Are Daily and Weekly Load and Spot Price Dynamics in Australia's National Electricity Market Governed by Episodic Nonlinearity?

By

Phillip Wild*

School of Economics, University of Queensland,

St Lucia, QLD, 4072, Australia

Phone: +61 7 3346 7058

Email: p.wild@uq.edu.au

Melvin. J. Hinich

Applied Research Laboratories, University of Texas at Austin,

Austin, TX 78712-1087

Phone: +1 512 232 7270

Email: hinich@austin.utexas.edu

and

John Foster

School of Economics, University of Queensland,

St Lucia, QLD, 4072, Australia

Phone: +61 7 3365 6780

Email: j.foster@uq.edu.au

* Correspondence to Phillip Wild, School of Economics, University of Queensland, St Lucia, QLD, 4072, Australia; Email: p.wild@uq.edu.au, Phone: +61 7 3346 7058, fax: +61 7 3365 7299.

ABSTRACT

In this article, we use half hourly spot electricity prices and load data for the National Electricity Market (NEM) of Australia for the period from December 1998 to June 2009 to test for episodic nonlinearity in the dynamics governing daily and weekly cycles in load and spot price time series data. We apply the portmanteau correlation, bicorrelation and tricorrelation tests introduced in Hinich (1996) to the time series of half hourly spot prices and load demand from 7/12/1998 to 30/06/2009 using a FORTRAN 95 program. We find the presence of significant third and fourth order (non-linear) serial dependence in the weekly load and spot price data in particular, but to a much more marginal extent, in the daily data.

JEL Classification: C12, C14, C15.

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

1 INTRODUCTION

The Australian electricity market encompasses generation, transmission, distribution and retail sale activities. The key part of the market in Australia is the wholesale National Electricity Market (NEM) which is structured as a gross pool arrangement. It commenced operation as a de-regulated wholesale market in New South Wales, Victoria, Queensland, the Australian Capital Territory (ACT) and South Australia in December 1998. In 2005, Tasmania joined as a sixth region. There are six interconnected regions that broadly follow state boundaries (AEMO (2009)).

There are a number of 'stylised' facts that are widely accepted as applying to load demand and spot price dynamics in the market. First, observed load demand patterns in the market tend to vary from region to region, depending upon such factors as population, temperature and industrial and commercial needs, and there are well-defined cyclical variations in electricity demand over the year.

The second 'stylised' fact is that the load curve has both a weekly and a daily cycle. The peak hourly load in Australia has two distinct peaks, in early morning and early evening, that are generated by domestic activity and these vary at weekends.

The third 'stylised' fact is that spot electricity prices exhibit both the properties of high volatility (i.e. a lot of price spikes) and strong mean-reverting behaviour (volatility clustering followed by sustained periods of 'normality'). Spot price spikes are outliers producing significant deviations from the Gaussian distribution. In fact, the spot price data displays the same predominant empirical 'leptokurtosis' feature found in most high frequency asset price data – the tails of the empirical distribution functions are much fatter than those associated with the normal distribution implying large fourth order cumulants.

In Foster et al (2008), the extent of and stability of daily and weekly cycles in both load and spot price time series data was investigated using the Randomly Modulated Periodicity (RMP) Model introduced in Hinich (2000) and Hinich and Wild (2001). A major finding was that the mean properties of both the load and spot price data for the NEM States considered were periodic.[1] The most important periodicities for both datasets were found to contain significant but imperfect signal coherence suggesting that some 'wobble' existed in the waveforms of the load and spot price data. It was originally postulated in Hinich (2000) and Hinich and Wild (2001) that the generating mechanism for an RMP process would be nonlinear. Therefore, a natural research question is whether the mechanism responsible for generating both daily and weekly electricity load and price data exhibits some type of nonlinearity, and if so, whether this nonlinearity is 'episodic' in character. It is likely that the existence of episodic nonlinearity is necessary for the strong mean reversion observed in spot electricity prices. If nonlinearity exists, it also rules out many classes of linear models as candidates for explaining both load and spot price dynamics.

The article is organized as follows. In Section 2 we briefly discuss the data used and highlight some transformations that were made to the spot price electricity data in order to implement the tests considered. In Section 3 we outline the portmanteau correlation, bicorrelation and tricorrelation tests employed. These tests were used to test for second-order (linear), third- and fourth-order (nonlinear) serial dependence, respectively. In

[1] Serletis (2007, Ch 19) has applied the same type of RMP analysis to the Canadian Alberta market.

Section 4 we briefly state the well-known Engle LM ARCH test that was used to test for the presence of ARCH and GARCH structures in the daily and weekly waveforms. In Section 5, the empirical results for both the daily and the weekly waveforms are presented. In Section 6, the implications of our findings for modeling load and spot price dynamics and implications for risk management practices within the industry are discussed. Finally, in Section 7, some concluding comments are offered.

2 DATA AND ASSOCIATED TRANSFORMATIONS

In this article, we use half hourly spot electricity prices and load data for the period from 7/12/1998 to 30/06/2009.[2] This produced a sample size of 185,162 observations. We apply the tests to time series load and spot price data from New South Wales (NSW), Queensland (QLD), Victoria (VIC) and South Australia (SA).

In applying the various tests outlined in this article, we convert all data series to continuous compounded returns by applying the relationship

$$r(t) = \ln\left(\frac{y(t)}{y(t-1)}\right) * 100\,, \tag{1}$$

where:

- $r(t)$ is the continuous compounded return for time period 't'; and
- $y(t)$ is the source price or load time series data.

In order to apply (1), $y(t)$ cannot take negative or zero values. However, it was evident that for Queensland, Victoria and South Australia, there was the occasional occurrence of negative spot prices.

[2] 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.

In the presence of negative prices, some transformations had to be made to the respective price series to remove negative prices before we were able to apply (1) to convert the data to returns. This transformation involves two steps. First, any values which were negative or zero are set to the previous non-negative value using the following decision rule:

$$if \quad y(t) = \begin{cases} \leq 0, & x(t) = y(t-1) \\ else, & x(t) = y(t) \end{cases}, \tag{2}$$

where $y(t)$ is the source time series data and $x(t)$ is the transformed data series. The second step involved applying a linear interpolation routine to the transformed series $x(t)$ obtained by using the following decision rule:

$$if \quad y(t) = \begin{cases} \leq 0, \; z(t) = \left\{ \dfrac{\left[x(t-1) + x(t+1)\right]}{2} \right\} \\ else, \qquad z(t) = x(t) \; \left[= y(t)\right] \end{cases}, \tag{3}$$

where $z(t)$ is the new transformed data (see Foster, Hinich and Wild (2008)).

3 THE PORTMANTEAU CORRELATION, BICORRELATION AND TRICORRELATION TEST STATISTICS IN MOVING TIME WINDOWS FRAMEWORK

We utilize the framework originally proposed in Hinich and Patterson (1995), (now published as Hinich and Patterson (2005)) which seeks to detect epochs of transient serial dependence in a discrete-time pure white noise process (i.e. *i.i.d* random variates).[3] A common approach to processing time series with a periodic structure is to partition the

[3] Also see Czamanski, Dormaar, Hinich and Serletis (2007) who apply a similar testing methodology to the Canadian Alberta market.

observations into non-overlapping frames where there is exactly one waveform in each sample (data) frame. This methodology involves computing the portmanteau correlation, bicorrelation and tricorrelation test statistics (termed as C, H and $H4$ statistics) for each frame to detect linear and nonlinear serial dependence respectively.

