The Use of Trimming to Improve the Performance of Tests for Nonlinear Serial Dependence with Application to the Australian National Electricity Market

Phillip Wild*, Melvin J. Hinich** and John Foster***, School of Economics Discussion Paper No. 367, June 2008, School of Economics, The University of Queensland. Australia.

Full text available as:
PDF- Requires Adobe Acrobat Reader or other PDF viewer

Abstract

In this article, we build on the results reported in Wild, Hinich and Foster (2008) for the National Electricity Market (NEM) of Australia by testing for episodic nonlinearity in the dynamics governing weekly cycles in spot price time series data. We apply the portmanteau correlation, bicorrelation and tricorrelation tests introduced in Hinich (1996) and the Engle (1982) ARCH LM test to the time series of half hourly spot prices from 7/12/1998 to 29/02/2008. We use trimming to improve the finite sample performance of the various test statistics mentioned above given the presence of significant skewness and leptokurtosis in the source datasets which may adversely affect the convergence properties of the test statistics in finite samples. With trimming, we still find the presence of significant third and fourth order (non-linear) serial dependence in the weekly spot price data, pointing to the presence of 'deep' nonlinear structure in this data.

EPrint Type:	Departmental Technical Report
Keywords:	Bicorrelations, Tricorrelations, Episodic Nonlinearity, ARCH/GARCH, Australian Spot Electricity Prices
Subjects:	340000 Economics;
ID Code:	JEL Classification: C32, C53, C89.
Deposited By:	

Phillip Wild*
School of Economics & ACCS
The University of Queensland
St Lucia QLD 4072
Tel : +61 7 3346 9258
email : p.wild@uq.edu.au

** Applied Research Laboratories
University of Texas at Austin
Austin
TX 78712-1087
Tel : 1 512 232 7270
email : hinich@austin.utexas.edu

John Foster***
School of Economics,
The University of Queensland,
St Lucia. QLD 4072..Australia
Tel : +61 7 3365 6780
j.foster@uq.edu.au

.* Corresponding author.

The Use of Trimming to Improve the Performance of Tests for Nonlinear Serial Dependence with Application to the Australian National Electricity Market

Phillip Wild*, Melvin J. Hinich** and John Foster***

ABSTRACT

In this article, we build on the results reported in Wild, Hinich and Foster (2008) for the National Electricity Market (NEM) of Australia by testing for episodic nonlinearity in the dynamics governing weekly cycles in spot price time series data. We apply the portmanteau correlation, bicorrelation and tricorrelation tests introduced in Hinich (1996) and the Engle (1982) ARCH LM test to the time series of half hourly spot prices from 7/12/1998 to 29/02/2008. We use trimming to improve the finite sample performance of the various test statistics mentioned above given the presence of significant skewness and leptokurtosis in the source datasets which may adversely affect the convergence properties of the test statistics in finite samples. With trimming, we still find the presence of significant third and fourth order (non-linear) serial dependence in the weekly spot price data, pointing to the presence of 'deep' nonlinear structure in this data.

JEL Classifications: C32, C53, C89.

Keywords: Bicorrelations, Tricorrelations, Episodic Nonlinearity, ARCH/GARCH, Australian Spot Electricity Prices.

* Corresponding author. School of Economics and ACCS, University of Queensland, St Lucia, QLD, 4072, Australia, Tel: +61 7 3346 9258, email: p.wild@uq.edu.au.

** Applied Research Laboratories, University of Texas at Austin, Austin, TX 78712-1087, Tel: +1 512 232 7270, email: hinich@austin.utexas.edu.

*** School of Economics, University of Queensland, St Lucia, QLD, 4072, Australia, Tel: +61 7 3365 6780, email: j.foster@uq.edu.au.

1. INTRODUCTION

The Australian electricity market as a whole encompasses both supply and demand side interactions encompassing generation, transmission, distribution and retail sale activities. The predominant market in Australia is the National Electricity Market (NEM) which is structured as a gross pool institution. The NEM commenced operation as a de-regulated wholesale market in New South Wales, Victoria, Queensland, the Australian Capital Territory and South Australia in December 1998. In 2005, Tasmania joined as a sixth region. Operations are essentially based on six interconnected regions that broadly follow state boundaries (NEMMCO 2005, 4).

Two key 'stylised' facts are widely accepted as applying to spot price dynamics in this market. These relate to spot electricity price dynamics that exhibit 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'). The numerous spot price spikes act as outliers producing significant deviations in the empirical distribution from Gaussianity. In fact, the spot price data displays the same predominant empirical 'leptokurtosis' feature of most high frequency asset price data sets – the tails of the empirical distribution functions are much fatter than those associated with normal distribution implying large fourth order cumulants.

In Foster, Hinich and Wild (2008), the extent of and stability of a weekly cycle in spot price time series data was investigated. A major finding of that article was that the mean properties of the spot price data for the NEM states considered were periodic. 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 Model introduced in Hinich (2000) and Hinich and Wild (2001) to the data.

It was originally postulated in Hinich (2000) and Hinich and Wild (2001) that the generating mechanism for an RMP process would be nonlinear in character. Therefore, a naturally arising research question is whether the mechanism responsible for generating weekly data exhibits some type of nonlinearity, and if so, whether this nonlinearity is 'episodic' in character. The rationale for the likely existence of episodic nonlinearity is that this type of behaviour would seem to be required if the commonly accepted 'stylised' fact of strong mean reversion in spot electricity prices, in particular, is to eventuate.

Another reason why the finding of the presence of nonlinearity would be important is because this finding would effectively rule out many classes of linear models as candidates for modelling spot price dynamics. Instead, the finding would suggest that attempts to fully model weekly spot price dynamics would have to encompass models that could possibly generate nonlinear 'bursting' to model the episodic nonlinear serial dependence evident in the underlying data.

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 this article. In Section 3 we outline the portmanteau correlation, bicorrelation and tricorrelation tests proposed in Hinich (1996) that will be employed in this article. These tests will be used to test for second-order (linear), third- and fourth-order (nonlinear) serial dependence, respectively. In Section 3 we will also briefly state the well-known Engle (1982) ARCH LM test that will be used to test for the presence of pure ARCH and GARCH structures in the weekly spot price data. In Section 4, the rationale for and practical aspects of the trimming procedure utilized in this article will be outlined. In Section 5, the empirical results will be presented. In Section 6, concluding comments will be offered.

2. DATA AND ASSOCIATED TRANSFORMATIONS

In this article, we use half hourly spot electricity prices for the period from 7/12/1998 to 29/02/2008,[1] producing a sample size of 161786 observations. We apply the tests to time series 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

$$r(t) = \ln\left(\frac{y(t)}{y(t-1)}\right) * 100\,, \tag{2.1}$$

[1] The half hourly load and spot price data were sourced from files located at the following web addresses: http://www.nemmco.com.au/data/aggPD_1998to1999.htm#aggprice1998link, http://www.nemmco.com.au/data/aggPD_2000to2005.htm#aggprice2000link, and http://www.nemmco.com.au/data/aggPD_2006to2010.htm#aggprice2006link.

where:

- $r(t)$ is the continuous compounded return for time period “t”; and
- $y(t)$ is the ‘source’ spot price time series data.

