Chapter 2
Long-Range Dependence and ARFIMA Models

Abstract In this chapter, long-range dependence concept, Hurst phenomenon and ARFIMA models are introduced and the earlier work on these subjects are reviewed. Several methodologies are introduced for the estimation of long-range dependence index (Hurst number or fractional difference parameter).

Keywords Long-range dependence • Long memory • ARFIMA models • Hurst phenomenon

2.1 Long-Range Dependence

Long-range dependence or long memory has drawn the attention of scientists since 1960s when the so called Hurst phenomenon (Hurst 1951) was discussed and explained by Mandelbrot and Van Ness (1965), and Mandelbrot and Wallis (1968, 1969). Hurst (1951) investigated the water levels of Nile River for optimum dam sizing. The Hurst number, named after Harold Edwin Hurst, is an index of long memory. Hurst number $H = 0$ represents processes that have independent increments while $0.5 < H < 1$ indicates long-range dependence.

The Hurst phenomenon has been utilized in literature extensively to asses variability of climatic and hydrologic quantities including wind power resources (Haslett and Raftery 1989), global mean temperatures (Bloomfield 1992), river flows (Eltahir 1996; Montanari et al. 1997; Vogel et al. 1998), porosity and hydraulic conductivity in subsurface hydrology (Molz and Boman 1993), indexes of North Atlantic Oscillation (Stephenson et al. 2000), tree-ring widths (Koutsoyiannis 2002), temperature anomalies in Northern Hemisphere (Koutsoyiannis 2003). In addition, long-range dependence is reported for sea levels by a power spectrum analysis (Hsui et al. 1993) and by wavelet analysis (Barbosa et al. 2006). According to Koutsoyiannis (2003), climate changes are closely related to the Hurst phenomenon, which is stochastically equivalent to a simple scaling behavior of climate variability through time.

A. Ercan et al., *Long-Range Dependence and Sea Level Forecasting*, SpringerBriefs in Statistics, DOI: 10.1007/978-3-319-01505-7_2,

A process has long memory if its autocovariances are not absolutely summable (Palma 2007). The slowly decaying autocorrelations and unbounded spectral density near zero frequency are characteristics of the long memory signals (Beran 1994). A stationary process X_t has long memory (Beran and Terin 1996) if, as $|k| \to \infty$.

$$r(k) \sim L_1(k)|k|^{2H-2}, H \in (0.5,1) \tag{2.1}$$

where $r(k) = \mathrm{cov}(X_t, X_{t+k})$ and $L_1(k)$ is a slowly varying function as $|k| \to \infty$. In other words, $L_1(ta) / L_1(t) \to 1$ as $t \to \infty$ for any $a > 0$. This implies that the correlations are not summable and the spectral density near zero frequency is unbounded. Fractional Gaussian noise (Mandelbrot and Van Ness 1968 ; Mandelbrot and Wallis 1969 ; Mandelbrot 1971, Koutsoyiannis 2002) and ARFIMA (Granger and Joyeux 1980; Hosking 1981; and Geweke and Porter-Hudak 1983) models are among the best known long memory models.

Several methodologies are available for the estimation of long-range dependence index (taken as Hurst number or fractional difference parameter which will be discussed in ARFIMA models section) such as the Rescaled Range Method (Hurst 1951; Mandelbrot and Taqqu 1979; Lo 1991), Aggregated Variance Method, Differencing the Variance Method, Absolute Moments Method, Detrended Fluctuation Analysis (Peng et al. 1994), Regression Method based on the periodogram (Geweke and Porter-Hudak 1983) and Whittle Estimator (Whittle 1951; Fox and Taqqu 1986; Dahlhaus 1989). Taqqu et al. (1995) analyzed the performance of nine different estimators. Estimation methods for long-memory models are reviewed in detail in Beran (1994), Palma (2007), and Box et al. (2008).

Minimum water levels of Nile river, as reported in Beran (1994), are depicted in Fig. 3.1 Hurst (1951) estimated the Hurst Number as 0.93 for Nile river's historical water levels. Sample ACF, periodogram, and the logarithm of the periodogram of minimum water levels of Nile River are depicted in Fig. 3.2 The calculated Hurst number and the Fig. 3.2 clearly show the long memory behavior of the historical Nile river levels.

The ARFIMA models are generalization of the linear stationary ARMA and linear nonstationary ARIMA models. The autoregressive (AR) model of first order emerged as the dominant model for background climate variability for over three decades; however, some aspects of the climate variability are best described by the long memory models (Vyushin and Kushner 2009).

2.2 ARFIMA Models

According to Beran (1994), stochastic processes may be utilized to model the behavior of observed time series solely by the statistical approach without a physical interpretation of the process parameters. Long-range memory models, ARFIMA processes in particular, have been used extensively in different fields such as astronomy, economics, geosciences, hydrology and mathematics

(Beran 1994). Early applications of the ARFIMA model were performed by Granger and Joyeux (1980), Hosking (1981) and Geweke and Porter-Hudak (1983). The process (X_t) is said to be an ARFIMA(p,d,q) process if it is a solution to the following difference equation:

$$\phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t \tag{2.2}$$

where $\phi(z) = 1 - \sum_{j=1}^{p} \phi_j z^j$ and $\theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k$ are the autoregressive and moving-average operators, and d is the fractional difference parameter.$\varepsilon_t$is the white noise process with $E(\varepsilon_t) = 0$ and variance σ_ε^2. B is the backward shift operator such that $BX_t = X_{t-1}$. For any real number $d > -1$, the difference operator $(1-B)^d$ can be defined by means of a binomial expansion (Brockwell and Davis 1987).