Let the sequence $\{x(t)\}$ denote the sampled (and transformed) data process in (3), where the time unit 't' is an integer. The test procedure employs non-overlapped time frames (windows), thus if n is the frame length, then the k-th window is defined as $\{x(t_k), x(t_k + 1), ..., x(t_k + n - 1)\}$. The next non-overlapped window is $\{x(t_{k+1}), x(t_{k+1} + 1), ..., x(t_{k+1} + n - 1)\}$, where $t_{k+1} = t_k + n$. Define $Z(t)$ as the sequence of standardized observations given by

$$Z(t) = \frac{x(t) - m_x}{s_x}, \qquad (4)$$

for each $t = 1,2,...,n$ where m_x and s_x are the sample mean and standard deviation of the sample frame. So the data in each sample is standardised 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 $C_{ZZ}(r) = E[Z(t)Z(t+r)] = 0,\ \forall\ r \neq 0$, the bicorrelations $C_{ZZZ}(r,s) = E[Z(t)Z(t+r)Z(t+s)] = 0,\ \forall\ r, s$ except when $r = s = 0$, and the tricorrelations $C_{ZZZZ}(r,s,v) = E[Z(t)Z(t+r)Z(t+s)Z(t+v)] = 0,\ \forall\ r, s, \mathit{and}\ v$ except when $r = s = v = 0$. The alternative hypothesis is that the process in the sample frame has some non-zero 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. In other words, if there exists second-order (linear) or third- or fourth-order (nonlinear) serial dependence in the data generating process, then $C_{ZZ}(r) \neq 0, C_{ZZZ}(r,s) \neq 0, \text{or } C_{ZZZZ}(r,s,v) \neq 0$ 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

$$C_{ZZ}(r) = \frac{1}{\sqrt{n-r}} \sum_{t=1}^{n-r} Z(t)Z(t+r). \tag{5}$$

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

$$C = \sum_{r=1}^{L} [C_{ZZ}(r)]^2 \approx \chi_L^2. \tag{6}$$

The (r,s) sample bicorrelation coefficient is

$$C_{ZZZ}(r,s) = \frac{1}{n-s} \sum_{t=1}^{n-s} Z(t)Z(t+r)Z(t+s), \text{ for } 0 \leq r \leq s. \tag{7}$$

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

$$H = \sum_{s=2}^{L} \sum_{r=1}^{s-1} G^2(r,s) \approx \chi^2_{L(L-1)/2} \tag{8}$$

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

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

$$C_{ZZZZ}(r,s,v) = \frac{1}{n-v} \sum_{t=1}^{n-v} Z(t)Z(t+r)Z(t+s)Z(t+v), \text{ for } 0 \leq r \leq s \leq v. \tag{9}$$

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

$$H4 = \sum_{v=3}^{L}\sum_{s=2}^{v-1}\sum_{r=1}^{s-1} T^3(r,s,v) \approx \chi^2_{L(L-1)(L-2)/3} \tag{10}$$

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

Since it is conceptually difficult to quantify how much of any ‘significant’ autocorrelation can be attributed to thin trading volume or spot price limits, this investigation focuses instead on whether load and spot price data contain predictable nonlinearities after removing all linear dependence. The autocorrelation structure in each sample frame is removed by an autoregressive AR(*p*) fit, where ‘ p ’ is the number of lags that is selected in order to remove significant C statistics at some pre-specified threshold level.[4] It is worth noting that the AR fitting is employed purely as a ‘pre-whitening’ operation and not in order to obtain a model of ‘best fit’. The portmanteau bicorrelation and tricorrelation tests are then applied to the residuals of the fitted AR(p) model of each sample frame, so that any rejections of the null hypothesis of pure white noise can be attributed to significant H or $H4$ statistics.

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, and where b is a parameter to be chosen by the user. Based on results of Monte Carlo simulations, Hinich and

[4] In the literature particularly dealing with long-term dependence, pre-filtering by means of an AR-GARCH procedure is often used to remove short-term autocorrelation and time-varying volatility. However, this procedure is unnecessary, in the current context, since the bicorrelation and tricorrelation tests rely on the property that the bicorrelation and tricorrelation coefficients equal zero for a pure noise process. As such, the null hypothesis is only rejected when there exists some non-zero bicorrelations or tricorrelations suggesting nonlinear serial dependence in the conditional mean (additive nonlinearity), and not the presence of conditional variance dependence (conditional heteroskedasticity).

Patterson (1995, 2005) recommended the use of $b = 0.4$ (in relation to the bicorrelation test) which is a good compromise between: (1) using the asymptotic result as a valid approximation for the sampling properties of the H statistic for moderate sample sizes; and (2) having enough sample bicorrelations in the statistic to have reasonable power against non-independent variates.

Another element that must be decided upon is the choice of the frame length. In principal, there is no unique value for the frame length. The larger the frame length, the larger the number of lags and hence the greater the power of the test, but at the 'expense' of increasing the uncertainty of the event time when the serial dependence 'episode' occurs. The data is split into a set of equal-length non-overlapped moving frames of 48 and 336 half hour observations corresponding to a frame of a day and a week's duration, respectively.[5] Our objective is to measure the extent to which any observed nonlinearity that is episodically present appears to be operating on a daily or weekly time scale.

We can also use the correlation, bicorrelation and tricorrelation tests to examine whether a symmetric GARCH or stochastic volatility model represent adequate characterisations of the data under investigation. We can define a $GARCH(p,q)$ process as

$$y_t = \varepsilon_t h_t,\ \ \varepsilon_t \approx NIID\left(0, h_t^2\right),\ \ h_t^2 = \alpha_0 + \sum_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}^2. \qquad (11)^{6}$$

We can similarly define a stochastic volatility model as

[5] In principle, this window length needs to be sufficiently long enough to validly apply the bicorrelation and tricorrelation tests and yet short enough for the data generating process to have remained roughly constant (see Monte Carlo results in Hinich (1996) and Hinich and Patterson (1995, 2005)).

[6] If we set the β_j's coefficients to zero in (11), we get an $ARCH(q)$ process.

$$y_t = \varepsilon_t \exp\left(h_t \big/ 2 \right), \;\; \varepsilon_t \approx NIID(0,1), \;\; h_{t+1} = \alpha_0 + \alpha_1 h_t + \eta_t, \;\; \eta_t \approx NIID\left(0, \delta_\eta^2\right) \tag{12}$$

(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 below removes the amplitude affects of the processes modelled by the h_t term in the above equations and yields a Bernoulli process given the assumption that the 'volatility' models in (11) and (12) are adequate characterisations of the data and provided and that the distribution of ε_t is symmetric.[7]

The binary data transformation, called 'hard clipping', is defined as

$$\{y(t)\} : \begin{array}{l} y(t) = 1, \;\; if \;\; Z(t) \geq 0 \\ y(t) = -1, if \;\; Z(t) < 0 \end{array} . \tag{13}$$

If $Z(t)$ is generated by a pure ARCH/GARCH or stochastic volatility process whose innovations are symmetrically distributed with zero mean, then the binary data set $\{y(t)\}$ will be a stationary pure noise (*i.i.d*) Bernoulli sequence. While $Z(t)$ is a martingale difference process, the binary transformation outlined in (13) converts it into a pure noise process (Lim, Hinich and Liew (2005)) 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 (13), this then signifies the presence of structure in the data that cannot be modelled by symmetric

---

[7] For our purposes, the crucial requirement is that ε_t is a pure (symmetric) white noise process, (i.e. *iid)*. 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 Generalised Error Distribution (Nelson (1991)).

