International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)
ISSN: 2150-7988 Vol.2 (2010), pp.187-211
http://www.mirlabs.org/ijcisim

Performance Analysis of Hybrid Forecasting models with Traditional ARIMA Models - A Case Study on Financial Time Series Data

P. Bagavathi Sivakumar [1], V. P. Mohandas [2]
[1] *Department of Computer Science and Engineering*
[2] *Department of Electronics and Communication Engineering*
School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore – 641 105, India
pbsk@cb.amrita.edu, vp_mohandas@cb.amrita.edu

Abstract

ARIMA and GARCH models in their various flavors are frequently used in modeling of real world financial time series. Often, those models do not produce the best possible results in terms of modeling and forecasting. Of late, researchers across the world have gone for hybrid models. In principle, hybrid models bring the best out of both worlds. The modeling and forecasting ability of ARFIMA-FIGARCH model is investigated in this study. It is widely agreed that financial time series data like stock index exhibit a pattern of long memory. Short term and long term influences are also observed. Empirical investigation has been made on ten such stock indices comprising of various segments of Indian stock data. The obtained results clearly illustrate the modeling power of ARFIMA-FIGARCH. The performance of this model is compared with traditional Box and Jenkins ARIMA models. The results obtained illustrate the need for hybrid modeling. ARFIMA-FIGARCH is compared with seven different flavors of ARIMA and ARFIMA-FIGARCH emerges as the clear winner.

Keywords:
Time Series Analysis, Long memory, ARFIMA-FIGARCH, ARIMA,

1. Introduction

1.1 Time series and time series analysis

A discrete-time signal or time series [21] is a set of observations taken sequentially in time, space or some other independent variable. Many sets of data appear as time series: a monthly sequence of the quantity of goods shipped from a factory, a weekly series of the number of road accidents, hourly observations made on the yield of a chemical process and so on. Examples of time series abound in such fields as economics, business, engineering, natural sciences, medicine and social sciences.

An intrinsic feature of a time series is that, typically, adjacent observations are related or dependent. The nature of this dependence among observations of a time series is of considerable practical interest. Time Series Analysis is concerned with techniques for the analysis of this dependence. This requires the development of models for time series data and the use of such models in important areas of application.

When successive observations of the series are dependent, the past observations may be used to predict future values. If the prediction is exact, the series is said to be deterministic. We cannot predict a time series exactly in most practical situations. Such time series are called random or stochastic, and the degree of their predictability is determined by the dependence between consecutive observations. The ultimate case of randomness occurs when every sample of a random signal is independent of all other samples. Such a signal, which is completely unpredictable, is known as White noise and is used as a building block to simulate random signals with different types of dependence. To properly model and predict a time series, it becomes important to fundamentally and thoroughly analyze the time series data or signal itself.

There are two aspects to the study of time series - analysis and modeling, the aim of analysis is to summarize the properties of a series and to characterize its salient features. This may be done either in the time domain or in the frequency domain. In the time domain attention is focused on the relationship between observations at different points in time, while in the frequency domain it is cyclical movements which are studied [27].

1.2 Financial time series and their characteristics

Financial time series analysis is concerned with theory and practice of asset valuation over time. It is a highly empirical discipline, but like other scientific fields theory forms the foundation for making inference. There is, however, a key feature that distinguishes financial time series analysis from other time series analysis. Both financial theory and its empirical time series contain an element of uncertainty. For example, there are various definitions of asset volatility, and for a stock return series, the volatility is not directly observable. As a result of the added uncertainty, statistical theory and methods play an important role in financial time series analysis [25].

Econometric models are designed to capture characteristics that are commonly associated with financial time series, including fat tails, volatility clustering, and leverage effects. Probability distributions for asset returns often exhibit fatter tails than the standard normal distribution. The fat tail phenomenon is called excess kurtosis. Time series that exhibit fat tails are often called leptokurtic. Financial time series also often exhibit volatility clustering or persistence. In volatility clustering, large changes tend to follow large changes, and small changes tend to follow small changes. The changes from one period to the next are typically of unpredictable sign. Large disturbances, positive or negative, become part of the information set used to construct the variance forecast of the next period's disturbance. In this way, large shocks of either sign can persist and influence volatility forecasts for several periods. Volatility clustering suggests a time series model in which successive disturbances are uncorrelated but serially dependent.

1.2.1 Conditional versus unconditional variance. The term conditional implies explicit dependence on a past sequence of observations. The term unconditional applies more to long-term behavior of a time series, and assumes no explicit knowledge of the past. Time series typically modeled by Econometrics Toolbox software have constant means and unconditional variances but non-constant conditional variances.

1.3 Prices, returns, and compounding

Generally rather than using the actual raw price series for analysis, returns are used. A price series is converted to a return series with either continuous compounding or simple periodic compounding.

If successive price observations made at times t and t+1 are denoted as P_t and P_{t+1}, respectively, continuous compounding transforms a price series $\{P_t\}$ into a return series $\{y_t\}$ using

$$y_t = \log \frac{P_{t+1}}{P_t} = \log P_{t+1} - \log P_t$$

Simple periodic compounding uses the transformation

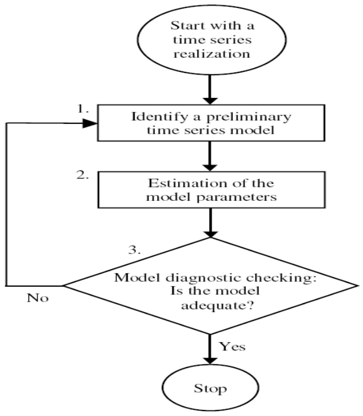
$$y_t = \frac{P_{t+1} - P_t}{P_t} = \frac{P_{t+1}}{P_t} - 1$$

Continuous compounding is the default compounding method, and is the preferred method for most of continuous-time finance. Since modeling is typically based on relatively high frequency data (daily or weekly observations), the difference between the two methods is usually small.

1.4 Time series modeling

One of the major tasks of a statistician is to come up with a probability model that can adequately describe his data. Hence, the often asked question, “which model describes the data best?” or equivalently, “which model provides the best fit to the data?” In the days of Karl Pearson where the emphasis was usually on whether the data are from a certain distribution family this question translates into testing the hypothesis that the common distribution $F(\cdot)$, of an independent identically distributed sample $X1, \ldots, Xn$ is equal to a family of distribution indexed by, say, a parameter θ. That is, we test the null hypothesis $H0$: $F(\cdot) = G(\cdot|\theta)$. This gives rise to Pearson’s 1900 paper on the classical chi-squared goodness-of-fit test. Since then there evolves a huge literature on goodness-of-fit tests. Modern statistics have developed many more tools than the chi-square tests in order to answer the question, “which model(s) describes the data more adequately?” [22].

Atkinson suggested that in regression “diagnostics is the name given to a collection of techniques for detecting disagreement between a regression model and the data to which it is fitted.” The same can be said about time series analysis. The same classical question “which model best describes the data?” is asked by both theorists and practitioners. The Box-Jenkins approach to time series modeling (Box and Jenkins, 1970; 1976) reflects both the influences of the classical goodness-of-fit and diagnostic approaches. Their approach can be described by the following flowchart [22].



In the first stage a preliminary autoregressive moving average (ARMA) model is suggested based on information on the sample path, sample moments: autocorrelations and partial autocorrelations. In the second stage, the estimation of stationary ARMA models is done. At present approximate or exact maximum likelihood procedures are often used for estimation once the autoregressive and moving average orders are specified. For pure autoregressive models there are at least two more choices in terms of methods of estimation: the least squares procedure and the Yule-Walker equations. The third stage in the Box-Jenkins approach is called model diagnostic checking which involves techniques like over fitting, residual plots, and more importantly, checking that the residuals are approximately uncorrelated. This makes good modeling sense since in the time series analysis a good model should be able to describe the dependence structure of the data adequately, and one important measurement of dependence is via the autocorrelation function. In other words, a good time series model should be able to produce residuals that are approximately uncorrelated, that is, residuals that are approximately white noise.

1.5. Financial time series forecasting

A financial time series can be treated as a sequence of random observations. This random sequence, or stochastic process, may exhibit a degree of correlation from one observation to the next. This correlation structure can be used to predict future values of the process based on the past history of observations. Exploiting the correlation structure, if any, allows the modeler to decompose the time series into the following components:

A deterministic component (the forecast)

A random component (the error, or uncertainty, associated with the forecast)

The following represents a univariate model of an observed time series:

$$y_t = f(t-1, X) + \varepsilon_t$$

In this model, f (t – 1, X) is a nonlinear function representing the forecast, or deterministic component, of the current return as a function of information known at time t – 1. The forecast includes:

Past disturbances $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$

Past observations $\{y_{t-1}, y_{t-2}, ...\}$

Any other relevant explanatory time series data X

$\{\varepsilon_t\}$ is a random innovations process. It represents disturbances in the mean of $\{y_t\}$. ε_t can also be interpreted as the single-period-ahead forecast error. An important branch of Econometrics is the analysis of time series. Here, it is assumed that a time series follows a certain pattern for which a model can be found. Such a model consists of a number of parameters which have to be estimated. Once a model has been chosen and estimated, there exist various tools to judge whether the model is appropriate.

2. Time series models

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the *autoregressive* (AR) models, the *integrated* (I) models, and the *moving average* (MA) models. These three classes depend linearly on previous data points. Combinations of these ideas produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. The autoregressive fractionally integrated moving average (ARFIMA) model generalizes the former three.

2.1. ARFIMA models

An important breakthrough in Time Series Analysis was the introduction of ARIMA models, by Box and Jenkins in 1970. ARIMA models are able to capture the short run dynamics of a time series and are also suitable for time series which are only stationary after taking first or even higher order differences. However, it is often difficult to judge whether a time series is integrated (i.e. non-stationary) or not. This can be unsatisfactory because the two classes of processes have substantially different theoretical properties. Here, the Fractionally Integrated ARMA of ARFIMA model, first developed by Hosking, Granger and Joyeux in the early 1980's, comes into the picture.

The ARFIMA model possesses theoretical properties which lie between the two worlds of ARMA and integrated processes, and it is able to model both the short and long run dynamics of a time series. Since they possess the latter property, ARFIMA models are often called long memory models.

In an ARIMA model, the *integrated* part of the model includes the differencing operator, in terms of the backspace operator B, as an integer power of $(1 - B)$. For example

$$(1-B)^2 = 1 - 2B + B^2,$$ Where $B^2 X_t = X_{t-2}$.

In a *fractional* model, the power is allowed to be fractional, with the meaning of the term identified using the following formal binomial series expansion

$$\begin{aligned}(1-B)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k \\ &= \sum_{k=0}^{\infty} \frac{\prod_{a=0}^{k-1}(d-a)\ (-B)^k}{k!} \\ &= 1 - dB + \frac{d(d-1)}{2}B^2 - \cdots\end{aligned}$$

2.2. Modeling with GARCH

GARCH stands for generalized autoregressive conditional heteroscedasticity. The word "autoregressive" indicates a feedback mechanism that incorporates past observations into the present. The word "conditional" indicates that variance has a dependence on the immediate past. The word "heteroscedasticity" indicates a time-varying variance (*volatility*). GARCH models, introduced by Bollerslev generalized Engle's earlier ARCH models to include autoregressive (AR) as well as moving average (MA) terms. GARCH is a mechanism that includes past variances in the explanation of future variances. More specifically, GARCH is a time series technique used to model the serial dependence of volatility. Whenever a time series is said to have *GARCH effects*, the series is heteroscedastic, meaning that its variance varies with time. If its variances remain constant with time, the series is homoscedastic.

