	9.0	.62	.84	.77	.87	.94	.99	.97	1.00	1.00
A3	2.0	9.0	.65	.86	.78	.87	.95	1.00	.97	.99	1.00
A4	.5	2.2	.30	.54	.16	.70	.83	.48	.99	1.00	.97
A5	1.5	7.5	.44	.70	.32	.76	.90	.85	.96	.99	1.00
A6	2.0	21.2	.19	.38	.15	.37	.56	.38	.69	.82	.86
A7	3.16	23.8	.48	.72	.54	.67	.84	.90	.86	.94	.99
A8	3.8	40.7	.46	.67	.52	.64	.81	.92	.82	.92	.98
$\rho = .8$											
S1	0	3.0	.02	.05	.01	.04	.06	.01	.03	.04	.02
S2	0	9.0	.04	.06	.02	.04	.06	.02	.04	.05	.03
S3	0	2.5	.02	.04	.01	.03	.05	.01	.05	.06	.03
S4	0	3.0	.02	.03	.01	.04	.05	.04	.05	.05	.04
S5	0	6.0	.02	.04	0	.03	.04	0	.04	.05	.02
S6	0	11.6	.04	.06	.03	.03	.06	.03	.06	.06	.04
S7	0	126.0	.04	.05	.02	.04	.04	.02	.04	.05	.03
A1	6.18	113.9	.07	.20	.06	.29	.55	.25	.64	.79	.81
A2	2.0	9.0	.01	.05	.04	.13	.28	.17	.73	.88	.77
A3	2.0	9.0	.01	.06	.05	.12	.27	.16	.73	.89	.78
A4	.5	2.2	.06	.18	.03	.12	.25	.06	.32	.47	.17
A5	1.5	7.5	.04	.16	0	.20	.41	.05	.64	.83	.51
A6	2.0	21.2	.06	.22	.04	.15	.33	.11	.45	.68	.42
A7	3.16	23.8	.02	.11	.01	.18	.43	.07	.72	.88	.82
A8	3.8	40.7	.04	.15	.01	.22	.49	.14	.72	.88	.89

* One-sided test. ** Two-sided test.

From the simulations, $\hat{\mu}_{35}$ has a smaller probability of type I error and higher power and dominates $\hat{\pi}_3$. In principle, a joint test of more moments is feasible. But the higher-order moments are difficult to estimate precisely. Thus the $\hat{\mu}_{35}$ test is to be recommended when symmetry is the main concern.

There exist moment-free symmetry tests that do not require the existence of high moments. For example, the test of Bai and Ng (2001) is based on empirical distribution functions, whereas that of Fan and Gencay (1995) is based on nonparametrically estimated density functions. Even though these symmetry tests may have better theoretical properties than moment-based tests, moment-based tests such as the skewness coefficient are widely used by applied researchers. There is thus merit to improving our understanding of the properties of moment-based symmetry tests in a time series setting.

3.2 Kurtosis and Normality

Table 2 reports results for $\hat{\pi}_4(\kappa)$, which tests the true population value of κ, and $\hat{\pi}_4(3)$, which tests $\kappa = 3$ as would be the case under normality. There are two notable results. First, there are large size distortions, so that a large type I error could be expected if one were to test $\kappa = \kappa^0$. Second, although the test has power to reject $\kappa = 3$, the power is very low for the sample sizes considered. In many cases, serial correlation in the data further reduces the power of the test. For example, consider case A3 with $\kappa = 9$. With $T = 1{,}000$, the power is .90 when $\rho = 0$, falls to .84 when $\rho = .5$, and is only .14 when $\rho = .8$. One needs more than 5,000 observations to reject the hypothesis that $\kappa = 3$ when $\rho = .8$. For this reason, we report results for sample sizes much larger than for $\hat{\tau}$ to highlight the problems with testing for kurtosis in finite samples.

