9 Chapter 9: Introduction to ARFIMA models

9.1 Long range dependence

The phenomenon of long memory had been known well before suitable stochastic models were developed. Scientists in diverse fields of statistical applications observed empirically that correlations between observations that are far apart (in time or space) may decay to zero at a slower rate than one would expect from independent data or data following classic ARMA or Markovian models. In 1951, Hurst showed that wet and drought periods from Nile River flow data exhibit quite substantial persistence, which was later called the Hurst phenomenon.

Then later, Mandelbrot and van Ness (1968) and Mandelbrot and Wallis (1969) constructed a weakly stationary stochastic process with a hyperbolically decaying autocorrelation function in continuous time, namely fractional Brownian motion (fBm)

$$B_d(t) = \frac{\sqrt{\Gamma(2d+2)\cos(\pi d)}}{\Gamma(d+1)} \int_{-\infty}^{\infty} [(t-s)^d I_{[t-s>0]} - (-s)^d I_{[-s>0]}] dB(s),$$

where I is an indicator function and $B(t)$ is a Brownian motion ($B(0) = 0$ a.s.; $B(t)$ has independent increments; $EB(t) = EB(s)$, $\mathrm{Var(B(t) - B(s)) = \sigma^2|t - s|}$); and showed that that persistence discovered by Hurst is compatible with weak stationarity; also it turns that the parameter d in fBm is related to the Hurst exponent H as $H = d + 0.5$. Since then, long memory (or long-range dependence) has become a rapidly developing subject. Because of the diversity of applications, the literature on the topic is broadly scattered in a large number of journals, including those in fields such as agronomy, astronomy, computer science, chemistry, economics, engineering, environmental sciences, finance, geosciences, hydrology, mathematics, physics and statistics. Finally, nowadays there exist two main approaches in modeling long memory processes: 1) **fractional Gaussian noise (gGn)** in continuous time, and **autoregressive fractionally integrated moving average (ARFIMA) models**. (ARFIMA was introduced independently by Granger and Joyeux (1980) and Hosking (1981)).

Let $\{X_t\}$ be a weakly stationary process with autocorrelation function $\rho(\cdot)$ and spectral density $f(\cdot)$. There exist two equivalent definitions of long memory in time and frequency domains.

Definition (time domain): $\{X_t\}$ is called a stationary process with long memory property if there exist a real number $H \in (0, 1)$ and a slowly varying function $c_k > 0$ such that

$$\lim_{k\to\infty} \frac{\rho(k)}{c_k k^{2(H-1)}} = \lim_{k\to\infty} \frac{\rho(k)}{c_k k^{2d-1}} = 1,$$

where H is called the Hurst parameter and $d = H - 0.5$ is called the long memory parameter or fractional differencing parameter in ARFIMA(p, d, q) processes.

Definition (Frequency domain): $\{X_t\}$ is called a stationary process with long memory property if there exits a slowly varying function $c_f > 0$ such that

$$\lim_{\nu\to 0} \frac{f(\nu)}{c_f|\nu|^{1-2H}} = \lim_{\nu\to 0} \frac{f(\nu)}{c_f|\nu|^{-2d}} = 1. \tag{9.1}$$

9.2 Definition of ARFIMA, causality and invertibility

Define an ARFIMA (p,d,q) process, where p, q are integers and d is real, to be a stochastic process $\{X_t\}$ represented as

$$\phi(B)(1 - B)^d X_t = \theta(B)\epsilon_t, \qquad \epsilon_t \sim WN(0, \sigma^2) \tag{9.2}$$

where $\phi(Z)$ and $\theta(Z)$ are polynomials of back-shift operator B of degree p and q respectively; $\phi(Z)$ and $\theta(Z)$ have no common factors.

The fractional differencing operator $(1 - B)^d$ can be expanded as

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = \sum_{k=0}^{\infty} b_k B^k \tag{9.3}$$

where $b_k = \frac{\Gamma(-d+k)}{\Gamma(-d)\Gamma(K+1)}$, if $d < \frac{1}{2}$, then $\sum\limits_{k=0}^{\infty} \binom{d}{k}^2 < \infty$. Thus, X_t is weakly stationary.

Theorem: Let X_t be an ARFIMA (p,d,q) process defined in (9.2) and $d \in (-1, \frac{1}{2})$. Then

(1) If all the roots of $\phi(\cdot)$ lie outside the unit circle, i.e. no roots along the unit circle,

then there exists a unique stationary solution of (9.2) given by

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j},$$

where $\psi(Z) = \frac{\theta(Z)}{(1-Z)^d \phi(Z)}$.

