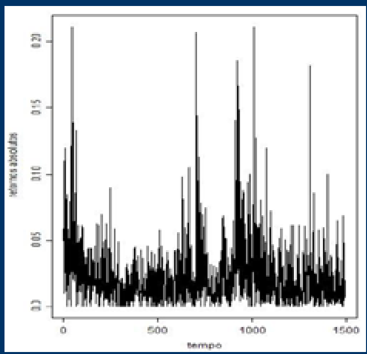
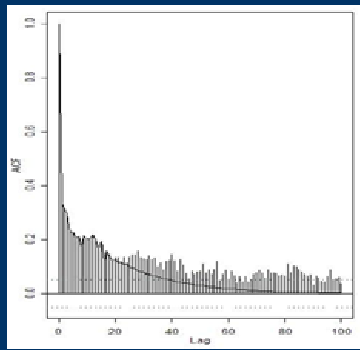
Modelos de memória longa





- ✓ Média próxima de zero (0.001)
- ✓ Retornos absolutos: medida de volatilidade
- ✓ Forte persistência até lags altos
- ✓ Modelo ARMA tradicional: número excessivo de parâmetros

$\rho_k \to C k^{-\alpha}$ quando $k \to \infty$

constante $C > 0$

$\alpha \in (0, 1)$

$$\sum_{k=-\infty}^{\infty} \rho_k = \infty$$

Coeficiente de Hurst → $H = 1 - \alpha/2 \in (0.5, 1)$ → Quanto maior H, decaimento mais lento

Processo integrado fracionário → $(1-L)^d (y_t - \mu) = u_t$ $\quad d = H - 1/2$ $\quad 0 < d < 1/2$

5.2 Long Memory ARMA and Fractional Differencing

The conventional ARMA(p, q) process is often referred to as a short-memory process because the coefficients in the representation

$$x_t = \sum_{j=0}^{\infty} \psi_j w_{t-j},$$

obtained by solving

$$\phi(z)\psi(z) = \theta(z),$$

are dominated by exponential decay. As pointed out in §3.3, this result implies the ACF of the short memory process $\rho(h) \to 0$ exponentially fast as $h \to \infty$.

Example 3.32 Fitting the Glacial Varve Series

Consider the series of glacial varve thicknesses from Massachusetts for $n = 634$ years, as analyzed in Example 2.6 and in Problem 2.8, where it was argued that a first-order moving average model might fit the logarithmically transformed and differenced varve series, say,

$$\nabla \log(x_t) = \log(x_t) - \log(x_{t-1}) = \log\left(\frac{x_t}{x_{t-1}}\right),$$