Table 2. Size and Power of the Test Kurtosis: Newey–West Kernel, No Prewhitening

	τ	κ	$T = 100$		$T = 200$		$T = 500$		$T = 1{,}000$		$T = 2{,}500$		$T = 5{,}000$	
			$\hat{\pi}_4(\kappa)$	$\hat{\pi}_4(3)$	$\hat{\pi}_4(\kappa)$	$\hat{\pi}_4(3)$	$\hat{\pi}_4(\kappa)$	$\hat{\pi}_4(3)$	$\hat{\pi}_4(\kappa)$	$\hat{\pi}_4(3)$	$\hat{\pi}_4(\kappa)$	$\hat{\pi}_4(3)$	$\hat{\pi}_4(\kappa)$	$\hat{\pi}_4(3)$
$\rho = 0$														
S1	0	3.0	.02	.02	.02	.02	.02	.02	.02	.02	.01	.01	.01	.01
S2	0	9.0	.68	.02	.74	.13	.69	.53	.64	.70	.61	.82	.58	.88
S3	0	2.5	.01	.23	.01	.44	.01	.80	.01	.96	0	1.00	0	1.00
S4	0	3.0	.02	.02	.03	.03	.01	.01	.01	.01	.01	.01	.01	.01
S5	0	6.0	.35	.03	.37	.16	.37	.66	.32	.87	.25	.94	.24	.99
S6	0	11.6	.60	.08	.69	.28	.66	.63	.58	.73	.53	.84	.46	.89
S7	0	126.0	.43	.10	.48	.28	.37	.51	.25	.58	.11	.67	.05	.70
A1	6.18	113.9	.15	.19	.32	.28	.69	.41	.85	.49	.78	.60	.75	.67
A2	2.0	9.0	.24	.12	.29	.38	.29	.75	.24	.92	.17	.98	.15	1.00
A3	2.0	9.0	.25	.13	.29	.41	.27	.73	.26	.90	.16	.97	.15	1.00
A4	.5	2.2	.02	.52	.01	.81	.01	.99	.02	1.00	.01	1.00	.02	1.00
A5	1.5	7.5	.36	.04	.38	.20	.34	.65	.31	.84	.23	.97	.18	.99
A6	2.0	21.2	.61	.07	.81	.25	.77	.52	.74	.63	.67	.75	.62	.81
A7	3.16	23.8	.34	.16	.57	.34	.66	.57	.58	.66	.51	.82	.46	.88
A8	3.8	40.7	.30	.16	.56	.34	.78	.51	.73	.60	.67	.75	.61	.79
$\rho = .5$														
S1	0	3.0	.01	.01	.01	.01	.01	.01	.01	.01	0	0	0	0
S2	0	9.0	.70	0	.84	.03	.85	.22	.81	.56	.84	.80	.81	.86
S3	0	2.5	0	.06	0	.07	0	.19	.02	.28	.10	.62	.40	.92
S4	0	3.0	.02	.02	.01	.01	.01	.01	.01	.01	0	0	0	0
S5	0	6.0	.35	0	.49	0	.59	.06	.61	.50	.69	.90	.78	.98
S6	0	11.6	.66	.01	.82	.07	.83	.43	.80	.68	.80	.84	.79	.88
S7	0	126.0	.48	.03	.62	.11	.60	.44	.54	.57	.37	.66	.28	.70
A1	6.18	113.9	.29	.09	.42	.24	.76	.39	.91	.49	.91	.60	.88	.68
A2	2.0	9.0	.46	.03	.58	.14	.67	.60	.70	.87	.78	.97	.85	1.00
A3	2.0	9.0	.44	.02	.56	.13	.63	.57	.71	.83	.78	.96	.86	.99
A4	.5	2.2	0	.12	0	.19	.01	.38	.13	.68	.76	.98	1.00	1.00
A5	1.5	7.5	.43	0	.52	.01	.61	.20	.65	.63	.72	.95	.78	.98
A6	2.0	21.2	.63	.01	.84	.05	.90	.33	.87	.59	.86	.75	.83	.80
A7	3.16	23.8	.42	.01	.65	.09	.83	.45	.80	.62	.80	.80	.79	.88
A8	3.8	40.7	.38	.03	.63	.11	.91	.44	.89	.59	.87	.75	.83	.80
$\rho = .8$														
S1	0	3.0	.01	.01	.01	.01	0	0	0	0	0	0	0	0
S2	0	9.0	.53	.01	.77	0	.96	0	.95	.02	.96	.19	.97	.59
S3	0	2.5	0	.01	0	.01	0	.01	.02	.01	.20	.02	.75	.01
S4	0	3.0	.01	.01	.01	.01	0	0	0	0	0	0	0	0
S5	0	6.0	.28	0	.50	0	.81	0	.92	0	.96	.02	.99	.20
S6	0	11.6	.50	0	.79	.01	.96	.01	.97	.09	.96	.53	.97	.82
S7	0	126.0	.41	0	.62	.01	.83	.05	.86	.22	.79	.59	.75	.68
A1	6.18	113.9	.28	0	.48	.04	.72	.22	.88	.43	.98	.57	.97	.66
A2	2.0	9.0	.23	0	.65	0	.94	.03	.97	.16	.99	.64	1.00	.94
A3	2.0	9.0	.23	0	.66	.01	.91	.03	.98	.14	.99	.67	1.00	.94
A4	.5	2.2	0	0	0	0	.08	.01	.49	.02	.99	.03	1.00	.06
A5	1.5	7.5	.36	0	.63	0	.88	0	.95	.01	.97	.09	.99	.52
A6	2.0	21.2	.52	0	.76	0	.96	.03	.97	.13	.97	.55	.96	.75
A7	3.16	23.8	.32	0	.60	0	.90	.01	.96	.16	.97	.67	.97	.83
A8	3.8	40.7	.26	0	.59	0	.91	.07	.96	.28	.98	.69	.98	.77

To understand the properties of the kurtosis tests, Table 3 reports the average estimates of κ and τ at the different sample sizes. Three results are noteworthy. First, both $\hat{\tau}$ and $\hat{\kappa}$ are generally downward biased, with biases increasing in ρ. However, the biases are substantially larger for $\hat{\kappa}$. Second, even with T as large as 5,000, κ cannot be estimated precisely from serially correlated data. In some cases, $\hat{\kappa}$ is severely biased for even iid data (see, e.g., case A1). This result has the important empirical implication that the sample kurtosis measure is generally unreliable and should always be viewed with caution. Third, the one exception when $\hat{\kappa}$ can be well estimated is when $\kappa = 3$. This is important in interpreting the results for testing normality.

Table 4 reports the results for testing normality. Except for the first row, which is based on the normal distribution and thus indicates size, all other rows indicate power. The $\hat{\mu}_{34}$ test is generally more powerful than the $\hat{\pi}_{34}$ test. Because the kurtosis test has such low power, the results for normality by and large reflect the results for the tests for skewness. The tests have low power when a distribution is symmetric. We also considered prewhitening, as discussed by Andrews and Monahan (1992). Prewhitening lowers the power of the tests when in fact the data are not serially correlated, but raises power when the data are genuinely serially correlated. These additional results are available in the working version of this article.