(2) if all the roots of $\phi(\cdot)$ lie outside the **closed** unit circle, then X_t is causal.

(3) if all the roots of $\theta(\cdot)$ lie outside the **closed** unit circle, then X_t is invertible.

The most significant feature of ARFIMA processes are their long range dependence. The reason for choosing d as a fractional value is that the effect of d on distant observations decays hyperbolically as the lag increases, while the effects of ϕ and θ decay exponentially. The long term behavior of an ARFIMA(p, d, q) process should be similar to that of an ARFIMA($0, d, 0$) process with the same value of d. Because the effects of ϕ and θ decay faster than that of d, for very distant observations, the effects of ϕ and θ are negligible. In particular,

- If $0 < d < \frac{1}{2}$, the process exhibits strong positive dependence (persistence) between distant observations.
- If $-1 < d < 0$, the process exhibits negative dependence (antipersistence) between distant observations.
- We say that the a weakly stationary process has intermediate memory if for a large lag k its acf $\rho(k) \sim c(k)|k|^{2d-1}$ with $d < 0$, where $c(k)$ is a slowly varying function. Hence, the acf decays to zero at a hyperbolic rate but it is absolutely summable, i.e.

$$\sum |\rho(k)| < \infty.$$

9.3 Estimation

Let $\{X_t\}$ be a zero-mean weakly stationary Gaussian process. Then, the log-likelihood function of this process is given by

$$L(\psi) = -\frac{1}{2}\ln \det\Gamma_\psi - \frac{1}{2}\mathrm{X}'\Gamma_\psi^{-1}\mathrm{X}, \tag{9.4}$$

where $X = (X_1, \ldots, X_n)'$, $\Gamma_\psi = \mathrm{Var}(\mathrm{X})$ and ψ is the parameter vector. Consequently, the **maximum-likelihood** (ML) estimate $\hat{\psi}$ is obtained by maximizing $L(\psi)$. The log-likelihood function (9.4) requires the calculation of the determinant and the inverse of the variance-covariance matrix Γ_ψ. Those calculation can be conducted using the Cholesky decomposition method, i.e. $\Gamma_\psi = U'U$ where U is an upper triangular matrix, and $\det\Gamma_\psi = \prod\prod_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{u}_{\mathrm{jj}}^2$. The Cholesky decomposition can be inefficient for long sample sizes, which is typically the case with long range dependence, so we can alternatively use the Durbin-Levinson algorithm which exploits the symmetric Toeplitz structure of Γ_ψ.

To simplify calculation of autocovariances in Γ_ψ, we can employ the **splitting** method. In particular, we can decompose ARFIMA(p, d, q) into its ARMA(p, q) and its fractionally integrated (FI($0, d, 0$)) parts. Let $\rho_1(k)$ be the acf of the ARMA (p, q) component and $\rho_2(k)$ be the acf of the FI component. Then, the acf of X_t is given by the convolution of these two functions

$$\rho(k) = \sum_{-\infty}^{\infty} \rho_1(j)\rho_2(j-k),$$

which can be truncated up to m terms.

Alternatively, we can employ the AR and MA approximations, which include state space techniques, the Haslett-Raftery and Beran approaches. In particular, ARFIMA (p, d, q) can be converted to an AR(∞) (or MA(∞)) model as follows. Taking into account (9.3), we can re-write ARFIMA as

$$\phi(B) \cdot \sum_{k=0}^{\infty} b_k B^k x_t = \theta(B)\epsilon_t,$$

where $b_0 = 1$, $b_1 = -d$, $b_2 = \frac{1}{2}d(1-d)$, $b_k = \frac{1}{k}b_{k-1}(k-1-d)$, $\quad k \geq 3$, $\phi(B) =$

$\sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = \sum_{j=1}^{q} \theta_j B^j$. Hence, the model (9.2) takes the form

$$\left(1 - \sum_{i=1}^{\infty} \delta_i L^i\right) X_t \;=\; \epsilon_t,$$

where $\sum \delta_i^2 < \infty$, for $d < \frac{1}{2}$.

In practice, we fit truncated AR(P) (or MA(Q)) models, with very large model orders P or Q, to approximate ARFIMA (p, d, q) process.

