
5.1 Summary

Below Proposition 3.5 we saw that the autocorrelation sequence of any stationary ARMA process dies out at exponential rate: $|\rho(h)| \leq c g^h$, see (3.14). This is too restrictive for many time series of stronger persistence, which display long memory in that the autocovariance sequence vanishes at a slower rate. In some fields of economics and finance long memory is treated as an empirical stylized fact.¹ Fractional integration as a model for long memory will be presented in this chapter. In the same paper where Granger (1981) introduced the Nobel prize winning concept of cointegration (see Chap. 16) he addressed the idea of fractional integration, too. For an early survey on fractional integration and applications see Baillie (1996).

5.2 Persistence and Long Memory

We have already briefly touched upon the so-called random walk, see Eq. (1.9). In terms of the difference operator this can be written as $\Delta x_t = \varepsilon_t$, i.e. the process has to be differenced once to obtain stationarity. Alternatively, the process is given by a cumulation or summation over the shocks, $x_t = \sum_{j=1}^t \varepsilon_j$, see (1.8), which is the reason to call the process $\{x_t\}$ integrated of order 1, see also Chap. 14. In this section, differencing or integration of order 1 will be extended by introducing non-integer orders of differencing and integration.

¹See e.g. the special issue edited by Maasoumi and McAleer (2008) in *Econometric Reviews* on “Realized Volatility and Long Memory”.

Persistence

By persistence we understand how strongly a past shock affects the presence of a stochastic process. We stick to the $MA(\infty)$ representation behind the Wold decomposition of a stationary process that we briefly touched upon in Chap. 3,

$$x_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad \sum_{j=0}^{\infty} c_j^2 < \infty,$$

with $\{\varepsilon_t\}$ forming a white noise sequence. The impulse responses coefficients $\{c_j\}$ measures the response of x_t on a shock j periods ago. With stationary processes, the shocks are transitory in that $\lim_{j \rightarrow \infty} c_j = 0$. In particular, for stationary ARMA processes we know that the impulse responses die out so fast that they are summable in absolute value, see (3.2). To model a stronger degree of persistence and **long memory**, we require a slower convergence to zero. The model of fractional integration of order d will impose the so-called hyperbolic decay rate,

$$c_j = c j^{d-1}, \quad c \neq 0.$$

Under $d < 1$, the sequence $\{c_j\}$ converges to zero. Clearly, the larger d , the stronger is the persistence in that j^{d-1} dies out more slowly. Hence, the parameter d measures the strength of persistence. Contrary to the exponential case characteristic of ARMA processes, hyperbolic decay is so slow for positive $d > 0$, that the impulse responses are not summable. In Problem 5.1 we will establish the following convergence result concerning the so-called (generalized) harmonic series, often also called p -series:

$$\sum_{j=1}^{\infty} j^{-p} < \infty \quad \text{if and only if } p > 1. \quad (5.1)$$

Moreover, we will show in Problem 5.2 that exponential decay to zero is faster than hyperbolic one:

$$\lim_{j \rightarrow \infty} \frac{g^j}{j^{d-1}} = 0, \quad 0 < g < 1, \quad |d| < 1.$$

In order to illustrate the different decay rates, we display in Figs. 5.1 and 5.2 sequences j^{d-1} and g^{j-1} , respectively; by construction they all have the value 1 at $j = 1$.

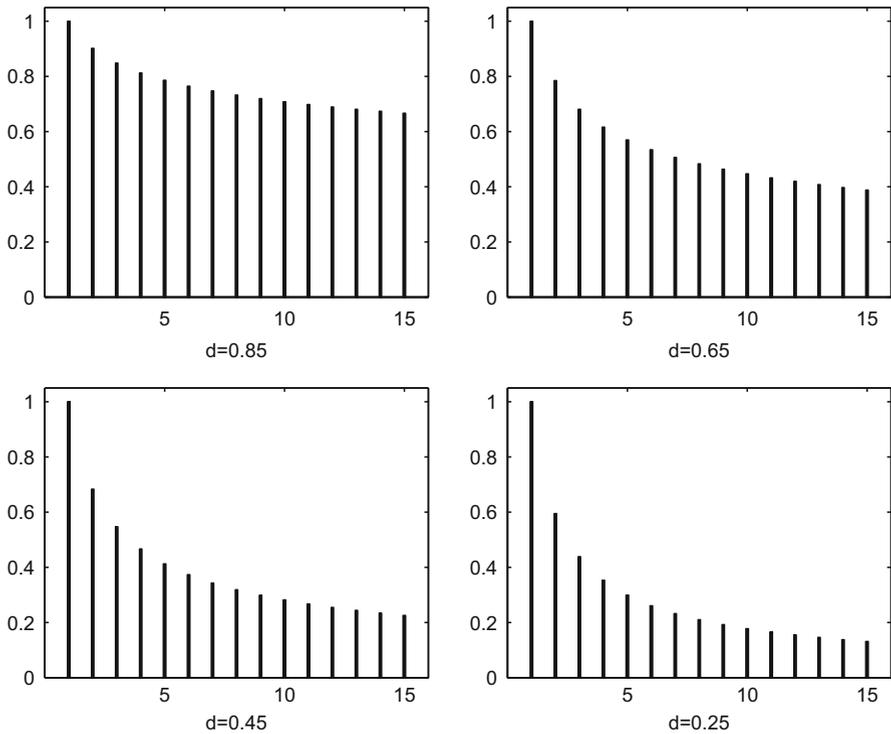


Fig. 5.1 j^{d-1} for $d = 0.85, 0.65, 0.45, 0.25$

For a process to be stationary, we require the impulse response sequence to be square summable, see Proposition 3.2. From (5.1) we learn that $\sum_{j=1}^{\infty} j^{2d-2}$ is finite if and only if $d < 0.5$, which hence turns out to be the stationarity condition for processes with impulse responses $\{j^{d-1}\}$. The model of fractional integration, however, does not directly assume $c_j = c j^{d-1}$; rather this power law will hold only true for large j ,

$$c_j \sim c j^{d-1}, \quad j \rightarrow \infty,$$

where “ \sim for $j \rightarrow \infty$ ” is to be understood as a proper limit in the following way:

$$a_j \sim b_j \iff \lim_{j \rightarrow \infty} \frac{a_j}{b_j} = 1, \quad b_j \neq 0. \tag{5.2}$$