In order to apply (2.1), $y(t)$ 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. 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 (2.1) to convert the data to returns. The transformation used in this article involves two steps. First, any values which were negative or zero are set to the previous non-negative value using the following decision rule:

$$if \quad y(t) = \begin{cases} \leq 0, & x(t) = y(t-1) \\ else, & x(t) = y(t) \end{cases}, \tag{2.2}$$

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:

$$if \quad y(t) = \begin{cases} \leq 0, \; z(t) = \left\{ \dfrac{\left[x(t-1) + x(t+1) \right]}{2} \right\}, \\ else, \qquad z(t) = x(t) \; \left[= y(t) \right] \end{cases} \tag{2.3}$$

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

3. THE PORTMANTEAU CORRELATION, BICORRELATION, TRICORRELATION AND ARCH LM 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.[2] 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. For each sample frame, we standardise the data using the relation

$$Z(t) = \frac{x(t) - m_x}{s_x} \tag{3.1}$$

for each $t = 1,2,...,n$ where m_x and s_x are the sample mean and standard deviation of the sample frame and $x(t)$ are the 'source' data observations comprising the sample frame determined from data generated by (2.3). As such, the data in each sample frame 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 $C_{ZZ}(r) = E[Z(t)Z(t+r)] = 0,\ \textit{for all } r \neq 0$, the bicorrelations $C_{ZZZ}(r,s) = E[Z(t)Z(t+r)Z(t+s)] = 0,\ \textit{for all } r, s$ except when $r = s = 0$, and the tricorrelations $C_{ZZZZ}(r,s,v) = E[Z(t)Z(t+r)Z(t+s)Z(t+v)] = 0,\ \textit{for all } r, s, \textit{ and } v$ 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. 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 $r, s \textit{ and } v$ values, respectively.

The r sample correlation coefficient is defined as

$$C_{ZZ}(r) = \frac{1}{\sqrt{n-r}} \sum_{t=1}^{n-r} Z(t)Z(t+r). \tag{3.2}$$

---

[2] Other references utilizing this framework include 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 Hinich and Serletis (2007).

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

$$C = \sum_{r=1}^{L} \left[C_{ZZ}(r)\right]^2 \approx \chi_L^2 . \tag{3.3}$$

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

$$C_{ZZZ}(r,s) = \frac{1}{n-s} \sum_{t=1}^{n-s} Z(t)Z(t+r)Z(t+s), \text{ for } 0 \le r \le s . \tag{3.4}$$

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

$$H = \sum_{s=2}^{L} \sum_{r=1}^{s-1} G^2(r,s) \approx \chi^2_{L(L-1)/2} \tag{3.5}$$

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

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

$$C_{ZZZZ}(r,s,v) = \frac{1}{n-v} \sum_{t=1}^{n-v} Z(t)Z(t+r)Z(t+s)Z(t+v), \text{ for } 0 \le r \le s \le v . \tag{3.6}$$

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

$$H4 = \sum_{v=3}^{L} \sum_{s=2}^{v-1} \sum_{r=1}^{s-1} T^3(r,s,v) \approx \chi^2_{L(L-1)(L-2)/3} \tag{3.7}$$

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

In principal, the tests can be applied to either the source returns data determined from application of (2.3)-(3.1) or to the residuals from frame based autoregressive AR(p) fits of this data, where 'p' is the number of lags that is selected in order to remove significant C statistics at some pre-specified threshold level. The latter is a 'pre-whitening' 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 removing all linear dependence. 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 - thereby, only significant H and H4 statistics will lead to the rejection of the null hypothesis of a pure 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, and 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).[3] In this article, the data is split into a set of equal-length non-overlapped moving frames of 336 half hour observations corresponding to a week's duration.

We can also use the correlation, bicorrelation and tricorrelation tests to examine whether a GARCH formulation represents an adequate characterisation of the data under investigation. This is accomplished by transforming the returns data into a set of binary data according to

$$\{y(t)\} : \begin{matrix} y(t) = 1, & if \;\, Z(t) \geq 0 \\ y(t) = -1, & if \;\, Z(t) < 0 \end{matrix} . \tag{3.8}$$

If $Z(t)$ is generated by a pure ARCH or GARCH process whose innovations are symmetrically distributed with zero mean, then the binary data set $\{y(t)\}$ will be a stationary pure noise Bernoulli sequence. In essence, while $Z(t)$ (a symmetric GARCH process) is a martingale difference process, the binary transformation outlined in (3.8) converts it into a pure noise process (Lim, Hinich and Liew (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 (3.8), this then signifies the presence of structure in the data that cannot be modelled by GARCH models.

---

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

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

$$x_t^2 = \beta_0 + \sum_{i=1}^{p} \beta_i x_{t-i}^2 + \xi_t \ , \tag{3.9}$$

where x_t^2 are typically squared residuals from a linear regression. Therefore, equation (3.9) 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 (3.9) 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 (3.9).

To implement the test procedures on a frame-by-frame basis, define a frame as significant with respect to the C, H, H4 or ARCH LM tests 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 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 in fact true.[4] 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 (or no ARCH structure) in 10 out of every 100 frames. In accordance with 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 and fourth order serial dependence or ARCH/GARCH structure in the data set.

[4] 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%.

4. SAMPLING PROPERTIES OF CORRELATION, BICORRELATION, TRICORRELATION AND ARCH LM TESTS AND THE USE OF TRIMMING TO IMPROVE FINITE SAMPLE PERFORMANCE OF VARIOUS TEST STATISTICS

The sampling properties of 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.[5] 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 will be 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.

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 McAliser (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 $(\kappa/2)\,\%$ values of the sample $\{x(t_1),\ldots,x(t_N)\}$. To accomplish this, we order the data and find the $(\kappa/200)$ quantile $x_{\kappa/200}$ and the $(1-\kappa/200)$ quantile $x_{1-\kappa/200}$ of the order statistics. Then set all sample values less than the $(\kappa/200)$ quantile to $x_{\kappa/200}$ and set all sample values greater than $(1-\kappa/200)$ quantile to $x_{1-\kappa/200}$. The remaining $(100-\kappa)\,\%$ data values are not transformed in any way.

[5] 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'.

5. EMPIRICAL RESULTS

In Table 1, the summary statistics of the NEM State spot price returns series are documented. It is apparent from inspection of this table that the mean of the series are very small in magnitude - the four spot price returns series listed in Table 1 were negative over the complete sample. The 'scale' of the spot price returns data appears quite large, being in the order of 490 to 610 percent. Moreover, the size of the sixth order cumulants listed in Table 1 are also quite large in magnitude, thus pointing to the large scale implicit in the spot price returns data.

It is also evident from inspection of Table 1 that the spot price returns are quite volatile with sizeable standard deviations being observed. Volatility in spot prices is slightly higher for SA than for the other three states considered - SA has the highest standard deviations for 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 significant leptokurtosis with excess kurtosis values in the ranges of 72 to 104 in magnitude. This implies that the tails of the empirical distributions of the spot prices returns taper down to zero much more gradually than would the tails of the normal distribution (Lim, Hinich, Liew (2005, p.270)). Not unexpectedly, the Jarques-Bera (JB) Normality Test for all of the returns series 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 non-zero skewness and excess kurtosis listed in the table.