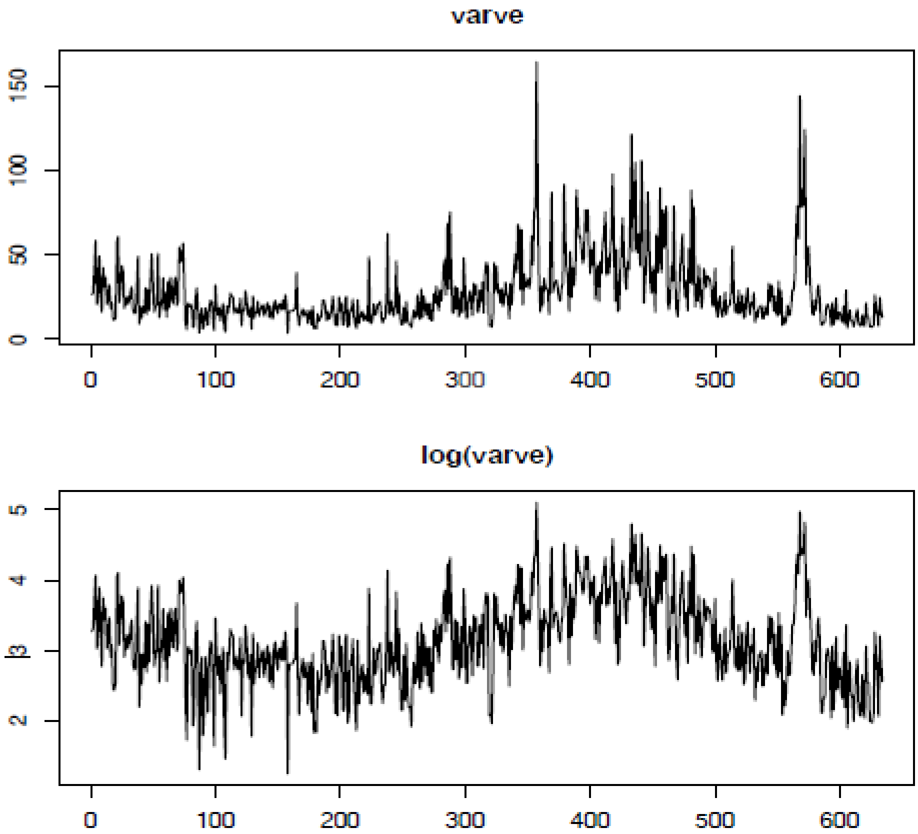


Fig. 2.6. Glacial varve thicknesses (top) from Massachusetts for $n = 634$ years compared with log transformed thicknesses (bottom).

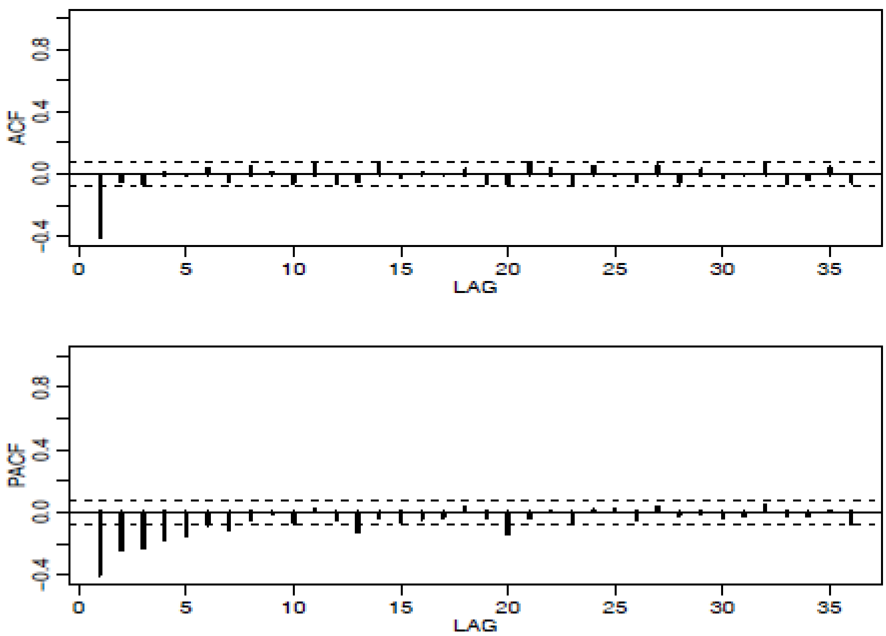


Fig. 3.7. ACF and PACF of transformed glacial varves.

leads to fitting an ARIMA(1, 1, 1) model,

$$\nabla x_t = \phi \nabla x_{t-1} + w_t + \theta w_{t-1},$$

where we understand x_t is the log-transformed varve series. In particular, the estimates of the parameters (and the standard errors) were $\widehat{\phi} = .23(.05)$, $\widehat{\theta} = -.89(.03)$, and $\widehat{\sigma}^2_w = .23$. The use of the first difference $\nabla x_t = (1 - B)x_t$ can be too severe a modification in the sense that the nonstationary model might represent an overdifferencing of the original process.

$$(1 - B)^d x_t = w_t, \tag{5.1}$$

where w_t still denotes white noise with variance σ^2_w. The fractionally differenced series (5.1), for $|d| < .5$, is often called *fractional noise* (except when d is zero). Now, d becomes a parameter to be estimated along with σ^2_w. Differencing the original process, as in the Box–Jenkins approach, may be thought of as simply assigning a value of $d = 1$. This idea has been extended to the class of fractionally integrated ARMA, or ARFIMA models, where $-.5 < d < .5$; when d is negative, the term antipersistent is used.