GARCH builds on advances in the understanding and modeling of volatility in the last decade. It takes into account excess kurtosis (fat tail behavior) and volatility clustering, two important characteristics of financial time series. It provides accurate forecasts of variances and covariances of asset returns through its ability to model time-varying conditional variances.

2.3. FIGARCH models

It is well known that the volatility of financial time series, such as stock returns, can vary greatly over time and GARCH models can account for this. Other types of non-linear time series models, comprises a wide variety of representations like TARCH, EGARCH, FIGARCH or Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, CGARCH, etc.

Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally-varying variability, where the variability might be modeled as being driven by a separate time-varying process, as in a doubly stochastic model. FIGARCH models and their ability to model real world data are explored in this work. Generally it is believed that FIGARCH models can account for short and long term influences on the conditional variance of a process.

3. Literature survey and related works

Over the years researchers have tried various time series models for modeling real world data. Recently, considerable research has been made on exploring the applications of ARFIMA-FIGARCH model. Some among them are discussed here.

Long memory in Istanbul Stock Exchange was examined by Emrah Ismail Cevic *et al.* [1]. GRANN_ARIMA model that integrates nonlinear Grey Relational Artificial Neural Network (GRANN) and linear ARIMA model was proposed in [2]. Genetic Algorithm based hybrid forecasting model was used in [3] to forecast real world time series data. Long memory in the Turkish stock market was explored in [4]. An attempt was made in [5] to forecast future trends of financial time series using ridge polynomial neural network. A linkage between inflation and output growth were examined in [6] and highlights the importance of modeling long memory not only in the conditional mean but also in conditional variance. Long memory models and their application were studied in [7]. Reference [8] introduced a new time series forecasting model based on the flexible neural tree. Way of combining a few long memory models was proposed in [9]. An attempt was made in [10] to forecast temperature indices. Reference [11] shows that some financial processes have long memory in mean or variance, which can be described by ARFIMA-FIGARCH model. Wllfredo Palma *et al.* have analyzed the correlation structure [12]. FIGARCH modeling was used in forecasting Great Salt Lake surface level in

[13]. Practical aspects of likelihood-based inference and forecasting of series with long memory are considered in [14]. For the very first time, these kinds of models were proposed and introduced in [15]. In the context of non-stationarity [16] analyzes the behavior of volatility for several international stock market indexes. Reference [17] reported that, during the last couple of years ARFIMA type modeling of high-frequency squared returns has proved very fruitful.

4. Empirical modeling and investigations

4.1. Data used in the study

Financial time series data used in this study is closing stock prices of various indices of National Stock Exchange (NSE) of India. (Source: www.nseindia.com). Totally ten such indices, denoted as SERIES A, B, C, D, E, F, G, H, I and J are explored. The details about this stock index data are listed in Table 1 and Table 2. Descriptive statistics of dependent variable that is being modeled (actual series from A to J) is shown in Table 3a and Table 3b. Plot of this is shown in Figure 1. Autocorrelation Function (ACF) of the actual series under study is shown in Figure 2. Histogram and Distribution function (CDF- Cumulative Distribution Function) is shown in Figure 3 and Figure 4 respectively.

4.2. Residual details

The various details of residuals like Descriptive statistics of residuals, ACF of residuals, Spectrum of residuals, Plot of residuals, Plot of conditional variance and Histogram of standardized residuals are shown in Table 4a and Table 4b, Figure 5, Figure 6, Figure 7, Figure 8 and Figure 9 respectively.

4.3. ARFIMA-FIGARCH modeling

The command for estimating ARFIMA (p1, d1, q1)-FIGARCH (p2, d2, q2) model has the following form

arfimafigarch! (p1, q1, p2, q2) <variable>

Various values for p1, q1, p2, q2 were tried and arfimafigarch! (1, 1, 1, 1) A ….J, [24] produced the best results in terms of modeling and fitting. The actual estimates and statistics (results) obtained are shown in Table 5a and Table5b.

4.4. Actual versus fitted values

Plot of actual and fitted values, Scatter of actual versus fitted values and Scatter of residuals versus fitted values are shown in Figure 10, Figure 11 and Figure 12 respectively. Descriptive statistics of fitted values is shown in Table 6a and 6b.

5. Findings and conclusion

5.1. Observations

The various observations are as follows [19] [20]

5.1.1 Durbin-Watson statistic (DW). The calculated value of the DW varies between 0 and 4, with a value of 2 denoting no autocorrelations and values of 0 and 4 denoting perfect positive and negative autocorrelations, respectively [18]. For this study the calculated DW values are shown below. (Refer Table 5a and Table 5b)

Series	Calculated DW value
A	1.9636
B	1.9676
C	1.9824
D	1.9811
E	1.9605
F	2.0013
G	1.9706
H	1.9332
I	1.9803
J	1.9775

In all the cases, the calculated DW values are almost 2 that denote there is no autocorrelations as desired.

5.1.2. R^2. Ideally the value of R^2, should be 100% or the adjusted R^2 should be close to 1. The obtained values for R^2 are shown below. (Refer Table 5a and Table 5b)

Series	R^2 value
A	99.887949822%
B	99.809085876%
C	99.764649409%
D	99.709459575%
E	99.764150147%
F	99.878723895%
G	99.862122729%
H	99.45162314%
I	99.864379717%
J	99.855963211%

In all the cases, R^2 values are close to 100% as desired.

5.1.3 SE versus standard deviation. The standard deviations of the dependent variables are shown in (biased estimate) Table 3a and Table 3b and also in Table 5a and Table 5b. Standard Error of Estimates (SE or RSE) are shown in Table 5a and Table 5b. For good modeling, SE should be low relative to the standard deviation of the estimate. In all these cases, SE is comparatively very low relative to standard deviations of the dependent variable as shown below.

Series	SE	Standard deviation
A	41.63620589	1243.8822783
B	102.5496103	2347.01468
C	62.810822572	1295.1401717
D	68.882461447	1278.1867239
E	80.500123031	1657.8146725
F	71.704150217	2058.9899951
G	95.317723473	2566.9074027
H	30.486067732	411.42991693
I	43.007419017	1167.8049523
J	37.206235377	980.30367359

5.1.4. Akaike and Schwarz Bayesian Criteria. Many techniques and methods have been suggested to add mathematical rigour to the search process of an ARMA model, including Akaike's information criterion (AIC), Akaike's final prediction error (FPE), and the Bayes information criterion (BIC). Often these criteria come down to minimizing (in-sample) onestep- ahead forecast errors, with a penalty term for overfitting [26]. Schwarz developed the Bayesian information criterion (BIC) and Akaike developed the information criterion (AIC), for selecting the models that trade off model complexity and the error in fitting so as to achieve the most accurate out-of-sample forecasts. They suggest going for models that have the lowest AIC or BIC respectively. In this study, the obtained values are low for ARFIMA-FIGARCH model, which denotes the case of exact modeling. [22] [23].

5.2. Performance comparison

The performance of the ARFIMA-FIGARCH model is compared with the traditional models like ARIMA (1, 1, 0), ARIMA (0, 0, 1), ARIMA (0, 1, 1), ARIMA (1, 0, 0), ARIMA (1, 0, 1), ARIMA (1, 1, 1), ARIMA (1, 2, 1) and is shown in Table 7. In all the cases of time series under analysis and investigation, the ARFIMA-FIGARCH model outperforms all other ARIMA models.

5.3. Conclusion

From this research study, it is strongly concluded that, all the financial time series under study, exhibits long memory and both short term and long term influences are observed. It is proven that, ARFIMA-FIGARCH model has the ability to model this accurately.

6. Acknowledgment

The authors would like to acknowledge [24] for some computations.

7. References

[1] Emrah Ismail Cevik and Nesrin Ozatac, "Testing for long memory in ISE using ARFIMA-FIGARCH model and structural break test", International Research Journal of Finance and Economics, ISSN 1450-2887, Issue 26,2009, pp. 186-191.

[2] Roselina Sallehuddin, Siti Mariyam Hj. Shamsuddin, Siti Zaiton Mohd. Hashim and Ajith Abraham, "Forecasting time series data using hybrid grey relational artificial neural network and auto regressive integrated moving average model", Neural Network World 6/07, 2007, pp-573-605.

[3] Ruhaidah Samsudin, Puteh Saad and Ani Shabri, "Combination of Froecasting using modified GMDH and Genetic Algorithm", International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Vol.1 (2009), pp. 170-176, ISSN: 2150-7988

[4] Adnan Kasman and Erdost Torun, "Long memory in the Turkish stock market return and volatility", Central Bank Review 2 (2007) pp. 13-27, ISSN 1303-0701 print / 1305-8800 online, Central Bank of the Republic of Turkey, 2007.

[5] R. Ghazali, N. Mohd Nawi and M. Z. Mohd. Salikon, "Forecasting the UK/EU and JP/UK trading signals using polynomial neural networks", International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Vol.1 (2009), pp. 110-117, ISSN: 2150-7988

[6] Christian Conrad and Menelaos Karanasos, "Dual Long Memory in Inflation Dynamics across Countries of the Euro Area and the Link between Inflation Uncertainty and Macroeconomic Performance", Studies in Nonlinear Dynamics & Econometrics, Volume 9, Issue 4, Article 5, 2005, pp. 1-36.

[7] Wolfgang Karl Hardle and Julius Mungo, "Long memory persistence in the factor of implied volatility dynamics", International Research Journal of Finance and Economics, ISSN 1450-2887 Issue 18, 2008, pp. 213-230.

[8] Yuehui Chen, Bo Yang, Jiwen Dong and Ajith Abraham, "Time series forecasting using flexible neural tree model", Information Sciences, Vol 174, Issues 3-4, 2005, pp. 219-235.

