Macroeconomic Dynamics, 2010, Page 1 of 23. Printed in the United States of America.
doi:10.1017/S1365100509090543

IDENTIFYING NONLINEAR SERIAL DEPENDENCE IN VOLATILE, HIGH-FREQUENCY TIME SERIES AND ITS IMPLICATIONS FOR VOLATILITY MODELING

PHILLIP WILD AND JOHN FOSTER
University of Queensland

MELVIN J. HINICH
University of Texas at Austin

In this article, we show how tests of nonlinear serial dependence can be applied to high-frequency time series data that exhibit high volatility, strong mean reversion, and leptokurtotis. Portmanteau correlation, bicorrelation, and tricorrelation tests are used to detect nonlinear serial dependence in the data. Trimming is used to control for the presence of outliers in the data. The data that are employed are 161,786 half-hourly spot electricity price observations recorded over nearly a decade in the wholesale electricity market in New South Wales, Australia. Strong evidence of nonlinear serial dependence is found and its implications for time series modeling are discussed.

Keywords: Nonlinearity, Portmanteau Tests, Bicorrelation, Tricorrelation, Trimming, Deep Structure

1. INTRODUCTION

In this article, we show how tests of nonlinear serial dependence can be applied to high-frequency time series data that exhibit high volatility, strong mean reversion, and leptokurtotis. These are features typically found in financial, currency exchange, and commodity market data. Portmanteau correlation, bicorrelation, and tricorrelation tests are applied to detect nonlinear serial dependence. In addition to using these tests to discover the presence and extent of nonlinear serial dependence, we also use them to examine whether "stochastic volatility" models provide adequate characterizations of the data.

Large positive and negative deviations from the mean are often deemed to be a principal source of deviations from the Gaussian ideal in finance and energy economics. The empirical distributions of most finance and energy time

Address correspondence to: Phillip Wild, School of Economics, University of Queensland, St. Lucia, QLD 4072, Australia; e-mail: p.wild@uq.edu.au.

 1365-1005/10

series data when transformed to continuous compound rates of return have much fatter tails than the normal (Gaussian) distribution—an empirical feature called "leptokurtosis." A key consequence of this empirical feature is that the returns data have large fourth-order cumulants. For example, many NYSE stock rates of return have an excess kurtosis value larger than one [Hinich and Patterson (1989, 1995, 2005)]. Foreign exchange rates are very leptokurtic [Brooks and Hinich (1998)].

From a statistical perspective, the principal effect of leptokurtosis is to slow down significantly the rate of convergence of finite sample tests whose asymptotic distributions are based upon asymptotic normality. Because economic time series have lagged time dependence in the mean and variance, standard bootstrapping is not appropriate to determine the variance of finite sample statistics based on some averaging of the data.

One classical approach designed to improve the rate of convergence of test statistics so that we can have greater confidence in the use of standard asymptotic theory for estimation and inference is to trim back very high positive and negative rates of return in the time series data. This operation improves the finite sample performance of the test statistics. Trimming also allows us to discover whether an observed nonlinear generating mechanism is a deep structure phenomenon. For our purposes, the concept of "deep structure" can be viewed as arising if observed nonlinear serial dependence continues to arise after the effects of outliers have been removed. We conduct several trimming experiments to examine the impacts on nonlinear serial dependence test results.

The data we employ in this study are half-hourly wholesale spot electricity prices for the state of New South Wales, Australia, over the period from 7 December 1998 to 29 February 2008, producing a sample size of 161,786 observations.[1] There are two key stylized facts concerning spot price dynamics in this market. First, there is high volatility (i.e., a lot of price spikes), and second, there is strong mean-reverting behavior (volatility clustering followed by sustained periods of normality). The numerous spot price spikes act as outliers that produce significant deviations from the Gaussian distribution, producing the predominant empirical leptokurtotic feature of most high-frequency asset price data.

In Foster, Hinich, and Wild (2008), the extent and stability of a weekly cycle in this time series data were investigated. A major finding was that the mean properties of the spot price data had a weak weekly and daily periodicity. The most important periodicities were found to contain significant but imperfect signal coherence, suggesting that some wobble existed in the waveforms. This was determined by applying the randomly modulated periodicity (RMP) model introduced in Hinich (2000) and Hinich and Wild (2001) to the data. They suggested that the generating mechanism for an RMP process is likely to be nonlinear. Thus, a research question was posed: is the mechanism responsible for generating weekly data nonlinear, and if so, is it episodic in character? The rationale for episodic nonlinearity is that this type of behavior would seem to be required to generate the aforementioned stylized fact of strong mean reversion in spot electricity prices.

The finding that nonlinearity is present can rule out many classes of linear models as candidates for modeling spot price dynamics. This leads to the question of how spot price dynamics can be modeled in a way that can capture observed nonlinearity.

In the energy and finance literature, the notion of nonlinearity has been most commonly associated with *multiplicative nonlinearity* or *nonlinearity in variance*. In principle, such models can be defined as either *observation-driven* or *parameter-driven*. The main class of parametric models associated with the first category is autoregressive conditional heteroskedasticity (ARCH/GARCH) models [Engle (1982), Bollerslev (1986), and Taylor (1986)].[2] In these models, the conditional variance is postulated to depend on the variability of recent observations. The main class of parametric models that are associated with the second category is stochastic volatility models. In these models, volatility is postulated to be a function of some unobserved or latent stochastic process [Shephard (1996, pp. 6–7)].[3]

A key aspect of both ARCH/GARCH and stochastic volatility modeling frameworks is that the time series is assumed to be a zero-mean process. This implies that the mean of the source time series has to be removed, typically by some linear time series model. The residuals from this model then constitute the zero-mean data series that underpins theoretical discussion of volatility models. A potential problem emerges when the mean of the process is nonlinear. In this case, erroneous conclusions of nonlinearity in variance might emerge when the prime source of serial dependence is nonlinear structure in residuals that could not successfully be extracted by conventional linear-based time series models. This nonlinearity-in-mean structure would end up in the residuals of the fitted model and might subsequently trip ARCH tests even though the serial dependence structure was associated with the mean properties of the data-generating process.[4] Thus, volatility modeling might lead to situations involving the acceptance of linear specifications (with conditionally heteroscedastic disturbances) that actually constitute a misspecification of the actual process in statistical terms. Thus, it seems essential to discover as much as possible about the nature of any nonlinearity present in the time series data before techniques such as GARCH are applied.

The article is organized as follows. In Section 2 we outline the portmanteau correlation, bicorrelation, and tricorrelation tests proposed in Hinich (1996). These tests are used to test for second-order (linear) and third- and fourth-order (nonlinear) serial dependence, respectively. In this section, we also outline the test for the presence of pure ARCH and GARCH structures in the weekly spot price data using the well-known Engle (1982) ARCH LM test. In Section 3, the rationale for and practical aspects of the trimming procedure are outlined. In Section 4, the empirical results are presented, and concluding comments are offered in Section 5.

2. THE TESTING METHODOLOGY

In this section, it is explained how the testing methodology proposed in Hinich and Patterson (1995, 2005) is applied in the context of spot price dynamics. The objective is to detect epochs of transient serial dependence in a discrete-time pure white noise process.[5] This methodology involves computing the portmanteau correlation, bicorrelation, and tricorrelation test statistics (denoted as C, H, and $H4$ statistics, respectively) for each sample frame to detect linear and nonlinear serial dependence, respectively.

Before applying the various tests outlined in this article, we convert the source spot price data series to continuous compounded returns by applying the relationship

$$r(t) = \ln\left[\frac{x(t)}{x(t-1)}\right] \times 100, \tag{1}$$

where $r(t)$ is the continuous compounded return for time period t; and $x(t)$ is the source spot price time series data.