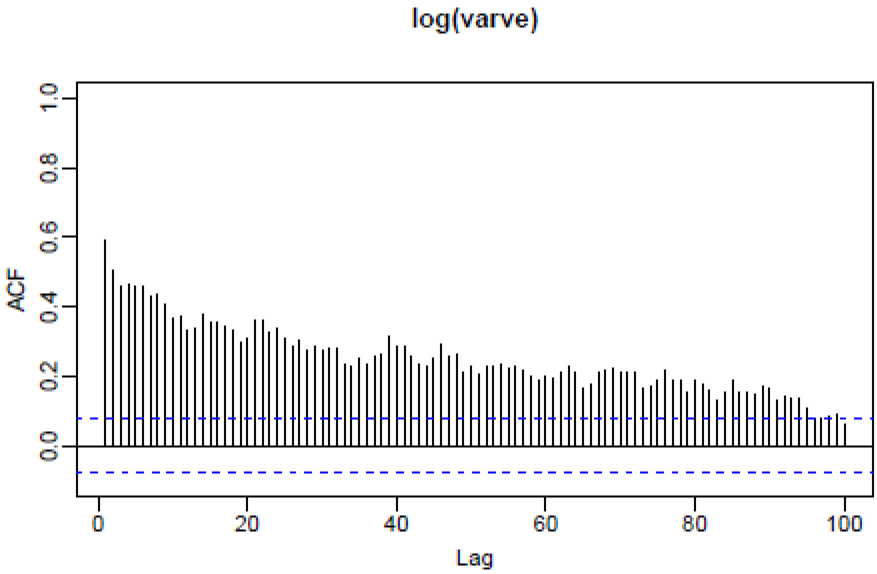


Fig. 5.1. Sample ACF of the log transformed varve series.

To investigate its properties, we can use the binomial expansion ($d > -1$) to write

$$w_t = (1-B)^d x_t = \sum_{j=0}^{\infty} \pi_j B^j x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j} \tag{5.2}$$

where

$$\pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \tag{5.3}$$

with $\Gamma(x+1) = x\Gamma(x)$ being the gamma function. Similarly ($d < 1$), we can write

$$x_t = (1-B)^{-d} w_t = \sum_{j=0}^{\infty} \psi_j B^j w_t = \sum_{j=0}^{\infty} \psi_j w_{t-j} \tag{5.4}$$

where

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}. \tag{5.5}$$

When $|d| < .5$, the processes (5.2) and (5.4) are well-defined stationary processes (see Brockwell and Davis, 1991, for details). In the case of fractional differencing, however, the coefficients satisfy $\sum \pi_j^2 < \infty$ and $\sum \psi_j^2 < \infty$ as opposed to the absolute summability of the coefficients in ARMA processes.

Using the representation (5.4)–(5.5), and after some nontrivial manipulations, it can be shown that the ACF of x_t is

$$\rho(h) = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)} \sim h^{2d-1} \tag{5.6}$$

for large h. From this we see that for $0 < d < .5$

$$\sum_{h=-\infty}^{\infty} |\rho(h)| = \infty$$

and hence the term *long memory*.

Gamma function

From Wikipedia, the free encyclopedia

For the gamma function of ordinals, see Veblen function. For the gamma distribution in statistics, see Gamma distribution.

In mathematics, the **gamma function** (represented by the capital Greek alphabet letter Γ) is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers. That is, if n is a positive integer:

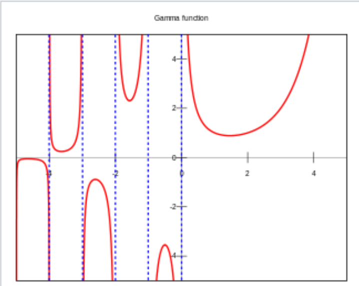
$$\Gamma(n) = (n-1)!$$

The gamma function is defined for all complex numbers except the non-positive integers. For complex numbers with a positive real part, it is defined via a convergent improper integral:

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x}\, dx$$

This integral function is extended by analytic continuation to all complex numbers except the non-positive integers (where the function has simple poles), yielding the meromorphic function we call the gamma function. It has no zeroes, so its reciprocal $1/\Gamma(z)$ is a holomorphic function. In fact the gamma function corresponds to the Mellin transform of the negative exponential function:

$$\Gamma(z) = \{\mathcal{M} e^{-x}\}(z)$$



The gamma function along part of the real axis

The gamma function is a component in various probability-distribution functions, and as such it is applicable in the fields of probability and statistics, as well as combinatorics.