Recall from the discussion in Section 4 that the finite sample properties of the various test statistics considered in this article might be expected to deviate substantially from their theoretical distributions when the underlying data also substantially deviates from normality as in the present case. Specifically, the evidence cited in Table 1 pointing to 'significant' non-zero shewness, leptokurtosis and large 6^{th} order cumulants together with rejections of 'Jarques-Bera' normality tests all point to substantial deviations from Gaussianity in the underlying spot price returns data.

To ascertain whether the empirical distributions of the various test statistics derivable from the bootstrap process utilized in this article deviate from the expected theoretical distribution, we calculated the empirical distribution for each test statistic across a wide assortment of quantiles and have displayed these results graphically via a series a 'QQ Plots' depicted in Figures 1-7, respectively. These plots are derived from the NSW spot price returns data but are representative of the results more generally obtained for the other NEM states.

The bootstrap process used to enumerate the empirical distributions of the various test statistics was implemented in the following way. Given the (possibly trimmed) 'global' sample of 161785 returns for each respective spot price series, 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 500000 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 (threshold) variable associated with each test statistic was transformed to a uniform variate which means, for example, that the 10% threshold corresponds to 0.90, the 5% threshold is 0.95, and the 1% threshold is 0.99 and the 'transformed' test statistic threshold values are now in the interval $(0,1)$. As such, the theoretical distribution can be represented graphically by forty-five degree line. The arrays containing the bootstrap 'threshold' values for each respective test statistic from the bootstrap process was then sorted in ascending order and associated with a particular quantile scale producing the empirical distribution for each test statistic.[6]

Recall further from Section 4 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 increase the rate of convergence of the tests towards 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 above-mentioned 'QQ Plots'.

The results obtained for the C statistic applied to the NSW spot price returns data is documented in Figure 1. In deriving all results for each state, an 'AR(10) pre-whitening fit' was applied to each bootstrap frame of 336 half hourly bootstrap observations which produced 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

---

[6] In the current context, the term 'quantile' can be interchanged with the term 'percentile' which is used on the horizontal axis of Figures 1-7.

is substantially different from the theoretical distribution (corresponding to a forty-five degree line).[7] It is also apparent from Figure 1 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 NSW 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 had fallen outside of the '10%-90%' quantile range. For example, for the NSW spot price returns, the '10%-90%' trimming 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 will provide the 'global population' underpinning the bootstrap process outlined above.

The 'QQ Plots' for the H and H4 statistics are documented in Figure 2 and Figure 3. It is evident from inspection of both figures 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 are 'worse' than the associated performance of the C Statistic (as depicted in Figure 1) because the variance of the third and four 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 Gaussianity 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 while trimming rates of at least '20%-80%' (and perhaps

---

[7] 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 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 of the distributions.

even as much as '30%-70% for the H4 statistic) would seem to be required to obtain a good approximation to the theoretical distribution.[8]

The 'QQ Plots' for the ARCH LM tests are depicted in Figure 4. It is apparent from inspection of this figure that the empirical distribution for the 'no trimming' scenario once again deviates substantially from the theoretical distribution. The pattern is similar in nature to the pattern observed in Figures 2 and 3 in relation to the H and H4 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 with those required for the C statistic outlined in Figure 1, i.e. trimming in the range of '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 H4 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 Gaussianity.

It should also be noted that the empirical distribution associated with the 'no trimming' scenario's listed in Figures 1-4 all generally lie below the theoretical distribution except in the upper tail regions where the H, H4 and ARCH LM tests distributions, in particular, lie above the theoretical distributions. This result points to conservative test outcomes for all four tests at the appropriate rejection regions in the upper tails of the distributions of the test statistics in the case of the 'no trimming' scenario.[9]

The 'QQ Plots' for C, H and H4 Statistics for the bootstrap sample frame based hard clipping is documented in Figures 5-7. The data that underpins these results are the same set of spot price returns that underpinned the results in Figures 1-4 except that prior to applying the test statistics, the data in each bootstrap sample frame is hard clipped using the binary data transformation outlined in equation (3.8) in Section 3. It is apparent from inspection of Figures 5-7 that the 'no trimming' scenario produces

[8] 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.

[9] In this context, the 'no trimming' scenario corresponds to the framework underpinning the results reported in Wild, Hinich and Foster (2008). Given the conservative character of the tests at the confidence levels considered in that article (i.e. at 0.90, 0.95 and 0.99), then any frame based rejections reported in that article are believable in statistical terms.

conservative empirical distributions – that is, the empirical distributions of the various test statistics generally lie above their theoretical distributions. However, it is also apparent that the binary transformation implied in (3.8) produces an underlying data set that is more well behaved to the extent that 'minimal' trimming associated with '10%-90%' trimming rates would appear to be sufficient to produce good approximation to the theoretical distributions. In fact, the closeness of the 'no trimming' empirical distributions cited in Figures 6 and 7 for the H and (particularly) the H4 statistic provides quite strong support for the proposition that the binary transformation does converts the martingale difference process into a pure noise process having moments that are well behaved with respect to asymptotic theory that was made in 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 would seem to 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 source returns data. Inspections of Figures1, 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 H4 tests when the returns are hard clipped using trimming rates of the order '10%-90%'. For the H and H4 statistics applied to spot price returns data, the empirical performance appeared to be more sensitive to deviations from Gaussianity and trimming rates of at least '20%-80%' appeared to be necessary in order 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, H4 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 they were conservative in the upper tail regions. As such, this would make the conclusions in Wild, Hinich and Foster (2008) conservative in nature and thus believable at 10%, 5% and 1% levels of significance (corresponding to 'false alarm' thresholds of 0.90, 0.95 and 0.99, respectively). For the hard clipped data and results reported in Figures 5-7, the empirical distribution functions of the C, H and H4 statistics tended to be conservative – that is, they tended to lie above the theoretical distribution function.

As such, the conclusions reached in Wild, Hinich and Foster (2008) in relation to frame based hard clipping were also conservative in nature and hence believable.

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 Gaussianity in the source returns data. Essentially, trimming produced underlying data that was more consistent with Gaussianity by reducing the impact of outliers.

Trimming can also be used to see if any observed nonlinear serial dependence can be viewed as a 'deep structure' phenomenon which arises when the '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 produces 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 threshold values associated with a user specified 'false alarm' threshold.[10] 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 the user specified 'false alarm' threshold.[11] For example, if the user set the 'false alarm' threshold to 0.90, the bootstrap threshold value would be the '90% quantile' of the empirical distribution of the relevant test statistic determined from the bootstrap process.

[10] This bootstrap framework mirrors the framework used in Wild, Hinich and Foster (2008).

[11] Recall from the discussion in Section 3 that the 'false alarm' threshold is used to control the probability of a TYPE I error.

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 (possibly trimmed) returns data.[12] 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' 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. If the actual number of frame rejections (significantly) exceeds 10%, 5% and 1% of the total number of frames, then this points to the presence of (significant) linear and/or nonlinear serial dependence, thus confirming the presence of a nonlinear generating mechanism in the latter case.