The fact that the kurtosis test has large size distortions might appear problematic for the normality test. Interestingly, however, this is not the case. This is because $\hat{\kappa}$ precisely estimates κ when $\kappa = 3$, and size distortion is not an issue. Thus, although the kurtosis test is itself not very useful per se, we can still use it to test normality, as was done by Bera and Jarque (1981). It is interesting to note that although the fourth moment

Table 3. Sample Estimates of τ and κ

	τ	100	200	500	1,000	2,500	5,000	κ	100	200	500	1,000	2,500	5,000
$\rho = 0$														
S1	0	.01	−0	−0	0	0	−0	3.0	2.95	2.95	3.00	2.99	3.00	3.00
S2	0	−.01	.02	−.02	.01	−.02	−.02	9.0	5.38	6.11	6.66	7.88	7.87	8.12
S3	0	.01	0	0	0	0	0	2.5	2.47	2.49	2.49	2.50	2.50	2.50
S4	0	−.01	−0	0	−0	−0	−0	3.0	2.96	2.97	2.99	2.99	3.00	3.00
S5	0	−.02	−.02	0	.01	−0	0	6.0	4.95	5.39	5.63	5.75	5.90	5.88
S6	0	.03	.02	.02	−0	−.01	−.01	11.6	6.70	7.66	8.92	9.80	10.71	11.09
S7	0	.03	−.04	.06	−.01	−.03	−.08	126.0	8.96	12.49	17.62	19.64	28.66	35.93
A1	6.18	3.15	3.81	4.44	4.82	5.34	5.71	113.9	17.24	25.71	37.63	46.55	63.72	79.20
A2	2.0	1.78	1.90	1.95	1.97	1.99	1.99	9.0	7.14	8.05	8.47	8.70	8.91	8.93
A3	2.0	1.76	1.89	1.98	1.97	2.00	1.99	9.0	6.98	7.94	8.85	8.70	9.02	8.91
A4	.5	.48	.50	.51	.51	.51	.51	2.2	2.19	2.21	2.22	2.22	2.22	2.22
A5	1.5	1.33	1.43	1.47	1.49	1.51	1.52	7.5	5.86	6.53	6.90	7.16	7.30	7.42
A6	2.0	1.34	1.49	1.72	1.80	1.87	1.97	21.2	8.30	10.00	12.66	14.14	15.74	18.68
A7	3.16	2.33	2.60	2.84	3.01	3.11	3.13	23.8	10.81	13.64	16.98	19.83	21.92	22.58
A8	3.8	2.58	2.89	3.21	3.56	3.66	3.72	40.7	12.65	16.43	20.86	27.93	31.22	32.50
$\rho = .5$														
S1	0	0	−0	−0	0	0	0	3.0	2.91	2.94	2.98	2.99	2.99	3.00
S2	0	−.02	.02	−.02	.01	−.01	−.01	9.0	4.31	4.81	5.16	5.91	5.88	6.09
S3	0	0	.01	0	0	−0	−0	2.5	2.63	2.67	2.68	2.70	2.70	2.70
S4	0	−.01	−.01	0	0	−0	−0	3.0	2.91	2.94	2.98	2.99	3.00	3.00
S5	0	−.02	−.01	−0	0	−0	−0	6.0	4.05	4.34	4.56	4.64	4.75	4.73
S6	0	.03	.01	.01	−0	−.01	−.01	11.6	5.06	5.72	6.51	7.08	7.61	7.87
S7	0	.02	−.05	.05	−0	−.03	−.06	126.0	6.43	8.57	11.65	12.91	18.27	22.72
A1	6.18	2.23	2.77	3.27	3.57	3.95	4.24	113.9	10.92	16.11	23.49	29.03	39.22	48.85
A2	2.0	1.21	1.35	1.42	1.45	1.47	1.47	9.0	5.14	5.80	6.18	6.36	6.51	6.52
A3	2.0	1.21	1.35	1.45	1.45	1.48	1.47	9.0	5.09	5.76	6.39	6.38	6.57	6.52
A4	.5	.35	.36	.37	.38	.38	.38	2.2	2.43	2.47	2.51	2.52	2.53	2.53
A5	1.5	.96	1.05	1.09	1.10	1.12	1.13	7.5	4.53	5.04	5.32	5.45	5.55	5.65
A6	2.0	.99	1.09	1.27	1.34	1.39	1.46	21.2	6.00	7.10	8.72	9.65	10.68	12.42
A7	3.16	1.65	1.90	2.10	2.22	2.31	2.32	23.8	7.31	9.17	11.29	13.01	14.33	14.68
A8	3.8	1.85	2.09	2.36	2.62	2.71	2.75	40.7	8.44	10.67	13.54	17.74	19.83	20.60
$\rho = .8$														
S1	0	−0	0	−0	−0	0	0	3.0	2.80	2.90	2.95	2.99	2.98	2.99
S2	0	−.01	.01	−.01	.01	−.01	−.01	9.0	3.21	3.53	3.72	4.02	4.03	4.12
S3	0	0	.01	0	0	0	0	2.5	2.71	2.79	2.84	2.88	2.88	2.89
S4	0	−.01	−.02	0	0	−0	0	3.0	2.78	2.89	2.95	2.98	2.99	3.00
S5	0	−.01	0	−.01	0	−0	0	6.0	3.14	3.35	3.50	3.57	3.63	3.63
S6	0	.02	0	0	0	−0	−.01	11.6	3.48	3.81	4.18	4.44	4.64	4.78
S7	0	−.01	−.05	.03	.01	−.02	−.04	126.0	3.92	4.80	6.03	6.53	8.49	10.13
A1	6.18	.84	1.36	1.81	2.06	2.31	2.51	113.9	5.01	6.87	9.86	12.09	15.84	19.67
A2	2.0	.20	.49	.70	.79	.85	.86	9.0	3.71	3.90	4.07	4.19	4.25	4.27
A3	2.0	.20	.48	.73	.79	.85	.86	9.0	3.72	3.86	4.15	4.19	4.26	4.26
A4	.5	.32	.27	.25	.24	.23	.23	2.2	2.77	2.79	2.82	2.82	2.83	2.82
A5	1.5	.42	.52	.61	.64	.66	.67	7.5	3.17	3.46	3.73	3.82	3.90	3.97
A6	2.0	.56	.62	.76	.78	.83	.87	21.2	3.67	4.23	4.95	5.29	5.77	6.45
A7	3.16	.53	.88	1.14	1.27	1.35	1.37	23.8	4.07	4.85	5.76	6.47	7.09	7.20
A8	3.8	.67	.98	1.29	1.50	1.59	1.63	40.7	4.40	5.19	6.46	8.08	9.03	9.37