$$\pi_{j+1}(d) = \frac{(j-d)\pi_j(d)}{(j+1)}, \tag{5.7}$$

for $j = 0, 1, \ldots$, with $\pi_0(d) = 1$. Maximizing the joint likelihood of the errors under normality, say, $w_t(d)$, will involve minimizing the sum of squared errors

$$Q(d) = \sum w_t^2(d).$$

The usual Gauss–Newton method, described in §3.6, leads to the expansion

$$w_t(d) = w_t(d_0) + w_t'(d_0)(d - d_0),$$

where

$$w_t'(d_0) = \left.\frac{\partial w_t}{\partial d}\right|_{d=d_0}$$

and d_0 is an initial estimate (guess) at to the value of d. Setting up the usual regression leads to

$$d = d_0 - \frac{\sum_t w_t'(d_0) w_t(d_0)}{\sum_t w_t'(d_0)^2}. \tag{5.8}$$

The derivatives are computed recursively by differentiating (5.7) successively with respect to d: $\pi_{j+1}'(d) = [(j-d)\pi_j'(d) - \pi_j(d)]/(j+1)$, where $\pi_0'(d) = 0$. The errors are computed from an approximation to (5.2), namely,

$$w_t(d) = \sum_{j=0}^{t} \pi_j(d) x_{t-j}. \tag{5.9}$$

It is advisable to omit a number of initial terms from the computation and start the sum, (5.8), at some fairly large value of t to have a reasonable approximation.

Example 5.1 Long Memory Fitting of the Glacial Varve Series

We consider analyzing the glacial varve series discussed in Examples 2.6 and 3.32. Figure 2.6 shows the original and log-transformed series (which we denote by x_t). In Example 3.40, we noted that x_t could be modeled as an ARIMA(1, 1, 1) process. We fit the fractionally differenced model, (5.1), to the mean-adjusted series, $x_t - \bar{x}$. Applying the Gauss–Newton iterative procedure previously described, starting with $d = .1$ and omitting the first 30 points from the computation, leads to a final value of $d = .384$, which implies the set of coefficients $\pi_j(.384)$, as given in Figure 5.2 with $\pi_0(.384) = 1$. We can compare roughly the performance of the fractional difference operator with the ARIMA model by examining the autocorrelation functions of the two residual series as shown in Figure 5.3. The ACFs of the two residual series are roughly comparable with the white noise model.

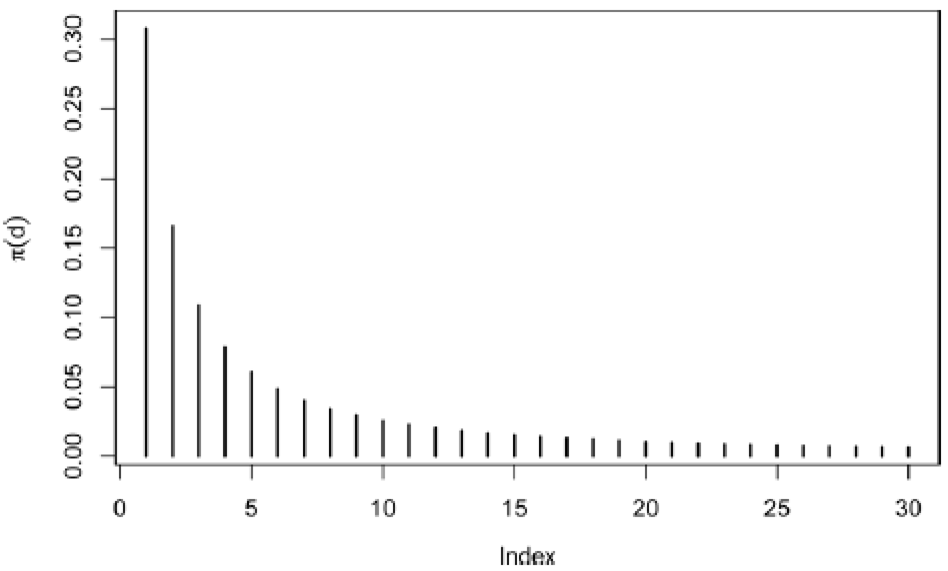


Fig. 5.2. Coefficients $\pi_j(.384), j = 1, 2, \ldots, 30$ in the representation (5.7).