ARCH/GARCH or stochastic volatility models. Moreover, while the rejections might be because of the presence of serial dependence in the innovations, this outcome still violates a critical assumption underpinning the formulation of these 'volatility' models. Specifically, if the innovations are dependent (not *i.i.d*), then the statistical properties of the parameter estimates of ARCH/GARCH processes, for example, are unknown (Bonilla, Meza and Hinich (2007)).

To implement the test procedures on a frame-by-frame basis, define a frame as significant with respect to the C, H or H4 tests if the null of pure noise is rejected by each of the respective tests for that particular sample frame at some pre-specified (false alarm) threshold. This threshold controls the probability of a TYPE I error, - that of falsely rejecting the null hypothesis when it is true.[8] For example, if we adopt 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 in 10 out of every 100 frames. In a similar way, false alarm thresholds of 0.95 and 0.99 would signify that false rejections of the null hypothesis in 5 out of 100 frames and 1 out of 100 frames respectively could be attributed to random chance.

Thus, according to the above criteria, if we secure 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 or fourth order serial dependence in the data set.

---

[8] 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 as 1 minus the threshold value. Therefore, for a threshold of 0.9, we get a corresponding significance level 0.1 – that is, a significance level of 10%.

In principal, the tests can be applied to either the source returns data determined from application of (3)-(4) or to residuals from frame based autoregressive fits of this data. Recall that the latter can be viewed as a 'pre-whitening' operation and can be used to effectively remove second order (linear) serial dependence, thus producing no significant C frames. In this case, any remaining serial dependence left in the residuals must be a consequence of nonlinearity that is episodically present in the data - thereby, only significant H and H4 statistics will lead to the rejection of the null hypothesis of a pure noise process.

4 ENGLE LM ARCH TEST

In this article, we also investigate the issue of parameter instability of GARCH models and the transient nature of ARCH effects. The well-known Engle LM test for Autoregressive and Conditional Heteroscedasticity (ARCH) in residuals of a linear model was originally proposed in Engle (1982). This test should have power against more general GARCH alternatives, see Bollerslev (1986). The test statistic is based on the R^2 of the following auxiliary regression

$$x_t^2 = \beta_0 + \sum_{i=1}^{p} \beta_i x_{t-i}^2 + \xi_t \ , \tag{14}$$

where x_t^2 are typically squared residuals from a linear regression. Therefore, equation (14) 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 (14) 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 in (14).

The ARCH testing procedure that is applied in this article involves applying the LM test to the squared data in each sample frame. As in the case of the application of C, H and H4 statistics on a frame-by-frame basis, this data will typically be the (squared) residuals from a frame-by-frame 'pre-whitening' AR(p) fit in the case of the ARCH LM test. One key aspect of interest with this test procedure will be to determine whether there is a strong ARCH effect over all time periods (i.e. all sample frames) or whether ARCH is present only for short periods of time, for example, in a relatively small number of sample frames. It should also be noted that the same arguments made in the previous section in relation to false alarm thresholds and extent of rejections that can be attributed to random chance will continue to hold in this current case.

The ARCH test is only applied to the spot price data. The load data does not exhibit any 'volatility clustering' affects that generate the conditional variance dependence (conditional heteroskedasticity) that the ARCH test is designed to identify. The spot price data, on the other hand, does display the type of patterns conventionally associated with conditional heteroskedasticity.

5 EMPIRICAL RESULTS

In Table 1 and Table 2 the summary statistics of the NEM State load and spot price returns series are documented. It is apparent from inspection of both tables that the mean of the series are very small in magnitude. In Table 1, the mean returns for the load data are all positive while the average returns for the four spot price returns series listed in

Table 2 were negative over the complete sample. A difference in scale can also be observed from an investigation of the maximum and minimum values of the respective returns series. For the load data, the maximum and minimum returns are in the range between 30 and 50 percent in absolute terms while the corresponding results for the spot price returns are of the order of 480 to 610 percent. Moreover, the differences in the values of the sixth order cumulants listed in both tables also reinforce the obvious difference in scale of the different series.

It is also evident from inspection of both tables that the spot price returns are more volatile when compared with the load data as indicated by the higher standard deviations documented in Table 2 compared to those listed in Table 1. This indicates that the likely ‘risk profile’ of the load and spot price returns is quite different. Furthermore, volatility in both load and spot prices is slightly higher for SA than for the other States considered - SA has the highest standard deviations for both load and spot price returns data.

All of the series, except for SA spot price returns, display positive or right skewness. All of the series also display evidence of leptokurtosis although this is a much more prominent feature in the case of the spot price returns data with excess kurtosis values in the range of 69 to 104 in magnitude. This implies that the tails of the empirical distribution functions of the spot price returns in particular taper down to zero much more gradually than would the tails of the normal distribution (Lim, Hinich, Liew (2005)). Not unexpectedly, the Jarques-Bera (JB) Normality Test for all of the returns series listed in both tables 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 non-zero skewness and excess kurtosis listed in both tables.

The 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 is finite since all measurements have an upper bound to their magnitudes.[9] However, if the data is leptokurtic, as is typically the case for stock returns, exchange rate and energy spot prices, than 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 a more rapid convergence to the asymptotic (theoretical) distribution.

In Wild et al (forthcoming), it was demonstrated in context of weekly NSW spot price rates of returns that trimming in the order of 10%-90% provides acceptable results in terms of the empirical distribution of the tests, closely tracking their desired theoretical distributions. In order to improve the finite sample approximation of the theoretical distribution of all four statistics considered in this article, the spot price returns data was subsequently trimmed using a '10%-90%' scheme. This means that data values that either exceeded the 90% quantile or were less than the 10% quantile of the empirical distribution of the spot price returns data series were set to those two particular quantile values, respectively.

[9] In the current context, a maximum spot price that can be bid by wholesale market participants is $10000/ 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 AEMO (2009)). Thus, the range of the spot price data is finite ensuring that all moments are finite.

In the case of the load returns data, departures from Gaussianity were much less pronounced as indicated, in particular, by the much smaller excess kurtosis values listed in Table 1. So, for the load returns data, we employed '1%-99%' trimming.

The summary statistics of the 'trimmed' load and spot price returns data series are documented in Table 3 and Table 4 respectively. The effect of the trimming operations in reducing the scale of the data is clear. For example, in Table 4, the corresponding results for the spot price returns are now of the order of 15.2 to 18.5 percent in absolute terms which can be compared against the range of 480 to 610 percent originally cited in Table 2. These reductions are also evident in the much smaller values of the sixth order cumulants listed in both Tables 3 and 4, compared with those listed in Tables 1 and 2.

The skewness and excess kurtosis values cited in Tables 3 and 4 have also been reduced in magnitude (especially in the case of the spot price returns) in line with a movement in the empirical distributions functions towards the Gaussian ideal. However, in all cases, the trimmed data still 'trips' the Jarques-Bera Normality test, leading to strong rejections at the 1% level of significance. Furthermore, the volatility findings continue to hold with the trimmed spot price returns remaining more volatile. The other noticeable feature is that the mean of the spot price returns has increased in magnitude when compared with the results cited in Table 2. The practical implications of this are marginal however as the data is standardized before the tests are applied to the trimmed returns data series.