In order to investigate the issue of whether nonlinear serial dependence could be viewed as a 'deep structure' phenomenon, a number of different trimming based 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;[13]
- 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 that were then used to underpin the global population from which bootstrap sample frames were constructed and tests applied.

---

[12] In the results reported below (in Tables 2-9), trimming is incorporated in Scenarios B-F and no trimming is used in Scenario A only.

[13] Note that the results corresponding to Scenario A are the same set of results reported in Wild, Hinich and Foster (2008).

The results for the C, H, H4 and ARCH LM tests applied to the spot price returns for the NEM states of NSW, QLD, VIC and SA are documented in Tables 2-5, respectively. Inspection of Table 2 (associated with NSW) shows 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).[14] This finding signifies the existence of statistical significant third-order and fourth-order (nonlinear) serial dependence. Note further that the 'AR(10) pre-whitening' 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 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. Similar results also occur for the other NEM States considered as indicated in Tables 3-5.

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 an above what can be reasonably attributed to random chance for all trimming scenarios considered, including the more stringent 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. As such, the rejections rates are believable and 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.

[14] The results for the H4 and ARCH LM tests also point to significant structure (see columns 6 and 7 of Table 2) because the rejections 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.

The conclusions determined from inspection of Table 2 for NSW can be broadly extended to the other states – the results in Tables 3-5 match the results cited in Table 2 in qualitative terms. In Table 4 (column 6), the results for the H4 test for VIC appears to be slightly more prominent than is the case for the other two states which more closely approximate the results listed in Table 2 that were obtained for NSW.

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 Tables 6-9 for the NEM states of NSW, QLD, VIC and SA, respectively. Note that these residuals are the same set of data that underpins the results cited in Tables 2-5 except that the transformation in (3.8) was applied to the residuals prior to applying the three above-mentioned portmanteau tests. Recall further that the intention of this particular test framework is to see if 'non-GARCH' generating mechanisms are in operation in explaining weekly spot price returns dynamics.

It is evident from inspection of Table 6 (for NSW) that the number 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 'non-GARCH' generating mechanisms. This conclusion holds for all trimming scenarios considered although the underlying rejection rates discernible from Table 6 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 tailor off somewhat over Scenarios B to D but then become more prominent for Scenarios E and F that correspond to the most stringent trimming conditions considered in this article. This is interesting to the extent that ARCH/GARCH processes are driven by volatility clustering associated with the 'episodic' presence of outliers in the data. The more stringent trimming conditions increasing abstract from this type of generating mechanism while at the same time the results cited in Table 6 point increasingly 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 is the same set of residuals that produced very few or no significant C frames in Tables 2-5. Moreover, the results discernible from Table 6 continue to hold in Tables 7-9 in qualitative terms for the other three NEM states.

These conclusions seem to indicate the non-trivial presence of a nonlinear generating mechanism that is operating over the 'central' quantile ranges of the empirical distribution of the spot price returns data that cannot be explained by a symmetric ARCH/GARCH process. 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 predominantly generated by outliers.

6. CONCLUDING COMMENTS

In this article, we have investigated whether nonlinear serial dependence is present in NEM State weekly spot price returns data and attempted to discover whether the nonlinear serial dependence is being driven by outliers in the data or by some other generating mechanism. This task was accomplished by applying the portmanteau correlation, bicorrelation and tricorrelation tests introduced in Hinich (1996) and the Engle (1982) ARCH LM test to the time series of half hourly spot prices data from 7/12/1998 to 29/02/2008. The data corresponds to spot price time series data for the NEM states of New South Wales (NSW), Queensland (QLD), Victoria (VIC) and South Australia (SA).

These tests have been used previously 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 was also investigated by applying the LM ARCH test and, additionally, 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 in the transformed data.

The finite sample properties of the empirical distribution of the various tests were investigated using a bootstrap framework. This framework allowed an assessment to be made of how close the empirical properties of the tests applied to the source spot price returns data were given the significant deviations from Gaussianity implied in this data. Under this particular circumstance, it was likely that the empirical distributions of the tests would deviate substantially from the theoretical distributions – an outcome that was subsequently confirmed from inspection of the QQ Plots contained in Figures 1-4, in particular. It was also demonstrated that the empirical properties of the tests could be improved substantially through the use of trimming. Trimming permits an investigator to control the impact of outliers on the sample performance of the test statistics and essentially reduced the range (scale) of the data by moderating the size of outliers, thus producing an underlying data set that was closer to Gaussianity.

Analysis indicated that the sample performance of the C and ARCH LM tests (applied to 'source' returns data) and C, H and H4 statistics (applied to 'hard clipped' data) could be improved substantially by employing trimming in the range of 10%-90%. The H and H4 data applied to source returns data appeared to be more sensitive to departures from Gaussianity and trimming in the range of 20% - 80% seemed to be required to get a close approximation to the theoretical distribution.

It was also demonstrated how trimming could be used to investigate whether the nonlinear generating mechanism was a 'deep structure' phenomenon – that is, the nonlinear serial dependence was generated by more than the presence of outliers. The results reported in Section 5 indicated that the 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 strong H and H4 rejections associated with hard clipped data for trimming scenarios associated with Scenarios E and F reported in Tables 6-9 of Section 5 also indicated that the generating mechanism was not consistent with a pure ARCH-GARCH process.

The findings of nonlinearity have implications for modeling weekly spot price dynamics. Given the prevalence of both third and fourth-order nonlinear serial dependence in the data, it seems that time series models that employ a linear structure or assume a pure noise input such as Geometric Brownian Motion (GBM) stochastic diffusion model would be problematic. In particular, the dependence structure would violate both the normality and Markovian assumptions underpinning conventional GBM models.

Furthermore, the finding that the nonlinear serial dependence can be categorized as a 'deep structure' phenomenon poses questions about the validity of jump diffusion models which employ the Poisson Process in order to model the probability of the occurrence of outliers (i.e. jump events). In the case where confirmatory evidence points to the presence of 'deep structure' in the nonlinear generating mechanism, than it is no longer appropriate to simply equate the presence of nonlinearity with the presence of outliers.

Any modeling process that does this by attempting to model nonlinearity purely in terms of the probability of the occurrence of outliers will not fully or adequately capture all of the salient features of the mechanism that is generating the nonlinearity present in the underlying data.

This observation will potentially have 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' that are employed in finance more generally and in the electricity industry in particular.

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.

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

Bonilla, C.A., Romero-Meza, R., 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.

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

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

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

Hinich, M.J., and D.M. Patterson (2005): "Detecting Epochs of Transient Dependence in White Noise," in Belongia, M. and J. Binner, ed., *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, 68-74.

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., 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, 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: 5-18.

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, 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 (NEMMCO) (2005): "An Introduction to Australia's National Electricity Market," National Electricity Market Management Company Ltd, June 2005. (http://www.nemmco.com.au/nemgeneral/000-0187.pdf).

Wild, P., Hinich, M.J., and J. Foster (2008): "Are Daily and Weekly Load and Spot Price Dynamics in Australia's National Electric Market Governed by Episodic Nonlinearity?" *Mimeo*, School of Economics, University of Queensland, Australia.