[9] Chin Wen Cheong, "A generalized discrete-time long memory volatility model for financial stock exchange", American Journal of Applied Sciences 4 (12): pp. 970- 976, ISSN 1546-9239, Science Publications, 2007.
[10] Massimiliano Caporin and Juliusz Pres, "Forecasting temperature indices with time varying long-memory models", Marco Fanno working paper, N.88, January 2009, pp. 1-58.
[11] Piotr Fiszeder, "Modeling financial processes with long memory in mean and variance", Dynamic econometric models, Vol. 7, Nicolaus Copernicus university, 2006, pp. 133-142.
[12] Wilfredo Palma and Maurico Zevallos, "Analysis of the correlation structure of square time series", Journal of Time Series Analysis, vol. 25, No 4, 2004, pp. 529-550.
[13] Qianru Li, Christophe Tricaud, Rongtao Sun and YangQuan Chen, "Great salt lake surface level forecasting using FIGARCH modeling", Proc. ASME 2007, International Design Engineering Technical Conferences & Computers and Information in Engineering Conference IDETC/CIE 2007 September 4-7, 2007, Las Vegas, Nevada, USA
[14] Jurgen A Doornik and Marius Ooms, "Inference and forecasting for ARFIMA models with an application to US and UK inflation", Studies in Nonlinear Dynamics & Econometrics, Volume 8, Issue 2, Article 14, 2004 (Linear and Nonlinear Dynamics in Time Series, Estella Bee Dagum and Tommaso Proietti, Editors)
[15] Tim Bollerslev and Hans Ole Mikkelsen, "Modeling and pricing long memory in stock market volatility", Elsevier Journal of Econometrics 73, 1996, pp. 151-184.
[16] Andreia Dionisio, Rui Menezes and Diana A Mendes, "On the integrated behaviour of non-stationarity volatility in stock markets", Physica A 382, 2007, pp. 58-65.
[17]Anders Wilhelmsson, "GARCH forecasting performance under different distribution assumptions", Journal of Forecasting, J. Forecast. 25, 2006, pp. 561–578, DOI: 10.1002/for.1009
[18] Stephen A. Delurgio, Forecasting Principles and Applications, (McGraw-Hill International Editions, 1998)
[19] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Time Series Analysis Forecasting and Control (Pearson Education, Inc. 2004)
[20] Ruey S. Tsay, Analysis of Financial Time Series (John Wiley & Sons, Inc., 2002)
[21] P. Bagavathi Sivakumar, V. P. Mohandas, "Evaluating the predictability of financial time series A case study on Sensex data", Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, Sobh, Tar(Ed.), 2007, XVIII, ISBN: 978-1-4020-6267-4
[22] Li, Wai Keung, *Diagnostic checks in time series*, (Monographs on statistics and applied probability; 102, Chapman & Hall/CRC, 2004)
[23] Dimitris G. Manolakis, Vinay K. Ingle, Stephen M. Kogon, *Statistical and Adaptive Signal Processing, Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing* (McGraw-Hill International Editions, 2000)
[24] Matrixer econometric program
[25] Ruey S. Tsay, *Analysis of Financial Time Series (*John Wiley & Sons, Inc., 2002)
[26] Jan G. De Gooijer, Rob J. Hyndman, "25 years of time series forecasting", International Journal of Forecasting 22 (2006) 443– 473, Elsevier.
[27] Andrew C. Harvey, *Time Series Models*, (The MIT Press, Cambridge, Massachusetts, second edition)

Table 1. Time series data used in the study

Index	Denoted as SERIES	Period (Data range)	Number of Data/observations
S & P CNX NIFTY	A	03-07-1990 to 26-06-2009	4530
BANK NIFTY	B	01-01-2000 to 26-06-2009	2370
CNX 100	C	01-01-2003 to 26-06-2009	1619
CNX INFRASTRUCTURE	D	01-01-2004 to 26-06-2009	1366
CNX IT	E	01-01-1996 to 26-06-2009	3347
CNX MIDCAP	F	01-01-2001 to 26-06-2009	2119
CNX NIFTY JUNIOR	G	04-10-1995 to 26-06-2009	3429
CNX REALTY	H	01-01-2007 to 26-06-2009	612
S & P CNX 500	I	07-06-1999 to 26-06-2009	2513
S & P CNX DEFTY	J	03-08-1990 to 26-06-2009	4511

Table 2. Description of the Time series data (stock market price indices) used in the study

Index	Details
S & P CNX NIFTY	50 stock index accounting for 22 sectors of the economy
BANK NIFTY	Benchmark of Indian banking sector comprising 12 stocks from banking sector
CNX 100	Combined portfolio of two indices viz Nifty and Nifty junior
CNX INFRASTRUCTURE	Portfolio of infrastructure
CNX IT	Captures the performance of IT segment
CNX MIDCAP	Medium capitalized segment of stock market; attractive investment segment with high growth potential
CNX NIFTY JUNIOR	50 stock index for 23 sectors; NIFTY and NIFTY Junior are disjoint
CNX REALTY	Realty sector
S & P CNX 500	Broad based benchmark of Indian capital market; represents 92.66% of total capitalization; 72 indices
S & P CNX DEFTY	Measuring returns on equity investments in dollar terms

Table 3a. Descriptive statistics of dependent variable (data being modeled)

Variable (Series)	A	B	C	D	E
Minimum	279.02	743.7	863.15	841.11	76.247
Maximum	6287.85	10698.35	6205.1	6260.66	9550.155
Mean	1697.3863068	3357.0011224	2850.3381779	2595.3034407	2373.8843116
Median	1159.5	2818.49	2703.25	2390.61	2088.15
Variance (biased estimate)	1547243.1222	5508477.9082	1677388.0644	1633761.3011	2748349.4882
Variance (unbiased estimate)	1547584.7523	5510803.1416	1678424.769	1634958.1958	2749170.8718
Standard deviation (biased estimate)	1243.8822783	2347.01468	1295.1401717	1278.1867239	1657.8146725
Standard deviation (unbiased estimate)	1244.0195948	2347.5099875	1295.5403386	1278.6548384	1658.0623848
Asymmetry	1.5969270596	0.7855499006	0.4149319638	0.6985509961	0.5299225754
Excess kurtosis	1.7008895242	-0.195540225	-0.698152882	-0.173304246	-0.0767285951
Coefficient of variation	0.7329030462	0.6992878173	0.4545216244	0.4926802848	0.6984596413
Sum	7689159.97	7956092.66	4614697.51	3545184.5	7945390.791
Sum of squares about mean	7009011343.4	13055092642	2715691276.2	2231717937.3	9198725737.2
Sum of squares	20060486188	39763704632	15869139768	11432547468	28060164286
1-st order autocorrelation	0.9987839191	0.9981799539	0.9978271971	0.9977210006	0.9984214372

Table 3b. Descriptive statistics of dependent variable (data being modeled)

Variable (Series)	F	G	H	I	J
Minimum	608.43	912.89	154.94	545.85	526.9
Maximum	9655.45	13069.45	1798.65	5502.6	5548.5
Mean	3181.6461067	3420.265331	793.9879902	1879.750386	1442.411605
Median	2964.27	2439.7	823.305	1484.2	1060.1
Variance (biased estimate)	4239439.8	6589013.6139	169274.57655	1363768.4067	960995.29245
Variance (unbiased estimate)	4241441.4241	6590935.7298	169551.62168	1364311.3082	961208.37344
Standard deviation (biased estimate)	2058.9899951	2566.9074027	411.42991693	1167.8049523	980.30367359
Standard deviation (unbiased estimate)	2059.4760072	2567.2817784	411.76646498	1168.0373745	980.41234868
Asymmetry	0.6087783077	1.2277766785	0.1545818454	0.8790629412	1.961704381
Excess kurtosis	-0.469511158	0.8399077834	-0.761202249	-0.228155515	3.3646939354
Coefficient of variation	0.6472988944	0.7506089528	0.5186054072	0.6213789784	0.6797035918
Sum	6741908.1	11728089.82	485920.65	4723812.72	6506718.75
Sum of squares about mean	8983372936.2	22593727682	103596040.85	3427150006.1	4335049764.2
Sum of squares	30433738594	62706906692	489411201.13	1230673879	13720416399
1-st order autocorrelation	0.9988320993	0.9987277145	0.9963577362	0.9987450266	0.9988784199

Table 4a. Descriptive statistics of residuals

Variable (Series)	A	B	C	D	E
Minimum	-464.9911104	-654.9602152	-491.4493285	-595.3636392	-608.08525506
Maximum	0.7925711664	1043.8262615	591.74229955	548.6654894	537.53108458
Mean	0.7925711664	2.4491875811	0.1638891707	0.0931598598	0.9434729418
Median	0.5630494957	1.0027926842	2.2871117235	1.7780310891	0.3479792646
Variance (biased estimate)	1732.9454719	10510.424053	3945.1725725	4744.7848163	6479.3796669
Variance (unbiased estimate)	1733.3281895	10514.862577	3947.6123824	4748.2633975	6481.3167011
Standard deviation (biased estimate)	41.628661663	102.52035921	62.810608758	68.882398451	80.494594023
Standard deviation (unbiased estimate)	41.633258214	102.54200396	62.830027713	68.90764397	80.5066252
Asymmetry	0.4422083276	0.2932207303	0.0733862725	-0.115265792	-0.3214331484
Excess kurtosis	27.805324758	11.899789776	11.419255537	11.991788564	10.462459847
Coefficient of variation	52.529362634	41.8677625	383.36900145	739.67097086	85.330083811
Sum	3589.5548127	5802.1253796	265.17267816	127.16320868	3156.8604634
Sum of squares about mean	7848510.0421	24899194.581	6383289.2224	6476631.2743	21680004.365
Sum of squares	7851355.0197	24913405.075	6383332.6813	6476643.1208	21682982.778
1-st order autocorrelation	-0.065861838	0.0271642234	-0.038888361	-0.035701347	0.1189082943

Table 4b. Descriptive statistics of residuals

Variable (Series)	F	G	H	I	J
Minimum	-870.4906609	-1281.558409	-192.4648746	-460.6186320	-433.03447633
Maximum	613.03728782	905.71542097	171.21726128	451.61286331	525.3144236
Mean	1.578133658	1.7066579695	-0.682917277	0.8945493937	0.0950696281
Median	2.871885567	2.1863138659	-0.215886713	1.6448600012	0.2409383263
Variance (biased estimate)	5138.9946525	9082.5557266	928.93394977	1848.8378719	1384.2949127
Variance (unbiased estimate)	5141.4221417	9085.20602	930.45679231	1849.5741673	1384.6019198
Standard deviation (biased estimate)	71.686781574	95.302443445	30.47841777	42.998114748	37.206113916
Standard deviation (unbiased estimate)	71.703710794	95.316347077	30.503389849	43.006675846	37.210239448
Asymmetry	-0.938907353	-0.750663106	-0.146187704	-0.189250428	0.2996251822
Excess kurtosis	21.779047224	24.213998327	5.7396764024	18.120870515	25.848962151
Coefficient of variation	45.435765487	55.849706724	-44.66630273	48.076356821	391.39986348
Sum	3342.4870877	5850.4235195	-417.2624563	2247.108077	428.7640226
Sum of squares about mean	10884390.674	31135001.031	567578.64331	4644280.7341	6243170.0563
Sum of squares	10889665.565	31144985.703	567863.59905	4646290.8833	6243210.8188
1-st order autocorrelation	-0.038704025	-0.015064607	-0.032612247	-0.043102634	-0.0636702056

Table 5a. The actual estimates and statistics (results) obtained

ARFIMA-FIGARCH MODEL					
	A	B	C	D	E
DW	1.9636	1.9676	1.9824	1.9811	1.9605
R^2	99.887949822%	99.809085876%	99.764649409%	99.709459575%	99.764150147%
Standard deviation **(dependent variable)**	1243.8822783	2347.01468	1295.1401717	1278.1867239	1657.8146725
S.E	41.63620589	102.5496103	62.810822572	68.882461447	80.500123031
AIC	9.0484111802	10.70189112	10.212820249	10.361884507	10.063969891
BIC	9.0597483521	10.721376888	10.239465223	10.392471523	10.078591579