is less reliably estimated, the higher odd moments of symmetric distributions, which are 0, if they exist, appear to be well estimated. This is why the symmetry test $\hat{\mu}_{35}$ performs well even for random variables that do not have finite fifth moment. For nonsymmetric distributions, the bias in high moments often translates into better power for $\hat{\mu}_{35}$. In summary, whereas the performance of the kurtosis test raises concerns about the use of high moments in statistical testing, this concern is less important when the objective is symmetry. As also pointed out earlier, the behavior of the kurtosis test does not undermine the performance of the normality test.

4. EMPIRICAL APPLICATIONS

We applied our tests to 21 macroeconomic time series. Data for GDP, the GDP deflator, the consumption of durables, final sales, the consumption of nondurables, residential investment, and nonresidential investment are taken from the national accounts and are quarterly data. The unemployment rate, employment, M2, and CPI are monthly series. The 30-day interest rate and M2 are weekly data. (All data are taken from the Economic Time Series Page, available at www.economagic.com.) With the exception of the interest rate and the unemployment rate (for which we do not take logs), we take the first difference of the logarithm of the data before applying the tests. We also considered three exchange rates (in logged first differences), and the value as well as the equally weighted CRSP daily stock returns. These data are not transformed. The sample skewness and kurtosis coefficients for the 21 series are also reported, along with tests for skewness, kurtosis, and normality.

The first column of Table 5 reports tests for symmetry, that is, testing $\tau = 0$, with $\hat{\tau}$ given in the fourth column. Several aspects of the results are of note. Following Delong and Summers (1985), we fail to reject symmetry in output and industrial pro-

Table 4. Size and Power of the Test Normality: Newey–West Kernel, No Prewhitening

			$T = 100$		$T = 200$		$T = 500$		$T = 1{,}000$		$T = 2{,}500$		$T = 5{,}000$	
	τ	κ	$\hat{\pi}_{34}$	$\hat{\mu}_{34}$	$\hat{\pi}_{34}$	$\hat{\mu}_{34}$	$\hat{\pi}_{34}$	$\hat{\mu}_{34}$	$\hat{\pi}_{34}$	$\hat{\mu}_{34}$	$\hat{\pi}_{34}$	$\hat{\mu}_{34}$	$\hat{\pi}_{34}$	$\hat{\mu}_{34}$
$\rho = 0$														
S1	0	3.0	.05	.02	.09	.03	.08	.03	.09	.03	.07	.02	.06	.03
S2	0	9.0	.03	.02	.11	.09	.39	.32	.62	.57	.81	.76	.89	.80
S3	0	2.5	.27	.16	.58	.35	.90	.72	.98	.94	1.00	1.00	1.00	1.00
S4	0	3.0	.03	.02	.08	.03	.07	.03	.07	.03	.06	.03	.06	.02
S5	0	6.0	.06	.03	.13	.12	.53	.43	.82	.74	.94	.89	.98	.96
S6	0	11.6	.07	.06	.22	.16	.50	.45	.67	.61	.85	.77	.91	.83
S7	0	126.0	.09	.07	.20	.17	.37	.35	.55	.46	.64	.56	.71	.61
A1	6.18	113.9	.84	.48	.94	.55	.99	.69	1.00	.80	1.00	.84	1.00	.90
A2	2.0	9.0	.99	.71	1.00	.88	1.00	.98	1.00	1.00	1.00	1.00	1.00	1.00
A3	2.0	9.0	.99	.74	1.00	.88	1.00	.96	1.00	1.00	1.00	1.00	1.00	1.00
A4	.5	2.2	1.00	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
A5	1.5	7.5	.81	.60	1.00	.83	1.00	.98	1.00	.99	1.00	1.00	1.00	1.00
A6	2.0	21.2	.22	.27	.52	.53	.87	.76	.97	.86	1.00	.94	1.00	.94
A7	3.16	23.8	.95	.59	.99	.75	1.00	.86	1.00	.92	1.00	.97	1.00	.98
A8	3.8	40.7	.94	.56	.99	.68	1.00	.82	1.00	.88	1.00	.94	1.00	.98
$\rho = .5$														
S1	0	3.0	.03	.01	.05	.01	.09	.02	.09	.02	.08	.01	.05	.02
S2	0	9.0	.02	.02	.05	.04	.20	.13	.48	.36	.76	.72	.87	.80
S3	0	2.5	.06	.02	.21	.05	.49	.11	.67	.20	.93	.44	1.00	.79
S4	0	3.0	.03	.01	.06	.02	.09	.02	.09	.02	.07	.02	.05	.01
S5	0	6.0	.01	.01	.04	.02	.22	.07	.59	.23	.90	.78	.98	.95
S6	0	11.6	.04	.03	.09	.07	.34	.24	.59	.55	.82	.74	.89	.83
S7	0	126.0	.04	.04	.11	.08	.33	.27	.50	.43	.62	.57	.70	.62
A1	6.18	113.9	.73	.48	.93	.57	.97	.69	.99	.80	.99	.85	1.00	.90
A2	2.0	9.0	.71	.51	1.00	.83	1.00	.97	1.00	1.00	1.00	1.00	1.00	1.00
A3	2.0	9.0	.71	.54	1.00	.84	1.00	.97	1.00	.99	1.00	1.00	1.00	1.00
A4	.5	2.2	.35	.21	.90	.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
A5	1.5	7.5	.21	.20	.83	.65	1.00	.95	1.00	.99	1.00	1.00	1.00	1.00
A6	2.0	21.2	.10	.11	.35	.36	.79	.73	.96	.84	1.00	.94	1.00	.95
A7	3.16	23.8	.45	.41	.97	.67	1.00	.85	1.00	.91	1.00	.97	1.00	.98
A8	3.8	40.7	.51	.41	.97	.64	1.00	.82	1.00	.88	1.00	.94	1.00	.98
$\rho = .8$														
S1	0	3.0	.01	0	.02	.01	.04	.01	.07	.02	.09	.01	.07	.01
S2	0	9.0	.01	0	.02	.01	.03	.02	.09	.03	.26	.12	.64	.36
S3	0	2.5	.01	.01	.04	.01	.08	.01	.14	.03	.23	.02	.31	.03
S4	0	3.0	.01	.01	.03	.01	.06	.02	.08	.02	.10	.02	.08	.02
S5	0	6.0	0	0	0	0	.02	.01	.06	.01	.25	.04	.67	.11
S6	0	11.6	.01	.01	.02	.02	.06	.04	.14	.08	.55	.32	.78	.67
S7	0	126.0	.01	.01	.03	.02	.09	.05	.22	.14	.53	.42	.64	.58
A1	6.18	113.9	.02	.03	.16	.30	.88	.66	.97	.77	.99	.85	.99	.90
A2	2.0	9.0	.01	0	.07	.08	.56	.60	.99	.95	1.00	1.00	1.00	1.00
A3	2.0	9.0	.02	0	.06	.06	.60	.61	.99	.96	1.00	1.00	1.00	1.00
A4	.5	2.2	.01	0	.06	.03	.31	.13	.70	.38	.97	.85	1.00	.99
A5	1.5	7.5	0	0	.04	.04	.46	.40	.93	.85	1.00	1.00	1.00	1.00
A6	2.0	21.2	.01	.01	.06	.05	.36	.35	.73	.69	.98	.92	1.00	.94
A7	3.16	23.8	0	0	.03	.07	.73	.64	1.00	.85	1.00	.95	1.00	.98
A8	3.8	40.7	0	.01	.05	.11	.86	.71	.99	.85	1.00	.94	1.00	.97