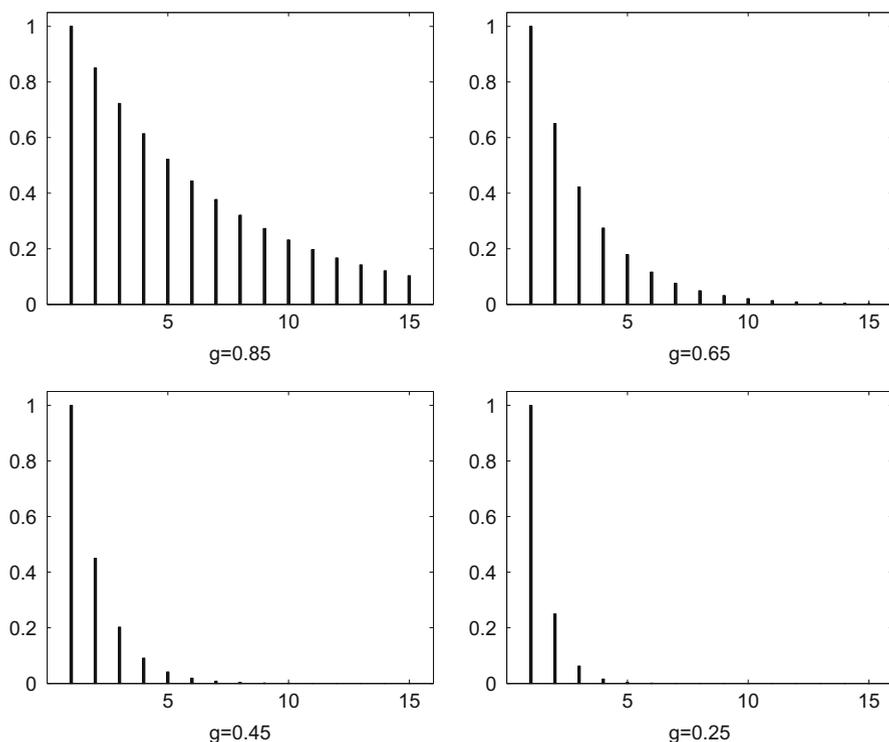


Fig. 5.2 g^{j-1} for $g = 0.85, 0.65, 0.45, 0.25$

Fractional Differencing and Integration

With the usual difference operator $\Delta = (1 - L)$, see Example 3.2, we define **fractional differences** by binomial expansion²:

$$\begin{aligned} \Delta^d &= (1 - L)^d = 1 - dL - \frac{d(1-d)}{2}L^2 - \frac{d(1-d)(2-d)}{6}L^3 - \dots \\ &= \sum_{j=0}^{\infty} \pi_j L^j, \quad d > -1. \end{aligned}$$

²For the rest of this chapter we maintain $d > -1$, which guarantees that $\{\pi_j\}$ converges to 0 with growing j , making the infinite expansion meaningful.

Readers not be familiar with binomial series for $d \notin \mathbb{N}$ may wish to consult e.g. Trench (2013, Sect. 4.5). The binomial series results from a Taylor expansion of $(1 - z)^d$ about $z = 0$, hence also called Maclaurin series. The coefficients $\{\pi_j\}$ are given in terms of binomial coefficients

$$\binom{d}{j} = \frac{d(d-1) \cdots (d-j+1)}{j!},$$

yielding the recursion

$$\pi_j = \binom{d}{j} (-1)^j = \frac{j-1-d}{j} \pi_{j-1}, \quad j \geq 1, \quad \pi_0 = 1. \quad (5.3)$$

For natural numbers d one has the more familiar finite expansions,

$$(1-L)^1 = 1-L, \quad (1-L)^2 = 1-2L+L^2,$$

while the expansion in (5.3) holds for non-integer (or fractional) values of d , too. In Problem 5.3 we derive the behavior for large j ,

$$\pi_j \sim \frac{j^{-d-1}}{\Gamma(-d)}, \quad j \rightarrow \infty, \quad d \neq 0, \quad (5.4)$$

where $\Gamma(\cdot)$ is the so-called **Gamma function** introduced at greater detail below its definition in (5.18) in the Problem section.

Similarly to fractional differences, we may define the **fractional integration** operator upon inversion,

$$\Delta^{-d} = (1-L)^{-d} = \sum_{j=0}^{\infty} \psi_j L^j, \quad \psi_0 = 1,$$

where the coefficients are given by simply replacing d by $-d$ in (5.3):

$$\psi_j = \binom{-d}{j} (-1)^j = \frac{j-1+d}{j} \psi_{j-1}, \quad j \geq 1. \quad (5.5)$$

The same arguments establishing (5.4) hence show

$$\psi_j \sim \frac{j^{d-1}}{\Gamma(d)}, \quad j \rightarrow \infty, \quad d \neq 0. \quad (5.6)$$

Fractional integration thus imposes the hyperbolic decay rate discussed above, where the speed of convergence varies with d . From (5.1) we observe with (5.6) that $\{\psi_j\}$ is summable if and only if $d < 0$, in which case

$$\sum_{j=0}^{\infty} \psi_j = (1-z)^{-d} \Big|_{z=1} = 0, \quad d < 0. \quad (5.7)$$

Further, $\{\psi_j\}$ is square summable if and only if $d < 0.5$. These are the ingredients to define a fractionally integrated process.

5.3 Fractionally Integrated Noise

We now apply the above findings to the simplest case of fractional noise (which is short for: fractionally integrated noise), define long memory in the time domain in (5.9), and translate it into the frequency domain.