Table 5b. The actual estimates and statistics (results) obtained

ARFIMA-FIGARCH MODEL					
	F	G	H	I	J
DW	2.0013	1.9706	1.9332	1.9803	1.9775
R^2	99.878723895%	99.862122729%	99.45162314%	99.864379717%	99.855963211%
Standard deviation **(dependent variable)**	2058.9899951	2566.9074027	411.42991693	1167.8049523	980.30367359
S.E	71.704150217	95.317723473	30.486067732	43.007419017	37.206235377
AIC	10.037214988	10.496036667	9.3630596184	9.2134284479	8.9172588996
BIC	10.058586152	10.510365097	9.4208677619	9.2319916152	8.9286363762

Table 6a. Descriptive statistics of fitted values

Variable (Series)	A	B	C	D	E
Minimum	277.81642247	738.76327442	865.10125695	826.73800192	76.207358387
Maximum	6287.1578437	10707.100247	6215.8329286	6264.2369413	9605.3782389
Mean	1696.90691	3355.5468698	2851.3178846	2596.3790013	2373.6204216
Median	1158.9047501	2818.2724818	2704.8196348	2393.4638809	2088.7695488
Variance (biased estimate)	1544893.1196	5503013.5935	1676261.3596	1632596.4285	2749360.183
Variance (unbiased estimate)	1545234.3062	5505337.5012	1677298.0086	1633793.3467	2750182.1143
Standard deviation (biased estimate)	1242.9372951	2345.8502922	1294.7051246	1277.7309687	1658.1194719
Standard deviation (unbiased estimate)	1243.0745377	2346.345563	1295.1054044	1278.1992594	1658.3673038
Asymmetry	1.5986514397	0.7871950924	0.4168140755	0.70091958	0.5314639298
Excess kurtosis	1.7092962058	-0.190143395	-0.694800879	-0.169173877	-0.0715984852
Coefficient of variation	0.7325531709	0.6992438652	0.4542129138	0.4923007229	0.6986657549
Sum	7685291.3952	7949290.5346	4613432.3373	3544057.3368	7942133.9305
Sum of squares about mean	6996820938.5	13036639203	2712190879.9	2228494124.8	9199359172.3
Sum of squares	20038045012	39710856174	15866553013	11430210174	28050970461
1-st order autocorrelation	0.9986668492	0.9980052389	0.9975844774	0.9973374182	0.9982273867

Table 6b. Descriptive statistics of fitted values

Variable (Series)	F	G	H	I	J
Minimum	598.24187011	909.55711444	154.22908027	542.05955021	518.18486803
Maximum	9659.7300842	13084.148728	1803.2933101	5513.0790835	5550.1495821
Mean	3181.0282592	3419.1982428	794.33373561	1879.290789	1442.4874137
Median	2967.7500549	2433.781367	823.31208872	1481.0595127	1060.9471955
Variance (biased estimate)	4239500.2587	6587135.7811	169562.6151	1362821.4347	960423.93784
Variance (unbiased estimate)	4241502.8569	6589057.91	169840.5866	1363364.1752	960636.93938
Standard deviation (biased estimate)	2059.0046767	2566.5415993	411.77981386	1167.3994324	980.01221311
Standard deviation (unbiased estimate)	2059.4909218	2566.916031	412.11720008	1167.6318663	980.12087998
Asymmetry	0.6108264601	1.2295803898	0.1563986597	0.8808960711	1.9636326665
Excess kurtosis	-0.463662929	0.8477138353	-0.755138098	-0.222222971	3.3750797723
Coefficient of variation	0.6474293071	0.7507362395	0.5188212229	0.6213151648	0.6794658107
Sum	6737417.8529	11721011.576	485337.91246	4720778.4619	6505618.236
Sum of squares about mean	8979261548	22580701458	103602757.83	3423407444	4331511959.7
Sum of squares	30411178132	62657163644	489123034.86	12295122924	13715784384
1-st order autocorrelation	0.9985291195	0.9985173575	0.9953687588	0.9985461891	0.9987072151

Table 7. Performance comparison of ARFIMA_FIGARCH

Time Series	AIC/ BIC	ARFIMA FIGARCH	ARIMA (0, 0, 1)	ARIMA (0, 1, 1)	ARIMA (1, 0, 0)	ARIMA (1, 0, 1)	ARIMA (1, 1, 0)	ARIMA (1, 1, 1)	ARIMA (1, 2, 1)
A	AIC	9.0484111802	10.2926	10.2924	10.2982	10.2926	10.2930	10.2927	10.2926
	BIC	9.0597483521	10.2969	10.2967	10.3024	10.2969	10.2986	10.2984	10.2997
B	AIC	10.70189112	12.1280	12.7004	12.1290	12.1022	12.1013	12.1010	12.0997
	BIC	10.721376888	12.1329	12.1077	12.1363	12.1095	12.1110	12.1107	12.1119
C	AIC	10.212820249	11.1268	11.1208	11.1283	11.1209	11.1222	11.1220	11.1228
	BIC	10.239465223	11.1335	11.1308	11.1383	11.1308	11.1355	11.1353	11.1394
D	AIC	10.361884507	11.3178	11.3070	11.3193	11.3079	11.3083	11.3072	11.3085
	BIC	10.392471523	11.3254	11.3185	11.3307	11.3193	11.3236	11.3225	11.3276
E	AIC	10.063969891	11.6489	11.6057	11.6493	11.5995	11.6060	11.5979	11.5976
	BIC	10.078591579	11.6526	11.6112	11.6548	11.6050	11.6134	11.6053	11.6067
F	AIC	10.037214988	11.4316	11.3849	11.4328	11.3850	11.3859	11.3857	11.3861
	BIC	10.058586152	11.4369	11.3929	11.4408	11.3930	11.3966	11.3964	11.3995
G	AIC	10.496036667	11.9790	11.9534	11.9797	11.9538	11.9541	11.9540	11.9546
	BIC	10.510365097	11.9826	11.9588	11.9850	11.9592	11.9612	11.9612	11.9636
H	AIC	9.3630596184	9.6969	9.6830	9.6992	9.6807	9.6849	9.6829	9.6829
	BIC	9.4208677619	9.7113	9.7047	9.7209	9.7024	9.7138	9.7113	9.7190
I	AIC	9.2134284479	10.3756	10.3610	10.3765	10.3608	10.3619	10.3616	10.3624
	BIC	9.2319916152	10.3802	10.3679	10.3834	10.3677	10.3712	10.3768	10.3740
J	AIC	8.9172588996	10.0739	10.0681	10.0743	10.0680	10.0685	10.0684	10.0688
	BIC	8.9286363762	10.0767	10.0724	10.0786	10.0722	10.0742	10.0740	10.0759

