Further Evidence on the Presence of Non-Linearity in the Dow Jones Industrial Average (DJIA).

Josiah Baker [a] Ossama Mikhail [a,*]

[a] *Department of Economics, College of Business Administration, University of Central Florida, 4000 Central Florida blvd., Orlando, FL 32816*

Abstract

In this paper, we report further evidence on the presence of non-linearity in the DJIA. Using a corrected for bias simple non-parametric test (SNT), we reject the null hypothesis of a linear structure in the DJIA, even after accounting for the presence of a GARCH. The results suggest that non-linearity should be a main component of the modeling and the forecasting of DIJA and that modeling the DJIA as a GARCH process is not sufficient to capture all the non-linearity.

JEL classification: C12, C22.

Keywords: Dow Jones Industrial Average (DJIA), GARCH, Simple Non-Paramteric Test (SNT), U-Statistics, Correlation Integral, Non-Linear Specification.

* Corresponding Author. Phone: (407) 823-4258, Fax: (407) 823-3269.

Email addresses: jbaker@bus.ucf.edu (Josiah Baker), omikhail@bus.ucf.edu (Ossama Mikhail).

1 Introduction

Research interests in equity returns led to the adoption of the modeling assumption that the returns should have independent increments because the ideal market of economic theory is the one in which speculation reduces the expected economic to zero. This framework led the empirical research agenda of securities returns into the arena of identifying the statistical process that generated the returns for the purpose of forecasting it.

Using the simple non-parameteric test (SNT), this article tests for the presence of non-linearity in the DJIA returns after accounting for the widely used class of non-linear models, namely the GARCH. The results strongly challenge the belief that the DJIA returns can be viewed as [linearly] independent random variables.

Logically, and prior to investigating the nature and characteristics of non-linearities as a theoretical possibility in financial modeling, one has to investigate and examine the empirical evidence of non-linearity - if any. The appropriate reduced-form model for the DJIA is a subject open for debate and depends - mostly - on the stylized facts. Of interest to researchers and practitioners, whenever evidence of non-linearity is detected, then further search for the appropriate modeling assumptions is needed. Once found, these assumptions unfold the complex dynamics into a model with relatively successful predictability.

Earliest to investigate, Hinich and Patterson (1985) reported evidence of the presence of non-linearity in the daily stock returns. With aim to detect, estimate and properly model this non-linearity and its implied trading rules, a spectrum of studies followed.

In this paper, we argue that the omnipresent persistence mechanism in the DJIA is often inaccurately modeled as a linear specification [approximation]. This implies that many common linear models of financial processes (e.g., Soysemir and Petrie 2003 or Taylor 2000) are missing an important part of the [non-linear] market dynamics. The presence of this non-linearity should be further exploited for modeling and prediction purposes.

The recent surge in interest in detecting nonlinearity in financial time series, coupled with the advancement of time series techniques, led researchers to investigate, test and challenge previously upheld and widely accepted assumptions (Hsieh 1991, Teräsvirta and Anderson 1992, Granger and Teräsvirta 1993, Teräsvirta 1994, Martens et al. 1998, Leung et al. 2000, Taylor 2000, McMillan 2001, 2003 and 2005, Maasoumi and Racine 2002, Soysemir and Petrie 2003, Shively 2003, Small and Tse 2003, Hinich, Mendes, and Stone 2005). Of interest are the studies which focus on the widely accepted class of ARMA-ARCH models that describe the returns data generating process[1].

A wide range of tests to detect, and models to investigate, non-linearities in financial time series have been proposed in the literature. We focus on the simple non-parametric test (SNT) devised by Mizrach (1991 and 1994). Devised to discriminate between linear and non-linear time series models, the SNT is carried on the residuals of a model. If non-linearities are ubiquitous, models such as ARCH, GARCH, switching Markov, bilinear time series and others are suggested to capture the non-linearities.[2]

The plan of this paper is as follows; Section 2 introduces the U- and the SNT statistics, Section 3 reports the data and discusses the results, and finally, Section 4 concludes.