Table 5 presents the results for the correlation [C], bicorrelation [H] and tricorrelation [H4] test statistics for the trimmed load returns data for a weekly sample frame of 336 (half hourly) observations. In all results reported, bootstrapped threshold values were used because the sample properties of the test statistics for very small frame lengths do not necessarily approximate theoretical thresholds, especially when the underlying sample data contains both significant non-zero skewness and excess kurtosis. Given the 'trimmed' global sample of 185,162 returns for each respective series, a bootstrap sample frame was constructed by randomly sampling 336 observations from the larger global population and the various test statistic outcomes 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 entail application of the chi-square distribution and for each bootstrap replication, the chi square levels variable associated with each test statistic was transformed to a uniform variate which means, for example, that the 10% significance threshold corresponds to 0.90, the 5% significance threshold is 0.95, and the 1% significance threshold is 0.99. The arrays containing the bootstrap 'confidence thresholds' for each respective test statistic (containing 500,000 elements) from the bootstrap process was then sorted in ascending order and the bootstrap confidence threshold was calculated as the quantile value of the empirical distribution function of the various test statistics associated with a user specified 'false alarm' threshold value. For example, if the user set the false alarm threshold value to 0.90, the bootstrap threshold value would be the 90% quantile of the empirical distribution function of the relevant test statistic determined from the bootstrap process.[10]

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 threshold when the tests are applied on a sequential frame by frame basis to the actual returns data. As such, the rejections rates determined from the actual returns are size

[10] We used three particular user specified false alarm threshold values corresponding to 0.9, 0.95 and 0.99, giving 10%, 5% and 1% levels of significance.

adjusted with the empirical sizes being determined from the bootstrap method mentioned above. A frame based rejection is secured if, for an actual frame, the calculated threshold value exceeds the bootstrap determined false alarm threshold.[11] 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.

The results for the weekly load returns data presented in Table 5 were determined after applying a 'global AR(340)' fit to the complete sample data.[12] This operation was employed purely to remove second order serial dependence. The AR lag length of 340 was chosen to exceed the weekly frame length of 336 observations. This regression can be viewed as essentially a type of weekly detrending operation and operates to remove the mean weekly periodicity from the underlying data series. The residuals from this global AR fit are then used to determine the bootstrap thresholds and underpin other empirical results obtained for the load returns data. To further eliminate second order serial dependence, an 'AR(10)' fit is applied on a frame by frame basis. The success of these combined prewhitening operations is evidenced by the fact that no significant C frames were found (see Column 4 of Table 5) in contrast to the much greater number of H and H4 frames found to be significant – see Columns 5 and 6 of Table 5).

Recall that, for the false alarm thresholds of 0.90, 0.95 and 0.99 respectively, we expect only 10%, 5% and 1% of the total number of frames to secure rejections that can be reasonably attributed to random chance. The fact that the actual number of rejections are significantly higher than 10%, 5% and 1% of the total number of frames for both the

[11] We term such frames 'significant' frames with respect to the relevant test statistic.
[12] For the frame length of 336, the number of lags employed for the C and H statistics were determined to be 10 and the number of lags for the H4 test was determined to be 6. The number of bicovariances and tricovariances used were determined to be 45 and 20 respectively.

H and H4 tests signify the existence of statistical significant third-order and fourth-order (nonlinear) serial dependence in the load returns data, supporting the presence of a nonlinear generating mechanism in weekly load dynamics. The fact that we do not secure rejections for all frames points to the higher order nonlinear serial dependence being somewhat episodic - there are frames where the null hypothesis of pure noise cannot be rejected. Similar interpretations can be given to all other test results cited in Table 5. It is also apparent from Table 5 that fourth-order nonlinear serial dependence seems to be a more prevalent feature in the data than third-order nonlinear serial dependence – the number of frame based rejections for the H4 test (column 6) generally exceeds the number of frame based rejections for the H test (column 5) at all three bootstrap false alarm thresholds reported.

In Table 6, the results for the three portmanteau tests, and additionally the LM ARCH test, are presented for the spot price returns. In this case, no global prewhitening was undertaken (in contrast to the load returns) although the frame by frame based 'AR(10)' prewhitening fit continued to be employed, thus suggesting a different type of dynamic driving the mean periodicity of the spot price returns data. It is evident from Table 6 (Column 4) that the prewhitening operation has been successful – no significant C frames are evident. However, there is a lot of evidence of significant H and H4 based frame rejections reported in Columns 5 and 6. The nature of the rejections indicates that both third- and fourth-order nonlinear serial dependence is much more prominent in the spot price returns data – the extent of the frame based H and H4 rejections are in the range of 55%-98% for all States and all bootstrapped false alarm thresholds considered. This can be compared with the corresponding 20%-85% range for the load returns data displayed in Table 5. Furthermore, relatively more H statistic rejections are reported in Table 6 when compared to the corresponding H4 rejections, unlike in Table 5. The frame by frame LM ARCH tests signify the presence of pure ARCH/GARCH structure in the spot price returns data. However, the order of magnitude, while significant, is lower than that associated with H statistic based frame rejections, although it is often of a similar order of magnitude to the H4 statistic based frame rejections rates.

The results associated with the 'hard clipping' transformation of the residuals from the frame by frame 'AR(10)' fits of the spot price returns are documented in Table 7. These are the same set of residuals that underpins the results in Table 6 except that the transformation in (13) was subsequently applied to the residuals, prior to applying the portmanteau tests, with the ARCH LM test being dropped. It is evident that the number of frame based rejections for the H and H4 statistics applied to the binary data sets are greater than the 10%, 5% and 1% rates associated with random chance, thus pointing to structures that cannot be modeled as pure ARCH/GARCH or stochastic volatility models. The relatively larger number H statistic rejections in Column 5 suggest that third-order nonlinear serial dependence is the most prominent type of nonlinear serial dependence.

We also investigate the presence and nature of any nonlinear serial dependence evident in the dynamics of the daily load and spot price returns. This is accomplished by choosing an underlying frame length of a day (48 half hours). The resulting analysis proceeds as before with the frame length set to 48 instead of 336. The total number of frames under investigation increases from 551 to 3,857 for the daily returns data.[13]

[13] For the frame length of 48, the number of lags employed for the C and H statistics were determined to be 5 and the number of lags for the H4 test was determined to be 3. The number of bicovariances and tricovariances used were determined to be 10 and 1 respectively.

The daily load returns results are reported in Table 8. Once again, we employ a ‘global AR(340)’ prewhitening fit to remove the mean weekly periodicity. In performing this operation, we also remove the mean daily periodicity because this periodicity is a harmonic of the weekly periodicity. We also adopt a frame-by-frame ‘AR(5)’ prewhitening fit. These combined prewhitening operations ensure that the number of significant C frames is very small – less than 0.05 of one percent of the total number of frames considered (see Column 4). This mirrors the results obtained in Table 5 in relation to the weekly data. As such, second-order (linear) serial dependence has been removed through the combined prewhitening process so any further rejections of the null hypothesis of pure white noise must be attributable to either H or H4 based rejections, indicating the presence of third- or fourth-order (nonlinear) serial dependence. As in the case of the weekly returns results in Table 5, there is evidence of nonlinear serial dependence but at a much lower order of magnitude - the frame based rejection for the H and H4 test statistics now occur at rates in the range of 2%-25% compared against the 20%-85% range associated with the weekly load returns. Moreover, inspection of Table 8 also indicates that third-order nonlinear serial dependence is slightly more prominent.