$$(1-B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} B^j \tag{2.3}$$

where $\Gamma(\cdot)$is the gamma function,

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \text{ for } z > 0 \tag{2.4}$$

$\Gamma(z) = \infty$ for $z = 0$, and $\Gamma(z) = z^{-1}\Gamma(1+z)$ for $z < 0$. The process is stationary and invertible when $\phi(z)$ and $\theta(z)$ have all their roots outside the unit circle, have no common roots, and $-0.5 < d < 0.5$ (Crato and Ray 1996). The process has long memory when $0 < d < 0.5$ and is nonstationary for $d \geq 0.5$ (Beran 1994). When $d = 0$, the model is referred to as autoregressive moving average (ARMA) model of order (p,q), and is capable of modeling linear stationary processes. On the other hand, when d is a positive integer, the model is referred to as autoregressive integrated moving average (ARIMA) process, and is capable of modeling linear nonstationary processes (Box and Jenkins 1976).

Box and Jenkins (1976) provide the details on the forecasting, confidence band estimation and forecast updating for ARIMA models. In this study, as described in detail in the next chapter, the governing equations, provided in Box and Jenkins (1976), are utilized and extended for ARFIMA models that use the non-integer differencing parameter d.

References

Barbosa SM, Fernandes MJ, Silva ME (2006) Long-range dependence in North Atlantic sea level. Phys A 371(2):725–731

Beran J (1994) Statistics for long-memory processes. Chapman and Hall, New York

Beran J, Terrin N (1996) Testing for a change of the long-memory parameter. Biometrika 83(3):627–638

Bloomfield P (1992) Trends in global temperature. Clim Change 21(1):1–16

Box, G.E.P, and Jenkins, G.M. (1976). Time series analysis: forecasting and control. Holden-Day, San Fransisco

Box GEP, Jenkins GM, Reinsel GC (2008) Time series analysis: forecasting and control. Wiley, Hoboken

Brockwell PJ, Davis RA (1987) Time series: theory and methods. Springer, New York

Crato N, Ray BK (1996) Model selection and forecasting for long-range dependent processes. J Forecast 15:107–125

Dahlhaus R (1989). Efficient parameter estimation for self-similar processes. Ann Stat 17, 1749–1766

Eltahir EAB (1996) El Nino and the natural variability in the flow of the Nile River. Water Resour Res 32(1):131–137

Fox R, Taqqu MS (1986) Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. The Ann Stat 14(2):517–532

Geweke J, Porter-Hudak S (1983) The estimation and application of long memory time series models. J Time Ser Anal 4(4):221–238

Granger CWJ, Joyeux R (1980) An introduction to long memory time series models and fractional differencing. J Time Ser Anal 1:15–29

Haslett J, Rafterv AE (1989) Space-time modelling with long-memory dependence: assessing Ireland’s wind power resource. Appl Statist 38(1):1–50

Hosking JRM (1981) Fractional differencing. Biometrika 68:165–176

Hsui, AT., Rust, K A., Klein, G D. (1993). A fractal analysis of Quaternary, Cenozoic- Mesozoic, and Late Pennsylvanian sea level changes. J Geophys Res 98 (B12), 21963–21967

Hurst HE (1951) Long-term storage capacity of reservoirs. Trans Am Soc Civ Eng 116:77–779

Koutsoyiannis D (2002) The hurst phenomenon and fractional Gaussian noise made easy. Hydrol Sci J 47(4):573–595

Koutsoyiannis D (2003) Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol Sci J 48(1):3–27

Lo A (1991) Long-term memory in stock market prices. Econometrica 59:1279–1313

Mandelbrot BB (1971) A fast fractional Gaussian noise generator. Water Resour Res 7(3):543–553

Mandelbrot BB, Van Ness JW (1968) Fractional Brownian motions, fractional noises and application. Soc Ind Appl Math Rev 10:422–437

Mandelbrot BB, Wallis JR (1968) Noah, Joseph and operational hydrology , Water Resour. Res. 4:909–920

Mandelbrot BB, Wallis JR (1969) Computer experiments with fractional Gaussian noises. Water Resour Res 5:228–267

Mandelbrot BB, Taqqu MS (1979) Robust R/S analysis of long run serial correlation. 42nd Session of the International Statistical Institute. Manila, Book 2:69–99

Molz FJ, Boman GK (1993) A fractal-based stochastic interpolation scheme in subsurface hydrology. Water Resour Res 29(11):3769–3774

Montanari A, Rosso R, Taqqu MS (1997) Fractionally differenced ARIMA models applied to hvdrologic time series. Water Resour Res 33(5):1035–1044

Palma W (2007) Long-memory time series: theory and methods. Wiley, Hoboken

Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL (1994) Mosaic organization of DNA nucleotides. Phys Rev E 49:1685–1689

Stephenson, D B.,Pavan, V., Bojariu, R (2000). Is the North Atlantic Oscillation a random walk? Int J Clim 20(1), 1-18

Taqqu, M.S., Teverovsky, V., Willinger, W. (1995). Estimators for long-range dependence: An empirical study. Fractals 3(4), 785-798

Vogel RM, Tsai Y, Limbrunner JF (1998) The regional persistence and variability of annual streamflow in the United States. Water Resour Res 34(12):3445–3459

Vyushin DI, Kushner PJ (2009) Power-law and long-memory characteristics of the atmospheric general circulation. J Clim 22(11):2890–2904

Whittle P (1951) Hypothesis testing in time series analysis. Hafner, New York