2 U-Statistics and The Simple Nonparametric Test (SNT)

U-statistics[3] are generalizations of sample averages. The components of a U-statistic include a kernel, a symmetric measurable function $h : R^m \to R$, and a permutation operator, $\sum_{n,m}$ that sums over the $\binom{n}{m}$ distinct combinations of m-elements in a sample space of size n. Let $\{x_i\}$ be a strictly stationary stochastic process with a distribution function F, and let $\{X_1, ..., X_n\}$ be a sample of size n. Define the canonical mapping,

$$U_n = U(X_1, ..., X_n) \equiv \binom{n}{m}^{-1} \sum_{n,m} h(X_1, ..., X_n) \tag{1}$$

Two examples follow to show how this U statistic relates to the sample moments. If $m \equiv 1$, $h(x_i) \equiv x_i$, $\binom{n}{1}^{-1} = \frac{1}{n}$, then $U(X_1, ..., X_n) = \overline{X}$, i.e., the sample mean. If $m \equiv 2$, $h(x_i - x_j) \equiv \frac{(x_i - x_j)^2}{2}$, $\binom{n}{2}^{-1} = \frac{2}{n(n-1)}$, then $U(X_1, ..., X_n)$ equals the sample variance. Now, consider the vector valued version of Equation (1). Let $x_t^m \in R^m$ be a random vector in R^m, and let $F(x_t^m)$ be its joint distribution. Define the kernel as, $h : R^m \times R^m \to R$,

$$h(x_t^m, x_s^m) = I[\parallel x_t^m - x_s^m \parallel < \varepsilon] \equiv I(x_t^m, x_s^m, \varepsilon) \tag{2}$$

where I is the indicator function, and $\parallel . \parallel$ denotes the max norm,

$$I(x_t^m, x_s^m, \varepsilon) = I[\max_i \prod_{i=0}^{m-1} |x_{t+j} - x_{s+j}| < \varepsilon] \tag{3}$$

The correlation integral is given by,

$$C(m, \varepsilon) = \int_X \int_X I(x_t^m, x_s^m, \varepsilon) dF(x_t^m) dF(x_s^m) \tag{4}$$

A consistent estimator of the correlation integral is,

$$C(m, N, \varepsilon) \equiv \frac{2}{N(N-1)} \sum_{t=1}^{N-1} \sum_{s=t+1}^{N} I(X_t^m, X_s^m, \varepsilon) \tag{5}$$

where $N = n - m + 1$. Note that $C(m, N, \varepsilon)$ is the expected number of m-vectors less than ε away from any given m-vector. In other words, $C(m, \varepsilon)$ measures the probability that any particular pair in the time series are 'ε-close'. The Brock, Dechert and Scheinkman (1987) (BDS hereafter) test statistic is computed as,

$$\sqrt{N}\frac{C(m, N, \varepsilon) - C(1, N, \varepsilon)^m}{\sqrt{var(C(m, N, \varepsilon) - C(1, N, \varepsilon)^m)}} \rightarrow^d N(0, 1) \tag{6}$$

If the series is linear but exhibits autocorrelation, then the BDS will reject the null. Therefore in practice, the BDS is usually applied to the residual of a model.

The SNT has advantages over the BDS. It involves simpler computation at the order of N rather than N^2. The variance of SNT is similar to that of a binomial random variable. And most importantly, the SNT is properly sized in small samples. The BDS high rate of error of Type I, led Mizrach (1991) to propose the SNT test, by replacing the kernel in the BDS test statistic by $h : R \rightarrow R$, where $h(.)$ is defined as follows,

$$h(x_t) = I[x_t < \varepsilon] = \left\{ \begin{array}{l} 1, \text{ if } x_t < \varepsilon \\ 0, \text{ otherwise} \end{array} \right\} \equiv I(x_t, \varepsilon) \tag{7}$$

Due to the choice of this kernel, the correlation integral (CI) at dimension m sums m-independent[4] events under the assumption of $i.i.d.$

$$\theta(m, \varepsilon) = \int_{x \in X} \prod_{i=1}^{m} I(x_{t+i}, \varepsilon) dF(x_{t+i}) \tag{8}$$

A consistent estimator for the CI is,

$$\theta(m, N, \varepsilon) = \sum_{i=1}^{N} \prod_{i=0}^{m-1} I(x_{t+i}, \varepsilon)/N \tag{9}$$

The expected number of m-chains with a value of 1 in a sample of size N is,

$$\mu = N\theta(m, \varepsilon) = \sum_{x=0}^{N} x \binom{N}{x} \theta(m, \varepsilon)^x (1 - \theta(m, \varepsilon))^{N-x} \tag{10}$$