The results for the daily spot price returns are reported in Table 9. We adopt the same prewhitening scheme that was adopted for the weekly spot price returns – no global prewhitening but a frame by frame based ‘AR(5)’ prewhitening fit – in order to remove second order (linear) serial dependence. Again, we get a very low number of significant C frames (Column 4) – less than 0.05 of one percent of the total number of frames. There is also evidence of the presence of nonlinear serial dependence – the number of significant H and H4 frames significantly exceeds the 10%, 5% and 1% rates that can be reasonably attributed to random chance. The order of magnitude of the frame based rejections for H and H4 are in the range of 5%-35% which is smaller than the corresponding range in Table 6 of 55%-98%. Thus, the presence of nonlinear serial dependence is a less prominent feature of the daily spot price returns data. Inspection of the last column of Table 9 also indicates the presence of 'marginally' significant GARCH structure but at a level that is much lower compared to the weekly returns results.

The results associated with the 'hard clipped' transformation applied to the residuals of the frame-by-frame based 'AR(5)' fits are reported in Table 10. It is apparent that we cannot secure rates of rejection that point to the presence of 'non-GARCH' alternatives at the accepted significance levels.

Overall, the results suggest that nonlinear serial dependence plays a much less prominent role in explaining the evolution of daily load and spot price return dynamics when compared to weekly returns. The ARCH LM test results reported in Table 9 indicate that GARCH effects play a much more marginal role in explaining nonlinearity evident in daily spot price returns. These conclusions are further reinforced by the hard clipping results reported in Table 10. Therefore, a definite type of 'time scale' effect appears to be in operation. Nonlinear serial dependence appears to play a much greater role in explaining dynamics in both load and spot price returns dynamics over a weekly time scale rather than a daily time scale. This backs up the results reported in Brooks and Hinich (1998) and Ammermann and Patterson (2003) in relation to the application of LM ARCH test on a frame-by-frame basis. Specifically, what we are seeing is that for very small frame lengths (i.e. of a day), there is increasingly longer periods of time during which there is no evidence of linear or non-linear serial dependence, including ARCH effects, in spot price returns. Nonlinear serial dependence is very episodic at this particular time scale. However, as the frame length is aggregated (i.e. increased to a week), these episodic effects are assimilated into both linear and nonlinear structures with increased incidence of frame based rejections of H, H4 and ARCH LM tests.

However, unlike the findings in Brooks and Hinich (1998) and Ammermann and Patterson (2003), the extent of aggregation from a day to a week is not large within the context of the overall sample being considered and the extent of the relatively large number of frame based rejections cited in Tables 5-7, in particular, do indicate the statistically significance presence of nonlinear serial dependence operating on a weekly time scale. This has not been observed in other studies utilizing the test methods employed here, for example, see Hinich and Patterson (1989, 2005), Brooks (1996), Brooks and Hinich (1998), Ammermann and Patterson (2003), Lim, Hinich and Liew (2003, 2004, 2005), Lim and Hinich (2005a, 2005b), Bonilla, Romero-Meza and Hinich (2007) and Czamanski, Dormaar, Hinich and Serletis (2007). Interestingly, the extent and pattern of rejections documented here significantly exceed the rejection patterns documented in Czamanski, Dormaar, Hinich and Serletis (2007) in relation to the Alberta market.

6 KEY IMPLICATIONS OF OUR FINDINGS FOR THE MODELING OF SPOT PRICE DYNAMICS AND GENERAL RISK MANAGEMENT CONSIDERATIONS WITHIN GROSS POOL PRIVATISED WHOLESALE ELECTRICITY MARKETS

Risk management practices within privatised gross pool wholesale electricity markets have historically been based upon bi-lateral contracts between demand and supply side participants. However, more recently, options based financial instruments have begun to penetrate the Australian market.[14] The pricing of these particular instruments have been primarily based upon the options pricing model initiated by Black and Scholes (1973) and Merton (1973). A central feature of this type of model is the use of Geometric Brownian Motion (GBM) models to capture the time evolution of asset price rates of return. The discrete-time version of the GBM model is a Gaussian random walk.[15] The first difference of this process can be modelled as a sequence of non-Gaussian, identically and independently distributed random variables with finite moments that are passed though a symmetric, absolutely summable band-limited filter. As such, this process is linear.

Given the prevalence of observed price spikes in electricity markets, the GBM model was subsequently augmented to include additive jump processes modeled typically by Poisson processes that model the probability of a jump occurring. These models can be viewed as encompassing a smooth continuous sample path process [the diffusion (or GBM) part] and a much less persistent discontinuous jump component [see Merton (1976), Bunn and Karakatsani (2003) and Chan, Gray and van Campen (2008)].

A key property of the diffusion component of a jump-diffusion model is that it is both Gaussian and linear. Any departures are associated with the jump component. This, in turn, is associated with the presence of conditional variance dependence (alternatively called conditional heteroskedasticity or multiplicative non-linearity). One stylized fact that the conventional jump-diffusion models have had trouble modeling is the observed

[14] The key energy derivative (futures) market in Australia is that operated by d-cypha Trade. Information about these products is located at: http://d-cyphatrade.com.au/.
[15] A detailed account outlining the derivation and use of GBM and its extensions can be found in Clewlow and Strickland (2000, Ch 2 and 3), Eydeland and Wolyniec (2003, Ch 11) and Geman (2005, Ch 3).

time variation in volatility. In many respects, the development and use of ARCH/GARCH and stochastic volatility models have represented attempts to devise statistical models that are able to accommodate time varying volatility in a statistically parsimonious way [see Leon and Rubia (2004), Higgs and Worthington (2005) and Solibakke (2006)]. However, a limitation of these models is the widespread tendency to equate non-linear serial dependence solely with conditional heteroskedasticity.

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 recourse to linear time series models [e.g., see Worthington and Higgs (2003), Hadsell, Marathe and Shawky (2004), Leon and Rubia (2004), Garcia and Contreras (2005), Higgs and Worthington (2005), Solibakke (2006), Bowden and Payne (2008) and Higgs (2008)]. The residuals from this model then constitute the 'zero mean' process that underpins theoretical discussion of these 'volatility' models

However, a potential problem emerges when the mean of the process is nonlinear. Erroneous conclusions of 'nonlinearity-in-variance' can emerge when the prime source of serial dependence in the residuals is nonlinear structure that could not be successfully extracted by conventional linear-based time series models. From a diagnostic testing perspective, 'nonlinear-in-mean' structure in the residuals would be detected using higher order analogues of conventional Portmanteau test for serial correlation such as the H and H4 test statistics utilized in this article.

The use of trimming allows us to directly control for the affects of outliers (i.e. jumps) in the data. As such, we are focusing attention upon the inter-quartile range of the empirical distribution function of the returns data that would be associated with the diffusion component of the jump-diffusion process, for example. While the mean (e.g. diffusion component) of the GBM framework used in the Black-Scholes-Merton option pricing framework is Gaussian and linear, our findings, on the other hand, indicate that the mean properties of the actual returns data are, in fact, additively non-linear – that is, 'non-linear in mean'.

If linear models are used to remove the mean of the process, any 'nonlinearity in mean' structure will end up in the residuals of the fitted model and can subsequently 'trip' ARCH tests.[16] However, in this case, equating non-linear serial dependence with conditional heteroskedasticity is erroneous. Moreover, 'volatility' modeling that led to the acceptance of linear specifications (with conditionally heteroscedastic disturbances) would represent a misspecification of the actual process in statistical terms.