Fractional Noise and Long Memory

In case of fractionally integrated noise the fractional differencing filter Δ^d has to be applied to $\{x_t\}$ in order to obtain white noise $\{\varepsilon_t\}$ with variance σ^2 : $\Delta^d x_t = \varepsilon_t$. Equivalently, we write after inverting the differences

$$\begin{aligned} x_t &= (1-L)^{-d} \varepsilon_t \\ &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \quad d < 0.5, \end{aligned} \quad (5.8)$$

with $\{\psi_j\}$ from (5.5) being the sequence of impulse response coefficients measuring the effect of a past shock on the presence. As we have discussed above the impulse responses die out the more slowly the larger d is. In that sense we interpret d as measure of persistence or memory. The impulse responses die out so slowly that they are not absolutely summable for positive memory parameter d . Consequently for $d > 0$, $\{x_t\}$ from (5.8) does not belong to the class of processes with absolutely summable autocovariances characterized in (3.2), while all stationary ARMA processes belong to this class. Hence, fractional integration models for $d > 0$ a feature that is not captured by traditional ARMA processes, which we call **strong persistence**; it is defined in the time domain by

$$\sum_{j=1}^J \psi_j \rightarrow \infty, \quad J \rightarrow \infty, \quad \text{if } d > 0.$$

Consequently, *CIR* or *VR* defined in (3.3) and (4.6), respectively, do not exist. Even though the MA coefficients are not summable for positive d , the process is still stationary as long as $d < 0.5$ because $\sum_j \psi_j^2 < \infty$ due to (5.6) and (5.1). Further, the process is often called invertible for $d > -0.5$ since an autoregressive representation exists that is square summable³:

$$\sum_{j=0}^{\infty} \pi_j x_{t-j} = \varepsilon_t \quad \text{with} \quad \sum_{j=0}^{\infty} \pi_j^2 < \infty.$$

Note that the existence of an autoregressive representation in the mean square sense does not require square summability; in fact, Bondon and Palma (2007) extend the range of invertibility in the mean square to $d > -1$. Given the existence of the $MA(\infty)$ representation the following properties of fractionally integrated noise are proven in the Problem section.

Proposition 5.1 (Fractional noise, time domain) *For fractionally integrated noise from (5.8) it holds with $-1 < d < 0.5$ that*

(a) *the variance equals*

$$\gamma(0) = \gamma(0; d) = \sigma^2 \frac{\Gamma(1 - 2d)}{(\Gamma(1 - d))^2},$$

with $\Gamma(\cdot)$ being defined in (5.18), and $\gamma(0; d)$ achieves its minimum for $d = 0$;

(b) *the autocovariances equal*

$$\begin{aligned} \gamma(h) &= \frac{h - 1 + d}{h - d} \gamma(h - 1), \quad h = 1, 2, \dots, \\ &\sim \gamma_d \sigma^2 h^{2d-1}, \quad h \rightarrow \infty, \quad d \neq 0, \end{aligned}$$

with

$$\gamma_d = \frac{\Gamma(1 - 2d)}{\Gamma(d)\Gamma(1 - d)},$$

where $\gamma_d < 0$ if and only if $d < 0$;

(c) *the autocorrelations $\rho(h) = \rho(h; d)$ grow with d for $d > 0$.*

Let us briefly comment those results. First, since $\Gamma(1) = 1$, the minimum variance obtained in the white noise case ($d = 0$) is of course $\gamma(0; 0) = \sigma^2$. Second,

³A more technical exposition can be found in Brockwell and Davis (1991, Thm. 13.2.1) or Giraitis, Koul, and Surgailis (2012, Thm. 7.2.1), although they consider only the range $|d| < 0.5$.

from the hyperbolic decay of the autocovariance sequence we observe that $\gamma(h)$ converges to zero with h as long as $d < 0.5$, but for $d > 0$ so slowly, that we have long memory defined as

$$\sum_{h=0}^H |\gamma(h)| \rightarrow \infty, \quad H \rightarrow \infty \quad \text{if } d > 0. \quad (5.9)$$

In particular, the autocovariances die out the more slowly the larger the memory parameter d is. Obviously, the same feature can be rephrased in terms of autocorrelations. The recursion carries over to the autocorrelations, and Proposition 5.1 (b) yields

$$\rho(h) \sim \frac{\gamma_d \sigma^2}{\gamma(0)} h^{2d-1}, \quad h \rightarrow \infty.$$

For a numerical and graphical illustration see Fig. 5.3. The asymptotic constant γ_d has the same sign as d , meaning that in case of long memory the autocovariances converge to zero from above, and vice versa from below zero for $d < 0$, see again

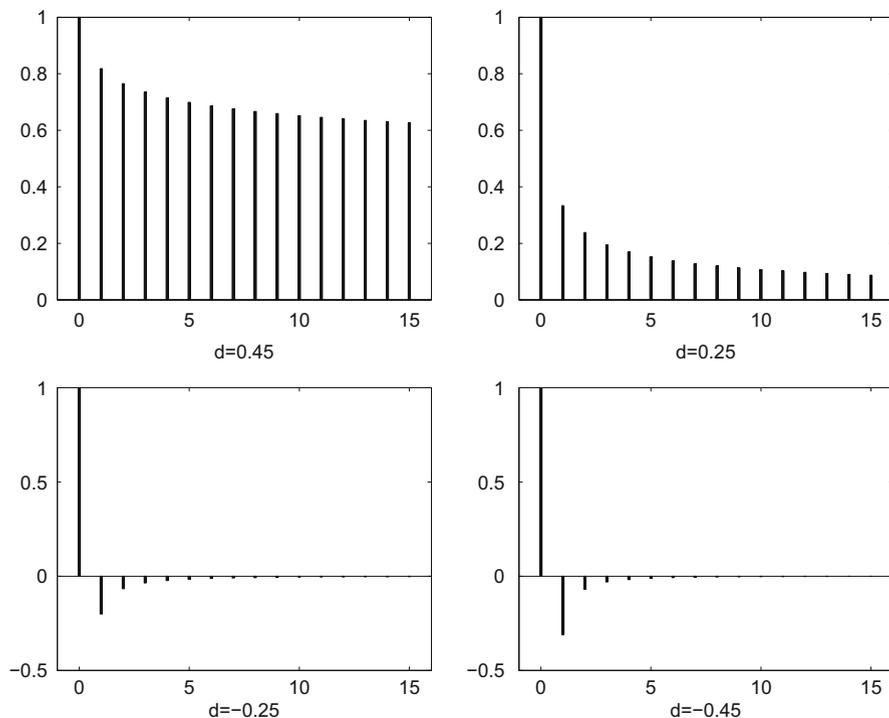


Fig. 5.3 $\rho(h)$ from Proposition 5.1 for $d = 0.45, 0.25, -0.25, -0.45$

Fig. 5.3. Note, however, that γ_d collapses to zero as $d \rightarrow 0$, simply meaning that the hyperbolic decay rate does not hold for $d = 0$. Third, a similar effect that d has on $\rho(h)$ at long lags, holds true for finite h . More precisely, Proposition 5.1 (c) says for each finite h that the autocorrelation grows with d (for $d > 0$), which reinforces the interpretation of d as measure of persistence or long memory.⁴

The case of negative d results in short memory in that the autocovariances are absolutely summable, which is clear again from the p -series in (5.1). This case is sometimes called antipersistent, the reason for that being

$$\sum_{j=1}^{\infty} \psi_j = 0, \quad \text{if } d < 0.$$

This property translates into a special case of short memory, namely

$$\sum_{h=-\infty}^{\infty} \rho(h) = 0, \quad \text{if } d < 0,$$

as we will become obvious from the spectrum at frequency zero.