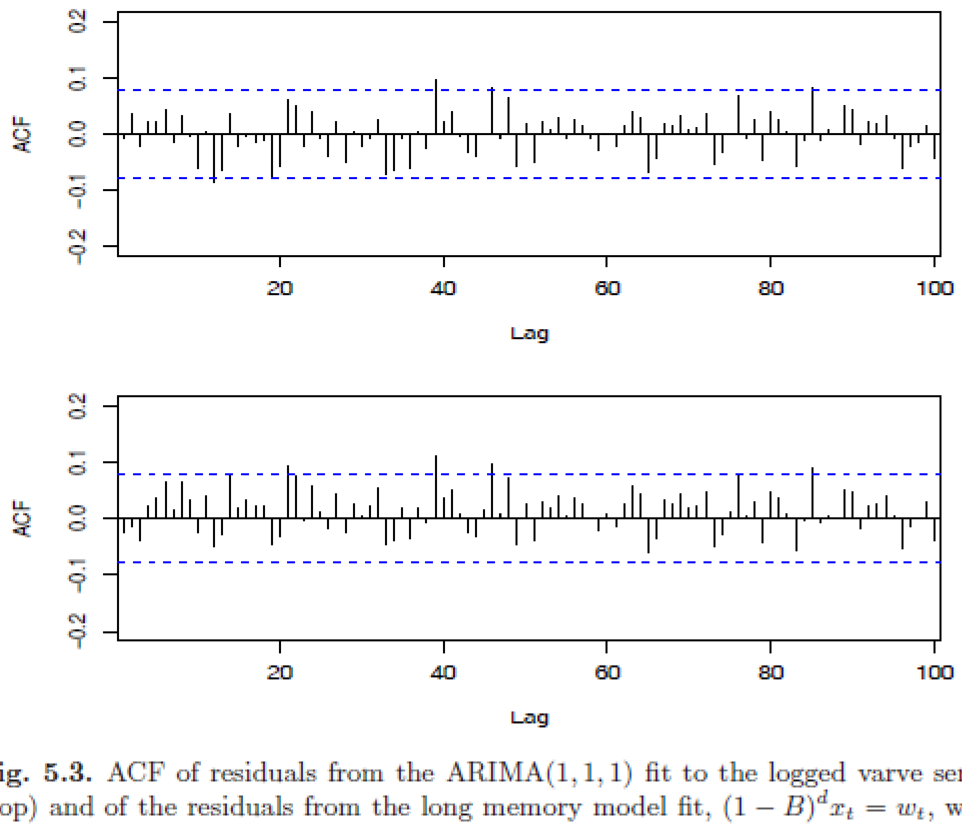


Fig. 5.3. ACF of residuals from the ARIMA(1, 1, 1) fit to the logged varve series (top) and of the residuals from the long memory model fit, $(1 - B)^d x_t = w_t$, with $d = .384$ (bottom).

No obvious short memory ARMA-type component can be seen in the ACF of the residuals from the fractionally differenced varve series shown in Figure 5.3. It is natural, however, that cases will exist in which substantial short memory-type components will also be present in data that exhibits long memory. Hence, it is natural to define the general ARFIMA(p, d, q), $-.5 < d < .5$ process as

$$\phi(B)\nabla^d(x_t - \mu) = \theta(B)w_t, \tag{5.13}$$

where $\phi(B)$ and $\theta(B)$ are as given in Chapter 3. Writing the model in the form

$$\phi(B)\pi_d(B)(x_t - \mu) = \theta(B)w_t \tag{5.14}$$

Forecasting long memory processes is similar to forecasting ARIMA models. That is, (5.2) and (5.7) can be used to obtain the truncated forecasts

$$\widetilde{x}^n_{n+m} = -\sum_{j=1}^{n} \pi_j(\widehat{d})\, \widetilde{x}^n_{n+m-j}, \tag{5.10}$$

for $m = 1, 2, \ldots$. Error bounds can be approximated by using

$$P^n_{n+m} = \widehat{\sigma}^2_w \left(\sum_{j=0}^{m-1} \psi^2_j(\widehat{d}) \right) \tag{5.11}$$

where, as in (5.7),

$$\psi_j(\widehat{d}) = \frac{(j + \widehat{d})\psi_j(\widehat{d})}{(j+1)}, \tag{5.12}$$

with $\psi_0(\widehat{d}) = 1$.

Testes para memória longa

✓ Estatística R/S (*Range Over Standard Deviation*)

Estatística R/S ⇒

$$Q_T = \frac{1}{s_T}\left[max_{1\le k\le T}\sum_{j=1}^{k}(y_j-\bar{y}) - min_{1\le k\le T}\sum_{j=1}^{k}(y_j-\bar{y})\right]$$

DP amostral (s_T); média amostral ($\bar{y}$)

Estatística R/S modificada ⇒

$$\tilde{Q}_T = \frac{1}{\hat{\sigma}_T(q)}\left[max_{1\le k\le T}\sum_{j=1}^{k}(y_j-\bar{y}) - min_{1\le k\le T}\sum_{j=1}^{k}(y_j-\bar{y})\right]$$

estimativa de Newey-West para a variância de longo prazo ($\hat{\sigma}_T(q)$)