9.4 Tests for Long Range Dependence

9.4.1 Periodogram method

The periodogram method is based on the equation 9.1. In particular, the power spectral density of a long memory process obeys a power law near the origin, i.e. $f(\nu) \sim c_f|\nu|^{-2d}$, as $\nu \to 0$. Thus, by taking logarithm on both sides, we get $\log f(\nu) \sim -2d\log(|\nu|)$, as $\nu \to 0$.

Since the spectral density $f(\nu)$ is the Fourier transform of the autocorrelation function, an estimate of the spectral density can be obtained by taking the inverse Fourier transform of the estimate of the autocorrelation function. This estimator is referred to as a periodogram $I(\nu)$.

Therefore, the long memory parameter d can be estimated from the least squares regression

$$\log(I(\nu_j)) = c - 2d\log(\nu_j) + \eta_j,\; j = 1, 2, \ldots, n$$

where $\nu_j = 2\pi j/T$, $j = 1, \ldots, T-1$, $n = g(T) \ll T$, and T is the sample size. $I(\nu_j)$ is the periodogram of the series at frequency ν_j defined by

$$I(\nu) = \frac{1}{2\pi T}|\sum_{t=1}^{T}(X_k - \bar{X})e^{ik\nu}|^2.$$

The periodogram plot is the graph of $\{\log(\nu_j), \log I(\nu_j)\}$, $j = 1, 2, \ldots, n$. The typical threshold value utilized in detection of d is $n = T^{0.5}$. Theoretically the log-log plot should provide a straight line with a slope of $-2d$.

9.4.2 Aggregated-variance methods

A characteristic trait of long-memory processes is that the variance of an N-member sample mean decreases more slowly than N^{-1} (Beran (1989)). In the same paper, Beran showed that given N data points X_i, $i = 1, \ldots, N$

$$Var(\frac{1}{N}\sum_{i=1}^{N} X_i) = N^{2d-1}, \text{ as } N \to \infty.$$

This suggests the following method for estimating d. Divide the series into $k = N/m$ blocks of size m and compute the mean for each block

$$x_k(m) = \frac{1}{m} \sum_{(k-1)m+1}^{km} X_i, \text{ where } k = 1, \ldots, N/m.$$

Variance of the block means

$$s^2(m) = \frac{1}{N/(m-1)} \sum_{k=1}^{N/m} (x_k(m) - \bar{x})^2, \text{ where } \bar{x} \text{ is the overall mean.}$$

Now a log-log plot of $s^2(m)$ against m should yield a straight line with a slope of $2d - 1$. This is known as aggregated variance method.

A drawback of this method is that inhomogeneity in the data can produce a positive value of d even in the absence of long memory. A modification of the above method is called the differenced variance method, which avoids this problem.

9.4.3 Differenced-variance method

The main idea of the differenced variance method is to study the first-order difference of the above variances

$$\nabla s^2(m) = s^2(m+1) - s^2(m).$$

Teverovsky and Taqqu (1997) show that a log-log plot of this quantity against m will again asymptotically produce a straight line with slop $2d - 1$ and the value of d is not affected by the inhomogeneity of the data.

9.4.4 Simulated Example

```
library(fArma)
y=farimaSim(n = 1000, model = list(ar =0.6, d = 0.3, ma =0.2),r
ts.plot(y)
acf(y, lag.max=100)
pacf(y, lag.max=100)
```