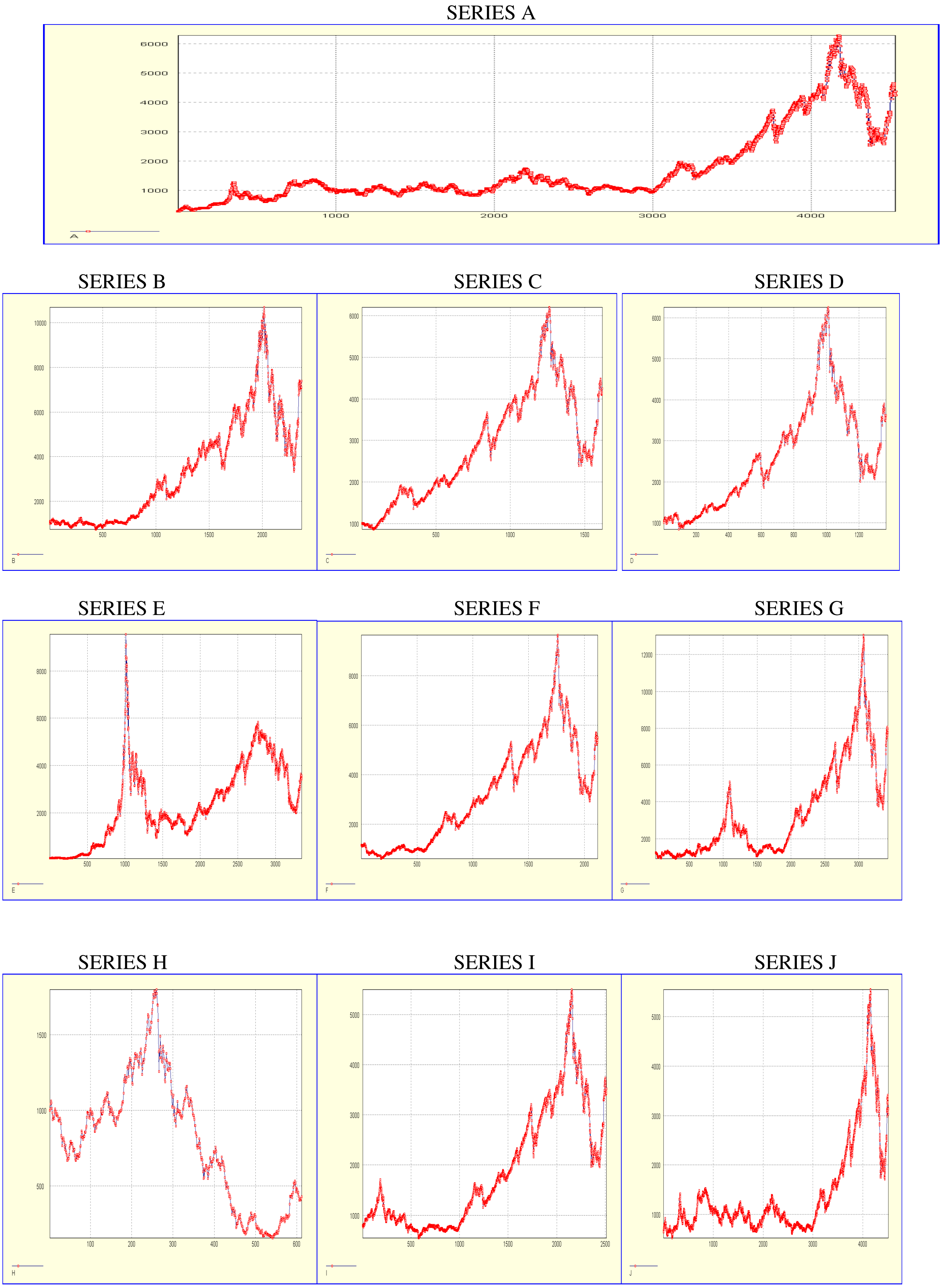


Figure 1. Plot of time series data under study

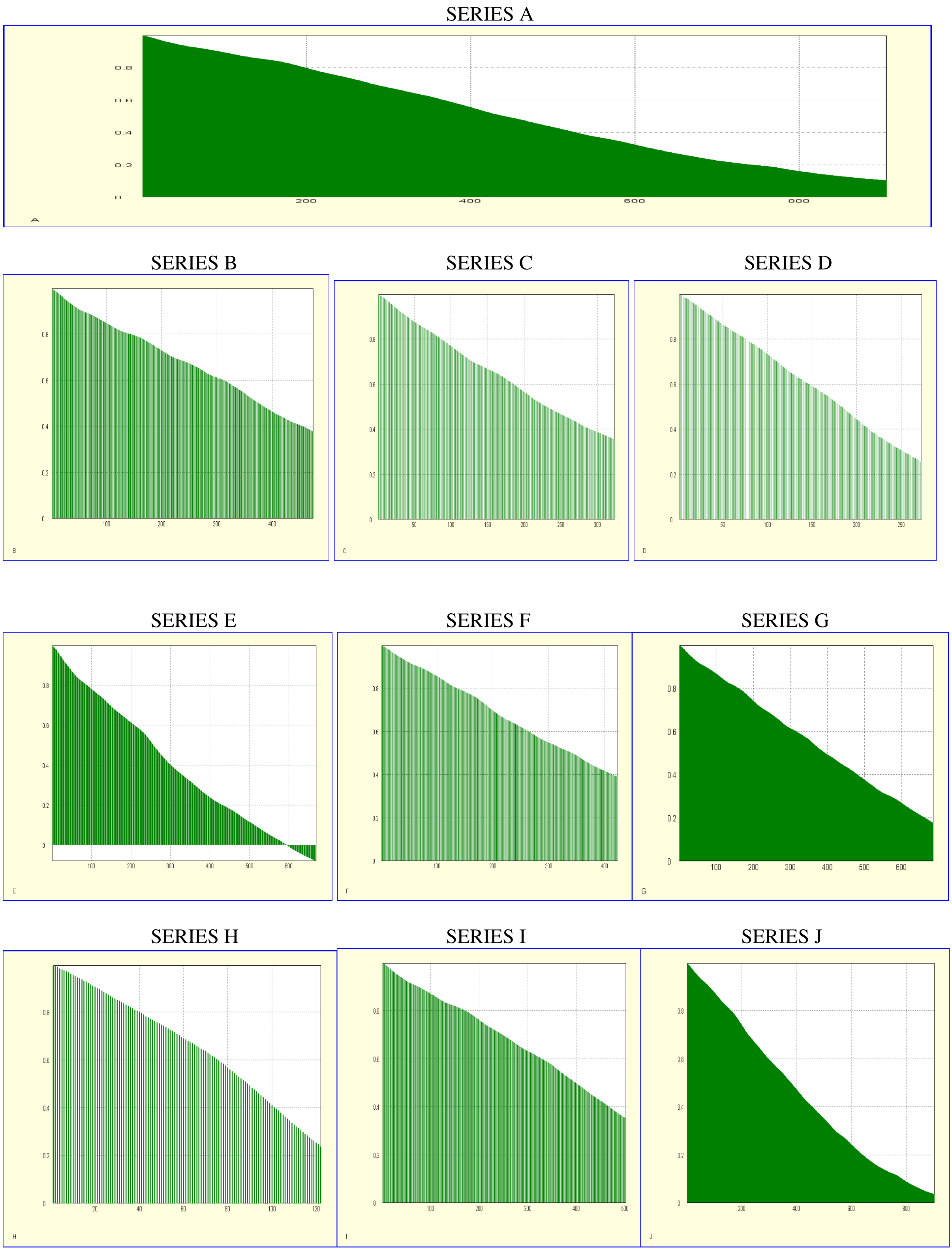


Figure 2. Autocorrelation function (ACF)