For each sample frame, the data are standardized using the relationship

$$Z(t) = \frac{r(t) - m_r}{s_r} \tag{2}$$

for each $t = 1, 2, \ldots, n$, where m_r and s_r are the sample mean and standard deviation of the sample frame and $r(t)$ are the returns data composing a sample frame of n observations. As such, the returns data in each sample frame are standardized on a frame-by-frame basis.

The null hypothesis for each sample frame is that the transformed data $\{Z(t)\}$ are realizations of a stationary pure white noise process. Therefore, under the null hypothesis, the correlations $C_{ZZ}(r) = E[Z(t)Z(t+r)] = 0$, for all $r \neq 0$, the bicorrelations $C_{ZZZ}(r, s) = E[Z(t)Z(t+r)Z(t+s)] = 0$, for all r, s except where $r = s = 0$, and the tricorrelations $C_{ZZZZ}(r, s, v) = E[Z(t)Z(t+r)Z(t+s)Z(t+v)] = 0$, for all $r, s,$ *and* v except where $r = s = v = 0$. The alternative hypothesis is that the process in the sample frame has some nonzero correlations, bicorrelations, or tricorrelations in the set $0 < r < s < v < L$, where L is the number of lags associated with the length of the sample frame. That is, either $C_{ZZ}(r) \neq 0$, $C_{ZZZ}(r, s) \neq 0$, *or* $C_{ZZZZ}(r, s, v) \neq 0$ for at least one r value or one pair of r and s values or one triple of r, s and v values, respectively.

The r sample correlation coefficient is defined as

$$C_{ZZ}(r) = \frac{1}{\sqrt{n-r}} \sum_{t=1}^{n-r} Z(t)Z(t+r). \tag{3}$$

The C statistic is designed to test for the existence of nonzero correlations (i.e., second-order linear dependence) within a sample frame, and its distribution is

$$C = \sum_{r=1}^{L} [C_{ZZ}(r)]^2 \approx \chi^2_L. \qquad \textbf{(4)}$$

The (r, s) sample bicorrelation coefficient is defined as

$$C_{ZZZ}(r,s) = \frac{1}{n-s} \sum_{t=1}^{n-s} Z(t)Z(t+r)Z(t+s), \quad \text{for } 0 \le r \le s. \qquad \textbf{(5)}$$

The H statistic is designed to test for the existence of nonzero bicorrelations (i.e., third-order nonlinear serial dependence) within a sample frame, and its corresponding distribution is

$$H = \sum_{s=2}^{L} \sum_{r=1}^{s-1} G^2(r,s) \approx \chi^2_{L(L-1)/2}, \qquad \textbf{(6)}$$

where $G(r,s) = \sqrt{n-s}\, C_{ZZZ}(r,s)$.

The (r, s, v) sample tricorrelation coefficient is defined as

$$C_{ZZZZ}(r,s,v) = \frac{1}{n-v} \sum_{t=1}^{n-v} Z(t)Z(t+r)Z(t+s)Z(t+v), \quad \text{for } 0 \le r \le s \le v. \qquad \textbf{(7)}$$

The $H4$ statistic is designed to test for the existence of nonzero tricorrelations (i.e., fourth-order nonlinear serial dependence) within a sample frame, and its corresponding distribution is

$$H4 = \sum_{v=3}^{L} \sum_{s=2}^{v-1} \sum_{r=1}^{s-1} T^3(r,s,v) \approx \chi^2_{L(L-1)(L-2)/3}, \qquad \textbf{(8)}$$

where $T(r,s,v) = \sqrt{n-v} \times C_{ZZZZ}(r,s,v)$.

In principle, the tests can be applied to either the source returns data determined from application of (2) or to the residuals from frame-based autoregressive AR(p) fits of these data, where p is the number of lags that is selected in order to remove significant C statistics at some prespecified threshold level. The latter is a prewhitening operation and can be used to effectively remove second-order (linear) serial dependence producing no significant C frames, thus allowing the investigator to focus on whether spot price data contain predictable nonlinearities after all linear dependence is removed. As such, the portmanteau bicorrelation and tricorrelation tests are applied to the residuals of the fitted AR(p) model of each sample frame. Any remaining serial dependence left in the residuals must be a consequence of nonlinearity that is episodically present in the data—thus, only significant H and $H4$ statistics will lead to the rejection of the null hypothesis of a pure white noise process.

The number of lags L is defined as $L = n^b$ with $0 < b < 0.5$ for the correlation and bicorrelation tests and $0 < b < 0.33$ for the tricorrelation test, where b is a parameter to be chosen by the user. Based on results of Monte Carlo simulations, Hinich and Patterson (1995, 2005) recommended the use of $b = 0.4$ (in relation to the bicorrelation test).[6] In this article, the data are split into a set of equal-length nonoverlapped moving frames of 336 half-hour observations, corresponding to a week's duration.

The correlation, bicorrelation, and tricorrelation tests can also be used to examine whether GARCH or stochastic volatility models represent adequate characterization of the data under investigation. We can define a GARCH(p, q) process as

$$y_t = \varepsilon_t h_t,\, \varepsilon_t \approx \mathrm{NIID}\big(0, h_t^2\big),\, h_t^2 = \alpha_0 + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}^2. \qquad \textbf{(9)}^7$$

We can similarly define a stochastic volatility model as

$$\begin{aligned} y_t &= \varepsilon_t \exp\left(h_t/2\right), \varepsilon_t \approx \mathrm{NIID}(0, 1), \\ h_{t+1} &= \alpha_0 + \alpha_1 h_t + \eta_t, \eta_t \approx \mathrm{NIID}\big(0, \delta_\eta^2\big) \end{aligned} \qquad \textbf{(10)}$$

[Shephard (1996, pp. 6–7)]. In both cases, the h_t term acts to model volatility of the observed process y_t by multiplicatively changing the amplitude of the NIID process ε_t. The binary transformation defined next removes the amplitude affects of the processes modeled by the h_t term in the above equations and yields a Bernoulli process, given the assumption that the volatility models in (9) and (10) are adequate characterizations of the data and provided that the distribution of ε_t is symmetric.[8]

The binary data transformation, called "hard clipping" in the signal processing literature, is defined as follows:

$$[y(t)] : \begin{matrix} y(t) = 1, & \text{if } Z(t) \geq 0 \\ y(t) = -1, & \text{if } Z(t) < 0 \end{matrix}. \qquad \textbf{(11)}$$

If $Z(t)$ is generated by a pure ARCH, GARCH, or stochastic volatility process whose innovations ε_t are symmetrically distributed with zero mean, then the hard clipped time series $\{y(t)\}$ will be a stationary pure-noise Bernoulli sequence. In essence, whereas $Z(t)$ is a special parameterized martingale difference process, the hard clipping defined in (11) converts it into a Bernoulli process [Lim et al. (2005, pp. 269–70)], which has moments that are well behaved with respect to asymptotic theory [Hinich (1996)]. Therefore if the null of pure noise is rejected by the C, H, or $H4$ tests when applied to binary data determined from (11), this then signifies the presence of structure in the data that cannot be modeled by ARCH, GARCH, or stochastic volatility models defined by (9) and (10).

The issue of parameter instability of GARCH models and the transient nature of ARCH effects can be examined by utilizing the Engle LM test for autoregressive and conditional heteroscedasticity (ARCH) in residuals of a linear model. This was originally proposed in Engle (1982) and should have power against more general GARCH alternatives [Bollerslev (1986)]. The test statistic is based on the R^2 of the auxiliary regression

$$x_t^2 = \beta_0 + \sum_{i=1}^{p} \beta_i x_{t-i}^2 + \xi_t, \qquad \textbf{(12)}$$

where x_t^2 are typically squared residuals from a linear regression. Therefore, equation (12) involves regressing the squared residuals on an intercept and its own p lags. Under the null hypothesis of a linear generating mechanism for x_t, (NR^2) from the regression outlined in (12) is asymptotically distributed as χ_p^2, where N is the number of sample observations and R^2 is the coefficient of multiple correlation from the regression outlined in (12).