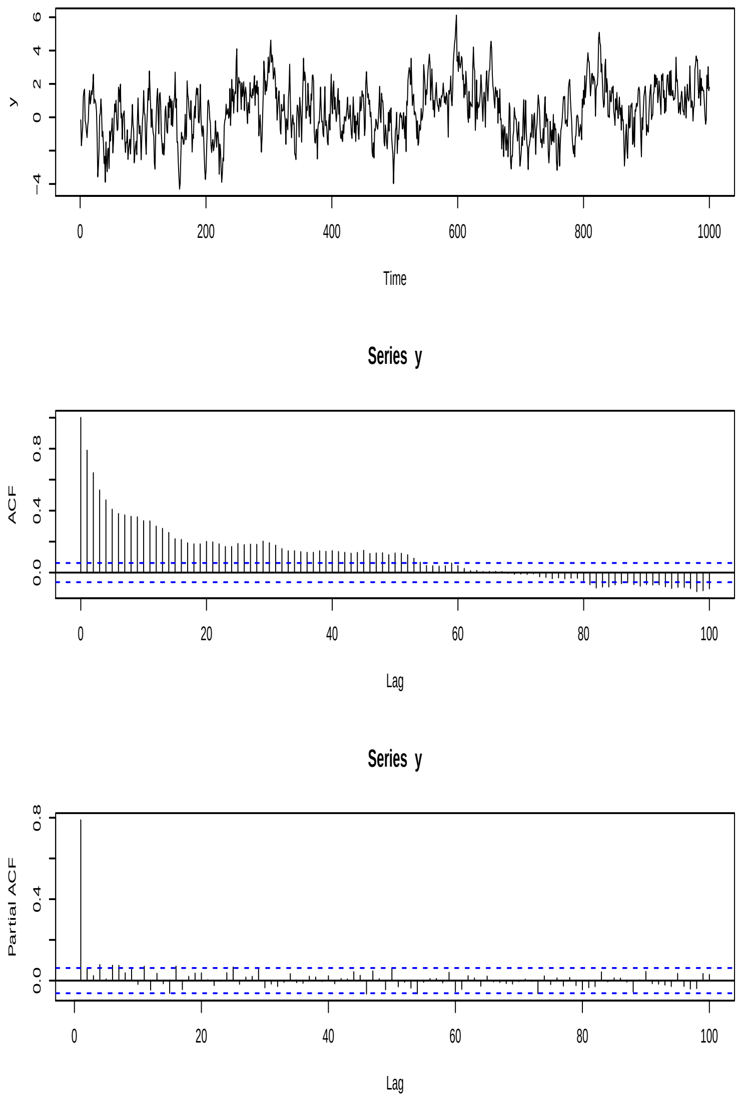

Figure 9.1: Time series, acf and pacf plots of ARFIMA(0.6,0.3, 0.2), 1000 observations.

Now let us apply various tests for long range dependence. We start from periodogram test:

```
> perFit(y, doplot=TRUE)

Title:
 Hurst Exponent from Periodgram Method

Call:
 perFit(x = y, doplot = TRUE)

Method:
 Periodogram Method

Hurst Exponent:
           H       beta
   0.9362142 -0.8724284

Hurst Exponent Diagnostic:
    Estimate    Std.Err  t-value     Pr(>|t|)
X 0.9362142 0.07007682 13.35983 1.030412e-23

Parameter Settings:
      n cut.off
   1000      10
```

Now we apply the aggregated variance test:

```
> aggvarFit(y, doplot=TRUE)

Title:
 Hurst Exponent from Aggregated Variances
```

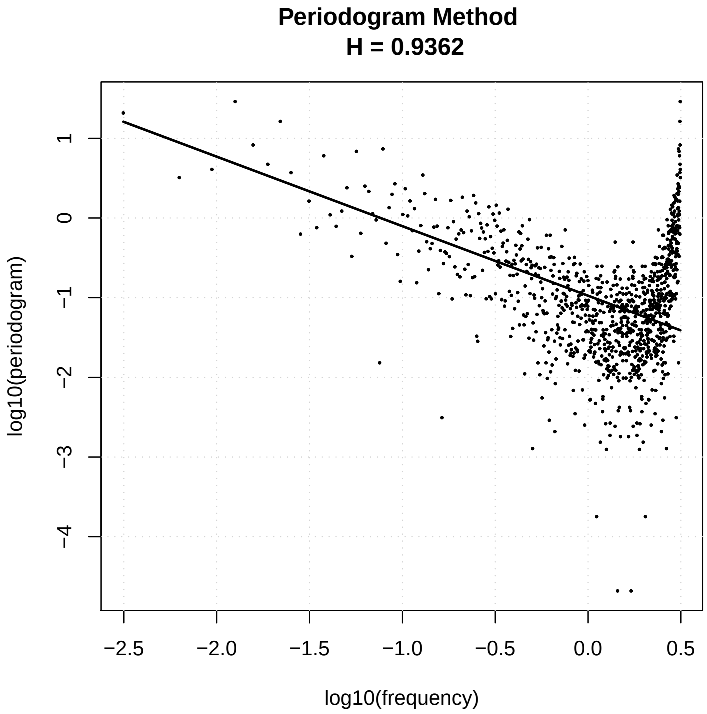

Figure 9.2: Periodogram fit of ARFIMA(1,0.3, 0.2)

```
Call:
 aggvarFit(x = y, doplot = TRUE)

Method:
 Aggregated Variance Method

Hurst Exponent:
          H       beta
   0.7033396 -0.5933208

Hurst Exponent Diagnostic:
    Estimate    Std.Err  t-value    Pr(>|t|)
X 0.7033396 0.03521754 19.97129 6.73822e-25
```

```
Parameter Settings:
       n   levels  minnpts cut.off1 cut.off2
    1000       50        3        5      316
```