Some of these problems have been recognized in the literature and this has led to attempts to employ linear operators, such as step or seasonal dummy variables[17] or 'seasonal' differencing operators[18] to model deterministic and stochastic daily, weekly and seasonal patterns. However, these cannot account for the 'higher-order' nonlinear dependencies that we have observed in both the load and spot price returns data series. Furthermore, the bi-correlations and tri-correlations encompass third and four order moment products. As such, the order of magnitude implied in the nonlinear structure being detected by the H and H4 statistics is of a higher order than that being detected and

[16] Ashley, Patterson and Hinich (1986) formally prove that if the data generating mechanism is nonlinear and a linear filter (i.e. time series model) is fitted to the nonlinear data, then the nonlinear structure will end up in the residuals of the fitted linear model.
[17] Consult Hadsell, Marathe and Shawky (2004), Chan, Gray and van Campen (2008), Higgs and Worthington (2008), Karakatsani and Bunn (2008) and Weron and Misiorek (2008).
[18] For example, see Bowden and Payne (2008), and Weron and Misiorek (2008).

modeled by GARCH and stochastic volatility processes which entail second order products implied in the squaring of fitted residuals.

The existence of statistically significant bi-correlation and tri-correlation structure in the returns data also pose problems for focusing purely on the second order volatility measure encompassed in the conventional standard deviation based measure of volatility. Not only will the higher order structure significantly contribute to the time evolution of the returns data itself, but, in amplitude terms, can dominate the contribution of the conventional second order standard deviation measure emphasized in the broader finance literature.

Finally, the testing procedure associated with the 'hard clipping' procedure outlined in equation (13) constitutes a direct test of the adequacy of symmetric ARCH/GARCH and stochastic volatility models. The results cited in Table 7 in relation to weekly spot price dynamics, in particular, indicate that these models cannot adequately capture the statistically significant non-linear dependence structure contained in the spot price returns data.

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 modeling are applied. These matters are of considerable importance in managing risk in electricity markets, particularly in option pricing. Only recently, we have observed the damage that has been done by incorrect option pricing based upon related applications of the Black, Scholes and Merton modeling methodology in financial markets that had questionable, if any, empirical basis.

7 CONCLUDING COMMENTS

In this article, we have tested for the presence of nonlinear serial dependence in NEM State daily and weekly load and spot electricity price data. This task was accomplished by applying the portmanteau correlation, bicorrelation and tricorrelation tests introduced in Hinich (1996) to the time series of half hourly spot prices. These tests were used to detect epochs of transient serial dependence in a discrete-time pure white noise process. The test framework involves partitioning the time series data into non-overlapping 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 were also investigated by applying the Engle LM ARCH test and, additionally, using a detection framework based upon converting a martingale difference process into a pure noise process and then testing for the presence of linear and nonlinear serial dependence.

Nonlinear serial dependence was found to be present in both daily and weekly load and spot price returns data. However, a 'time scale' effect was found to be present. Specifically, nonlinear serial dependence was found to be a much more prominent feature in both the load and spot price returns dynamics over a weekly time scale rather than a daily time scale. At the daily time scale, we found increasingly long periods during which there is no evidence of linear or non-linear serial dependence, including ARCH effects in load or spot price dynamics, followed by episodes of nonlinear dependence of limited duration. GARCH effects appeared to be a more prominent in the weekly dynamics of spot price returns than was the case with the daily dynamics. At the weekly time scale, there is significant evidence of nonlinear serial dependence. This finding most likely reflects the strong weekly periodicities found in both the load and spot price returns data which were identified in Foster et al (2008) using the RMP model. The finding of nonlinearity provides some added support for the proposition made in Hinich (2000) and Hinich and Wild (2001) that the generating mechanism for an RMP process is essentially nonlinear. The added finding of episodic nonlinearity is in line with the commonly accepted 'stylised' fact of strong mean reversion in spot electricity prices.

The finding of nonlinearity has important implications for modeling weekly and daily load and spot price dynamics and, therefore, for risk management. Because of the prevalence of both third- and fourth-order nonlinear serial dependence in the data, time series models that are linear in construction or assume a pure noise input, such as GBM stochastic diffusion models, are problematic. Their dependence structure violates both normality and Markovian assumptions that underpin conventional GBM models. Furthermore, the trimming methodology used in this article, indicates that observed nonlinear serial dependence is being generated by more than just the presence of outliers. Strong evidence of third- and fourth-order nonlinear serial dependence was found in all of the 'trimmed' weekly spot price returns data. So, from a risk management perspective, particularly in the context of option pricing models, serious questions arise about the ability of jump diffusion models, in particular, to adequately capture nonlinearity. Errors in pricing, induced by the use of such models, could prove very costly to participants in a market that is characterized by high levels of price volatility.

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**(2), 175-195.

Ashley, R.A., Patterson, D.M., and M.J. Hinich. 1986. A diagnostic test for nonlinear serial dependence in time series filtering errors. *Journal of Time Series Analysis* **7**(3), 165-178.

Australian Energy Market Operator (AEMO). 2009. An Introduction to Australia's National Electricity Market. Australian Energy Market Operator, July 2009.Available at: (http://www.aemo.com.au/corporate/0000-0006.pdf).

Black, F. and M. Scholes. 1973. The Pricing of Options & Corporate Liabilities. *Journal of Political Economy*, **81,** 637-59.

Bollerslev, T. 1986. Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3), 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.

Bonilla, C. A., Romero-Meza, R., and M. J. Hinich. 2007. GARCH inadequacy for modeling exchange rates: empirical evidence from Latin America. *Applied Economics* **39**(19), 2529-2533.

Bowden, N., and J. E. Payne 2008. Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA-EGARCH models. *Energy Economics **30***, 3186-3197.

Brooks, C. 1996. Testing for non-linearity in daily sterling exchange rates. *Applied Financial Economics* **6**(4), 307-317.

Brooks, C. and M. J. Hinich. 1998. Episodic nonstationarity in exchange rates. *Applied Economics Letters* **5**(11), 719-722.

Bunn, D. W., and N. V. Karakatsani. 2003. Forecasting Electricity Prices. *London Business School, Working Paper,* v1.

Chan, K. F., Gray, P., and B. van Campen. 2008. A new approach to characterizing and forecasting electricity price volatility. *International Journal of Forecasting* **24**, 728-743.

Clewlow, L., and C. Strickland. 2000. Energy Derivatives: Pricing and Risk Management. London: Lacima.

Czamanski, D., Dormaar, P. Hinich, M. J., and A. Serletis 2007. Episodic Nonlinearity and Nonstationarity in Alberta's Power and Natural Gas Markets. *Energy Economics* **29**, *94-104.*

Engle, R. F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4), 987-1007.

Eydeland, A., and K. Wolyniec. 2003. Energy and Power Risk Management: New Developments in Modeling, Pricing and Hedging. New Jersey: John Wiley & Sons.

Foster, J., Hinich, M. J. and P. Wild. 2008. Randomly Modulated Periodic Signals in Australia's National Electricity Market. *The Energy Journal* **29**(3), 105-130.

Garcia, R. C., Contreras, J., van Akkeren, M., and J. B. C. Garcia. 2005. A GARCH Forecasting Model to predict Day-Ahead Electricity Prices. *IEEE Transaction on Power Systems 2**0***(2), 867-874.

Geman, H. 2005. Commodities and Commodity Derivatives – Modeling and Pricing for Agriculturals, Metals and Energy. Chichester: John Wiley & Sons.

Hadsell, L., Marathe, A., and H. A. Shawky. 2004. Estimating the Volatility of Wholesale Electricity Spot Prices in the US. *The Energy Journal* **25**(4), 23-40.