To implement the test procedures on a frame-by-frame basis, a frame is defined as significant with respect to the $C, H, H4$, or ARCH LM test if the null of pure noise or no ARCH structure is rejected by each of the respective tests for that particular sample frame at some prespecified (false alarm) threshold. This threshold controls the probability of a TYPE I error, that of falsely rejecting the null hypothesis when it is in fact true.[9] For example, if we adopted a false alarm threshold of 0.90, this would signify that we would expect random chance to produce false rejections of the null hypothesis of pure noise (or no ARCH structure) in 10 of every 100 frames. In accordance with the above criteria, if we secured rejections of the test statistics at rates (significantly) exceeding 10%, 5%, and 1% of the total number of sample frames examined, then this would signify the presence of statistical structure, thus pointing to the presence of (significant) second-, third-, and fourth-order serial dependence or ARCH/GARCH structure in the data.

3. THE TRIMMING PROCEDURE

Correlation, bicorrelation, tricorrelation, and ARCH LM tests are large-sample results based on the asymptotic normal distribution's mean and variance. The validity of any asymptotic result for a finite sample is always an issue in statistics. In particular, the rate of convergence to normality depends on the size of the cumulants of the observed process. All data are finite because all measurements have an upper bound to their magnitudes.[10] However, if the data are leptokurtic, as is typically the case for stock returns, exchange rates, and energy spot prices, then the cumulants are large and the rate of convergence to normality is slow. Trimming the tails of the empirical distribution of the data is an effective statistical method to limit the size of the cumulants in order to get more rapid convergence to the asymptotic (theoretical) distribution.

Trimming data to make sample means less sensitive to outliers has been used in applied statistics for many years. Trimming is a simple data transformation that makes statistics based on the trimmed sample more normally distributed. Transforming data is a technique with a long pedigree, dating back at least to Galton (1879) and McAlister (1879). Subsequently, Edgeworth (1898) and Johnson (1949), among others, have contributed to the understanding of this technique for examining data.

Suppose we want to trim the upper and lower $\kappa\%$ values of the sample$\{x(t_1), \ldots, x(t_N)\}$. To accomplish this, we order the data and find the $(\kappa/100)$ quantile $x_{\kappa/100}$ and the $(1-\kappa/100)$ quantile $x_{1-\kappa/100}$ of the order statistics. Then set all sample values less than the $(\kappa/100)$ quantile to $x_{\kappa/100}$ and set all sample values greater than the $(1-\kappa/100)$ quantile to $x_{1-\kappa/100}$. The remaining $(100-\kappa)\%$ data values are not transformed in any way.

4. THE EMPIRICAL RESULTS

In Table 1, the summary statistics of the spot price returns series are documented. It is apparent that the mean of the series is very small in magnitude—the spot price return is negative over the complete sample. The scale of the spot price returns data appears quite large. The size of the sixth-order cumulants is also quite large, thus pointing to the large scale implicit in the spot price returns data. It is also evident that the spot price returns are quite volatile, with a sizable standard deviation being observed, and the spot price return displays positive skewness. There is evidence of significant leptokurtosis, with an excess kurtosis value of 104. Not unexpectedly, the Jarques–Bera (JB) normality test listed in Table 1 indicates that the null hypothesis of normality is strongly rejected at the conventional 1% level of significance. This outcome reflects the strong evidence of both nonzero skewness and excess kurtosis reported in Table 1 implying substantial deviations

TABLE 1. Summary statistics for New South Wales spot price returns data

No. of observations	161,785
Mean	−0.002
Maximum	545.0
Minimum	−572.0
Std. dev.	19.2
Skewness	0.49
Excess kurtosis	104.0
Sixth-order cumulant	41,958.9
JB test statistic	72,900,000.0
JB normality *P*-value	0.0000

from Gaussianity in the underlying spot price returns data [Lim et al. (2005, p. 270)].

A bootstrap procedure was used to enumerate the empirical distributions of the various test statistics and was implemented in the following way. Given the (possibly trimmed) global sample of 161,785 spot price returns, a bootstrap sample frame was constructed by randomly sampling 336 observations from the larger global population and the various test statistics were calculated for that particular sample frame. This process was repeated 500,000 times and the results for each test statistic were stored in an array. All test statistics entailed the application of the χ^2 distribution and, for each bootstrap replication, the χ^2 levels (threshold) variable associated with each test statistic was transformed to a uniform variate. This means, for example, that the 10% significance level corresponds to a confidence threshold of 0.90, the 5% significance level corresponds to a confidence threshold of 0.95, and the 1% significance level corresponds to a confidence threshold of 0.99. The transformed test statistic confidence threshold values are in the interval (0, 1). Thus, the theoretical distribution can be represented graphically by a 45° line. The arrays containing the bootstrap confidence threshold values for each respective test statistic from the bootstrap process were then sorted in ascending order and associated with a particular quantile scale producing the empirical distribution for each test statistic.[11]

Recall from the discussion in Section 3 that we raised the possibility of improving the finite sample performance of the various tests in the presence of significant outliers by employing trimming, which allows us to improve the rate of convergence of the tests toward their theoretical levels. In this context, it should be noted that trimming is applied to the global spot price returns data and the improved finite sample performance can be discerned from inspection of the QQ plots in Figures 1–7.

The results obtained for the *C* statistic are documented in Figure 1. In deriving the results, an AR(10) prewhitening fit was applied to each bootstrap frame of 336 half-hourly bootstrap observations, producing a bootstrap sample frame of a week's duration. It is clear from inspection of Figure 1 that the no-trimming scenario produced an empirical distribution that was substantially different from the theoretical distribution (corresponding to a 45° line).[12] It is also apparent that all trimming scenarios considered produce empirical distributions that are very close to the theoretical distribution.

In general terms, the trimming scenarios can be interpreted in the following way. The 10%–90% trimming scenario would involve trimming the bottom 10% and top 90% of the empirical distribution function of the complete sample of spot price returns. If a spot price return is smaller than the 10% quantile or larger than the 90% quantile, than the corresponding data values are replaced by the 10% and 90% quantile values, respectively. This operation serves to reduce the range of the data by increasing the minimum value and decreasing the maximum value of the data set, thereby reducing the affect of outliers that fall outside of the 10%–90% quantile range. For example, for the spot price returns, the 10%–90% trimming

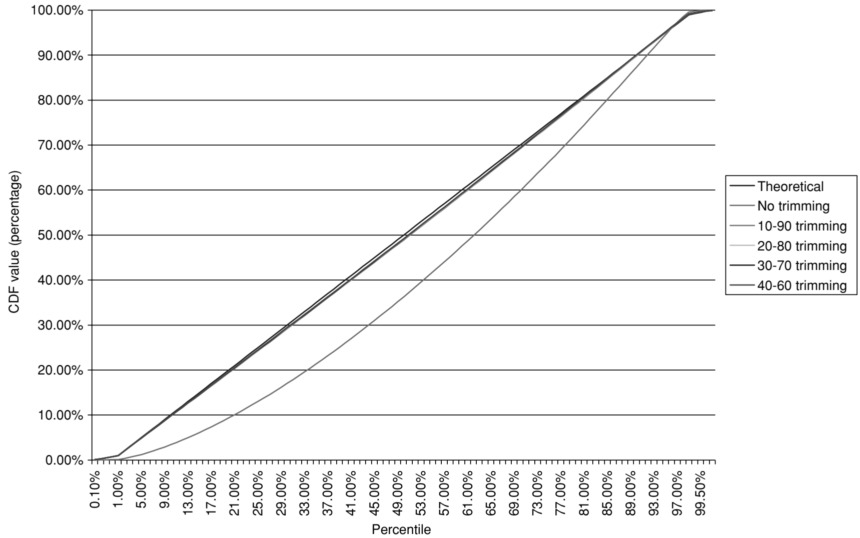

FIGURE 1. QQ plot for bootstrapped C statistic for New South Wales (weekly) spot price returns.

scenario increased the minimum value from -572.0 to -15.3 and decreased the maximum value from 545.0 to 16.1. The newly trimmed data set provides the global population concept underpinning the bootstrap procedure outlined above.