The variance is given by,

$$\sigma_{B_N}^2 = \mu_2^{'} - \mu^2 = \sum_{x=0}^{N} x^2 \binom{N}{x} \theta(m,\varepsilon)^x (1-\theta(m,\varepsilon))^{N-x} - N^2\theta(m,\varepsilon)^2 \tag{11}$$

$$= N\theta(m,\varepsilon)(1-\theta(m,\varepsilon)) \tag{12}$$

Using both moments, Mizrach (1991) constructed the following statistic and showed that it has an asymptotic normal distribution,

$$SNT \equiv \sqrt{N}\frac{\theta(m,N,\varepsilon) - \theta(m-1,N,\varepsilon)\theta(1,N,\varepsilon)}{\theta(m-1,N,\varepsilon)\theta(1,N,\varepsilon)(1-\theta(m-1,N,\varepsilon))(1-\theta(1,N,\varepsilon))} \rightarrow^d N(0,1) \tag{13}$$

The small sample properties of the SNT statistic are studied and reported in Mizrach (1994, pp. 381-382).

3 Data and Results

Daily observations of the Dow Jones Industrial Average are taken over the period 01/02/1930 to 06/01/2006. The data is transformed into continuously compounded returns. First, linear dependency in the data was removed by fitting a set of ARMA(p,q) models, where the choice of the orders was carried to minimize the Akaike (AIC) and Schwartz (SIC) Criteria. We experimented with different values for p and q to cover all possible combinations up to and including the orders $p \in \{0,1,2,3,4\}$ and $q \in \{0,1,2,3,4\}$. Often, the two criteria disagreed on the selection of the order. A closer inspection revealed the possibility of root canceling, and therefore, we choose the ARMA$(1,0)$ for the DJIA returns.[5] We used the squared residuals e_t^2 of this process - after being fitted into an ARMA $(1,0)$ - to identify the order of the GARCH and this led us to select an ARIMA$(1,1,0)$-GARCH$(1,1)$ for the DJIA. We then computed the residuals and tested for linearity using the simple non-parametric test.[6]

Table 1 reports the results.

[Insert Table 1 here]

At the 95 percent confidence level, the SNT rejects the null of i.i.d. for the DJIA returns. We also investigated the robustness of these results with respect to the outliers in the series and reached a similar conclusion. This conclusion also holds under alternative values for the embedding dimension m up to and including $m = 8$.

Contrary to the strong conclusion reached in Small and Tse (2003. p. 24), - wherein they rejected the random walk hypothesis for the DJIA returns using surrogate analysis - the conclusion reached here does not contradict market efficiency. It just emphasizes the missing [non-linear] dynamics in common financial models.

4 Conclusion

While rejecting the independent and identical distribution (IID) hypothesis of the Dow Jones Industrial Average (DJIA) returns does not contradict market efficiency, it surely emphasizes the need for more investigations into the nature of the non-linearity of the returns. After accounting for the general time-varying conditional variance, and using the simple non-parametric (SNT) test, we strongly reject the hypothesis of IID in the DJIA returns, i.e., it is not sufficient to model the DJIA returns as an ARMA-GARCH process. Other forms of non-linearities are well present in the data.

Notes

[1]Among many, Hsieh (1991) reported a weak evidence against the hypothesis that the residuals from such models are IID.

[2]Note that rejecting the null of i.i.d. on the residuals of fitted linear time series models does not imply that the alternative is only a non-linear time series model (Poirier 1997).

[3]For a detailed derivations, see Mizrach (1994) and Cromwell, Labys and Terraza (1994, pp. 32-36).

[4]To examine this 'spatial' correlation, the time series $x(t)$ must be embedded in m-space by constructing a vector. The choice of m for the dimensionality of the vectors is subjective. See Cromwell, Labys and Terraza (1994, p. 33) for details.

[5]Few of these ARMA orders pointed to the presence of a two (working) weeks memory in the process (lag=16). These were investigated and similar results were concluded. We also used the ARMASA automatic identification toolbox in Matlab to identify the orders. The ARMASA code used is available at <http://www.bus.ucf.edu/omikhail>.