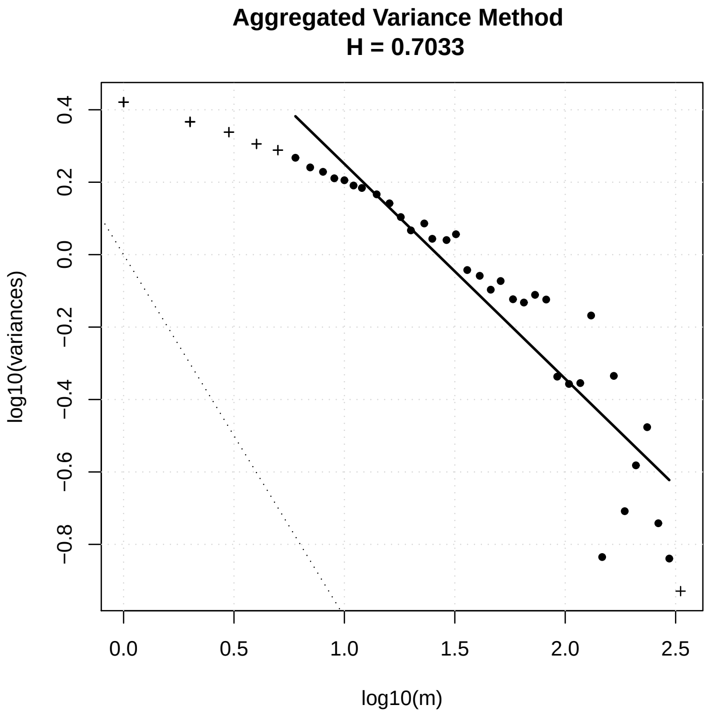

Figure 9.3: Aggregated variance fit of ARFIMA(1,0.3, 0.2)

Finally, we use the test of differenced aggregated variances:

```
> diffvarFit(y, doplot=TRUE)

Title:
 Hurst Exponent from Differenced Aggregated Variances

Call:
 diffvarFit(x = y, doplot = TRUE)

Method:
```

```
 Differenced Aggregated Variance

Hurst Exponent:
         H      beta
  1.0613639 0.1227278

Hurst Exponent Diagnostic:
   Estimate  Std.Err  t-value     Pr(>|t|)
X 1.061364 0.090066 11.78429 2.282599e-12

Parameter Settings:
      n   levels  minnpts cut.off1 cut.off2
    1000       50        3        5      316
```

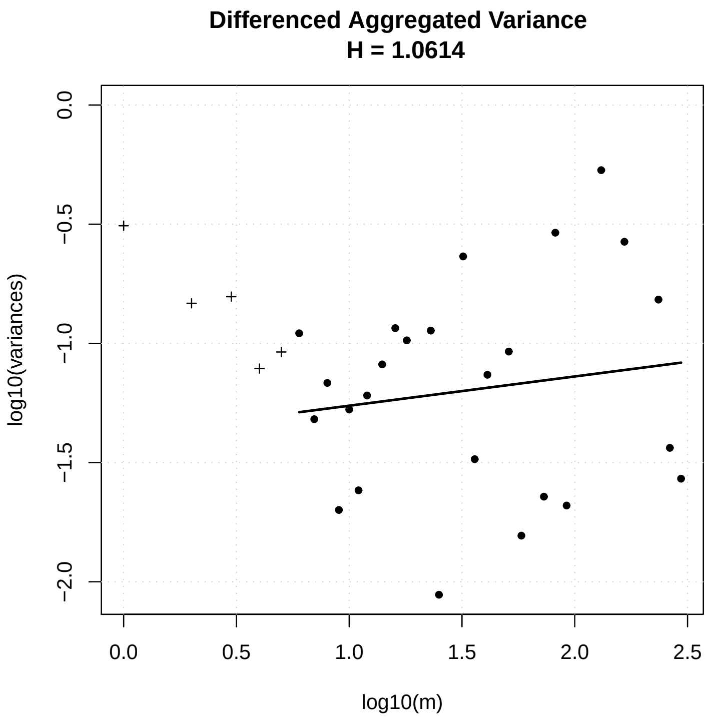

Figure 9.4: Differenced aggregated variance fit of ARFIMA(1,0.3, 0.2)