duction. However, although these authors find asymmetry in the unemployment rate, using a longer sample period and a different procedure to conduct inference, we find no evidence of skewness in the unemployment rate. The U.S.–Japan exchange rate, CPI inflation, and stock returns reject symmetry at 1% level. We also reject symmetry in manufacturing employment and the consumption of durable goods at the 10% level. Interestingly, series that exhibit skewness also failed our conditional symmetry test (Bai and Ng 2001).

The second column of Table 5 reports results for testing $\kappa = 3$ with $\hat{\kappa}$ given in the fifth column. We failed to find evidence of excess kurtosis in any of the real variables but found evidence of fat-tailed behavior in the two stock returns. These are financial series whose fat-tailed properties have been well documented. Our evidence is especially convincing in view of the lower power of the test for kurtosis reported earlier.

Results for the normality test are reported in the third column of Table 5. We reject normality in the U.S.–Japan exchange rate, unemployment rate, CPI inflation, 30-day interest rate, and two stock return series. With the exception of the 30-day interest rate, which failed the kurtosis test but not the skewness test, series that exhibit non-Gaussian behavior all failed the symmetry test. This accords with our observation that the power of the normality test is derived from asymmetry.

5. CONCLUDING COMMENTS

The goal of this article was to obtain tests for skewness, kurtosis, and a joint test of normality suited for time series observations. Our results depend only on stationarity and the existence of some moments. Monte Carlo simulations accord with our prior that tests for kurtosis will have low power because of the

Table 5. Macroeconomic Data

Sample	Series	$\hat{\pi}_3$	$\hat{\pi}_4$	$\hat{\pi}_{34}$	$\hat{\tau}$	$\hat{\kappa}$
71:1–97:12	Canada–U.S. exchange rate	1.505	.359	2.438	.226	3.139
71:1–97:12	German–U.S. exchange rate	−.682	.915	1.641	−.134	3.499
71:1–97:12	Japan–U.S. exchange rate	−2.532	1.567	7.444	−.481	3.905
48:1–97:12	Unemployment rate	.913	1.323	4.853	.308	8.621
46:1–97:12	Industrial production	1.195	1.417	2.305	.994	13.274
59:1–97:4	Inflation (GDP)	2.108	.188	4.529	.870	3.284
59:1–97:4	GDP	−1.420	1.335	2.356	−.561	4.717
47:1–97:12	Inflation (CPI)	2.618	1.484	6.858	.942	4.491
81:10:30–96:05:10	30-day interest rate	−.376	1.814	4.084	−.415	11.861
80:11:03–98:01:19	M2	−.096	.278	.202	−.017	3.116
59:3–96:4	Consumption durables	−1.858	1.722	3.499	−.791	5.023
59:3–96:4	Consumption nondurables	.596	1.341	3.623	.212	4.721
46:1–96:11	Employment	−1.590	1.659	3.761	−.280	3.733
49:3–97:4	Investment	−1.510	1.305	2.379	−.732	5.254
46:1–97:12	Manufacturing employment	−1.890	1.845	3.580	−1.644	10.543
46:1–97:12	Nonmanufacturing employment	.387	1.317	3.771	.117	5.857
59:3–97:4	Final sales	−.035	1.421	2.688	−.017	5.404
59:3–97:4	Nonresidential investment	−1.654	.828	2.775	−.409	3.680
59:3–97:4	Residential investment	−1.019	1.240	2.058	−.457	5.134
90:01:02–96:12:31	Stock returns (V)	−2.747	3.063	10.219	−.481	5.187
90:01:02–96:12:31	Stock returns (E)	−3.782	2.632	16.349	−.990	6.943

NOTE: The 5% critical values are 1.96 for π_3 and π_4 and 5.99 for π_{34}.

high moments involved. In finite samples, the test for kurtosis has poor size. The difficulty in estimating kurtosis does not pose a size problem for normality tests, however. Combining the coefficient of skewness with kurtosis as done by Bera and Jarque is still useful for time series data, once the limiting variance takes into account serial correlation in the data. Nonetheless, the primary source of power in the test for normality is derived from the test for skewness.