Long Memory in the Frequency Domain

It is obvious from the definition of the spectrum in (4.3) that it does not exist at the origin under long memory ($d > 0$), because the autocovariances are not summable. Still, the previous chapter has been set up sufficiently general to cover long memory, see (4.1). Given a singularity at frequency $\lambda = 0$, one still may determine the rate at which $f(\lambda)$ goes off to infinity as λ approaches 0. To determine f , we have to evaluate the power transfer function of $(1 - L)^{-d}$ from Proposition 4.1 and obtain⁵

$$T_{(1-L)^{-d}}(\lambda) = (2 - 2 \cos(\lambda))^{-d} = \left(4 \sin^2\left(\frac{\lambda}{2}\right)\right)^{-d},$$

⁴More complicated is the effect of changes in d if $d < 0$, see Hassler (2014).

⁵Readers not familiar with complex numbers, $i^2 = -1$, may skip the following equation, see also Footnote 6 in Chap. 4:

$$\begin{aligned} T_{(1-L)^{-d}}(\lambda) &= (1 - e^{i\lambda})^{-d}(1 - e^{-i\lambda})^{-d} \\ &= (1 - e^{i\lambda} - e^{-i\lambda} + 1)^{-d} \\ &= (2 - 2 \cos(\lambda))^{-d}. \end{aligned}$$

where the trigonometric half-angle formula was used for the second equality:

$$2 \sin^2(x) = 1 - \cos(2x). \quad (5.10)$$

We hence have the following result.

Proposition 5.2 (Fractional noise, frequency domain) *Under the assumptions of Proposition 5.1 it holds for the spectrum of fractional noise $x_t = (1 - L)^{-d} \varepsilon_t$ that*

$$f(\lambda) = \left(4 \sin^2 \left(\frac{\lambda}{2} \right) \right)^{-d} \frac{\sigma^2}{2\pi}, \quad \lambda > 0, \quad (5.11)$$

and

$$f(\lambda) \sim \lambda^{-2d} \frac{\sigma^2}{2\pi}, \quad \lambda \rightarrow 0. \quad (5.12)$$

The second statement in Proposition 5.2 is again understood to be asymptotic: Similarly to (5.2) we denote for two function $a(x)$ and $b(x) \neq 0$:

$$a(x) \sim b(x) \text{ for } x \rightarrow 0 \iff \lim_{x \rightarrow 0} \frac{a(x)}{b(x)} = 1. \quad (5.13)$$

Since $\lim_{x \rightarrow 0} \sin(x)/x = 1$ we write $\sin(x) \sim x$ for $x \rightarrow 0$. Consequently, (5.12) arises from (5.11).

From Proposition 5.2 we learn that long memory ($d > 0$) translates into a spectral singularity at frequency zero, and the negative slope is the steeper the larger d is. In other words: the longer the memory, the stronger is the contribution of the long-run trend to the variance of the process. The antipersistent case in contrast is characterized by the opposite extreme: $f(0) = 0$. For an illustration, have a look at Fig. 5.4.

Example 5.1 (Fractionally Integrated Noise) Although the fractional noise is dominated by the trend component at frequency zero (strongly persistent) for $d > 0$, the process is stationary as long as $d < 0.5$. Consequently, a typical trajectory can not drift off but displays somehow reversing trends. In Fig. 5.5 we see from simulated data that the deviations from the zero line are stronger for $d = 0.45$ than for $d = 0.25$. The antipersistent series ($d = -0.45$), in contrast, displays an oscillating behavior due to the negative autocorrelation. ■

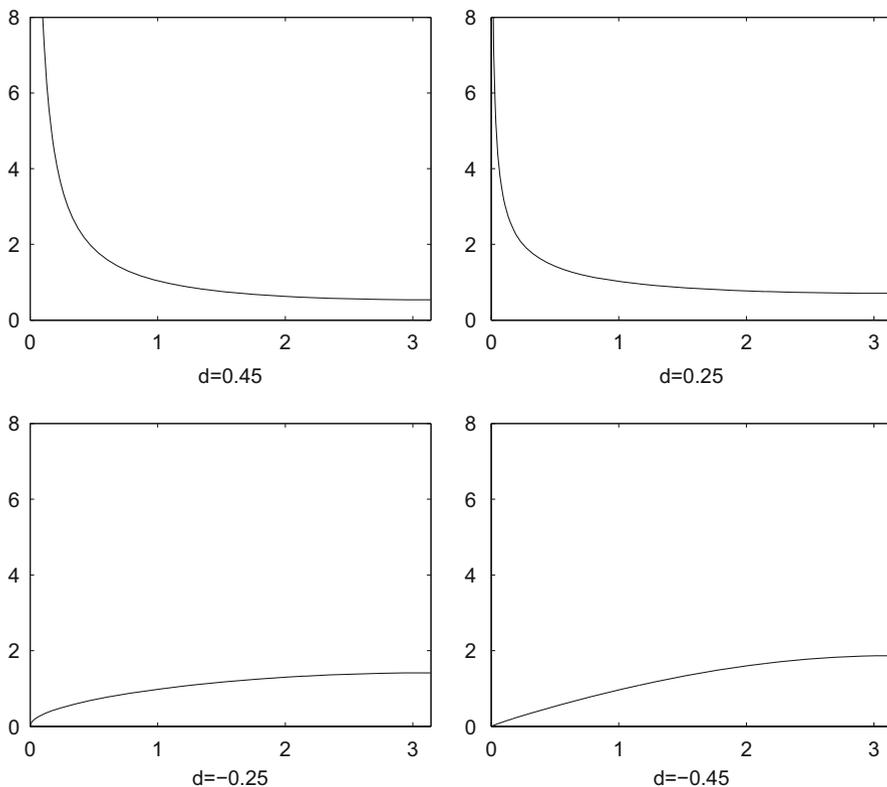


Fig. 5.4 $2\pi f(\lambda)$ from Proposition 5.2 for $d = 0.45, 0.25, -0.25, -0.45$