The QQ plots for the H and $H4$ statistics are documented in Figures 2 and 3. It is evident from inspection of both that the empirical distribution associated with the no-trimming scenario is very different from the theoretical distribution. In fact, the performance of both statistics under this scenario is worse than the associated performance of the C statistic (as depicted in Figure 1) because the variance of the third- and fourth-order products underpinning the H and $H4$ statistics depends more crucially on the higher order cumulants than does the variance of the C statistic, which operates at a lower order of magnitude. As such, the sample properties are much more sensitive to deviations from the Gaussian distribution than in the case of the C statistic. As a result, greater degrees of trimming appear to be needed to get the empirical distributions of the H and $H4$ statistics to closely approximate the theoretical distribution than was the case with the C statistic. Specifically, in Figure 1, trimming in the range of 10%–90% seems to be sufficient to achieve a good approximation to the theoretical distribution, whereas trimming rates of at least 20%–80% (and perhaps even as much as 30%–70% for the $H4$ statistic) would seem to be required to obtain a good approximation to the theoretical distribution.[13]

The QQ plots for the ARCH LM tests are depicted in Figure 4. It is apparent that the empirical distribution for the no-trimming scenario once again deviates substantially from the theoretical distribution. The pattern is similar to the pattern

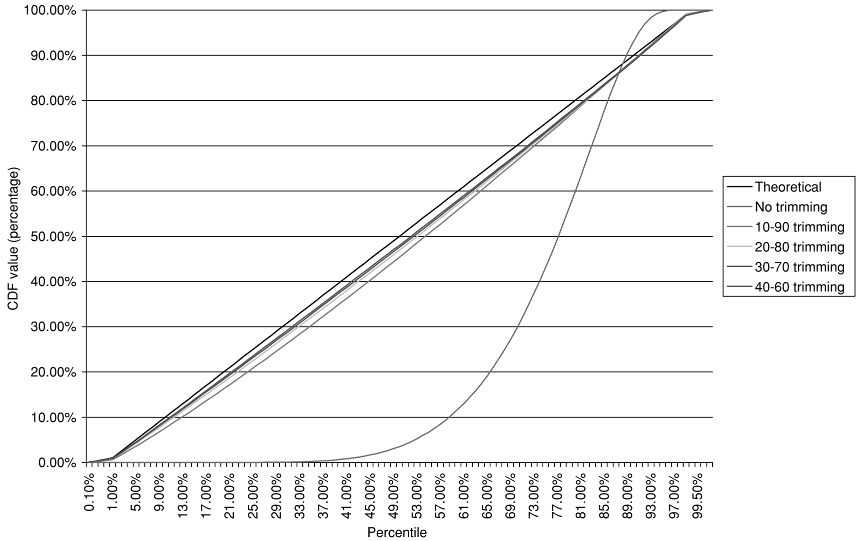

FIGURE 2. QQ plot for bootstrapped *H* statistic for New South Wales (weekly) spot price returns.

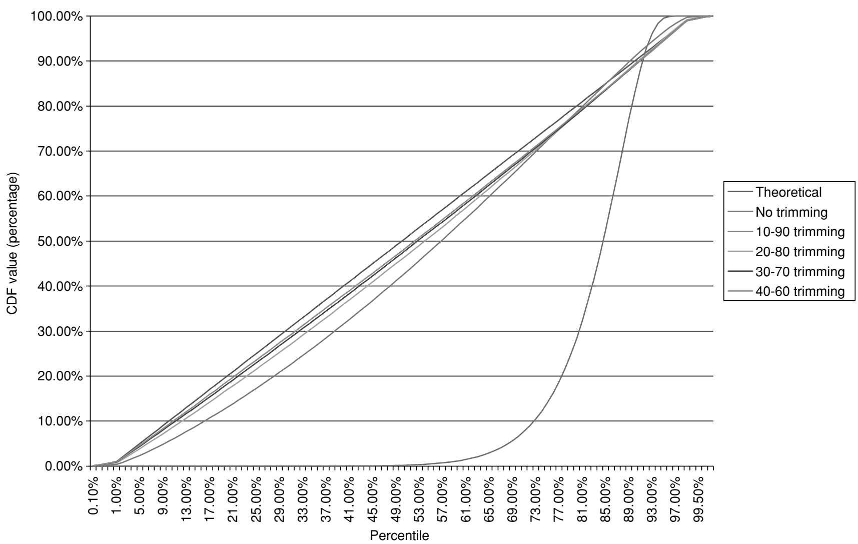

FIGURE 3. QQ plot for bootstrapped *H*4 statistic for New South Wales (weekly) spot price returns.

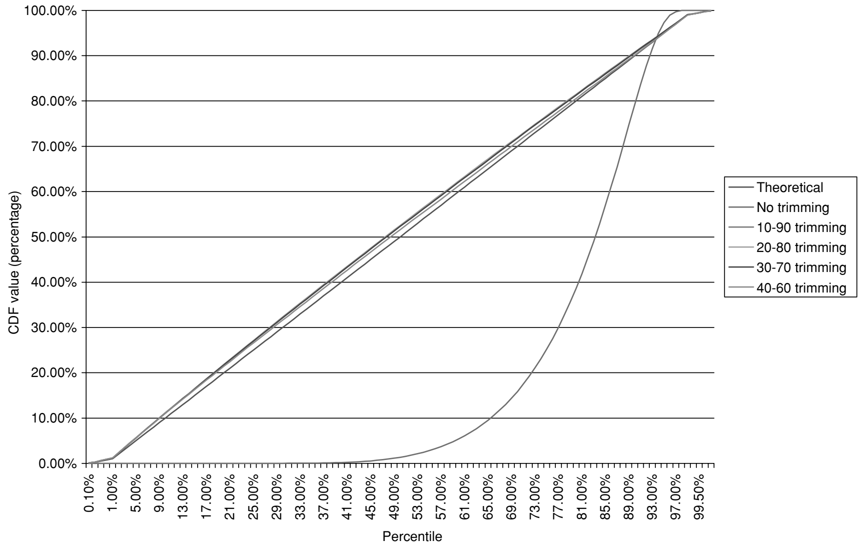

FIGURE 4. QQ plot for bootstrapped LM ARCH statistic for New South Wales (weekly) spot price returns.

observed in Figures 2 and 3 in relation to the *H* and *H*4 statistics. However, the level of trimming that appears to be required to get a good approximation to the theoretical distribution seems to be more closely aligned to that required for the *C* statistic as reported in Figure 1, i.e., trimming in the range 10%–90%. This might reflect the fact that the squaring of residuals involved in the construction of the ARCH LM test is of a similar order of magnitude to the product terms underpinning the *C* statistic, although it produces a different data scale, whereas the *H* and *H*4 statistics involving third- and fourth-order products involve a higher order of magnitude and, as such, are likely to be more sensitive to higher order cumulants and deviations from the Gaussian distribution.