We first presented tests for the unconditional moment properties of the data, and then showed that when the tests are applied to regression residuals, the limiting distributions of the conditional tests are the same as those of the unconditional tests. It should be clear that unconditional symmetry and conditional symmetry generally do not imply each other. Consider an extreme example, $X_t = \epsilon_t - \epsilon_{t-1}$, where the ϵ_t are iid. Whether or not ϵ_t has a symmetric distribution, X_t is always symmetric, because X_t and $-X_t$ have the same distribution. However, conditional on the information at $t-1$ (which include ϵ_{t-1}), the conditional distribution of X_t will be asymmetric provided that ϵ_t is asymmetric.

ACKNOWLEDGMENTS

We would like to thank two annonymous referees and the editor for constructive comments. The authors acknowledge financial support from the NSF (grants SBR 9896329, SES-0136923, and SES-0137084).

APPENDIX A: PROOFS

The following lemma is used in the proof of Theorem 1.

Lemma A.1. Suppose that X_t is stationary up to order r for some $r \geq 2$. Then

$$\hat{\mu}_r = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - \bar{X})^r$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - \mu)^r - r\mu_{r-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - \mu) + o_p(1).$$

Furthermore, by the delta method,

$$\sqrt{T}\big((\hat{\sigma}^2)^{r/2} - (\sigma^2)^{r/2}\big) = \frac{r}{2}(\sigma^2)^{r/2-1}\sqrt{T}[\hat{\sigma}^2 - \sigma^2] + o_p(1).$$

Proof. We show (without loss of generality) the derivations for $r = 3$. The generalization is immediate.

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - \bar{X})^3$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - \mu + \mu - \bar{X})^3$$

$$= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - \mu)^3 + 3\frac{1}{T} \sum_{t=1}^{T} (X_t - \mu)^2 \sqrt{T}(\mu - \bar{X})$$

$$+ 3(\mu - \bar{X})^2 \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (X_t - \mu) + \sqrt{T}(\mu - \bar{X})^3.$$

The last two terms are $o_p(1)$, because $\sqrt{T}(\bar{X} - \mu) = O_p(1)$. Finally, note that $\frac{1}{T}\sum_{t=1}^{T}(X_t - \mu)^2) = \mu_2 + o_p(1)$.

Proof of Theorem 1

$$\hat{\tau} - \tau = \frac{\hat{\mu}_3}{\hat{\sigma}^3} - \frac{\mu_3}{\sigma^3}$$

$$= \frac{\hat{\mu}_3 - \mu_3}{\hat{\sigma}^3} - \tau \frac{\hat{\sigma}^3 - \sigma^3}{\hat{\sigma}^3}$$

$$= \left[\frac{\frac{1}{T}\sum_{t=1}^{T}(X_t - \bar{X})^3 - \mu_3}{\hat{\sigma}^3}\right] - \tau\left[\frac{(\hat{\sigma}^2)^{3/2} - (\sigma^2)^{3/2}}{\hat{\sigma}^3}\right]$$

$$= \frac{1}{\hat{\sigma}^3}\left[\frac{1}{T}\sum_{t=1}^{T}\big((X_t - \mu)^3 - \mu_3\big)\right.$$

$$\left. - 3\left(\frac{1}{T}\sum_{t=1}^{T}(X_t - \mu)^2\right)(\bar{X} - \mu)\right]$$

$$
\begin{aligned}
&-\frac{3\tau\sigma(\hat\sigma^2-\sigma^2)}{2\hat\sigma^3}+o_p(1)\\
&=\frac{1}{\hat\sigma^3}\left[\frac{1}{T}\sum_{t=1}^T\big((X_t-\mu)^3-\mu_3\big)\right.\\
&\quad\left.-3\left(\frac{1}{T}\sum_{t=1}^T(X_t-\mu)^2\right)\left(\frac{1}{T}\sum_{t=1}^T X_t-\mu\right)\right]\\
&\quad-\frac{3\tau\sigma}{2T\hat\sigma^3}\left[\sum_{t=1}^T(X_t-\mu)^2-\sigma^2\right]+o_p(1)\\
&=\frac{1}{\hat\sigma^3}\left[1\quad -3\sigma^2\quad -\tfrac{3\sigma\tau}{2}\right]\frac{1}{T}\begin{bmatrix}\sum_{t=1}^T[(X_t-\mu)^3-\mu_3]\\ \sum_{t=1}^T(X_t-\mu)\\ \sum_{t=1}^T[(X_t-\mu)^2-\sigma^2]\end{bmatrix}\\
&\quad+o_p(1)\\
&\equiv\frac{\alpha}{\hat\sigma^3}\frac{1}{T}\sum_{t=1}^T Z_t+o_p(1).
\end{aligned}
$$

Proofs of Theorems 2 and 3

The proof follows from the same argument as Theorem 1.