5.4 Generalizations

On top of long memory as implied by fractional integration for $0 < d < 0.5$, we now want to allow for additional short memory. We assume that Δ^d has to be applied to $\{x_t\}$ in order to obtain a short memory process $\{e_t\}$: $\Delta^d x_t = e_t$. At the end of this section, the order of integration d will be extended beyond $d = 0.5$ to cover nonstationary processes, too. Thus we define general fractionally integrated processes of order d , in short $x_t \sim I(d)$.⁶

⁶The use of ' \sim ' with a differing meaning from that one in (5.2) should not be a source for confusion.

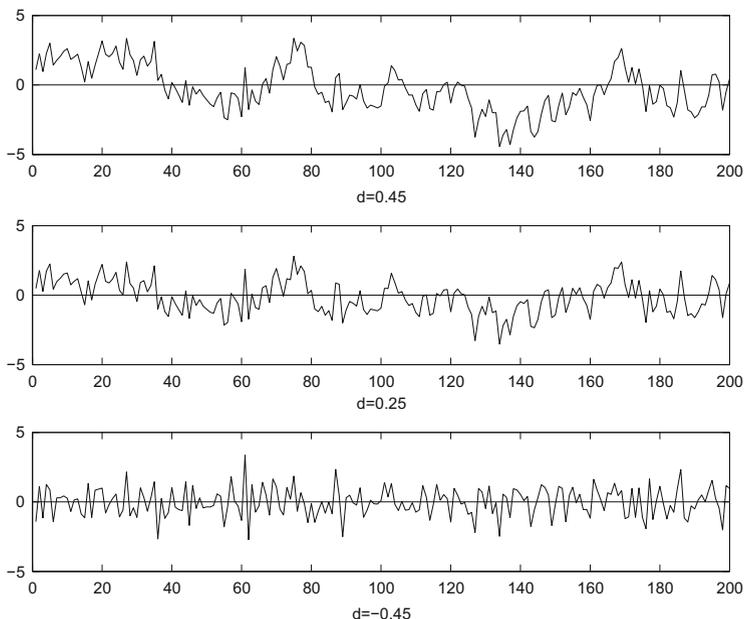


Fig. 5.5 Simulated fractional noise for $d = 0.45, 0.25, -0.45$

Fractionally Integrated ARMA Processes (ARFIMA)

Since the papers by Granger and Joyeux (1980) and Hosking (1981), it is often assumed that $\{e_t\}$ is a stationary and invertible ARMA(p, q) process, $A(L)e_t = B(L)\varepsilon_t$, with spectrum

$$f_e(\lambda) = \frac{T_B(\lambda)}{T_A(\lambda)} \frac{\sigma^2}{2\pi},$$

see Corollary 4.1. An ARFIMA(p, d, q) process is defined by replacing ε_t in (5.8) by e_t , such that

$$A(L)\Delta^d x_t = B(L)e_t.$$

With the expansion from (5.5) one obtains

$$x_t = (1 - L)^{-d} e_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}, \quad t \in \mathbb{Z}, \quad d < 0.5. \quad (5.14)$$

Under stationarity and invertibility of the ARMA process, the spectrum f_e of $\{e_t\}$ is bounded and bounded away from zero everywhere:

$$0 < f_e(\lambda) < \infty, \quad \lambda \in [0, \pi]. \quad (5.15)$$

The results from Propositions 5.1 and 5.2 carry over to the ARFIMA(p, d, q) process, see also Brockwell and Davis (1991, Sect. 13.2).

Proposition 5.3 (ARFIMA) *Let the ARMA process with $A(L)e_t = B(L)\varepsilon_t$ be stationary and invertible. Then the ARFIMA process with $A(L)\Delta^d x_t = B(L)\varepsilon_t$ and $-1 < d < 0.5$ is stationary, and it holds that*

(a) *the spectral density $f(\lambda)$ is given as*

$$\begin{aligned} f(\lambda) &= 4^{-d} \sin^{-2d} \left(\frac{\lambda}{2} \right) f_e(\lambda), \quad \lambda > 0 \\ &\sim \lambda^{-2d} f_e(0), \quad \lambda \rightarrow 0; \end{aligned}$$

(b) *the autocovariances satisfy*

$$\gamma(h) \sim \gamma_d f_e(0) 2\pi h^{2d-1}, \quad h \rightarrow \infty, \quad d \neq 0,$$

with γ_d from Proposition 5.1.

Hosking (1981, Thm. 2) and Brockwell and Davis (1991, Thm. 13.2.2) cover only $|d| < 0.5$, but their proof carries over to $-1 < d \leq -0.5$. Further, they state only $\gamma(h) \sim C h^{2d-1}$ for some constant $C \neq 0$; looking at the details of the proof of Brockwell and Davis (Thm. 13.2.2), however, it turns out that $C = \gamma_d f_e(0) 2\pi$, which of course covers the case of Proposition 5.1, too. In particular, we find again that long memory defined by a non-summable autocovariance sequence translates into a spectral peak at $\lambda = 0$. This feature occurs for $d > 0$.

Semiparametric Models

The parametric assumption that $\{e_t\}$ is an ARMA process is by no means essential for Proposition 5.3 to hold. More generally, we now define a stationary process $\{e_t\}$ to be integrated of order 0, $e_t \sim I(0)$, if

$$e_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}, \quad \text{with } \sum_{k=0}^{\infty} |b_k| < \infty \text{ and } \sum_{k=0}^{\infty} b_k \neq 0, \quad (5.16)$$

where $b_0 = 1$ and $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$. The absolute summability of $\{b_k\}$ rules out long memory (or $d > 0$) in that the autocovariances of $\{e_t\}$ are absolutely summable, see Proposition 3.2; the second condition that the sequence $\{b_k\}$ does not sum up to zero rules out that $\{e_t\}$ is integrated of order d with $d < 0$, see (5.7). This motivates to call processes from (5.16) integrated of order 0. Consequently, $\{x_t\}$ from (5.14) with $\{e_t\}$ from (5.16) is called integrated of order d , $x_t \sim I(d)$. The spectrum of $\{e_t\}$ is of course given by Proposition 4.2, see (4.5). With f_e being the spectrum of the $I(0)$ process, Proposition 5.3 continues to hold without changes, provided $0 < d$ (see Giraitis et al., 2012, Prop. 3.1.1).