✓ Teste semi-paramétrico GPH *(Geweke e Porter-Hudak)*

Estimativa de mínimos quadrados ⇒

$$\hat{d} \xrightarrow{D} N\left(d, \frac{\pi^2}{6\sum_{j=1}^{n_f}(U_j-\bar{U})^2}\right) \qquad U_j = ln\left[4sen^2\left(\frac{\omega_j}{2}\right)\right]$$

Estatística com distribuição normal padrão ⇒

$$t_{d=0} = \hat{d}\cdot\left(\frac{\pi^2}{6\sum_{j=1}^{n_f}(U_j-\bar{U})^2}\right)^{-1/2}$$

Estimação do parâmetro de memória longa

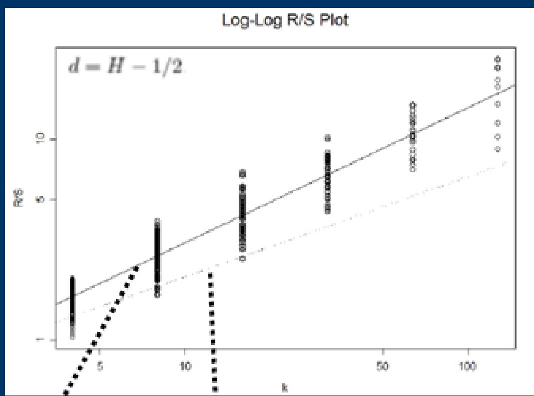
✓ Estimador baseado na estatística R/S
✓ Método do Periodograma
✓ Método de Whittle
✓ Método das Ondaletas
✓ Método de Máxima Verossimilhança

Gráfico na escala log-log da estatística R/S *versus* tamanho da amostra

Inclinação = H > 1/2

⇓

Memória longa



Reta ajustada

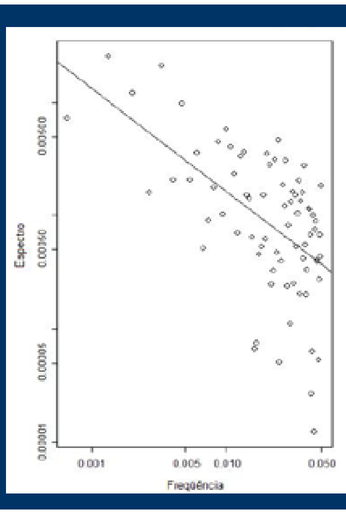
Ausência de memória longa

Gráfico na escala log-log do periodograma *versus* a freqüência

$f(\omega) \to C_f \omega^{1-2H}$ → Quando $\omega \to 0$

↓ Função densidade espectral

Inclinação para freqüências próximas a zero = 1 − 2H



Baseado na estimativa de máxima verossimilhança (domínio da freqüência) de um processo integrado fracionário

Estimador de Whittle para d minimiza esta função → $Q(\theta) = \int_{-\pi}^{\pi} \frac{I(\omega)}{f(\theta;\omega)} d\omega$

- $I(\omega)$ → Periodograma de y_t
- $f(\theta;\omega)$ → Densidade espectral teórica de y_t
- θ → Vetor de parâmetros desconhecidos

Ondaletas: sistema de funções usada como base para representar outras funções

Ondaleta Haar: mais simples e mais antiga. É igual a 1 se [0,1/2) e -1 se [1/2,1)

Coeficiente de ondaletas → $w_{j,k} = \langle x, \psi_{j,k} \rangle = 2^{j/2} \int x(t)\psi(2^j t - k)dt$ → Distribuição Normal $N(0, \sigma^2 2^{-2jd})$, $j \to \infty$

$R(j) = \sigma^2 2^{-2jd}$ → $\ln R(j) = \ln \sigma^2 - d \ln 2^{2j}$ → $R(j) = \frac{1}{2^j} \sum_{k=0}^{2^j - 1} w_{j,k}^2$