[6]For replication purposes: the data, the Eviews and the Fortran programs are available at <http://www.bus.ucf.edu/omikhail>. We acknowledge the support of Bruce Mizrach in making the Fortran code available. We modified the code to compute the SNT directly.

References

[1] Brock, William A., W. Dee Dechert, and Jose A. Scheinkman. (1987) A Test for Independence based on the Correlation Dimension, University of Wisconsin-Madison Workshop Paper 8702.

[2] Cromwell, Jeff B., Walter C. Labys, Michel Terraza. (1994) *Univariate Tests for Time Series Models*, New York: Sage Publications.

[3] Granger, Clive W. J. and Timo Teräsvirta. (1993) Modelling Nonlinear Economic Relationships, Oxford University Press, Oxford.

[4] Hinich, Melvin J., Eduardo M. Mendes, and Lewi Stone. (2005) Detecting Nonlinearity in Time Series: Surrogate and Bootstrap Approaches, *Studies in Nonlinear Dynamics & Econometrics* 9 (4), Article 3.

[5] Hinich, Melvin J. and Douglas M. Patterson. (1985) Evidence of Nonlinearity in Daily Stock Returns, *Journal of Business and Economic Statistics* 3 (1), 69-77.

[6] Hsieh, David A. (1991) Chaos and Nonlinear Dynamics: Application to Financial Markets, *Journal of Finance* 46 (5), 1839-1877.

[7] Leung, Mark T., Hazem Daouk and Chen, An-Sing. (2000) Forecasting stock indices: a comparison of classification and level estimation models, *International Journal of Forecasting* 16, 173–190.

[8] Maasoumi, Esfandiar and Racine, Jeff. (2002) Entropy and predictability of stock market returns, *Journal of Econometrics* 107, 291–312.

[9] Martens, Martin., Paul Kofman and Ton C. F. Vorst. (1998) A threshold error correction model for intraday futures and index returns, *Journal of Applied Econometrics* 13, 245–263.

[10] McMillan, David G. (2001) Non-linear predictability of stock market returns: evidence from non-parametric and threshold models, *International Review of Economics and Finance* 10, 353-68.

[11] McMillan David G. (2003) Non-linear predictability of UK stock market returns, *Oxford Bulletin of Economics and Statistics* 65, 557-73.

[12] McMillan, David G. (2005) Non-linear dynamics in inter-national stock market returns, *Review of Financial Economics* 14, 81-91.

[13] Mizrach, Bruce. (1994) Using U-statistics to Detect Business Cycle Nonlinearities, in Willi Semmler (Ed.), *Business Cycles: Theory and Empirical Investigation*, Boston: Kluwer Press, 369-388.

[14] Mizrach, Bruce. (1991) Nonparametric Estimation of the Correlation Exponent, (with E.S . Mayfield), *Physical Review* A 88 #4, 5298-5301.

[15] Poirier, Dale J. (1997) Comparing and Choosing Between Two Models with a Third Model in the Background, *Journal of Econometrics* 78, 139-151.

[16] Shively, Philip A. (2003) The nonlinear dynamics of stock prices, *The Quarterly Review of Economics and Finance* 43, 505–17.

[17] Small, Michael and Chi K. Tse. (2003) Determinism in Financial Time Series, *Studies in Nonlinear Dynamics & Econometrics* 7 (3), Article 5.

[18] Soydemir, Gökçe A. and A. George Petrie. (2003) Intraday Information Transmission Between DJIA Spot and Futures Markets, *Applied Financial Economics* 13 (11), 817-827.

[19] Taylor, Stephen J. (2000) Stock Index and Price Dynamics in the UK and the US: New Evidence from a Trading Rule and Statistical Analysis, *The European Journal of Finance* 6, 39-69.

[20] Teräsvirta, Timo. (1994) Specification, estimation and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* 89, 208-18.

[21] Teräsvirta, Timo and Anderson, Heather M. (1992) Characterising nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics* 7, S119-S136.

5 Table

Table 1: The SNT for the DJIA returns – ARIMA$(1,1,0)$**-GARCH**$(1,1)$

	SNT
$m = 1$	-181.449
$m = 2$	-646.218
$m = 3$	-628.486
$m = 4$	-485.897
$m = 5$	-329.766
$m = 6$	-303.553
$m = 7$	-204.696
$m = 8$	-289.836