It should also be noted that the empirical distributions associated with the no-trimming scenarios listed in Figures 1–4 all generally lie below the theoretical distribution, except in the upper tail regions, where the *H*, *H*4, and ARCH LM test distributions, in particular, lie above the theoretical distributions. This result points to conservative outcomes for all four tests in the appropriate rejection regions in the upper tails of the distributions of the test statistics, in the case of the no–trimming scenario.[14]

The QQ plots for *C*, *H*, and *H*4 statistics for the bootstrap sample frame–based hard clipping are documented in Figures 5–7. The data that underpin these results are the same data that underpinned the results in Figures 1–4, except that prior to the application of the test statistics, the data in each bootstrap sample frame are hard-clipped using the binary data transformation outlined in equation (11) in Section 2. It is apparent from inspection of Figures 5–7 that the no-trimming

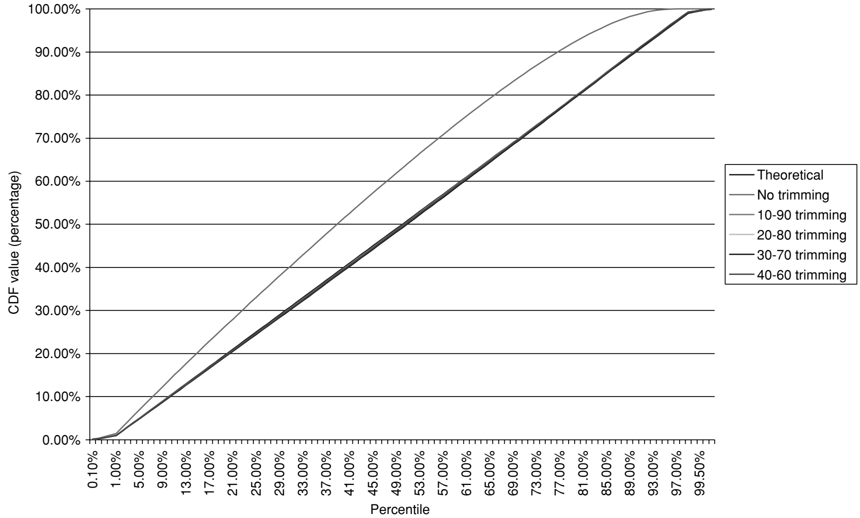

FIGURE 5. QQ plot for bootstrapped *C* statistic for hard-clipped New South Wales (weekly) spot price returns.

scenario produces conservative distributions—that is, the empirical distributions of the various test statistics generally lie above their theoretical distributions. It is also apparent that the binary transformation implied in (11) produces an underlying data set that is more well-behaved, to the extent that minimal trimming

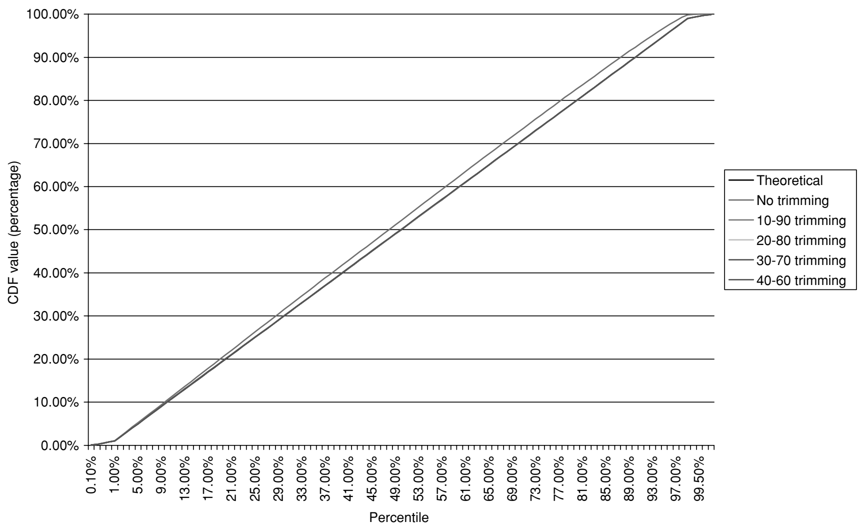

FIGURE 6. QQ plot for bootstrapped *H* statistic for hard-clipped New South Wales (weekly) spot price returns.

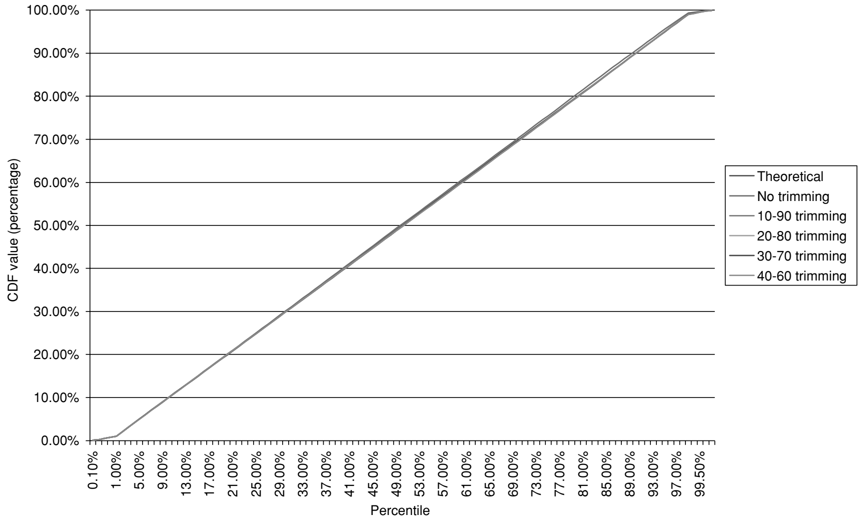

FIGURE 7. QQ plot for bootstrapped *H*4 statistic for hard-clipped New South Wales (weekly) spot price returns.

associated with 10%–90% trimming rates would appear to be sufficient to produce good approximations to the theoretical distributions for all three tests. In fact, the closeness of the no–trimming empirical distributions, reported in Figures 6 and 7, for the *H* and (particularly) the *H*4 statistic supports the theoretical result that the hard-clipping binary transformation converts the ARCH/GARCH and stochastic volatility processes into Bernoulli processes [Hinich (1996)]. Conversely, the statistic with the poorest performance now appears to be the *C* statistic (see Figure 5), although 10%–90% trimming appears to produce a good empirical approximation to the theoretical distribution.

Overall, the results from the QQ plots documented in Figures 1–7 indicate that the empirical distributions of all statistics tend to deviate substantially from the theoretical distribution in the case where no trimming is employed to control the convergence properties of the test statistics in the presence of substantial deviations from Gaussianity in the returns data. Inspections of Figures 1 and 4–7 appear to indicate that reasonable empirical performance can be obtained for the *C* and ARCH LM tests based on spot price returns and for the *C*, *H*, and *H*4 tests when the returns are hard-clipped using trimming rates on the order of 10%–90%. For the *H* and *H*4 statistics applied to spot price returns data, the empirical performance appears to be more sensitive to deviations from the Gaussian distribution, and trimming rates of at least 20%–80% appeared to be necessary to derive empirical distributions that closely approximated the theoretical distributions.

Furthermore, when applied to spot price returns (see Figures 1–4), the no-trimming empirical distributions of the *C*, *H*, *H*4, and ARCH LM tests appeared to generally lie below the theoretical functions, suggesting that the tests were anti-conservative for a wide assortment of quantiles but were conservative in the upper tail regions. For the hard-clipped data and results reported in Figures 5–7, the empirical distribution functions of the *C*, *H*, and *H*4 statistics tend to be conservative—that is, they tend to lie above the theoretical distribution function. The above analysis demonstrates how trimming can be used to improve the finite sample performance of the various test statistics in the presence of substantial deviations from the Gaussian distribution in the source returns data.