A further question is the behavior of the impulse responses of a general $I(d)$ process without parametric model: Does the decay rate j^{d-1} of $\{\psi_j\}$ from Δ^{-d} carry over? The answer is almost yes, but mild additional assumptions have to be imposed on $\{b_k\}$ from (5.16). Denote

$$x_t = \Delta^{-d} e_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j},$$

where the MA coefficients are given by convolution:

$$c_j = \sum_{k=0}^j b_k \psi_{j-k}, \quad j \geq 0.$$

Hassler and Kokoszka (2010) prove that a necessary and sufficient condition for

$$c_j \sim \frac{\sum_{k=0}^{\infty} b_k}{\Gamma(d)} j^{d-1}, \quad d > 0,$$

is under long memory

$$k^{1-d} b_k \rightarrow 0, \quad k \rightarrow \infty. \tag{5.17}$$

This is a very weak condition satisfied by all stationary ARMA models and most other processes of practical interest. Hassler (2012) proves that this condition remains necessary in the case of antipersistence, $d < 0$, and establishes a mildly stronger sufficient condition.

The statistical literature often refrains from a fractionally integrated model of the type $x_t = (1-L)^{-d} e_t$, and directly assumes for the corresponding spectral behavior in a vicinity of the origin:

$$f(\lambda) \sim \lambda^{-2d} g(\lambda), \quad \lambda \rightarrow 0.$$

When it comes to estimation of d , technical smoothness restrictions are imposed on g , including of course the minimum assumption

$$0 < g(0) < \infty,$$

which is required to identify d .

Nonstationary Processes

The simplest way to define a process that is integrated of a higher order than $d < 0.5$ is as follows. Consider a sequence $\{y_t\}$ that has to be differenced once in order to obtain an $I(\delta)$ process, $\Delta y_t = x_t$, $x_t \sim I(\delta)$ or $y_t = y_{t-1} + x_t$. Given a starting value $y_0 = 0$, the solution of this difference equation for $t \in \{1, 2, \dots, n\}$ is

$$y_t = \sum_{j=1}^t x_j, \quad t = 1, 2, \dots, n.$$

Since $\{y_t\}$ is given by integration over an $I(\delta)$ process, we say that $\{y_t\}$ is integrated of order d , $y_t \sim I(d)$, with $d = \delta + 1$. For $d \geq 0.5$, i.e. $\delta \geq -0.5$, the process $\{y_t\}$ is necessarily nonstationary. We illustrate this type of nonstationarity or drift by means of an example.

Example 5.2 (Nonstationary Fractional Noise) The middle graph in Fig. 5.6 displays a realization of a random walk ($d = 1$). It drifts off from the zero line for very long time spans and crosses only a few times. The $I(1.45)$ process drifts even more pronouncedly displaying a much smoother trajectory than the random walk. The $I(0.55)$ process does not drift as strongly, hitting the zero line much more often. In fact, comparing the $I(0.55)$ series with the $I(0.45)$ case from Fig. 5.5, one can imagine that it may be hard to tell apart stationarity and nonstationarity in finite samples. ■

The case of $\delta = 0$ or $y_t \sim I(1)$ is of particular interest in many financial and economic applications. Hence, one may wish to test whether a process is $I(1)$ or not,

$$H_0 : d = 1 \quad \text{vs.} \quad H_1 : d \neq 1.$$

One method to discriminate more specifically between $d = 1$ and $d = 0$ is the celebrated test by Dickey and Fuller (1979), see Chap. 15. In a fractionally integrated framework it is more generally possible to decide e.g. whether a process is nonstationary or not,

$$H_0 : d \geq 0.5 \quad \text{vs.} \quad H_1 : d < 0.5,$$

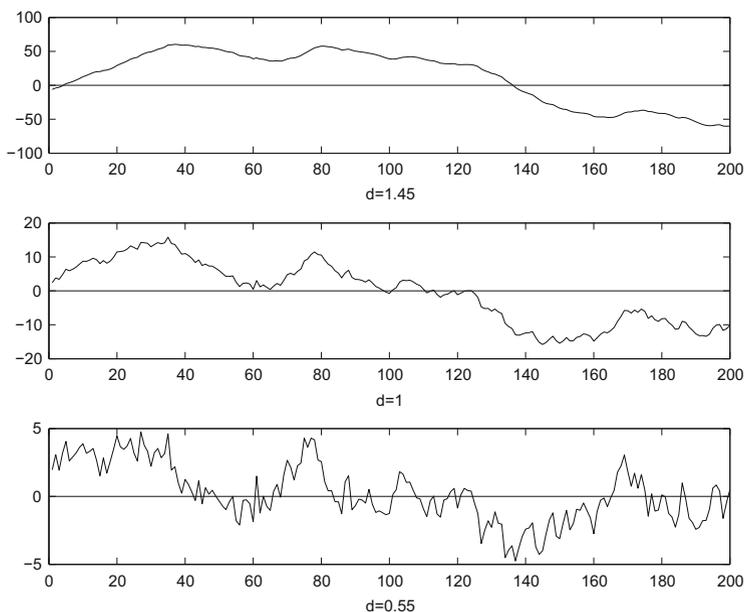


Fig. 5.6 Nonstationary fractional noise for $d = 1.45, 1.0, 0.55$

or whether a process has short memory or not,

$$H_0 : d \leq 0 \quad \text{vs.} \quad H_1 : d > 0.$$

Demetrescu, Kuzin, and Hassler (2008) suggested a simple procedure similar to the Dickey-Fuller test, to test for arbitrary values d_0 .

5.5 Problems and Solutions

Problems

5.1 Consider the p -series $\sum_{j=1}^J j^{-p}$ for $J \rightarrow \infty$. Show that the limit is finite if and only if $p > 1$, see (5.1).