Proof of Theorem 4

The convergence of $\hat\mu_{34}$ in distribution to χ_2^2 follows from the central limit theorem for Z_t as defined in (6). To see $\hat\pi_{34}\xrightarrow{d}\chi_2^2$, it suffices to argue that $\hat\pi_3$ and $\hat\pi_4$ are asymptotically independent, so that the squared value of $\hat\pi_3$ and $\hat\pi_4$ are also asymptotically independent. The limit of $\hat\pi_3$ is determined by the partial sums of $\mathbf{Z}_t$ [see (3)], whereas the limit of $\hat\pi_4$ is determined by the partial sums of $\mathbf{W}_t$ [see (5), with the second component of $\mathbf{W}_t$ dropped off because the second component of $\boldsymbol\beta$ is 0 under normality], where

$$
\mathbf{Z}_t=\begin{bmatrix}(X_t-\mu)^3\\(X_t-\mu)\end{bmatrix}\qquad\text{and}\qquad \mathbf{W}_t=\begin{bmatrix}(X_t-\mu)^4-\mu_4\\(X_t-\mu)^2-\sigma^2\end{bmatrix}.
$$

It is easy to verify that $E\mathbf{Z}_t\mathbf{W}_t'=0$ for all t. Moreover, $E\mathbf{Z}_{t-j}\mathbf{W}_t'=0$ for all j. To see this, using the projection theory for normal random variables, we can write $X_t-\mu=c(X_{t-j}-\mu)+\xi_t$ for a constant c, where ξ_t is a mean-0 normal random variable independent of X_{t-j}. Thus powers of $(X_t-\mu)$ can be expressed as those of $(X_{t-j}-\mu)$ and ξ_t. The desired result follows from $E\mathbf{Z}_{t-j}\mathbf{W}_{t-j}'=0$ for all j and from the property of ξ_t. The foregoing analysis shows that $\hat\pi_3$ and $\hat\pi_4$ are asymptotically uncorrelated. Because they are asymptotically normal, they are also asymptotically independent.

Proof of Theorem 5

It is sufficient to prove (7). By direct calculations, $\hat e_t=e_t-\bar e+(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)$. For $j=2$,

$$
\hat e_t^2=(e_t-\bar e)^2+2(e_t-\bar e)(\mathbf{x}_t-\bar{\mathbf{x}})(\hat{\boldsymbol\beta}-\boldsymbol\beta)+[(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)]^2.
$$

From $\sum_{t=1}^T\bar e(\mathbf{x}_t-\bar{\mathbf{x}})=0$, we have

$$
\begin{aligned}
&\frac{1}{\sqrt T}\sum_{t=1}^T\hat e_t^2\\
&\quad=\frac{1}{\sqrt T}\sum_{t=1}^T(e_t-\bar e)^2+2\frac{1}{T}\sum_{t=1}^T e_t(\mathbf{x}_t-\bar{\mathbf{x}})\sqrt T(\hat{\boldsymbol\beta}-\boldsymbol\beta)\\
&\qquad+\frac{1}{\sqrt T}\sum_{t=1}^T[(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)]^2.
\end{aligned}
$$

The last term is $O_p(T^{-1/2})$, because $\frac{1}{T}\sum_{t=1}^T\|\mathbf{x}_t\|^2=O_p(1)$ and $\|\hat{\boldsymbol\beta}-\boldsymbol\beta\|^2=O_p(T^{-1})$. The second term is also $O_p(T^{-1/2})$, because $\frac{1}{T}\sum_{t=1}^T e_t\mathbf{x}_t=O_p(T^{-1/2})$ and $\bar e\bar{\mathbf{x}}=O_p(T^{-1/2})$. For $j=3$,

$$
\begin{aligned}
\hat e_t^3&=(e_t-\bar e)^3-3(e_t-\bar e)^2(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)\\
&\quad+3(e_t-\bar e)[(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)]^2+[(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)]^3,
\end{aligned}
$$

and thus

$$
\begin{aligned}
\frac{1}{\sqrt T}\sum_{t=1}^T\hat e_t^3&=\frac{1}{\sqrt T}\sum_{t=1}^T(e_t-\bar e)^3\\
&\quad-3\frac{1}{T}\sum_{t=1}^T(e_t-\bar e)^2(\mathbf{x}_t-\bar{\mathbf{x}})'\sqrt T(\hat{\boldsymbol\beta}-\boldsymbol\beta)+r_T
\end{aligned}
$$

where r_T represents the remaining terms and it is easy to show that $r_T=O_p(T^{-1/2})$. The middle term requires extra argument. But

$$
\frac{1}{T}\sum_{t=1}^T(e_t-\bar e)^2(\mathbf{x}_t-\bar{\mathbf{x}})=\frac{1}{T}\sum_{t=1}^T e_t^2(\mathbf{x}_t-\bar{\mathbf{x}})+2\frac{1}{T}\sum_{t=1}^T e_t(\mathbf{x}_t-\bar{\mathbf{x}})\bar e.
$$

The second term is $O_p(T^{-1/2})$. The first term can be rewritten as $\frac{1}{T}\sum_{t=1}^T(e_t^2-\sigma^2)(\mathbf{x}_t-\bar{\mathbf{x}})$, which is also $O_p(T^{-1/2})$, given the assumptions on the series $(e_t^2-\sigma^2)$ and on $\mathbf{x}_t$. For $j=4$,

$$
\begin{aligned}
\hat e_t^4&=(e_t-\bar e)^4-4(e_t-\bar e)^3(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)\\
&\quad+6(e_t-\bar e)^2[(\mathbf{x}_t-\bar{\mathbf{x}})'(\hat{\boldsymbol\beta}-\boldsymbol\beta)]^2+O_p(T^{-3/2}),
\end{aligned}
$$