9.4.5 Temperature in Boston from January 1, 1970 to June 16, 1986

Here is the summary of periodogram fit:

```
> perFit(Model1_Residuals[1:6000], doplot=TRUE)

Title:
 Hurst Exponent from Periodgram Method

Call:
 perFit(x = Model1_Residuals[1:6000], doplot = TRUE)

Method:
 Periodogram Method

Hurst Exponent:
          H       beta
  0.6558462 -0.3116924

Hurst Exponent Diagnostic:
    Estimate    Std.Err  t-value     Pr(>|t|)
X 0.6558462 0.02890322 22.69112 1.146503e-82

Parameter Settings:
      n cut.off
   6000      10
```

the aggregated variance fit:

```
> aggvarFit(Model1_Residuals[1:6000], doplot=TRUE)
```

```
Title:
 Hurst Exponent from Aggregated Variances

Call:
 aggvarFit(x = Model1_Residuals[1:6000], doplot = TRUE)

Method:
 Aggregated Variance Method

Hurst Exponent:
          H       beta
   0.6036901 -0.7926198

Hurst Exponent Diagnostic:
    Estimate   Std.Err  t-value     Pr(>|t|)
X 0.6036901 0.0404211 14.93502 1.137483e-19

Parameter Settings:
       n   levels  minnpts cut.off1 cut.off2
    6000       50        3        5      316
```

and differenced aggregated variance fit:

```
> diffvarFit(Model1_Residuals[1:6000], doplot=TRUE)

Title:
 Hurst Exponent from Differenced Aggregated Variances

Call:
 diffvarFit(x = Model1_Residuals[1:6000], doplot = TRUE)
```

```
Method:
 Differenced Aggregated Variance

Hurst Exponent:
          H       beta
  0.6511117 -0.6977767

Hurst Exponent Diagnostic:
    Estimate    Std.Err  t-value     Pr(>|t|)
X 0.6511117 0.07828456 8.317242 5.399707e-10

Parameter Settings:
       n   levels  minnpts cut.off1 cut.off2
    6000       50        3        5      316
```