5.2 Show that exponential decay is faster than hyperbolic decay, i.e.

$$\lim_{j \rightarrow \infty} \frac{g^j}{j^{d-1}} = 0 \quad \text{for } 0 < g < 1, |d| < 1.$$

5.3 Show (5.4).

Hint: Use properties of the Gamma function

$$\Gamma(x) = \begin{cases} \int_0^\infty t^{x-1} e^{-t} dt, & x > 0 \\ \infty, & x = 0 \\ \Gamma(x+1)/x, & x < 0, x \neq -1, -2, \dots \end{cases}. \quad (5.18)$$

5.4 Establish the following expression for the autocovariance $\gamma(h)$ of fractionally integrated noise:

$$\gamma(h) = \sigma^2 \frac{\Gamma(1-2d)\Gamma(d+h)}{\Gamma(d)\Gamma(1-d)\Gamma(1-d+h)}. \quad (5.19)$$

Hint: Use Proposition 4.1 with (5.11), and apply the following identity from Gradshteyn and Ryzhik (2000, 3.631.8):

$$\int_0^\pi \sin^{\nu-1}(x) \cos(ax) dx = \frac{\pi \cos(\frac{a\pi}{2}) \Gamma(\nu+1)}{2^{\nu-1} \nu \Gamma(\frac{\nu+a+1}{2}) \Gamma(\frac{\nu-a+1}{2})}, \quad \text{where } \nu > 0.$$

5.5 Show Proposition 5.1 (a).

Hint: Use (5.19).

5.6 Show Proposition 5.1 (b).

Hint: Use (5.19).

5.7 Show Proposition 5.1 (c).

5.8 Consider the ARFIMA(0,d,1) model $\Delta^d x_t = B(L) \varepsilon_t$, $B(L) = 1 + bL$. Show for this special case that the proportionality constant from Proposition 5.3 (b) is indeed $\gamma_{df_e}(0) 2\pi$.

Solutions

5.1 We define the p -series of the first J terms,

$$S_J(p) = \sum_{j=1}^J \frac{1}{j^p}.$$

First, we discuss the case separating the convergence region from the divergent one ($p = 1$). For convenience choose $J = 2^n$, such that

$$\begin{aligned} S_{2^n}(1) &= \left(1 + \frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \cdots + \left(\frac{1}{2^{n-1} + 1} + \cdots + \frac{1}{2^n}\right) \\ &> \frac{1}{2} + \frac{2}{4} + \frac{4}{8} + \cdots + \frac{2^{n-1}}{2^n} = \frac{n}{2}. \end{aligned}$$

Since the lower bound of $S_{2^n}(1)$ diverges to infinity with n , $S_J(1)$ must diverge, too.

Second, consider the case where $p < 1$, such that $j^p < j$, or $j^{-p} > j^{-1}$. Hence, $S_J(p) > S_J(1)$, and divergence of $S_J(1)$ implies divergence for $p < 1$.

Third, for $p > 1$, we group the terms for $J = 2^n - 1$ as follows.

$$\begin{aligned} S_{2^n-1}(p) &= 1 + \left(\frac{1}{2^p} + \frac{1}{3^p}\right) + \cdots + \left(\frac{1}{(2^{n-1})^p} + \cdots + \frac{1}{(2^n-1)^p}\right) \\ &< 1 + \frac{2}{2^p} + \frac{4}{4^p} + \cdots + \frac{2^{n-1}}{(2^{n-1})^p} \\ &= 1 + \frac{1}{2^{p-1}} + \frac{1}{4^{p-1}} + \cdots + \frac{1}{(2^{n-1})^{p-1}}. \end{aligned}$$

We now abbreviate $g = \frac{1}{2^{p-1}}$ with $0 < g < 1$ since $p > 1$. Consequently,

$$\begin{aligned} S_{2^n-1}(p) &< 1 + g + g^2 + \cdots + g^{n-1} = \frac{1 - g^n}{1 - g} \\ &< \sum_{i=0}^{\infty} g^i = \frac{1}{1 - g} = \frac{2^{p-1}}{2^{p-1} - 1}, \end{aligned}$$

where we use the geometric series, see Problem 3.2. Hence, $S_{2^n-1}(p)$ is bounded for every n , while growing monotonically at the same time, which establishes convergence for $p > 1$. Hence, the proof of (5.1) is complete.

We want to add a final remark. While convergence is ensured for $p > 1$, an explicit expression for the limit is by no means obvious, and indeed only known for selected values. For example, for $p = 2$ one has the famous result by Leonhard Euler:

$$\lim_{J \rightarrow \infty} S_J(2) = \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}.$$

5.2 Since the limit of the ratio of interest is indeterminate, we apply L'Hospital's rule:

$$\begin{aligned}\lim_{j \rightarrow \infty} \frac{g^j}{j^{d-1}} &= \lim_{j \rightarrow \infty} \frac{j^{1-d}}{g^{-j}} = \lim_{j \rightarrow \infty} \frac{(1-d)j^{-d}}{g^{-j} \log(g)(-1)} \\ &= \frac{d-1}{\log(g)} \lim_{j \rightarrow \infty} \frac{g^j}{j^d} = 0, \quad d \geq 0.\end{aligned}$$

If $-1 < d < 0$, the expression g^j/j^d is still indeterminate. Application of L'Hospital's rule twice, however, yields:

$$\begin{aligned}\lim_{j \rightarrow \infty} \frac{g^j}{j^d} &= \lim_{j \rightarrow \infty} \frac{j^{-d}}{g^{-j}} = \lim_{j \rightarrow \infty} \frac{-dj^{-d-1}}{g^{-j} \log(g)(-1)} \\ &= \frac{d}{\log(g)} \lim_{j \rightarrow \infty} \frac{g^j}{j^{1+d}} = 0.\end{aligned}$$

This establishes the claim.