and thus

$$
\begin{aligned}
\frac{1}{\sqrt T}\sum_{t=1}^T\hat e_t^4&=\frac{1}{\sqrt T}\sum_{t=1}^T(e_t-\bar e)^4\\
&\quad-4\frac{1}{T}\sum_{t=1}^T(e_t-\bar e)^3(\mathbf{x}_t-\bar{\mathbf{x}})'\sqrt T(\hat{\boldsymbol\beta}-\boldsymbol\beta)+r_T,
\end{aligned}
$$

where $r_T=O_p(T^{-1/2})$ is very easy to argue. The middle term is also $O_P(T^{-1/2})$. This follows from expanding $(e_t-\bar e)^3$ and use $\frac{1}{T}\sum_{t=1}^T e_t^k(\mathbf{x}_t-\bar{\mathbf{x}})=o_p(1)$, for $k=1,2,3$. For example, from $\sum_{t=1}^T(\mathbf{x}_t-\bar{\mathbf{x}})=0$, $\frac{1}{T}\sum_{t=1}^T e_t^3(\mathbf{x}_t-\bar{\mathbf{x}})=\frac{1}{T}\sum_{t=1}^T(e_t^3-\mu_3)(\mathbf{x}_t-\bar{\mathbf{x}})=O_p(T^{-1/2})$, where $\mu_3=E(e_t^3)$ and $e_t^3-\mu_3$ is a mean-0 process.

APPENDIX B: DISTRIBUTIONS CONSIDERED

S1: $N(0, 1)$

S2: t_5

S3: $e_1 I(z \leq .5) + e_2 I(z > .5)$, where $z \sim U(0, 1)$, $e_1 \sim N(-1, 1)$, and $e_2 \sim N(1, 1)$

S4: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = .19754$, $\lambda_3 = .134915$, $\lambda_4 = .134915$

S5: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -.08$, $\lambda_4 = -.08$

S6: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = -.397912$, $\lambda_3 = -.16$, $\lambda_4 = -.16$

S7: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -.24$, $\lambda_4 = -.24$

A1: lognormal: $\exp(e)$, $e \sim N(0, 1)$

A2: χ^2_2

A3: exponential: $-\ln(e)$, $e \sim N(0, 1)$,

A4: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = 1.0$, $\lambda_3 = 1.4$, $\lambda_4 = .25$

A5: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -.0075$, $\lambda_4 = -.03$

A6: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -.1$, $\lambda_4 = -.18$

A7: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -.001$, $\lambda_4 = -.13$

A8: $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$, $\lambda_1 = 0$, $\lambda_2 = -1$, $\lambda_3 = -.0001$, $\lambda_4 = -.17$

Note that S1–S7 are symmetric distributions, and A1–A8 are asymmetric distributions.

[Received January 2003. Revised April 2004.]

REFERENCES

Andrews, D. W. K. (1991), "Heteroskedastic and Autocorrelation Consistent Matrix Estimation," *Econometrica*, 59, 817–854.

Andrews, D. W. K., and Monahan, J. (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, 60, 953–966.

Bai, J. (2003), "Testing Parameteric Conditional Distributions of Dynamic Models," *Review of Economics and Statistics*, 85, 531–549.

Bai, J., and Ng, S. (2001), "A Test for Conditional Symmetry in Time Series Models," *Journal of Econometrics*, 103, 225–258.

Bera, A., and Jarque, C. (1981), "Efficient Tests for Normality, Heteroskedasticty, and Serial Independence of Regression Residuals: Monte Carlo Evidence," *Economics Letters*, 7, 313–318.

Bontemps, C., and Meddahi, N. (2002), "Testing Normality: A GMM Approach," note, Université de Montreal.

Cabellero, R. J. (1992), "A Fallacy of Composition," *American Economic Review*, 82, 1279–1292.

Delong, J. B., and Summers, L. H. (1985), "Are Business Cycle Symmetrical," in *American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, pp. 166–178.

Dufour, J. M., Khalaf, L., and Beaulieu, M. C. (2002), "Exact Skewness Kurtosis Test for Multivariate Normality and Goodness of Fit in Multivariate Regression Regressions With Application to Asset Pricing Models," working paper, Cirano.

Fan, Y., and Gencay, R. (1995), "A Consistent Nonparametric Test of Symmetry in Linear Regression Models," *Journal of the American Statistical Association*, 90, 551–557.

Hamilton, J. (1989), "A New Approach to the Economic Analysis of Non-Stationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

Jarque, C., and Bera, A. (1980), "Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals," *Economics Letters*, 12, 255–259.

Kendall, M., and Stuart, A. (1969), *The Advanced Theory of Statistics*, New York: McGraw Hill.

Kilian, L., and Demiroglu, U. (2000), "Residuals-Based Tests for Normality in Autoregressions: Asymptotic Theory and Simulation Evidence," *Journal of Business & Economic Statistics*, 18, 40–50.

Lomnicki, Z. (1961), "Tests for Departure From Normality in the Case of Linear Stochastic Processes," *Metrika*, 4, 37–62.

Lutkepohl, H. (1993), *Introduction to Multiple Time Series*, Berlin: Springer-Verlag.

Neftci, S. N. (1984), "Are Economic Time-Series Asymmetric Over the Business Cycle," *Journal of Political Economy*, 92, 307–328.

Newey, W., and West, K. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Randles, R. H., Fligner, M. A., Policello, G. E., and Wolfe, D. A. (1980), "An Asymptotically Distribution-Free Test for Symmetry versus Asymmetry," *Journal of the American Statistical Association*, 75, 168–172.

Richardson, M., and Smith, T. (1993), "A Test for Multivariate Normality in Stock Returns," *Journal of Business*, 66, 295–321.

White, H., and Macdonald, G. M. (1980), "Some Large-Sample Tests for Nonnormality in the Linear Regression Model," *Journal of American Statistical Association*, 75, 16–28.