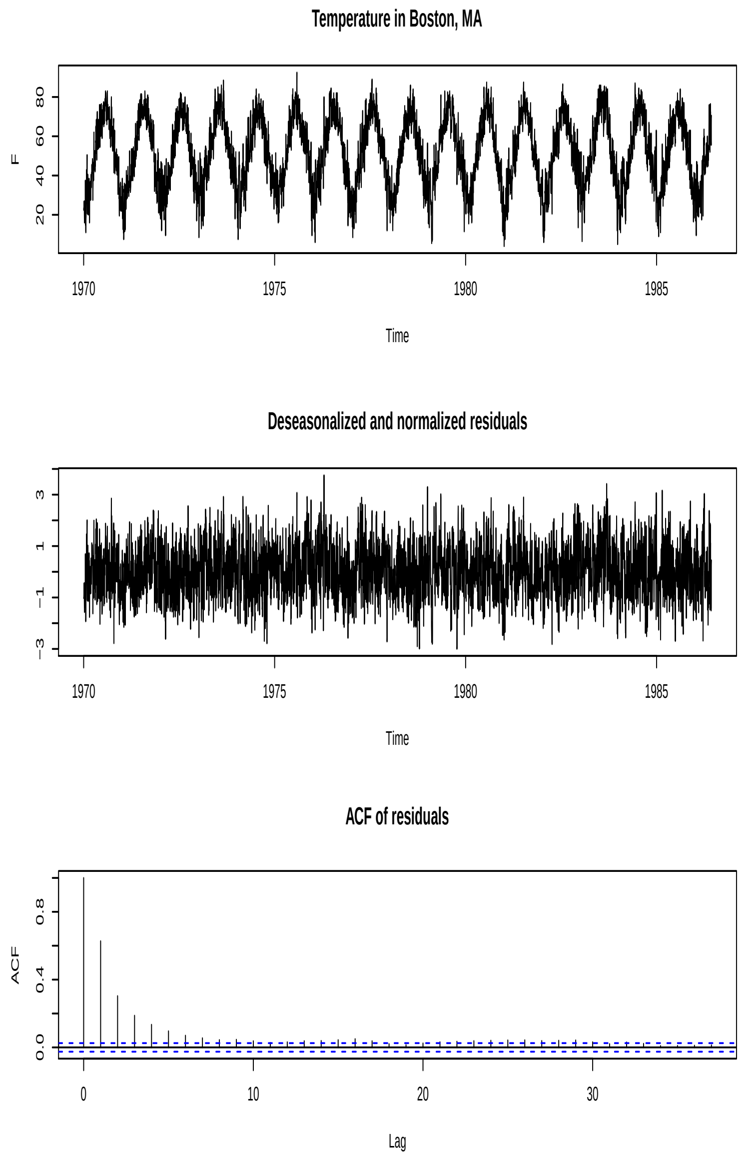

Figure 9.5: Time series, acf and pacf plots of temperature observations in Boston.

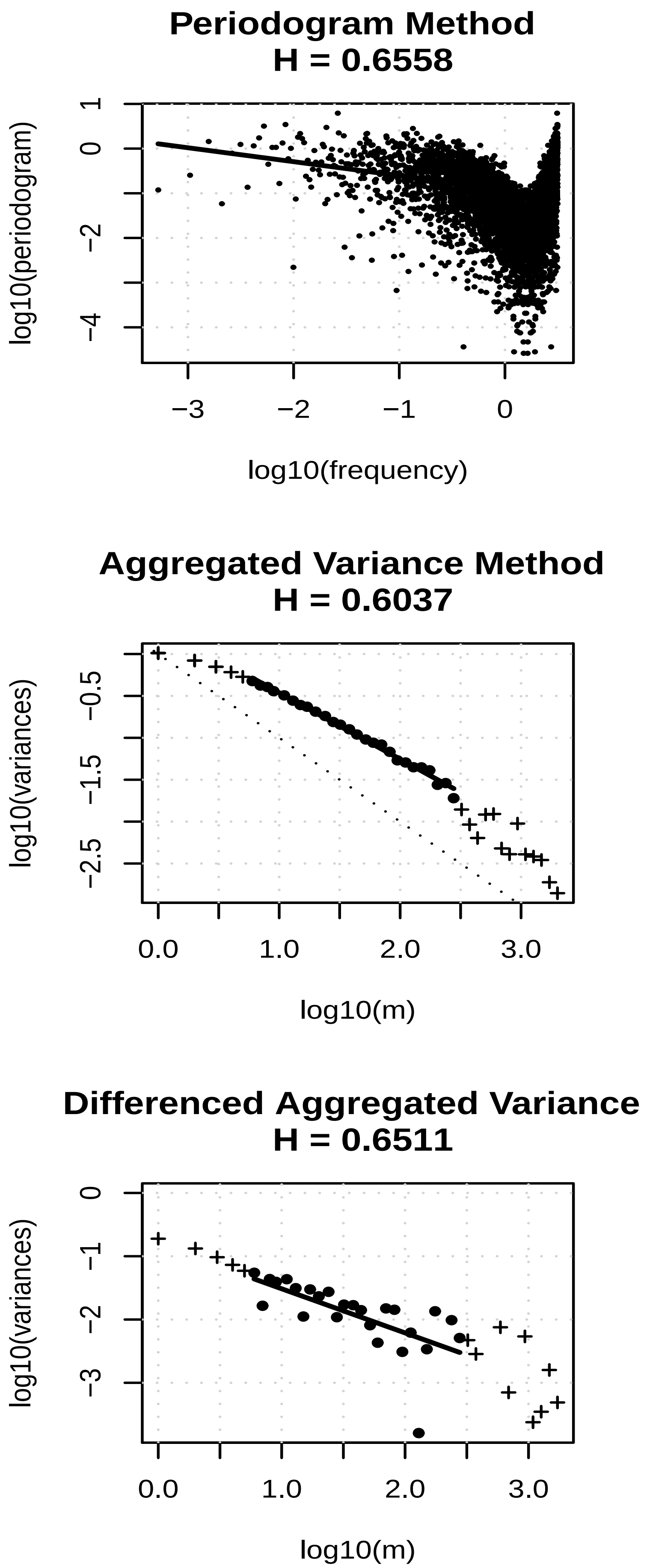

Figure 9.6: Periodogram, aggregated and differenced aggregated variance tests of deseasonlized and normalized temperature observations in Boston.