Trimming can also be used to see whether any observed nonlinear serial dependence can be viewed as a deep structure phenomenon that arises when nonlinearity is not generated purely by the presence of outliers in the data. This would arise, for example, when the structure associated with the 40%–60% or 30%–70% quantile range of the empirical distribution of the spot price returns data produced nonlinear structure, and not just the data outside of the 10%–90% quantile range, which would be more conventionally associated with outliers. As such, the presence of deep structure can be confirmed if the finding of nonlinear serial dependence continues to hold in the presence of increasingly stringent trimming operations.

The bootstrap procedure that is employed to address this issue differs slightly from that used above to enumerate the empirical properties of the various tests and is based on calculating specific confidence threshold values associated with a user-specified false alarm threshold.[15] The concepts of global sample, weekly bootstrap sample frame, number of bootstrap replications, and application of the various tests remain the same as outlined earlier in this section. Once again, the arrays containing the bootstrap thresholds for the test statistics from the bootstrap process are sorted in ascending order. However, the desired bootstrap thresholds are now calculated as the quantile values of the empirical distributions of the various test statistics associated with a user-specified false alarm confidence threshold.[16] For example, if the user set the false alarm confidence threshold to 0.90, the bootstrap confidence threshold value would be the 90% quantile of the empirical distribution of the relevant test statistic determined from the bootstrap process.

The number of frame-based rejections for each test statistic is calculated by summing the number of frames over which rejections were secured at the calculated bootstrap confidence threshold when the tests were applied on a sequential frame-by-frame basis to the actual (possibly trimmed) returns data.[17] Note in this context that a rejection is secured when the calculated threshold value for the various test statistics applied to actual data frames exceeds the bootstrap confidence threshold. For example, suppose the actual threshold value determined from application of one of the test statistics to an actual sample frame produced a value of 0.97. We would then secure a rejection for that particular frame if the bootstrap threshold value for the test at the user specified level of confidence was less than 0.97. The percentage of frame rejections for each test statistic is calculated as the total number of frame-based rejections computed as a percentage of the total number of frames. Recall that for false alarm confidence thresholds of 0.90, 0.95, and 0.99, we expect only 10%, 5%, and 1% of the total number of frames to secure rejections that can reasonably be attributed to random chance. If the actual number of frame rejections (significantly) exceeds 10%, 5%, or 1% of the total number of frames, then this points to the presence of (significant) linear and/or nonlinear serial dependence, confirming the presence of a nonlinear generating mechanism in the latter case.

In order to investigate whether nonlinear serial dependence could be viewed as a deep structure phenomenon, a number of different trimming scenarios were investigated. These scenarios involved the implementation of different degrees of trimming in order to ascertain whether any observed nonlinear serial dependence that had been observed under less stringent trimming conditions continued to arise, thus confirming the presence of deep nonlinear structure. Specifically, the following trimming scenarios were investigated:

Scenario A: No trimming;[18]
Scenario B: 1%–99% trimming;
Scenario C: 10%–90% trimming;
Scenario D: 20%–80% trimming;
Scenario E: 30%–70% trimming; and
Scenario F: 40%–60% trimming.

The trimming conditions were applied to the complete sample of spot price returns,which were then used to underpin the global population concept from which bootstrap sample frames were constructed and tests applied.

The results for the C, H, $H4$, and ARCH LM tests applied to the spot price returns are documented in Table 2. There are a significant number of frame-based rejections in excess of 10%, 5%, and 1% for H, $H4$, and ARCH LM tests that arise for all trimming scenarios considered and particularly for the H statistic [column (5)].[19] This finding signifies the existence of statistically significant third-order and fourth-order (nonlinear) serial dependence. Note further that the AR(10) prewhitening operation that was used to remove all linear dependence and significant C frames remained successful for all trimming scenarios considered. In fact, for Scenarios C–F, no significant C frames were detected.

It should also be noted that for the 0.99 confidence threshold for the H, $H4$, and ARCH LM statistics for Scenario A, we had to set the false alarm threshold to 0.9999 and 0.999999 because the bootstrapped values tended to be very high and crowded out actual applications to the data. This result seems to be driven by outliers in the data and disappears when trimming is employed to reduce the impact of outliers as indicated, for example, by the results for Scenarios C–F.

Overall, these conclusions confirm the presence of deep (nonlinear) structure because we secure a significant number of frame-based rejections (for H, $H4$, and ARCH LM tests) over and above what can reasonably be attributed to random chance for all trimming scenarios considered, including the more stringent

TABLE 2. Frame test results for New South Wales weekly spot price (returns) data

Scenario/ (state)	Total no. of frames	False alarm threshold	Significant *C* frames no. (%)	Significant *H* frames no. (%)	Significant *H*4 frames no. (%)	Significant ARCH frames no. (%)
Scenario A	481	0.90	1 (0.21)	464 (96.47)	452 (93.97)	395 (82.12)
	481	0.95	1 (0.21)	406 (84.41)	396 (82.33)	326 (67.78)
	481	0.99	1 (0.21)	358* (74.43)	358* (74.43)	186* (38.67)
Scenario B	481	0.90	3 (0.62)	464 (96.47)	422 (87.73)	415 (86.28)
	481	0.95	1 (0.21)	442 (91.89)	360 (74.84)	378 (78.59)
	481	0.99	0 (0.00)	368 (76.51)	280** (58.21)	303 (62.99)
Scenario C	481	0.90	0 (0.00)	462 (96.05)	427 (88.77)	401 (83.37)
	481	0.95	0 (0.00)	451 (93.76)	400 (83.16)	375 (77.96)
	481	0.99	0 (0.00)	403 (83.78)	308 (64.03)	305 (63.41)
Scenario D	481	0.90	0 (0.00)	446 (92.72)	378 (78.59)	356 (74.01)
	481	0.95	0 (0.00)	430 (89.40)	339 (70.48)	314 (65.28)
	481	0.99	0 (0.00)	374 (77.75)	257 (53.43)	217 (45.11)
Scenario E	481	0.90	0 (0.00)	428 (88.98)	304 (63.20)	281 (58.42)
	481	0.95	0 (0.00)	394 (81.91)	260 (54.05)	219 (45.53)
	481	0.99	0 (0.00)	331 (68.81)	173 (35.97)	135 (28.07)
Scenario F	481	0.90	0 (0.00)	376 (78.17)	220 (45.74)	165 (34.30)
	481	0.95	0 (0.00)	334 (69.44)	161 (33.47)	126 (26.20)
	481	0.99	0 (0.00)	239 (49.69)	76 (15.80)	65 (13.51)

Notes: No global AR prewhitening fit; applied frame by frame AR(10) prewhitening fit to remove linear dependence; various trimming scenarios.
*False alarm threshold arbitrarily set to 0.9999.
**False alarm threshold arbitrarily set to 0.999999.

scenarios associated with Scenarios E and F. The observed rejection rates are believable because the tests were demonstrated to be conservative for Scenario A and the empirical distribution of the tests were found to be quite close to the theoretical distributions for Scenarios C–F for the C and ARCH LM tests and for Scenarios D–F for the H and $H4$ tests. The rejection rates confirm the presence of nonlinear serial dependence under all trimming scenarios considered, pointing to the presence of deep nonlinear structure. This finding, in turn, implies that the observed nonlinear serial dependence is not purely determined by the presence of outliers in the data.