Table 1. Summary Statistics for Load Returns Data

	NSW	QLD	VIC	SA
No of Observations	161785	161785	161785	161785
Mean	-0.002	-0.0003	-0.002	-0.002
Maximum	545.0	591.0	497.0	597.0
Minimum	-572.0	-531.0	-488.0	-610.0
Std Dev	19.2	26.2	20.4	26.7
Skewness	0.49	0.29	0.36	-0.33
Excess Kurtosis	104.0	73.5	71.8	74.5
6^{th} Order Cumulant	41958.9	13650.2	21388.0	15187.0
JB Test Statistic	72900000.0	36300000.0	34700000.0	37300000.0
JB Normality P-Value	0.0000	0.0000	0.0000	0.0000

Table 2. Frame Test Results for NSW Weekly Spot Price (Returns) Data

Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(10) 'Prewhitening' Fit to Remove Linear Dependence
Various Trimming Scenarios

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 & (%)
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:

* - false alarm threshold arbitrarily set to 0.9999.
** - false alarm threshold arbitrarily set to 0.999999.

Table 3. Frame Test Results for QLD Weekly Spot Price (Returns) Data

Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(10) 'Prewhitening' Fit to Remove Linear Dependence
Various Trimming Scenarios

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 & (%)
Scenario A	481	0.90	4 (0.83%)	449 (93.35%)	478 (99.38%)	411 (85.45%)
	481	0.95	3 (0.62%)	380 (79.00%)	440 (91.48%)	327 (67.98%)
	481	0.99	2 (0.42%)	338** (70.27%)	401* (83.37%)	245* (50.94%)
Scenario B	481	0.90	3 (0.62%)	453 (94.18%)	443 (92.10%)	440 (91.48%)
	481	0.95	1 (0.21%)	422 (87.73%)	399 (82.95%)	425 (88.36%)
	481	0.99	0 (0.00%)	301 (62.58%)	365** (75.88%)	374 (77.75%)
Scenario C	481	0.90	0 (0.00%)	452 (93.97%)	420 (87.32%)	448 (93.14%)
	481	0.95	0 (0.00%)	435 (90.44%)	390 (81.08%)	424 (88.15%)
	481	0.99	0 (0.00%)	376 (78.17%)	316 (65.70%)	369 (76.72%)
Scenario D	481	0.90	0 (0.00%)	426 (88.57%)	366 (76.09%)	386 (80.25%)
	481	0.95	0 (0.00%)	403 (83.78%)	314 (65.28%)	348 (72.35%)
	481	0.99	0 (0.00%)	322 (66.94%)	223 (46.36%)	279 (58.00%)
Scenario E	481	0.90	0 (0.00%)	378 (78.59%)	271 (56.34%)	288 (59.88)
	481	0.95	0 (0.00%)	342 (71.10%)	214 (44.49%)	242 (50.31%)
	481	0.99	0 (0.00%)	252 (52.39%)	120 (24.95%)	174 (36.17%)
Scenario F	481	0.90	0 (0.00%)	302 (62.79%)	162 (33.68%)	198 (41.16)
	481	0.95	0 (0.00%)	249 (51.77%)	122 (25.36%)	157 (32.64%)
	481	0.99	0 (0.00%)	172 (35.76%)	53 (11.02%)	83 (17.26%)

Notes:

* - false alarm threshold arbitrarily set to 0.9999.
** - false alarm threshold arbitrarily set to 0.999999.

Table 4. Frame Test Results for VIC Weekly Spot Price (Returns) Data

Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(10) 'Prewhitening' Fit to Remove Linear Dependence
Various Trimming Scenarios

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 & (%)
Scenario A	481	0.90	1 (0.21%)	461 (95.84%)	461 (95.84%)	410 (85.24%)
	481	0.95	1 (0.21%)	438 (91.06%)	421 (87.53%)	334 (69.44%)
	481	0.99	1 (0.21%)	390* (81.08%)	386* (80.25%)	199* (41.37%)
Scenario B	481	0.90	0 (0.00%)	474 (98.54%)	452 (93.97%)	425 (88.36%)
	481	0.95	0 (0.00%)	460 (95.63%)	414 (86.07%)	397 (82.54%)
	481	0.99	0 (0.00%)	403 (83.78%)	324** (67.36%)	323 (67.15%)
Scenario C	481	0.90	0 (0.00%)	475 (98.75%)	455 (94.59%)	433 (90.02%)
	481	0.95	0 (0.00%)	464 (96.47%)	437 (90.85%)	400 (83.16%)
	481	0.99	0 (0.00%)	447 (92.93%)	386 (80.25%)	337 (70.06%)
Scenario D	481	0.90	0 (0.00%)	463 (96.26%)	429 (89.19%)	359 (74.64%)
	481	0.95	0 (0.00%)	452 (93.97%)	404 (83.99%)	316 (65.70%)
	481	0.99	0 (0.00%)	421 (87.53%)	336 (69.85%)	229 (47.61%)
Scenario E	481	0.90	0 (0.00%)	452 (93.97%)	370 (76.92%)	269 (55.93%)
	481	0.95	0 (0.00%)	430 (89.40%)	338 (70.27%)	214 (44.49%)
	481	0.99	0 (0.00%)	382 (79.42%)	242 (50.31%)	136 (28.27%)
Scenario F	481	0.90	0 (0.00%)	420 (87.32%)	273 (56.76%)	159 (33.06%)
	481	0.95	0 (0.00%)	384 (79.83%)	213 (44.28%)	115 (23.91%)
	481	0.99	0 (0.00%)	296 (61.54%)	129 (26.82%)	58 (12.06%)

Notes:

* - false alarm threshold arbitrarily set to 0.9999.
** - false alarm threshold arbitrarily set to 0.999999.

Table 5. Frame Test Results for SA Weekly Spot Price (Returns) Data

Specific Details: No Global AR 'Prewhitening' Fit;
Applied Frame by Frame AR(10) 'Prewhitening' Fit to Remove Linear Dependence
Various Trimming Scenarios

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 & (%)
Scenario A	481	0.90	2 (0.42%)	453 (94.18%)	462 (96.05%)	413 (85.86%)
	481	0.95	1 (0.21%)	385 (80.04%)	424 (88.15%)	332 (69.02%)
	481	0.99	1 (0.21%)	316** (65.70%)	376* (78.17%)	250* (51.98%)
Scenario B	481	0.90	0 (0.00%)	448 (93.14%)	434 (90.23%)	425 (88.36%)
	481	0.95	0 (0.00%)	416 (86.49%)	367 (76.30%)	400 (83.16%)
	481	0.99	0 (0.00%)	303 (62.99%)	317** (65.90%)	343 (71.31%)
Scenario C	481	0.90	0 (0.00%)	437 (90.85%)	393 (81.70%)	416 (86.49%)
	481	0.95	0 (0.00%)	410 (85.24%)	359 (74.64%)	385 (80.04%)
	481	0.99	0 (0.00%)	351 (72.97%)	271 (56.34%)	317 (65.90%)
Scenario D	481	0.90	0 (0.00%)	412 (85.65%)	335 (69.65%)	373 (77.55%)
	481	0.95	0 (0.00%)	390 (81.08%)	286 (59.46%)	320 (66.53%)
	481	0.99	0 (0.00%)	310 (64.45%)	190 (39.50%)	219 (45.53%)
Scenario E	481	0.90	0 (0.00%)	376 (78.17%)	244 (50.73%)	269 (55.93%)
	481	0.95	0 (0.00%)	331 (68.81%)	201 (41.79%)	220 (45.74%)
	481	0.99	0 (0.00%)	220 (45.74%)	119 (24.74%)	141 (29.31%)
Scenario F	481	0.90	0 (0.00%)	281 (58.42%)	169 (35.14%)	181 (37.63%)
	481	0.95	0 (0.00%)	239 (49.69%)	114 (23.70%)	139 (28.90%)
	481	0.99	0 (0.00%)	123 (25.57%)	54 (11.23%)	76 (15.80%)

Notes:

* - false alarm threshold arbitrarily set to 0.9999.
** - false alarm threshold arbitrarily set to 0.999999.