Higgs, H., and A. Worthington. 2005. Systematic Features of High-Frequency Volatility in Australian Electricity Markets: Intraday Patterns, Information Arrival and Calendar Effects. *The Energy Journal* **26**(4), 23-41.

Higgs, H. 2008. Modeling price and volatility inter-relationships in the Australian wholesale spot electricity markets. Working Paper, Department of Accounting, Finance and Economics, Griffith University, Nathan, Queensland, Australia.

Higgs, H., and A. Worthington. 2008. Stochastic price modeling of high volatility, mean reverting, spike prone commodities: The Australian wholesale spot electricity market. *Energy Economics* **30**, 3172-3185.

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. 1996. Testing for dependence in the input to a linear time series model. *Journal of Nonparametric Statistics* **6**(2-3), 205-221.

Hinich, M. J. 2000. A Statistical Theory of Signal Coherence. *Journal of Oceanic Engineering* **25**(2), 256-261.

Hinich, M. J. and P. Wild. 2001. Testing Time-Series Stationarity Against an Alternative Whose Mean is Periodic. *Macroeconomic Dynamics* **5**(3), 380-412.

Hinich, M. J. and D. M. Patterson. 2005. Detecting Epochs of Transient Dependence in White Noise. In: Belongia, M. and J. Binner, (Eds), *Money, Measurement and Computation*. 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**(477), 68-74.

Karakatsani N. V., and D. W. Bunn. 2008. Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. *International Journal of Forecasting* **24**, 764-785.

Leon, A., and A. Rubia. 2004. Forecasting Time-Varying Covariance Matrices in the Intradaily Spot Market in Argentina. In: D. W. Bunn, (Ed), *Modelling Prices in Competitive Electricity Markets*. Chichester: John Wiley & Sons.

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

Lim, K. P. and M. J. Hinich. 2005b. Non-linear market behavior: events detection in the Malaysian stock market. *Economics Bulletin* **7**(6), 1-5.

Lim, K. P., Hinich, M. J. 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**(2), 79-93.

Lim, K. P. 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), 5-18.

Lim, K-P., Hinich, M. J. and V. K. Liew. 2005. Statistical Inadequacy of GARCH Models for Asian Stock Markets: Evidence and Implications. *Journal of Emerging Market Finance* **4**(3), 263-279.

Merton, R.C. 1973. Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science* **4** (1), 141–183.

Merton, R.C. 1976. Option Pricing When Underlying Stock Returns are Discontinuous. *Journal of Financial Econometrics* **3**, 125–144.

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 Cox, D.R, Hinkley, D.V and Ole E. Barndorff-Nielsen (ed.), *Time Series Models in Econometrics, Finance and Other Fields,* pp. 1-67. London: Chapman & Hall.

Solibakke, P. B. 2006. Describing the Nordic Forward Electric-Power Market: A Stochastic Model Approach. *International Journal of Business* ***11***(4), 345-366.

Serletis, A. 2007. Quantitative and Empirical Analysis of Energy Markets. New Jersey: World Scientific.

Weron, R., and A. Misiorek. 2008. Forecasting spot electricity prices: A comparison of parametric and semi-parametric time series models. *International Journal of Forecasting* **24**, 744-763.

Wild, P., Foster, J. and M. J. Hinich. forthcoming. Identifying nonlinear serial dependence in volatile, high frequency time series and its implications for volatility modeling. *Macroeconomic Dynamics*.

Worthington, A., and H. Higgs. 2003. A multivariate GARCH analysis of the domestic transmission of energy commodity price and volatility: A comparison of the peak and off-peak periods in the Australian electricity spot market. Discussion Paper No 140, School of Economics and Finance, Queensland University of Technology, Brisbane, Australia.

Table 1. Summary Statistics for Load Returns Data

	NSW	QLD	VIC	SA
No of Observations	185162	185162	185162	185162
Mean	0.0002	0.0002	0.0002	0.0001
Maximum	36.80	42.3	40.2	32.5
Minimum	-30.90	-38.5	-41.3	-49.4
Std Dev	3.03	2.79	2.91	3.43
Skewness	1.01	0.84	0.94	0.34
Excess Kurtosis	1.52	1.71	1.34	1.25
6^{th} Order Cumulant	9.82	85.42	66.24	51.16
JB Test Statistic	49500.0	44300.0	41100.0	15600.0
JB Normality P-Value	0.0000	0.0000	0.0000	0.0000

Table 2. Summary Statistics for Spot Price Returns Data

	NSW	QLD	VIC	SA
No of Observations	185162	185162	185162	185162
Mean	-0.0003	-0.0001	-0.0004	-0.0005
Maximum	545.0	591.0	497.0	597.0
Minimum	-572.0	-531.0	-488.0	-610.0
Std Dev	18.9	25.7	20.1	26.2
Skewness	0.42	0.25	0.31	-0.45
Excess Kurtosis	104.0	78.8	69.8	80.0
6^{th} Order Cumulant	42740.5	15446.0	21020.3	17415.5
JB Test Statistic	84100000.0	47900000.0	37600000.0	49300000.0
JB Normality P-Value	0.0000	0.0000	0.0000	0.0000

Table 3. Summary Statistics for 'Trimmed' Load Returns Data

	NSW	QLD	VIC	SA
No of Observations	185162	185162	185162	185162
Mean	-0.0006	-0.0003	-0.001	0.0004
Maximum	2.12	1.98	2.30	3.32
Minimum	-1.98	-1.95	-2.21	-3.32
Std Dev	0.72	0.72	0.78	1.21
Skewness	0.08	0.02	0.06	0.01
Excess Kurtosis	0.58	0.31	0.73	0.41
6^{th} Order Cumulant	-3.62	-2.84	-3.85	-3.46
JB Test Statistic	2780.0	772.0	4270.0	1320.0
JB Normality P-Value	0.0000	0.0000	0.0000	0.0000

Table 4. Summary Statistics for 'Trimmed' Spot Price Returns Data

	NSW	QLD	VIC	SA
No of Observations	185162	185162	185162	185162
Mean	-0.316	-0.229	-0.287	-0.122
Maximum	16.1	16.1	18.5	18.5
Minimum	-15.2	-15.2	-16.9	-17.3
Std Dev	9.22	9.16	10.5	10.6
Skewness	0.16	0.15	0.20	0.14
Excess Kurtosis	-0.69	-0.66	-0.73	-0.70
6^{th} Order Cumulant	1.32	0.98	1.51	1.35
JB Test Statistic	4510.0	3980.0	5370.0	4360.0
JB Normality P-Value	0.0000	0.0000	0.0000	0.0000

Table 5. Frame Test Results for 'Trimmed' Weekly Spot Price (Returns) Data
Specific Details: Removed 'Weekly Mean' By Global AR(340) 'Prewhitening' Fit;
Applied Frame by Frame AR(10) 'Prewhitening' Fit to Remove Linear Dependence

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
NSW	551	0.90	0 (0.00%)	330 (59.89%)	442 (80.22%)
	551	0.95	0 (0.00%)	271 (49.18%)	398 (72.23%)
	551	0.99	0 (0.00%)	161 (29.22%)	281 (51.00%)
QLD	551	0.90	0 (0.00%)	246 (44.65%)	364 (66.06%)
	551	0.95	0 (0.00%)	189 (34.30%)	312 (56.62%)
	551	0.99	0 (0.00%)	103 (18.69%)	220 (39.93%)
VIC	551	0.90	0 (0.00%)	308 (55.90%)	446 (84.57%)
	551	0.95	0 (0.00%)	244 (44.28%)	419 (76.04%)
	551	0.99	0 (0.00%)	161 (29.22%)	313 (56.81%)
SA	551	0.90	0 (0.00%)	282 (51.18%)	377 (68.42%)
	551	0.95	0 (0.00%)	214 (38.84%)	311 (56.44%)
	551	0.99	0 (0.00%)	116 (21.05%)	188 (34.12%)