The results for the C, H, and $H4$ tests associated with hard-clipping transformation applied to the residuals from the frame by frame AR(10) fits are outlined in Table 3. Recall that these residuals are the same as those underpinning the results listed in Table 2, except that the transformation in (11) was applied to the residuals prior to the application of the three above-mentioned portmanteau tests. Recall further that the intention of this particular test framework is to see whether structure is present in the data that cannot be attributed to ARCH/GARCH or stochastic volatility models.

It is evident from inspection of Table 3 that the numbers of frame-based rejections for the C, H, and $H4$ statistics applied to the binary data sets are greater than the 10%, 5%, and 1% rates that we can reasonably attribute to random chance, thus pointing to the contributing presence of structure that cannot be explained by volatility models. This conclusion holds for all trimming scenarios considered, although the underlying rejection rates discernible from Table 3 suggest that third-order nonlinear serial dependence (associated with H-statistic rejections) is the most prominent form of nonlinear serial dependence. Interestingly, the rejection rates tail off somewhat over Scenarios B–D but then become more prominent for Scenarios E and F, which correspond to the most stringent trimming conditions considered. This is interesting because ARCH/GARCH processes, in particular, are seen as being driven by volatility clustering associated with the episodic presence of outliers in the data. The more stringent trimming conditions increasingly abstract from this type of generating mechanism whereas, at the same time, the results cited in Table 3 point to significant frame-based rejections, indicating the presence of linear as well as third- and fourth-order (nonlinear) serial dependence in the hard-clipped data. Finally, it should be noted that the data underpinning the significant C test outcomes are the same set of residuals that produced very few or no significant C frames in Table 2.

These conclusions indicate the nontrivial presence of a nonlinear generating mechanism operating over the central quantile ranges of the empirical distribution of the spot price returns data that cannot be explained by volatility models encompassing ARCH/GARCH and stochastic volatility models. The fact that the nonlinear serial dependence arises over this quantile range points to the presence of deep structure—that is, the presence of nonlinear structure that is not being generated predominantly by the presence of outliers.

TABLE 3. Frame test results for New South Wales weekly spot price (hard-clipped returns) data

Scenario/ (state)	Total no. of frames	False alarm threshold no. (%)	Significant C frames no. (%)	Significant H frames no. (%)	Significant $H4$ frames
Scenario A	481	0.90	176 (36.59)	272 (56.55)	149 (30.98)
	481	0.95	111 (23.08)	212 (44.07)	106 (22.04)
	481	0.99	52 (10.81)	98 (20.37)	55 (11.43)
Scenario B	481	0.90	129 (26.82)	280 (58.21)	113 (23.49)
	481	0.95	86 (17.88)	222 (46.15)	67 (13.93)
	481	0.99	34 (7.07)	114 (23.70)	24 (4.99)
Scenario C	481	0.90	38 (7.90)	297 (61.75)	112 (23.28)
	481	0.95	25 (5.20)	236 (49.06)	67 (13.93)
	481	0.99	8 (1.66)	127 (26.40)	18 (3.74)
Scenario D	481	0.90	57 (11.85)	320 (66.53)	133 (27.65)
	481	0.95	41 (8.52)	269 (55.93)	82 (17.05)
	481	0.99	17 (3.53)	169 (35.14)	28 (5.82)
Scenario E	481	0.90	197 (40.96)	353 (73.39)	144 (29.94)
	481	0.95	151 (31.39)	310 (64.45)	105 (21.83)
	481	0.99	85 (17.67)	214 (44.49)	49 (10.19)
Scenario F	481	0.90	378 (78.59)	391 (81.29)	192 (39.92)
	481	0.95	341 (70.89)	353 (73.39)	139 (28.90)
	481	0.99	277 (57.59)	277 (57.59)	85 (17.67)

Notes: No global AR prewhitening fit; applied frame by frame AR(10) prewhitening fit to remove linear dependence; frame by frame hard clipping of residuals; various trimming scenarios.

5. CONCLUDING COMMENTS

Before modeling techniques such as GARCH are applied, it is important to understand the nature of nonlinear serial dependence in the data under investigation. Using the example of wholesale electricity spot prices, we have demonstrated how an effective testing methodology can be applied. This involves partitioning time series data into nonoverlapping frames and computing the portmanteau correlation, bicorrelation, and tricorrelation test statistics for each frame to detect linear and nonlinear serial dependence, respectively. Furthermore, the presence of pure ARCH and GARCH effects in the spot price returns was also investigated by applying the LM ARCH test and, additionally, a detection framework based upon converting GARCH and stochastic volatility processes into a pure-noise process and then testing for the presence of linear and nonlinear serial dependence in the transformed data.

The finite sample properties of the empirical distribution of the various tests were investigated using a bootstrap procedure. This allowed assessment of how close the empirical properties of the tests were to their theoretical distributions given the significant observed deviations from the Gaussian distribution implied in the source returns data. It was also demonstrated that the empirical properties of the tests could be improved substantially through the use of trimming, which permits an investigator to control for the impact of outliers on the finite sample performance of the test statistics in order to obtain an empirical distribution closer to the Gaussian distribution.

It was also shown how trimming could be used to investigate whether the nonlinear generating mechanism was a deep structure phenomenon where observed nonlinear serial dependence was generated by more than the presence of outliers. In our example, it was found that nonlinear serial dependence did not disappear as more stringent trimming scenarios were adopted. This confirmed the presence of a deep structure in the generating mechanism. In particular, the results indicated that the generating mechanism was not consistent with the hypothesis that an ARCH, GARCH, or stochastic volatility process generated the returns.

The finding that nonlinearity is present has implications for modeling spot price dynamics. If there are both third- and fourth-order nonlinear serial dependence in the data, then time series models that employ a linear structure, or assume a pure noise input, such as the geometric Brownian motion (GBM) stochastic diffusion model, are problematic. In particular, the dependence structure violates both the normality and Markovian assumptions underpinning conventional GBM models. The finding that nonlinear serial dependence can be categorized as a deep structure phenomenon poses questions about the validity of jump diffusion models that employ the Poisson process in order to model the probability of the occurrence of outliers (i.e., jump events). It is no longer appropriate to simply equate the presence of nonlinearity with the presence of outliers. This finding has important implications for the use of GBM and jump diffusion models that currently underpin accepted risk management strategies based on the Black–Scholes option pricing model, which are employed in finance and energy economics.

NOTES

1. The half-hourly load and spot price data were sourced from files located at the following Web address: http://www.aemo.com.au/data/price_demand.html.

2. Also see Bollerslev et al. (1992), Bera and Higgins (1993), Bollerslev et al. (1995), and Diebold and Lopez (1995) for detailed surveys of these models.

3. We will demonstrate in the next section how a hard-clipping transformation can be used in conjunction with our nonparametric tests to falsify both types of models.

4. This nonlinear-in-mean structure in the residuals would be detected using higher order analogues of a conventional portmanteau test for serial correlation.

5. Other references utilizing this framework include Brooks (1996), Brooks and Hinich (1998), Ammermann and Patterson (2003), Lim et al. (2003, 2004, 2005), Lim and Hinich (2005a, 2005b), Bonilla et al. (2007), and Hinich and Serletis (2007).

6. In this article, we set $b = 0.4$ for the correlation and bicorrelation tests and $b = 0.3$ for the tricorrelation test.

7. If we set the β_j's coefficients to zero in (9), we get an ARCH(q) process.

8. For our purposes, the crucial requirement is that ε_t be a pure–white noise process (i.e., i.i.d.). The assumption of normality was made purely for convenience. Other distributional assumptions used in relation to ε_t in the literature include the t distribution [Bollerslev (1987)] and the generalized error distribution [Nelson (1991)]. For the hard-clipping technique to work successfully, whatever distributional assumption is made about ε_t, it must be a symmetric distribution.