Table 6. Frame Test Results for NSW Weekly Spot Price ('Hard Clipped' 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
Various Trimming Scenarios

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
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%)

Table 7. Frame Test Results for QLD Weekly Spot Price ('Hard Clipped' 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
Various Trimming Scenarios

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
Scenario A	481	0.90	192 (39.92%)	243 (50.52%)	190 (39.50%)
	481	0.95	171* (35.55%)	161 (33.47%)	151 (31.39%)
	481	0.99	123** (25.57%)	46 (9.56%)	102 (21.21%)
Scenario B	481	0.90	197 (40.96%)	241 (50.10%)	132 (27.44%)
	481	0.95	145 (30.15%)	184 (38.25%)	87 (18.09%)
	481	0.99	67 (13.93%)	101 (21.00%)	47 (9.77%)
Scenario C	481	0.90	50 (10.40%)	236 (49.06%)	89 (18.50%)
	481	0.95	31 (6.44%)	175 (36.38%)	47 (9.77%)
	481	0.99	14 (2.91%)	88 (18.30%)	15 (3.12%)
Scenario D	481	0.90	58 (12.06%)	248 (51.56%)	92 (19.13%)
	481	0.95	35 (7.28%)	194 (40.33%)	56 (11.64%)
	481	0.99	16 (3.33%)	114 (23.70%)	11 (2.29%)
Scenario E	481	0.90	182 (37.84%)	292 (60.71%)	105 (21.83%)
	481	0.95	132 (27.44%)	249 (51.77%)	61 (12.68%)
	481	0.99	78 (16.22%)	165 (34.30%)	22 (4.57%)
Scenario F	481	0.90	377 (78.38%)	359 (74.64%)	175 (36.38%)
	481	0.95	336 (69.85%)	318 (66.11%)	137 (28.48%)
	481	0.99	268 (55.72%)	245 (50.94%)	76 (15.80%)

Notes:

* - false alarm threshold arbitrarily set to 0.9999.
** - false alarm threshold arbitrarily set to 0.999999.

Table 8. Frame Test Results for VIC Weekly Spot Price ('Hard Clipped' 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
Various Trimming Scenarios

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
Scenario A	481	0.90	218 (45.32%)	311 (64.66%)	178 (37.01%)
	481	0.95	172 (35.76%)	270 (56.13%)	132 (27.44%)
	481	0.99	85 (17.67%)	162 (33.68%)	67 (13.93%)
Scenario B	481	0.90	153 (31.81%)	314 (65.28%)	144 (29.94%)
	481	0.95	107 (22.25%)	263 (54.68%)	96 (19.96%)
	481	0.99	45 (9.36%)	154 (32.02%)	41 (8.52%)
Scenario C	481	0.90	49 (10.19%)	342 (71.10%)	149 (30.98%)
	481	0.95	33 (6.86%)	297 (61.75%)	98 (20.37%)
	481	0.99	9 (1.87%)	192 (39.92%)	31 (6.44%)
Scenario D	481	0.90	66 (13.72%)	370 (76.92%)	156 (32.43%)
	481	0.95	40 (8.32%)	318 (66.11%)	94 (19.54%)
	481	0.99	11 (2.29%)	219 (45.53%)	40 (8.32%)
Scenario E	481	0.90	210 (43.66%)	400 (83.16%)	181 (37.63%)
	481	0.95	158 (32.85%)	366 (76.09%)	122 (25.36%)
	481	0.99	82 (17.05%)	274 (56.96%)	51 (10.60%)
Scenario F	481	0.90	374 (77.75%)	415 (86.28%)	230 (47.82%)
	481	0.95	339 (70.48%)	385 (80.04%)	176 (36.59%)
	481	0.99	264 (54.89%)	296 (61.54%)	92 (19.13%)

Table 9. Frame Test Results for SA Weekly Spot Price ('Hard Clipped' 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
Various Trimming Scenarios

Scenario / (State)	Total Num of Frames	False Alarm Threshold	Significant C Frames Num & (%)	Significant H Frames Num & (%)	Significant H4 Frames Num & (%)
Scenario A	481	0.90	248 (51.56%)	267 (55.51%)	193 (40.12%)
	481	0.95	190 (39.50%)	214 (44.49%)	158 (32.85%)
	481	0.99	136** (28.27%)	94 (19.54%)	113 (23.49%)
Scenario B	481	0.90	223 (46.36%)	248 (51.56%)	123 (25.57%)
	481	0.95	170 (35.34%)	204 (42.41%)	83 (17.26%)
	481	0.99	82 (17.05%)	117 (24.32%)	35 (7.28%)
Scenario C	481	0.90	46 (9.56%)	233 (48.44%)	96 (19.96%)
	481	0.95	23 (4.78%)	188 (39.09%)	50 (10.04%)
	481	0.99	8 (1.66%)	93 (19.33%)	13 (2.70%)
Scenario D	481	0.90	70 (14.55%)	272 (56.55%)	98 (20.37%)
	481	0.95	40 (8.32%)	221 (45.95%)	61 (12.68%)
	481	0.99	16 (3.33%)	137 (28.48%)	17 (3.53%)
Scenario E	481	0.90	191 (39.71%)	299 (62.16%)	112 (23.28%)
	481	0.95	149 (30.98%)	248 (51.56%)	75 (15.59%)
	481	0.99	93 (19.33%)	171 (35.55%)	26 (5.41%)
Scenario F	481	0.90	390 (81.08%)	331 (68.81%)	163 (33.89%)
	481	0.95	352 (73.18%)	280 (58.21%)	125 (25.99%)
	481	0.99	284 (59.04%)	219 (45.53%)	69 (14.35%)

Notes:

* - false alarm threshold arbitrarily set to 0.9999.
** - false alarm threshold arbitrarily set to 0.999999.