Estimador de mínimos quadrados ordinários → $\hat{d} = \left[\sum_{j=0}^{p-1} y_j^2 \right]^{-1} \left[\sum_{j=0}^{p-1} y_j \ln R(j) \right]$

$$y_j = \ln 2^{-2j} - \frac{1}{p} \sum_{j=0}^{p-1} \ln 2^{-2j}$$

Função de verossimilhança Processo ARFIMA(p,d,q)

$$\phi(L)(1-L)^d(y_t-\mu)=\theta(L)\epsilon_t$$

$$L(\eta,\sigma_a^2)=(2\pi\sigma_a^2)^{-n/2}(r_0\cdots r_{n-1})^{-1/2}\exp\left[-\frac{1}{2\sigma_a^2}\sum_{j=1}^{n}(Z_j-\hat{Z}_j)^2/r_{j-1}\right]$$

$$(d,\phi_1,\cdots,\phi_p,\theta_1,\cdots,\theta_q)$$

$$(\sigma_a^2)^{-1}E(Z_j-\hat{Z}_j)^2$$

Estimadores de máxima verossimilhança

$$\hat{\sigma}^2_{MV}=n^{-1}S(\hat{\eta}_{MV})$$

$$\sum_{j=1}^{n}(Z_j-\hat{Z}_j)^2/r_{j-1}$$

η_{MV} é o valor que minimiza a equação

$$l(\eta)=\ln(S(\eta)|n)+n^{-1}\sum_{j=1}^{n}\ln r_{j-1}$$

Aproximação alternativa para l(η)

$$l(\eta)\cong l_*(\eta)=\ln\frac{1}{n}\sum_{j}\frac{I_n(w_j)}{2\pi f(w_j;\eta)}$$

- $I_n(w_j)$: Periodograma de y_t
- $f(w_j;\eta)$: Densidade espectral do processo Z_t

5.3 Unit Root Testing

As discussed in the previous section, the use of the first difference $\nabla x_t = (1-B)x_t$ can be too severe a modification in the sense that the nonstationary model might represent an overdifferencing of the original process. For example, consider a causal AR(1) process (we assume throughout this section that the noise is Gaussian),

$$x_t = \phi x_{t-1} + w_t. \tag{5.26}$$

Applying $(1-B)$ to both sides shows that differencing, $\nabla x_t = \phi \nabla x_{t-1} + \nabla w_t$, or

$$y_t = \phi y_{t-1} + w_t - w_{t-1},$$

where $y_t = \nabla x_t$, introduces extraneous correlation and invertibility problems. That is, while x_t is a causal AR(1) process, working with the differenced process y_t will be problematic because it is a non-invertible ARMA$(1,1)$.

A unit root test provides a way to test whether (5.26) is a random walk (the null case) as opposed to a causal process (the alternative). That is, it provides a procedure for testing

$$H_0\colon \phi = 1 \quad \text{versus} \quad H_1\colon |\phi| < 1.$$

To examine the behavior of $(\widehat{\phi} - 1)$ under the null hypothesis that $\phi = 1$, or more precisely that the model is a random walk, $x_t = \sum_{j=1}^{t} w_j$, or $x_t = x_{t-1} + w_t$ with $x_0 = 0$, consider the least squares estimator of ϕ. Noting that $\mu_x = 0$, the least squares estimator can be written as

$$\widehat{\phi} = \frac{\frac{1}{n}\sum_{t=1}^{n} x_t x_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}^2} = 1 + \frac{\frac{1}{n}\sum_{t=1}^{n} w_t x_{t-1}}{\frac{1}{n}\sum_{t=1}^{n} x_{t-1}^2}, \tag{5.27}$$

where we have written $x_t = x_{t-1} + w_t$ in the numerator; recall that $x_0 = 0$ and in the least squares setting, we are regressing x_t on x_{t-1} for $t = 1, \ldots, n$. Hence, under H_0, we have that

$$\widehat{\phi} - 1 = \frac{\frac{1}{n\sigma_w^2}\sum_{t=1}^n w_t x_{t-1}}{\frac{1}{n\sigma_w^2}\sum_{t=1}^n x_{t-1}^2}. \tag{5.28}$$

Consider the numerator of (5.28). Note first that by squaring both sides of $x_t = x_{t-1} + w_t$, we obtain $x_t^2 = x_{t-1}^2 + 2x_{t-1}w_t + w_t^2$ so that

$$x_{t-1}w_t = \frac{1}{2}(x_t^2 - x_{t-1}^2 - w_t^2),$$

and summing,

$$\frac{1}{n\sigma_w^2}\sum_{t=1}^n x_{t-1}w_t = \frac{1}{2}\Big(\frac{x_n^2}{n\sigma_w^2} - \frac{\sum_{t=1}^n w_t^2}{n\sigma_w^2}\Big).$$