9. The false alarm threshold is to be interpreted as a confidence level; for example, a false alarm threshold of 0.90 is to be interpreted as a 90% confidence level. The level of significance associated with this confidence level is interpreted in the conventional way. Therefore, for a threshold of 0.9, we get a corresponding significance level of 10%.

10. In the current context, a maximum spot price that can be bid by wholesale market participants is $10,000/MWh, which corresponds to the value of lost load (VOLL) price limit that is triggered in response to demand–supply imbalances that trigger load shedding [see National Electricity Market Management Company Limited (2005)]. Thus, the range of the spot price data is finite, ensuring that all moments are finite.

11. In the current context, the term quantile can be interchanged with the term percentile, which is used on the horizontal axes of Figures 1–7.

12. It should be noted that all distributions represented graphically in the various QQ plots are plotted on a percentile basis. However, the extreme lower and upper tails of the distribution functions are defined at an interval of less than a percentile in order to enumerate the characteristics of the tails. This gives the slight dog-legged appearance at the start and end points of the plots.

13. This is especially the case in the upper tail region of the empirical distribution functions where the key rejection regions for the various test statistics in fact lie.

14. In this context, the no-trimming scenario corresponds to the framework underpinning the results reported in Wild et al. (2008). Given the conservative nature of the tests at the confidence levels considered in that article (i.e., at 0.90, 0.95, and 0.99), any frame-based rejections reported in that article are believable in statistical terms.

15. This bootstrap framework mirrors the framework used in Wild et al. (2008).

16. Recall from the discussion in Section 3 that the false alarm threshold is used to control for the probability of a Type I error.

17. In the results reported in Tables 2 and 3, trimming is incorporated in Scenarios B–F and no trimming is used in Scenario A only.

18. Note that the results corresponding to Scenario A are the same set of results reported in Wild et al. (2008).

19. The results for the *H*4 and ARCH LM tests also point to significant structure [see columns (6) and (7) of Table 2] because the rejection rates still exceed those that can be attributed to random chance (i.e., 10%, 5%, and 1%) but are still dominated by the *H*-test results, which involve much larger frame-based rejection rates.

REFERENCES

Ammermann, P.A. and D.M. Patterson (2003) The cross-sectional and cross-temporal universality of nonlinear serial dependencies: Evidence from world stock indices and the Taiwan Stock Exchange. *Pacific-Basin Finance Journal* 11, 175–195.

Bera, A.K. and M.L. Higgins (1993) ARCH models: Properties, estimation and testing. *Journal of Economic Surveys* 7, 305–362.

Bollerslev, T. (1986) Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Bollerslev, T. (1987) A conditional heteroscedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.

Bollerslev, T., R.Y. Chou, and K.F. Kroner (1992) ARCH modelling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52, 5–59.

Bollerslev, T., R.F. Engle, and D.B. Nelson (1995) ARCH models. In R.F. Engle and D. McFadden (eds.), *The Handbook of Econometrics*, Vol. 4, Chapter 49. Amsterdam: North-Holland.

Bonilla, C.A., R. Romero-Meza, and M.J. Hinich (2007) GARCH inadequacy for modelling exchange rates: Empirical evidence from Latin America. *Applied Economics* 39, 2529–2533.

Brooks, C. (1996) Testing for non-linearity in daily sterling exchange rates. *Applied Financial Economics* 6, 307–317.

Brooks, C. and M.J. Hinich (1998) Episodic nonstationarity in exchange rates. *Applied Economic Letters* 5, 719–722.

Diebold, F.X. and J.A. Lopez (1995) ARCH models. In K. Hoover (ed.), *Macroeconomics: Developments, Tensions and Prospects,* pp. 427–472. Boston: Kluwer Academic Press.

Edgeworth, F.Y. (1898) On the representation of statistics by mathematical formulae. *Journal of the Royal Statistical Society* 61, 670–700.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.

Foster, J., M.J. Hinich, and P. Wild (2008) Randomly modulated periodic signals in Australia's national electricity market. *The Energy Journal* 29, 105–130.

Galton, F. (1879) The geometric mean in vital and social statistics. *Proceedings of the Royal Society of London* 29, 365–367.

Hinich, M.J. (1996) Testing for dependence in the input to a linear time series model. *Journal of Nonparametric Statistics* 6, 205–221.

Hinich, M.J. (2000) A statistical theory of signal coherence. *Journal of Oceanic Engineering* 25, 256–261.

Hinich, M.J. and D.M. Patterson (1989) Evidence of nonlinearity in the trade-by-trade stock market return generating process. In W. Barnett, J. Geweke, and K. Shell (eds.), *Economic Complexity, Chaos, Sunspots, Bubbles, and Nonlinearity,* pp. 383–409. New York: Cambridge University Press.

Hinich, M.J. and D.M. Patterson (1995) Detecting Epochs of Transient Dependence in White Noise. Mimeo, University of Texas at Austin.

Hinich, M.J. and D.M. Patterson (2005) Detecting epochs of transient dependence in white noise. In M. Belongia and J. Binner (ed.), *Money, Measurement and Computation,* Part 2. London: Palgrave.

Hinich, M.J. and A. Serletis (2007) Episodic nonlinear event detection in the Canadian exchange rate. *Journal of the American Statistical Association*, *Applications and Case Studies* 102, 68–74.

Hinich, M.J. and P. Wild (2001) Testing time-series stationarity against an alternative whose mean is periodic. *Macroeconomic Dynamics* 5, 380–412.

Johnson, N.L. (1949) Systems of frequency curves generated by means of translation. *Biometrika* 36, 149–176.

Lim, K.P. and M.J. Hinich (2005a) Cross-temporal universality of non-linear dependencies in Asian stock markets. *Economics Bulletin* 7, 1–6.

Lim, K.P. and M.J. Hinich (2005b) Non-linear market behavior: Events detection in the Malaysian stock market. *Economics Bulletin* 7, 1–5.

Lim, K.P., M.J. Hinich, and V.K.S. Liew (2003) Episodic non-linearity and non-stationarity in ASEAN exchange rates returns series. *Labuan Bulletin of International Business and Finance* 1, 79–93.

Lim, K.P., M.J. Hinich, and V.K.S. Liew (2004) Non-linearity in financial markets: Evidence from ASEAN-5 exchange rates and stock markets. *ICFAI Journal of Applied Finance* 10, 5–18.

Lim, K-P., M.J. Hinich, and V.K. Liew (2005) Statistical inadequacy of GARCH models for Asian stock markets: Evidence and implications. *Journal of Emerging Market Finance* 4, 263–279.

McAlister, D. (1879) The law of the geometric mean. *Proceedings of the Royal Society of London* 29, 367–376.

National Electricity Market Management Company Limited (2005) An Introduction to Australia's National Electricity Market. National Electricity Market Management Company Ltd. Available at http://www.aemo.com.au/corporate/0000-0006.pdf.

Nelson, D.B. (1991) Conditional heteroscedasticity in asset pricing: A new approach. *Econometrica* 59, 347–370.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In D.R Cox, D.V. Hinkley, and Ole E. Barndorff-Nielsen (eds.), *Time Series Models in Econometrics, Finance and Other Fields,* pp. 1–67. London: Chapman & Hall.

Taylor, S.J. (1986) *Modelling Financial Time Series*, Chapter 3. Chichester. UK: John Wiley.

Wild, P., M.J. Hinich, and J. Foster (2008) Are Daily and Weekly Load and Spot Price Dynamics in Australia's National Electric Market Governed by Episodic Nonlinearity? Discussion Paper No. 368, School of Economics, University of Queensland, Australia. Available at http://www.uq.edu.au/economics/abstract/368.pdf.