Figure 1. QQ Plot for Bootstrapped C Statistic for NSW (Weekly) Spot Price Returns

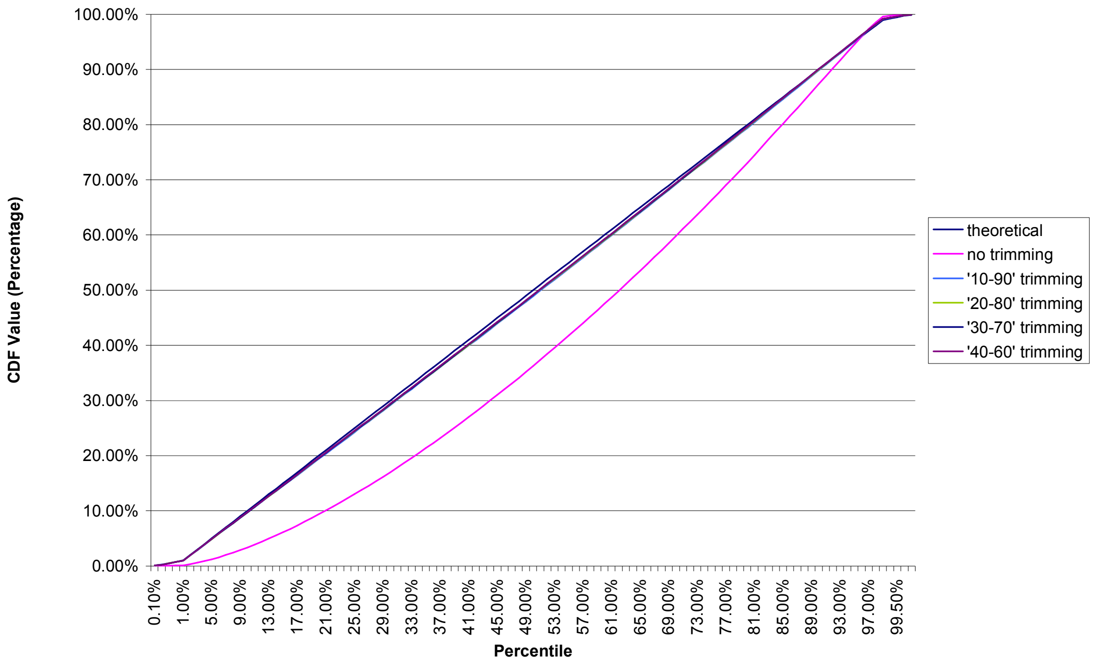

Figure 2. QQ Plot for Bootstrapped H Statistic for NSW (Weekly) Spot Price Returns

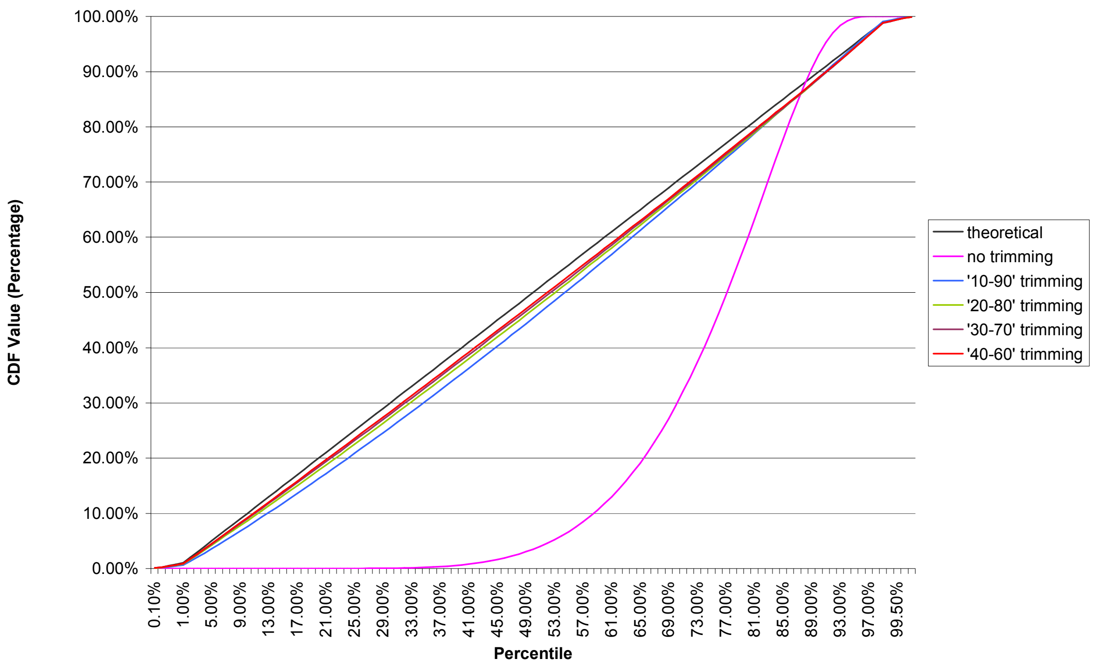

Figure 3. QQ Plot for Bootstrapped H4 Statistic for NSW (Weekly) Spot Price Returns

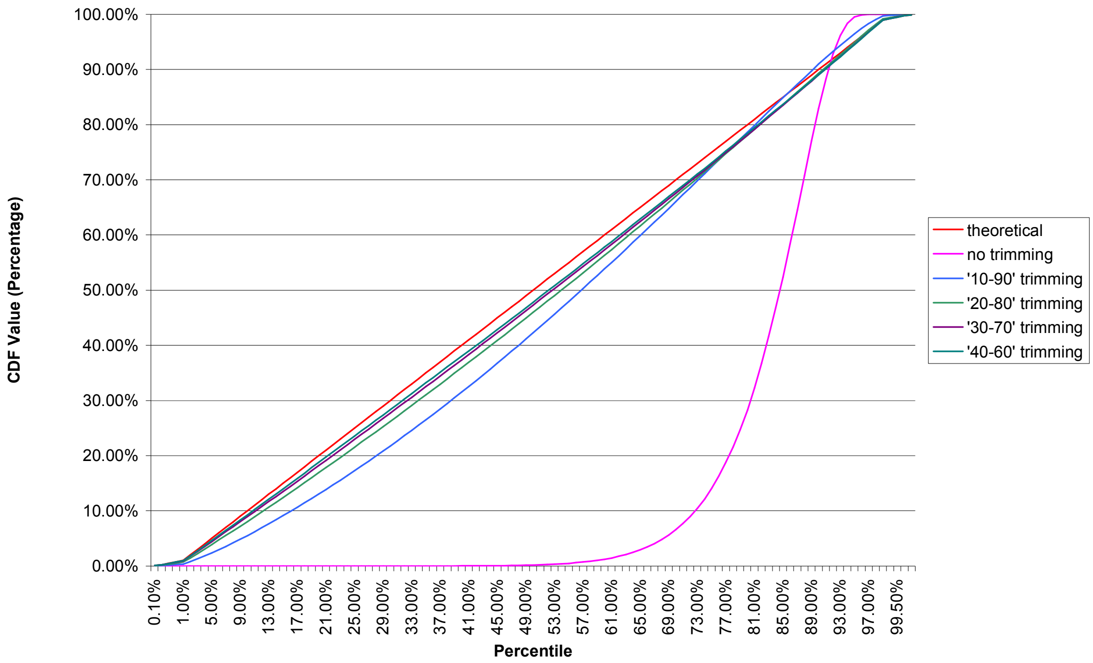

Figure 4. QQ Plot for Bootstrapped LM ARCH Statistic for NSW (Weekly) Spot Price Returns