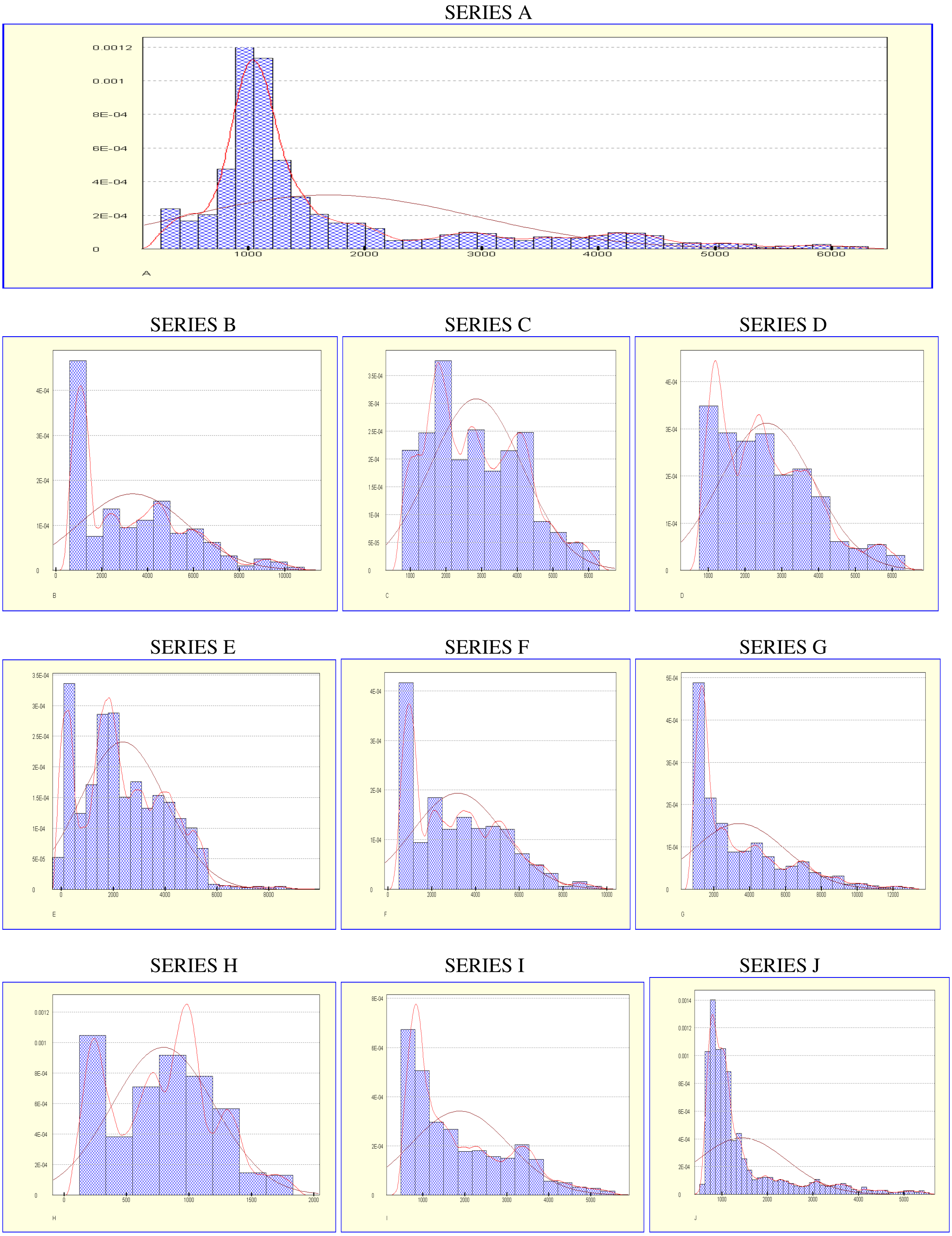


Figure 3. Histogram

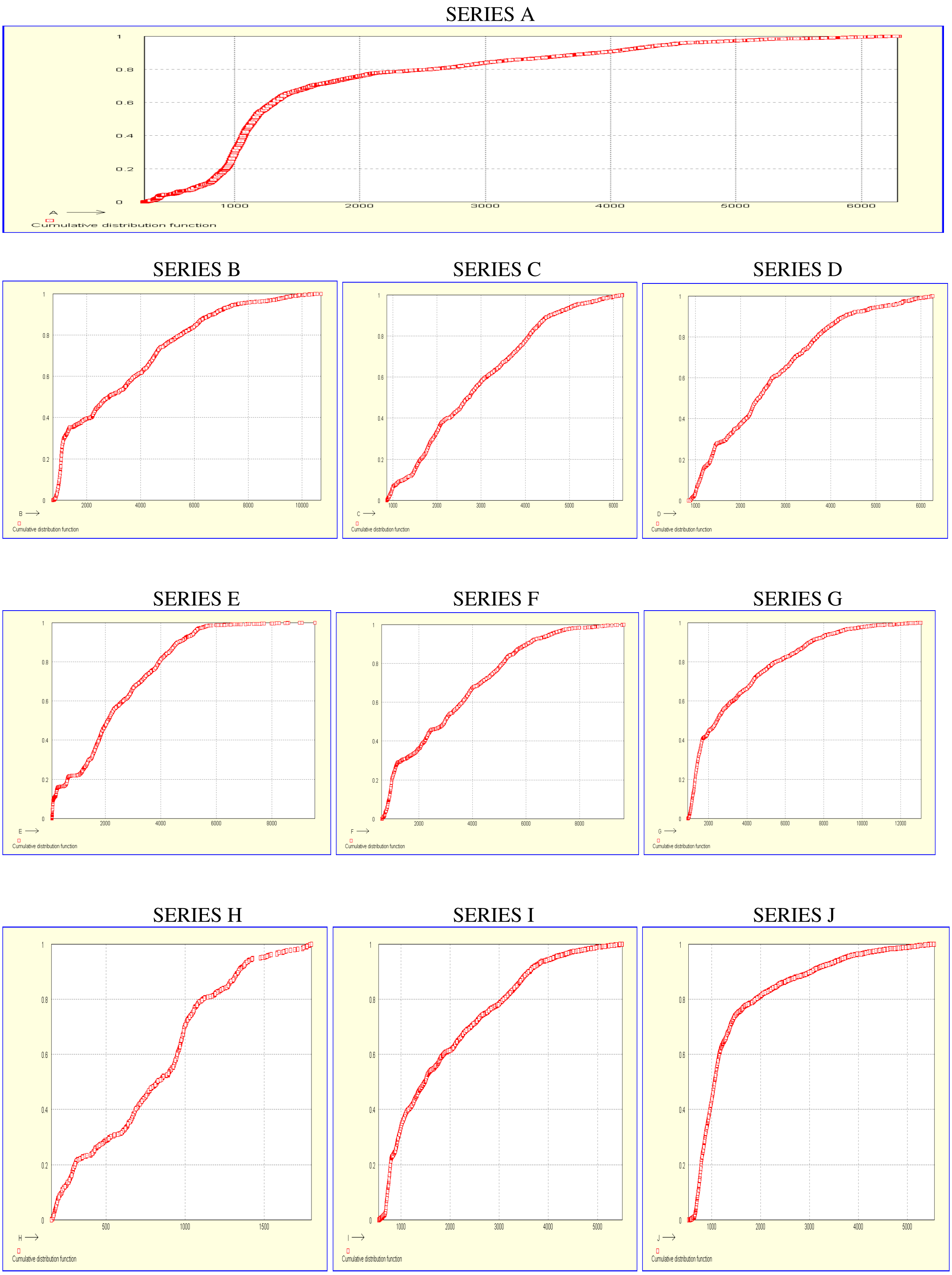


Figure 4. CDF- Cumulative Distribution Function

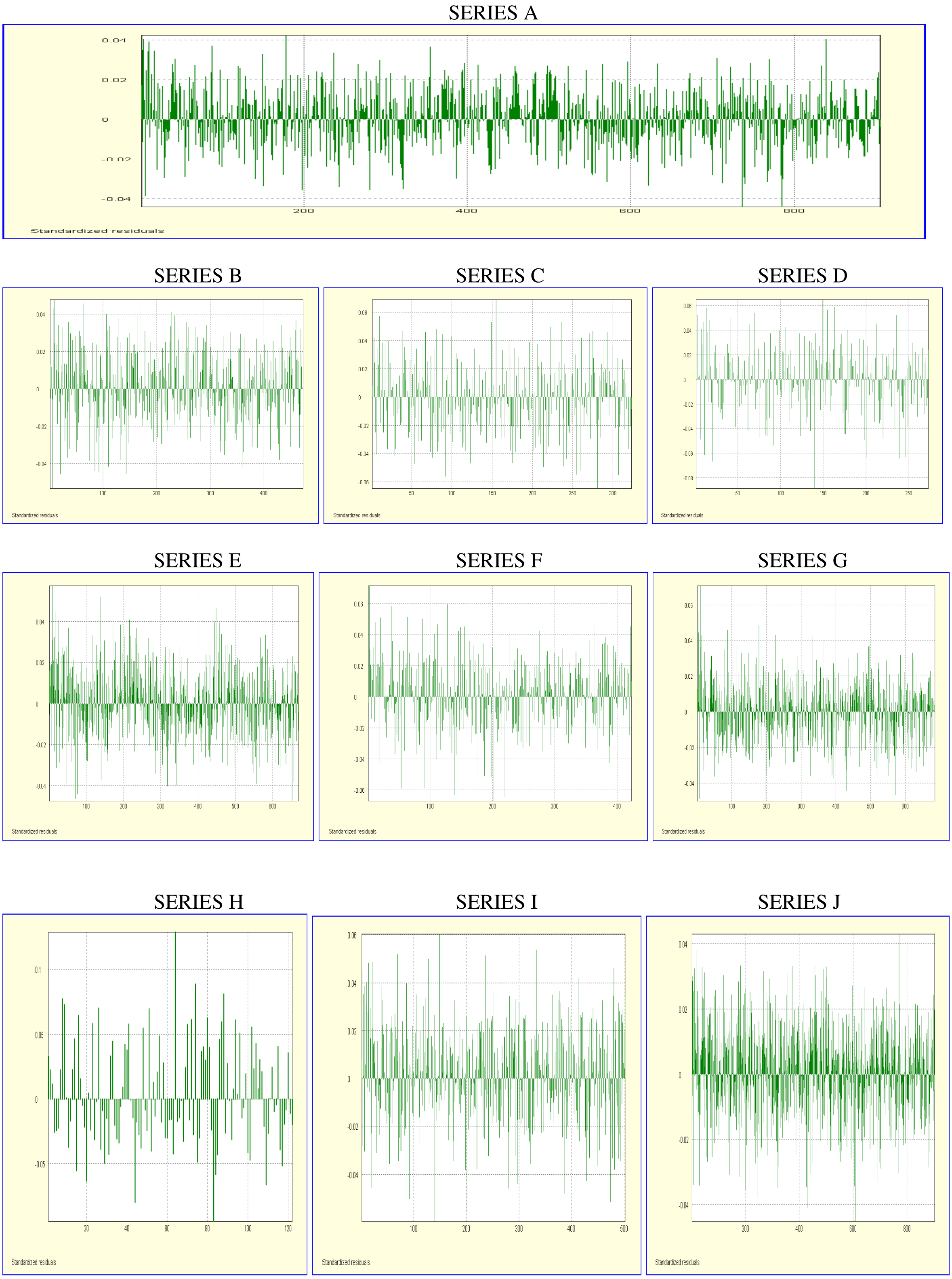


Figure 5. ACF of residuals

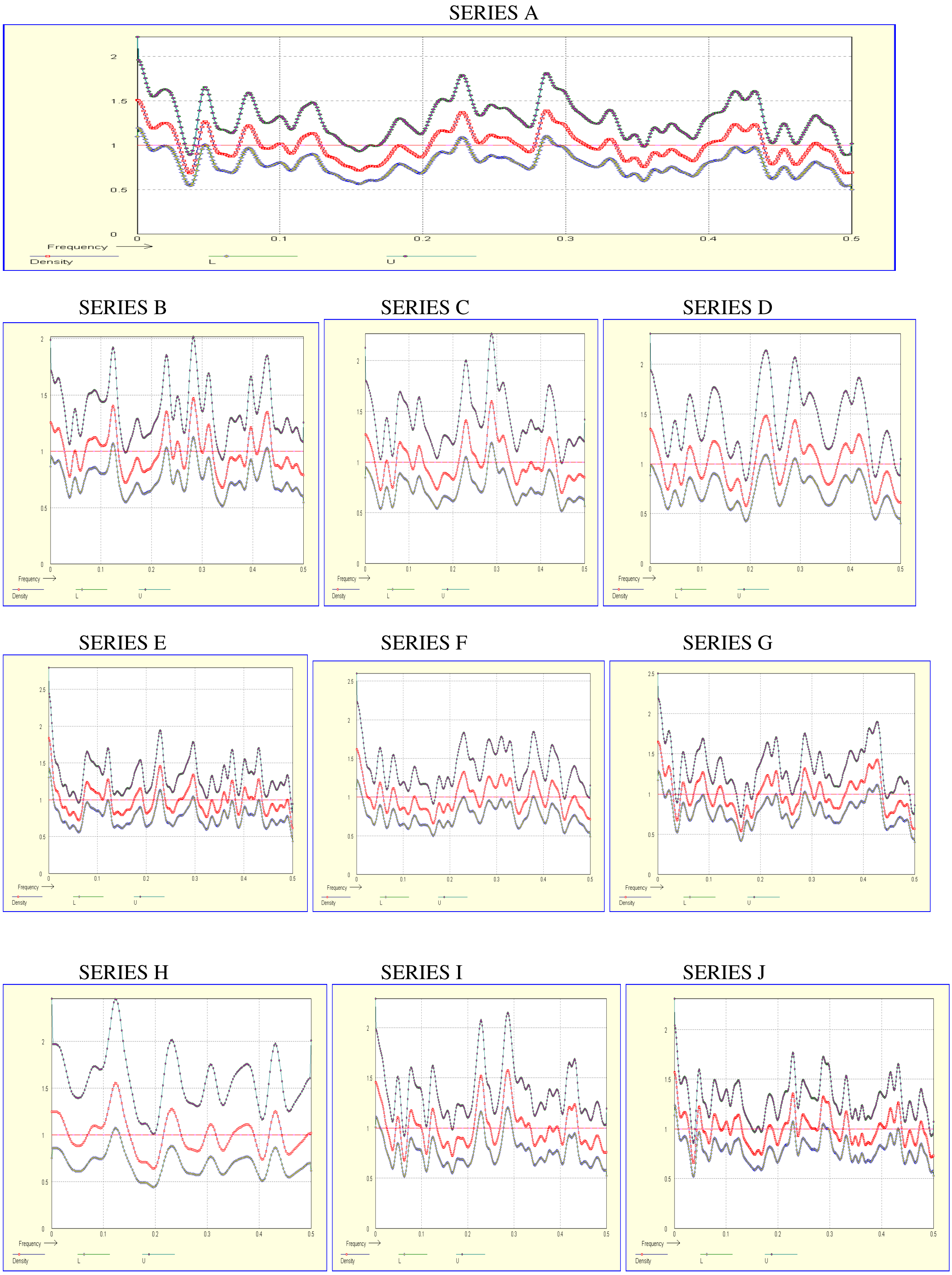


Figure 6. Spectrum of residuals

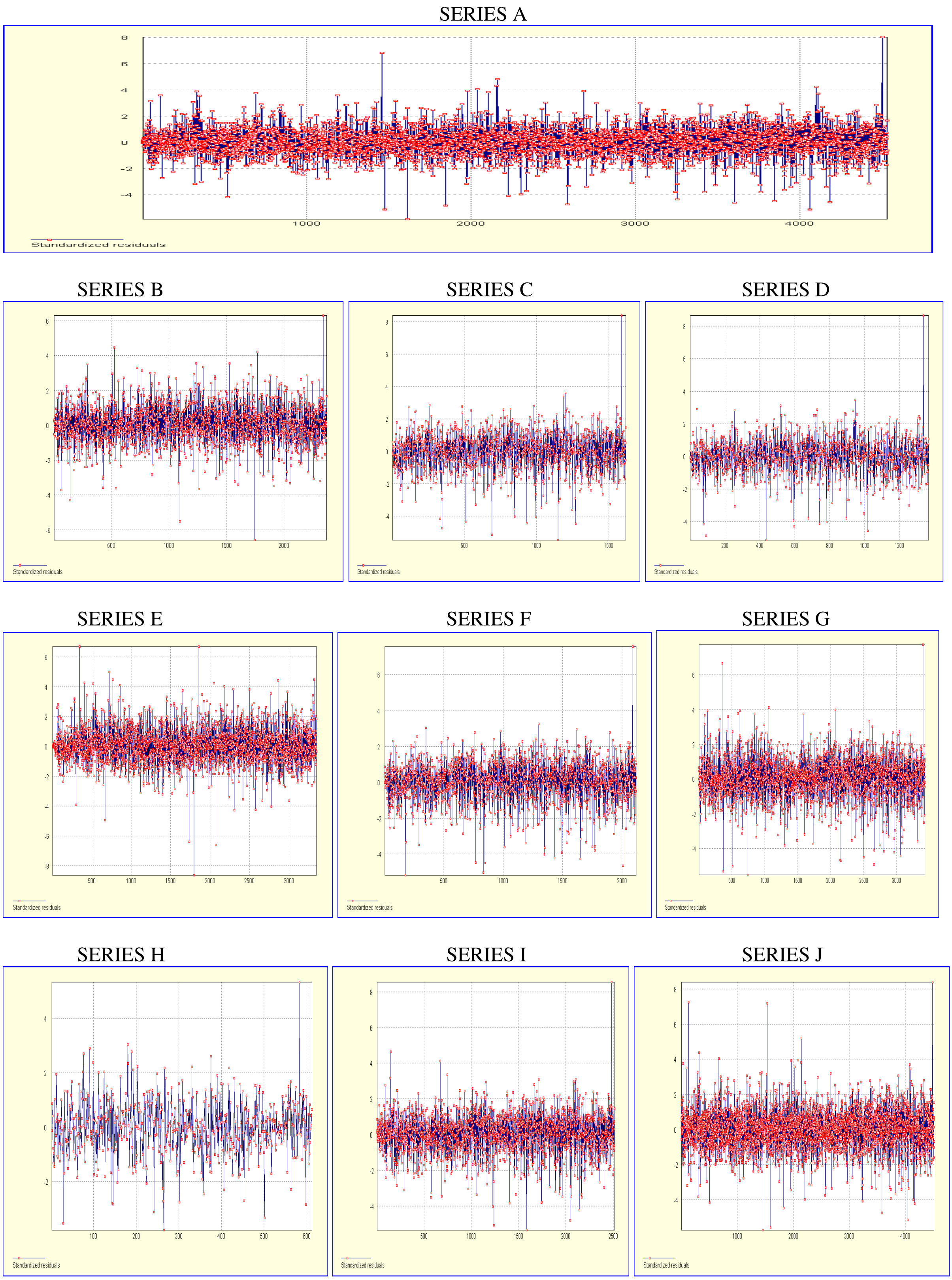


Figure 7. Plot of residuals

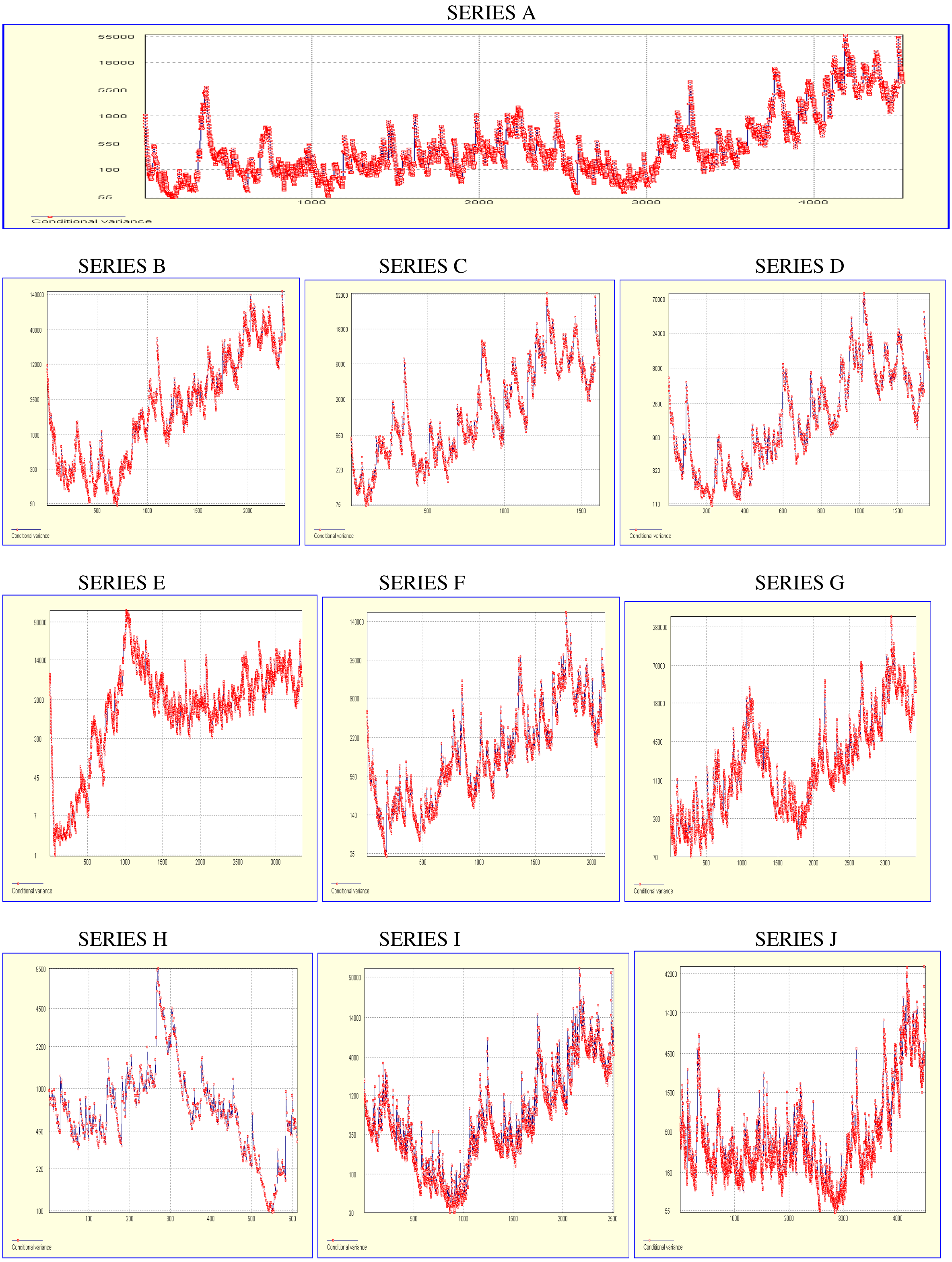


Figure. 8. Plot of conditional variance

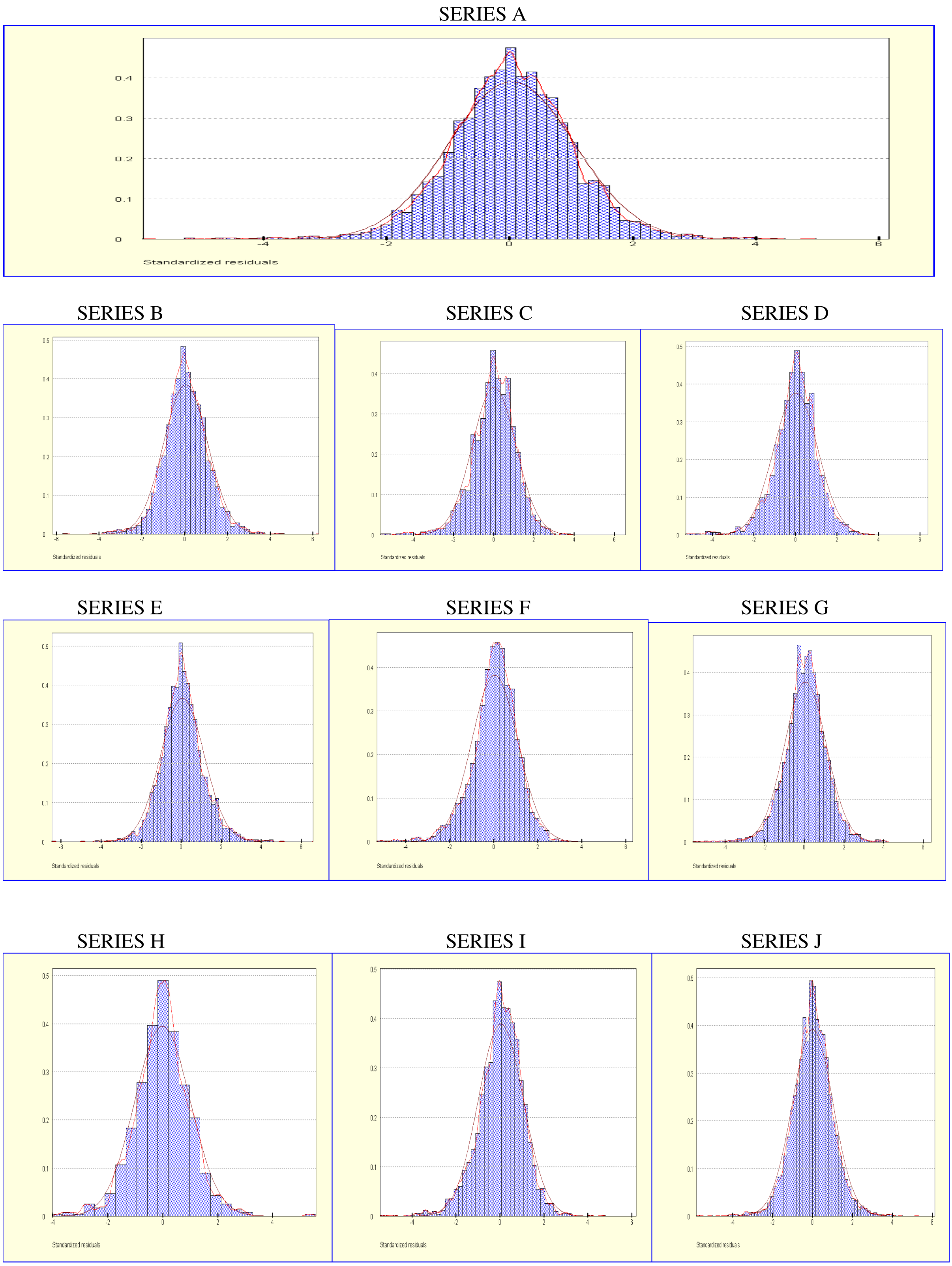


Figure 9. Histogram of standardized residuals

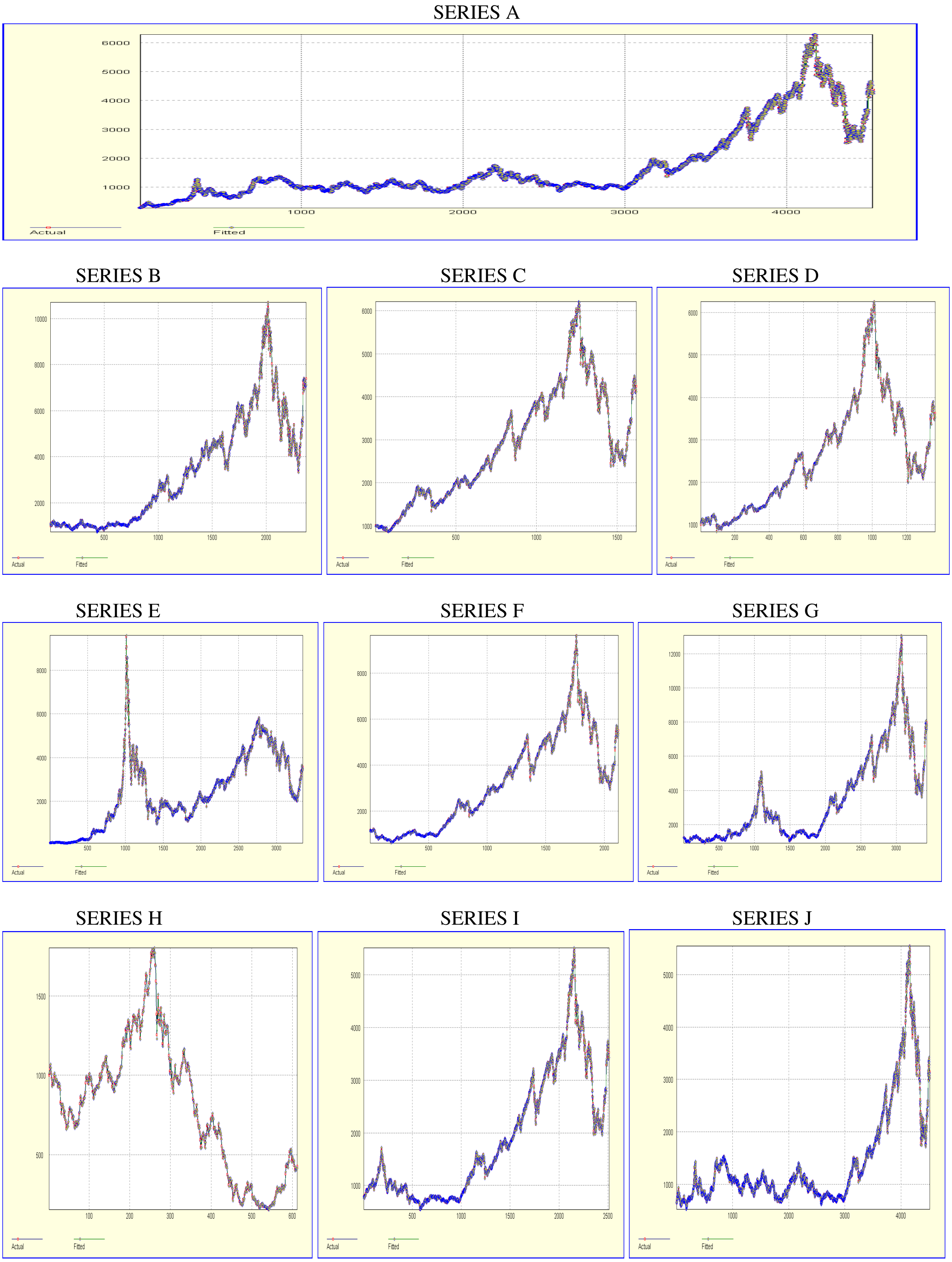


Figure 10. Plot of actual and fitted values

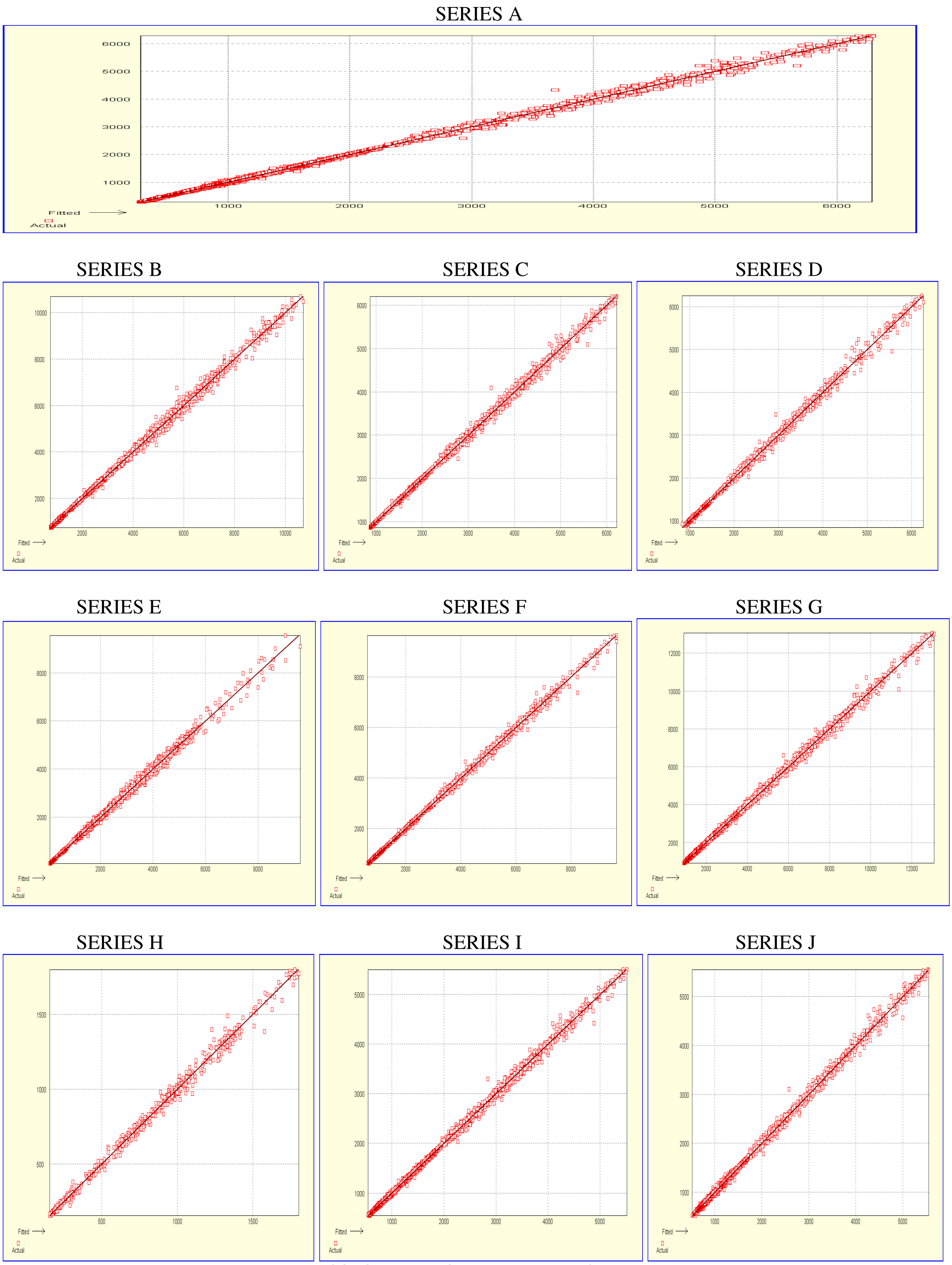


Figure 11. Scatter of actual versus fitted values

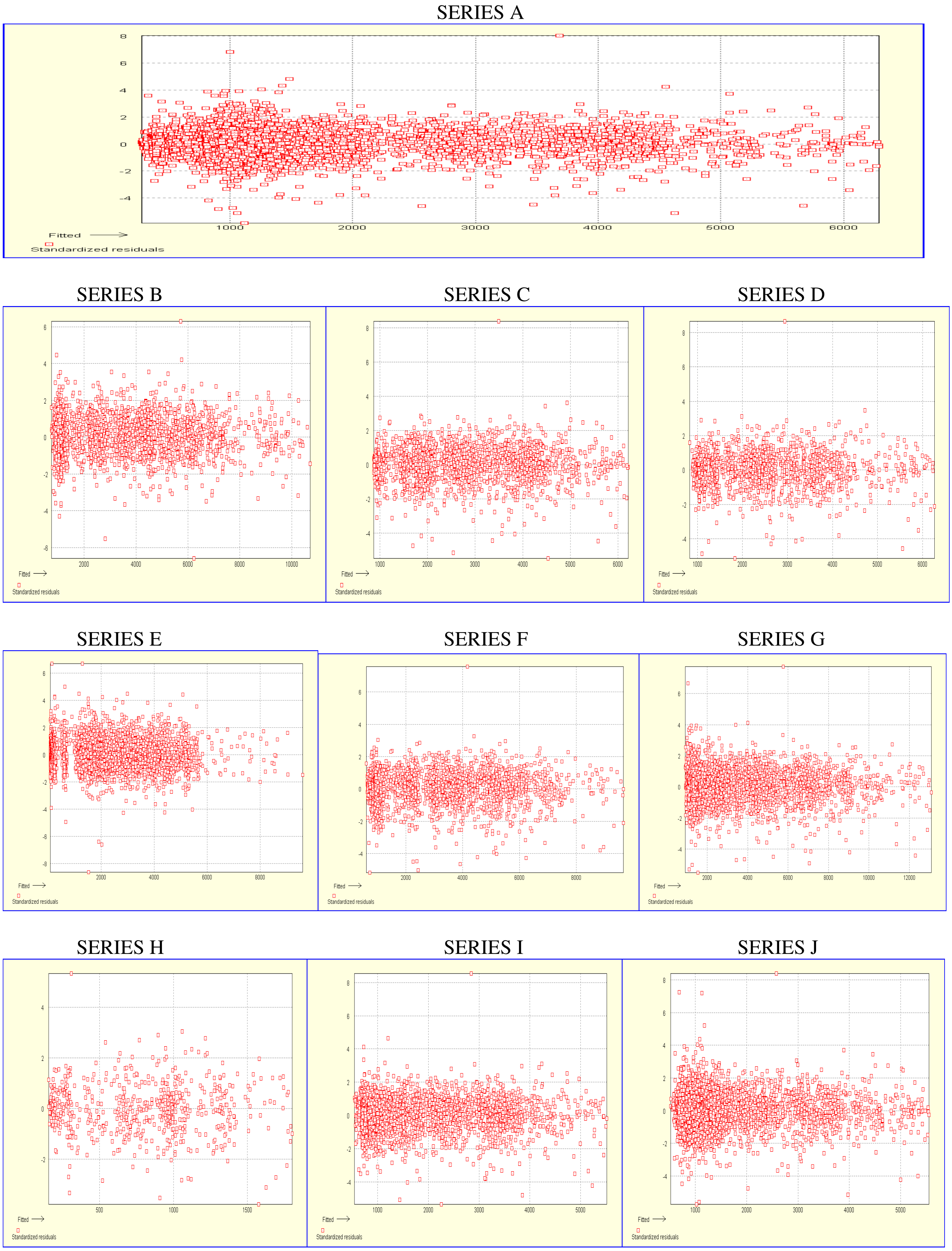


Figure12. Scatter of residuals versus fitted values

Palaniappan Bagavathi Sivakumar received his B.E. in Computer Science and Engineering from Madras University and M.S. from BITS, Pilani, India. He also holds Post Graduate Diplomas in Human Resource Management, Marketing Management, Operations Management and Financial Management. He has 18 years of Teaching and Research experience. He is currently with Coimbatore campus of Amrita Vishwa