Because $x_n = \sum_1^n w_t$, we have that $x_n \sim N(0, n\sigma_w^2)$, so that $\frac{1}{n\sigma_w^2}x_n^2 \sim \chi_1^2$, the chi-squared distribution with one degree of freedom. Moreover, because w_t is white Gaussian noise, $\frac{1}{n}\sum_1^n w_t^2 \to_p \sigma_w^2$, or $\frac{1}{n\sigma_w^2}\sum_1^n w_t^2 \to_p 1$. Consequently, $(n \to \infty)$

$$\frac{1}{n\sigma_w^2}\sum_{t=1}^n x_{t-1}w_t \xrightarrow{d} \tfrac{1}{2}(\chi_1^2 - 1). \tag{5.29}$$

Definition 5.1 *A continuous time process* $\{W(t);\, t \geq 0\}$ *is called* **standard Brownian motion** *if it satisfies the following conditions:*

(i) $W(0) = 0$;
(ii) $\{W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_n) - W(t_{n-1})\}$ *are independent for any collection of points,* $0 \leq t_1 < t_2 \cdots < t_n$, *and integer* $n > 2$;
(iii) $W(t + \Delta t) - W(t) \sim N(0, \Delta t)$ *for* $\Delta t > 0$.

The result for the denominator uses the functional central limit theorem, which can be found in Billlingsley (1999, §2.8). In particular, if $\xi_1, \ldots, \xi_n$ is a sequence of iid random variables with mean 0 and variance 1, then, for $0 \leq t \leq 1$, the continuous time process

$$S_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\xi_j \;\xrightarrow{d}\; W(t), \tag{5.30}$$

as $n \to \infty$, where $[\![\;]\!]$ is the greatest integer function and $W(t)$ is standard Brownian motion on $[0, 1]$. Note the under the null hypothesis, $x_s = w_1 + \cdots + w_s \sim \mathrm{N}(0, s\sigma_w^2)$, and based on (5.30), we have $\frac{x_s}{\sigma_w\sqrt{n}} \to_d W(s)$. From this fact, we can show that $(n \to \infty)$

$$\sum_{t=1}^n\Big(\frac{x_{t-1}}{\sigma_w\sqrt{n}}\Big)^2\frac{1}{n} \xrightarrow{d} \int_0^1 W^2(t)\,dt\,. \tag{5.31}$$

The denominator in (5.28) is off from the left side of (5.31) by a factor of n^{-1}, and we adjust accordingly to finally obtain ($n \to \infty$),

$$n(\widehat{\phi} - 1) = \frac{\frac{1}{n\sigma_w^2} \sum_{t=1}^n w_t x_{t-1}}{\frac{1}{n^2\sigma_w^2} \sum_{t=1}^n x_{t-1}^2} \xrightarrow{d} \frac{\frac{1}{2}(\chi_1^2 - 1)}{\int_0^1 W^2(t)\, dt}. \tag{5.32}$$

The test statistic $n(\widehat{\phi} - 1)$ is known as the unit root or Dickey-Fuller (DF) statistic (see Fuller, 1996), although the actual DF test statistic is normalized

Example 5.3 Testing Unit Roots in the Glacial Varve Series

In this example we use the R package `tseries` to test the null hypothesis that the log of the glacial varve series has a unit root, versus the alternate hypothesis that the process is stationary. We test the null hypothesis using the available DF, ADF and PP tests; note that in each case, the general regression equation incorporates a constant and a linear trend. In the ADF test, the default number of AR components included in the model, say k, is $[\![(n-1)^{\frac{1}{3}}]\!]$, which corresponds to the suggested upper bound on the rate at which the number of lags, k, should be made to grow with the sample size for the general ARMA(p, q) setup. For the PP test, the default value of k is $[\![.04n^{\frac{1}{4}}]\!]$.

```
library(tseries)
adf.test(log(varve), k=0)                # DF test
   Dickey-Fuller = -12.8572, Lag order = 0, p-value < 0.01
   alternative hypothesis: stationary
adf.test(log(varve))                     # ADF test
   Dickey-Fuller = -3.5166, Lag order = 8, p-value = 0.04071
   alternative hypothesis: stationary
```


```
library(tseries)
adf.test(log(varve), k=0)                # DF test
   Dickey-Fuller = -12.8572, Lag order = 0, p-value < 0.01
   alternative hypothesis: stationary
adf.test(log(varve))                     # ADF test
   Dickey-Fuller = -3.5166, Lag order = 8, p-value = 0.04071
   alternative hypothesis: stationary
```