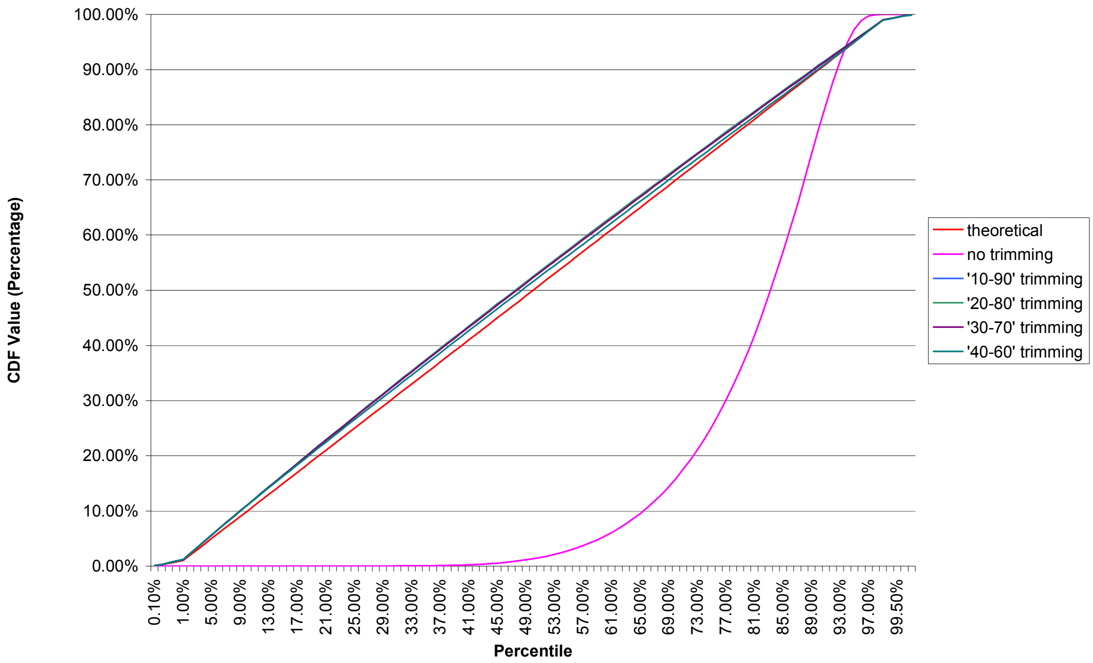

Figure 5. QQ Plot for Bootstrapped C Statistic for 'Hard Clipped' NSW (Weekly) Spot Price Returns

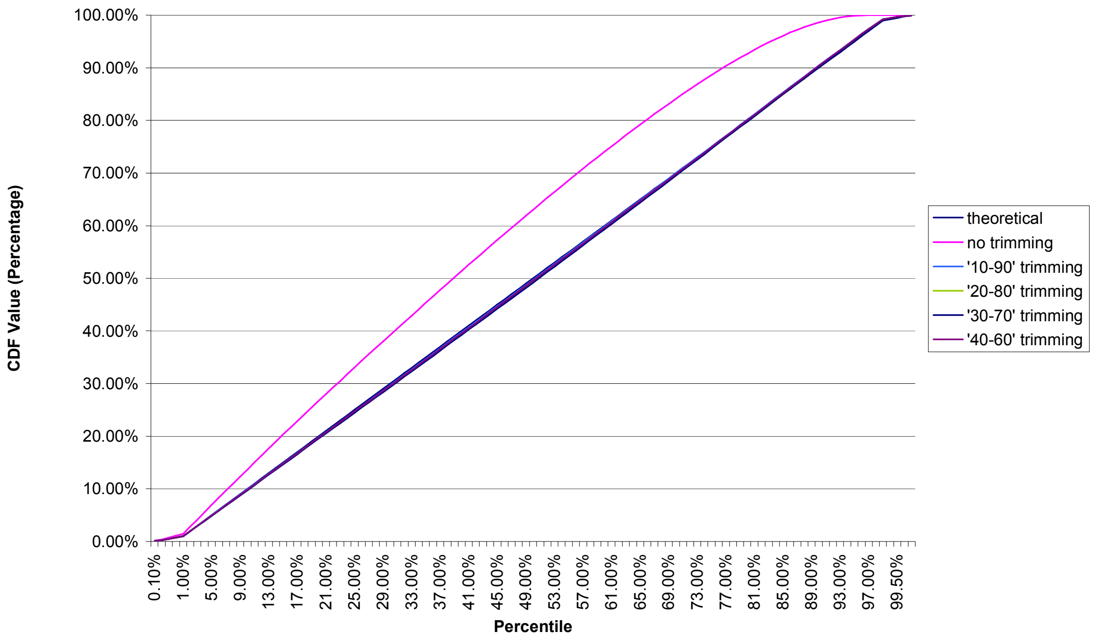

Figure 6. QQ Plot for Bootstrapped H Statistic for 'Hard Clipped' NSW (Weekly) Spot Price Returns

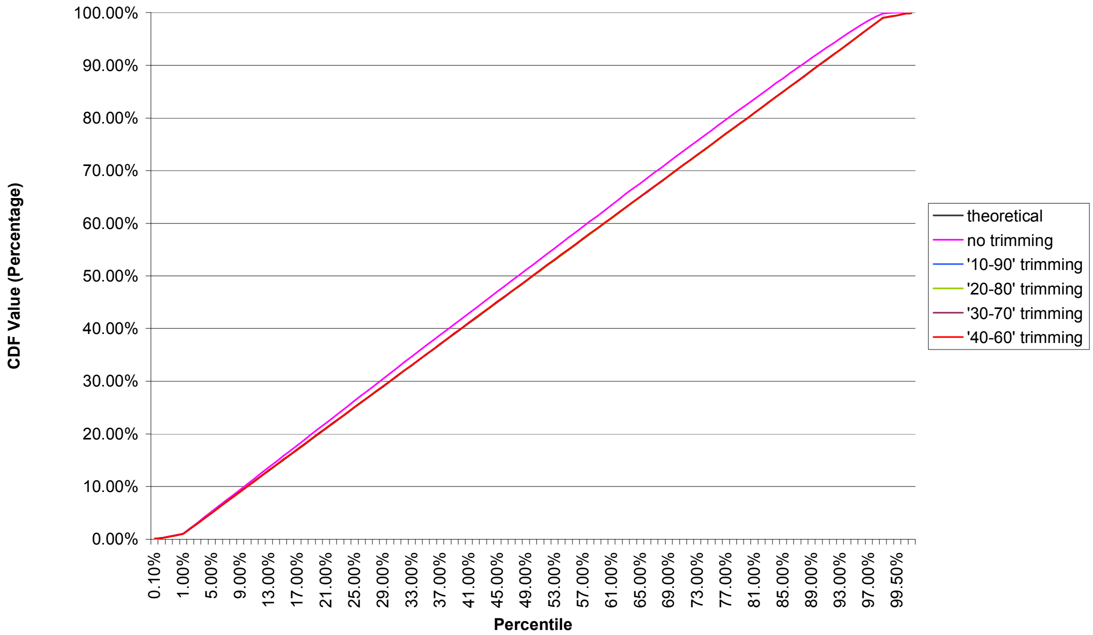

Figure 7. QQ Plot for Bootstrapped H4 Statistic for 'Hard Clipped' NSW (Weekly) Spot Price Returns