Vidyapeetham University. He has guided over 75 projects both at under graduate and graduate levels. He is serving as examiner for various universities and colleges. He has published technical papers and delivered talks at various international conferences and seminars. He has also visited University of Milan, Italy as part of a collaborative project. His name is included in the 2009 edition of Marquis who's who in the world. He has been an active member of the following technical societies IEEE, IEEE Computer Society, IEEE Computational Intelligence Society, Institution of Electronics and Telecommunications Engineers (IETE), Computer Society of India (CSI), Institution of Engineers, India (IE) and Indian Society for Technical Education (ISTE). His areas of interests include Time Series Analysis and Forecasting, Data Mining, Machine Learning, Pattern Recognition, Signal Analysis, Software Engineering and Artificial & Computational Intelligence.

V. P. Mohandas received his B.Sc. in Engineering from REC (Regional Engineering College) Calicut, M.Sc. in Engineering from College of Engineering Trivandrum and his Ph.D. from IIT Mumbai. He worked at the Small Industries Development Corporation, Khandelval, Ltd. and the Maharashtra State Electricity Board before joining N.S.S. College of Engineering in Palakkad in 1978, where he rose from the post of Lecturer in EEE to that of Principal. As Head of the Department at N.S.S., he built the department of Instrumentation and Control Engineering through many MHRD-funded projects. During his tenure as Principal, N.S.S. College of Engineering received NBA (AICTE) accreditation. In 2002, Dr. Mohandas joined the Department of Electronics and Communication Engineering of Amrita Vishwa Vidyapeetham University and now serves as Chair of that department in Coimbatore Campus. He has been an examiner for many doctoral dissertations and has held crucial positions in academic and administrative bodies of various universities. He is a Fellow of the Institution of Electronics and Telecommunications Engineers (IETE) and serves as Vice Chairman of the Coimbatore IETE Center. He has organized and chaired events of national and international standing and has many publications to his credit. His areas of interests include Dynamic System Theory, Signal Processing, Soft Computing and their application to socio-techno-economic and financial systems. He is guiding Ph.D. scholars in these areas.