Table 6. Frame Test Results for 'Trimmed' Weekly Spot Price (Returns) Data
Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(10) 'Prewhitening' Fit to Remove Linear Dependence

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)	Significant ARCH Frames Num & (%)
NSW	551	0.90	0 (0.00%)	530 (96.19%)	488 (88.57%)	461 (83.67%)
	551	0.95	0 (0.00%)	519 (94.19%)	458 (83.12%)	430 (78.04%)
	551	0.99	0 (0.00%)	464 (84.21%)	357 (64.79%)	352 (63.88%)
QLD	551	0.90	0 (0.00%)	519 (94.19%)	482 (87.48%)	509 (92.38%)
	551	0.95	0 (0.00%)	499 (90.56%)	444 (80.58%)	479 (86.93%)
	551	0.99	0 (0.00%)	430 (78.04%)	361 (65.52%)	417 (75.68%)
VIC	551	0.90	0 (0.00%)	544 (98.73%)	518 (94.01%)	500 (90.74%)
	551	0.95	0 (0.00%)	532 (96.55%)	494 (89.66%)	462 (83.85%)
	551	0.99	0 (0.00%)	512 (92.92%)	433 (78.58%)	386 (70.05%)
SA	551	0.90	0 (0.00%)	502 (91.11%)	451 (81.85%)	476 (86.39%)
	551	0.95	0 (0.00%)	475 (86.21%)	409 (74.23%)	437 (79.31%)
	551	0.99	0 (0.00%)	409 (74.23%)	298 (54.08%)	355 (64.43%)

Table 7. Frame Test Results for 'Trimmed' Weekly Spot Price (Returns) Data

Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(10) 'Prewhitening' Fit to Remove Linear Dependence
Frame by frame Hard Clipping of Residuals

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
NSW	551	0.90	42 (7.62%)	324 (58.80%)	124 (22.50%)
	551	0.95	26 (4.72%)	261 (47.37%)	75 (13.61%)
	551	0.99	10 (1.81%)	139 (25.23%)	22 (3.99%)
QLD	551	0.90	54 (9.80%)	259 (47.01%)	99 (17.97%)
	551	0.95	33 (5.99%)	193 (35.03%)	56 (10.16%)
	551	0.99	13 (2.36%)	96 (17.42%)	17 (3.09%)
VIC	551	0.90	53 (9.62%)	387 (70.24%)	164 (29.76%)
	551	0.95	34 (6.17%)	332 (60.25%)	108 (19.60%)
	551	0.99	9 (1.63%)	212 (38.48%)	35 (6.35%)
SA	551	0.90	50 (9.07%)	273 (49.55%)	93 (16.88%)
	551	0.95	26 (4.72%)	216 (39.20%)	53 (9.62%)
	551	0.99	9 (1.63%)	106 (19.24%)	12 (2.18%)

Table 8. Frame Test Results for 'Trimmed' Daily Load Demand (Returns) Data
Specific Details: Removed 'Weekly Mean' By Global AR(340) 'Prewhitening' Fit; Applied Frame by Frame AR(5) 'Prewhitening' Fit to Remove Linear Dependence

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
NSW	3857	0.90	0 (0.00%)	862 (22.35%)	694 (17.99%)
	3857	0.95	0 (0.00%)	475 (12.32%)	409 (10.60%)
	3857	0.99	0 (0.00%)	131 (3.40%)	107 (2.77%)
QLD	3857	0.90	2 (0.05%)	783 (20.30%)	609 (15.79%)
	3857	0.95	2 (0.05%)	446 (11.56%)	362 (9.39%)
	3857	0.99	1 (0.03%)	103 (2.67%)	98 (2.54%)
VIC	3857	0.90	0 (0.00%)	696 (18.05%)	623 (16.15%)
	3857	0.95	0 (0.00%)	364 (9.44%)	328 (8.50%)
	3857	0.99	0 (0.00%)	73 (1.89%)	83 (2.15%)
SA	3857	0.90	0 (0.00%)	957 (24.81%)	790 (20.48%)
	3857	0.95	0 (0.00%)	558 (14.47%)	474 (12.29%)
	3857	0.99	0 (0.00%)	157 (4.07%)	127 (3.29%)

Table 9. Frame Test Results for 'Trimmed' Daily Spot Price (Returns) Data
Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(5) 'Prewhitening' Fit to Remove Linear Dependence

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)	Significant ARCH Frames Num & (%)
NSW	3857	0.90	2 (0.05%)	1189 (30.83%)	937 (24.29%)	646 (16.75%)
	3857	0.95	2 (0.05%)	790 (20.48%)	583 (15.12%)	383 (9.93%)
	3857	0.99	0 (0.00%)	284 (7.36%)	219 (5.68%)	126 (3.27%)
QLD	3857	0.90	0 (0.00%)	1347 (34.92%)	967 (25.07%)	729 (18.90%)
	3857	0.95	0 (0.00%)	902 (23.39%)	635 (16.46%)	452 (11.72%)
	3857	0.99	0 (0.00%)	363 (9.41%)	223 (5.78%)	158 (4.10%)
VIC	3857	0.90	1 (0.03%)	1283 (33.26%)	1106 (28.68%)	655 (16.98%)
	3857	0.95	0 (0.00%)	859 (22.27%)	733 (19.00%)	392 (10.16%)
	3857	0.99	0 (0.00%)	348 (9.02%)	271 (7.03%)	111 (2.88%)
SA	3857	0.90	0 (0.00%)	1104 (28.62%)	934 (24.22%)	614 (15.92%)
	3857	0.95	0 (0.00%)	734 (19.03%)	580 (15.04%)	360 (9.33%)
	3857	0.99	0 (0.00%)	281 (7.29%)	214 (5.55%)	130 (3.37%)

Table 10. Frame Test Results for 'Trimmed' Daily Spot Price (Returns) Data

Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(5) 'Prewhitening' Fit to Remove Linear Dependence
Frame by frame Hard Clipping of Residuals

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
NSW	3857	0.90	134 (3.47%)	426 (11.04%)	307 (7.96%)
	3857	0.95	78 (2.02%)	199 (5.16%)	166 (4.30%)
	3857	0.99	27 (0.70%)	50 (1.30%)	33 (0.86%)
QLD	3857	0.90	177 (4.59%)	428 (11.10%)	342 (8.87%)
	3857	0.95	108 (2.80%)	252 (6.53%)	173 (4.49%)
	3857	0.99	37 (0.96%)	88 (2.28%)	33 (0.86%)
VIC	3857	0.90	107 (2.77%)	397 (10.29%)	315 (8.17%)
	3857	0.95	44 (1.14%)	193 (5.00%)	155 (4.02%)
	3857	0.99	6 (0.16%)	49 (1.27%)	27 (0.70%)
SA	3857	0.90	112 (2.90%)	361 (9.36%)	303 (7.86%)
	3857	0.95	47 (1.22%)	169 (4.38%)	139 (3.60%)
	3857	0.99	12 (0.31%)	28 (0.73%)	28 (0.73%)