5.3 Prior to solving the problem, we review some useful properties of the Gamma function that is often employed to simplify manipulations with binomial expressions, see e.g. Sydsæter, Strøm, and Berck (1999, p.52), and in much greater detail Gradshteyn and Ryzhik (2000, Sect. 8.31), or Rudin (1976, Ch. 8) containing proofs. For integer numbers, Γ coincides with the factorial,

$$\Gamma(n+1) = n(n-1) \cdots 2 = n!,$$

which implies a recursive relation holding in fact in general:

$$\Gamma(x+1) = x\Gamma(x). \tag{5.20}$$

Hence, obviously $\Gamma(1) = \Gamma(2) = 1$, and a further value often encountered is $\Gamma(0.5) = \sqrt{\pi}$. The recursive relation further yields the rate of divergence at the origin,

$$\Gamma(x) \sim x^{-1}, \quad x \rightarrow 0,$$

which justifies the convention $\Gamma(0)/\Gamma(0) = 1$. Finally, we want to approximate the Gamma function for large arguments. Remember Stirling's formula for factorials,

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

It generalizes to

$$\Gamma(x) \sim \sqrt{\frac{2\pi}{x}} \left(\frac{x}{e}\right)^x$$

for $x \rightarrow \infty$, which again has to be read as

$$\lim_{x \rightarrow \infty} \Gamma(x) / \sqrt{\frac{2\pi}{x}} \left(\frac{x}{e}\right)^x = 1.$$

Consequently, we have for finite x and y and large integer n that⁷

$$\frac{\Gamma(n+x)}{\Gamma(n+y)} \sim n^{x-y}, \quad n \rightarrow \infty. \quad (5.21)$$

Now, we turn to establishing (5.4). Repeated application of (5.20) gives

$$\frac{\Gamma(j-d)}{\Gamma(-d)} = (j-d-1)(j-d-2)\cdots(-d).$$

By definition of the binomial coefficients we conclude from (5.3) with $\Gamma(j+1) = j!$ that

$$\begin{aligned} \pi_j &= \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)}, \quad j \geq 0 \\ &\sim \frac{j^{-d-1}}{\Gamma(-d)}, \quad j \rightarrow \infty, \end{aligned}$$

where the approximation relies on (5.21). This is the required result.

5.4 With Proposition 4.1 and (5.11) we compute

$$\begin{aligned} \gamma(h) &= 2 \int_0^\pi f(\lambda) \cos(\lambda h) d\lambda \\ &= \frac{4^{-d}}{\pi} 2\sigma^2 \int_0^{\frac{\pi}{2}} \sin^{-2d}(x) \cos(2hx) dx, \end{aligned}$$

⁷Use $(1 + x/n)^n \rightarrow e^x$.

where we substituted $\frac{\lambda}{2} = x$. Both $\sin^2(x)$ and $\cos(2hx)$ are even symmetric around $\frac{\pi}{2}$, such that

$$\int_0^{\pi} \sin^{-2d}(x) \cos(2hx) dx = 2 \int_0^{\frac{\pi}{2}} \sin^{-2d}(x) \cos(2hx) dx.$$

Hence, the integration formula 3.631.8. from Gradshteyn and Ryzhik (2000) can be applied with $\nu = 1 - 2d > 0$ as long as $d < 0.5$ and $a = 2h$:

$$\begin{aligned} \gamma(h) &= \frac{4^{-d} \sigma^2}{\pi} \int_0^{\pi} \sin^{-2d}(x) \cos(2hx) dx \\ &= \frac{4^{-d} \sigma^2}{\pi} \frac{\pi \cos(h\pi) \Gamma(2 - 2d)}{2^{-2d} (1 - 2d) \Gamma(1 - d + h) \Gamma(1 - d - h)} \\ &= \sigma^2 \frac{(-1)^h \Gamma(2 - 2d)}{(1 - 2d) \Gamma(1 - d + h) \Gamma(1 - d - h)} \\ &= \sigma^2 \frac{(-1)^h \Gamma(1 - 2d)}{\Gamma(1 - d + h) \Gamma(1 - d - h)}. \end{aligned}$$

Using (5.20) once more, one can show that

$$\frac{\Gamma(d + h)}{\Gamma(d)} = (-1)^h \frac{\Gamma(1 - d)}{\Gamma(1 - d - h)}.$$

Therefore, we finally have

$$\gamma(h) = \sigma^2 \frac{\Gamma(1 - 2d) \Gamma(d + h)}{\Gamma(d) \Gamma(1 - d) \Gamma(1 - d + h)},$$

which is (5.19).

5.5 With (5.19) we obtain

$$\gamma(0; d) = \frac{\Gamma(1 - 2d)}{(\Gamma(1 - d))^2},$$

where we assumed $\sigma^2 = 1$ without loss of generality. Instead of the variance, we will equivalently minimize the natural logarithm thereof. We determine as derivative

$$\begin{aligned} \frac{\partial \log(\gamma(0; d))}{\partial d} &= -2 \left(\frac{\Gamma'(1 - 2d)}{\Gamma(1 - 2d)} - \frac{\Gamma'(1 - d)}{\Gamma(1 - d)} \right) \\ &= -2 (\psi(1 - 2d) - \psi(1 - d)), \end{aligned}$$

where the so-called psi function is defined as logarithmic derivative of the Gamma function,

$$\psi(x) = \frac{\partial \log(\Gamma(x))}{\partial x}.$$

The psi function is strictly increasing for $x > 0$, which can be shown in different ways, see e.g. Gradshteyn and Ryzhik (2000, 8.362.1). Consequently,

$$\psi(1-d) - \psi(1-2d) \begin{cases} > 0, & 0 < d < 0.5 \\ = 0, & d = 0 \\ < 0, & d < 0 \end{cases},$$

which proves that $\log(\gamma(0; d))$, and hence $\gamma(0; d)$, takes on its minimum for $d = 0$. This solves the problem.

5.6 The recursive relative for $\gamma(h)$ is obvious with (5.19) and (5.20) at hand.

For $h \rightarrow \infty$, the approximation in (5.21) yields

$$\gamma(h) \sim \sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} h^{2d-1},$$

which defines the constant from Proposition 5.1 (b):

$$\gamma_d = \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)}.$$

The Gamma function is positive for positive arguments and negative on the interval $(-1, 0)$. Hence, the sign of γ_d equals the sign of $\Gamma(d)$ since $d < 0.5$, which completes the proof of Proposition 5.1 (b).

5.7 In terms of autocorrelations the recursion from Proposition 5.1 (b) becomes

$$\rho(h; d) = f(h; d) \rho(h-1; d), \quad h \geq 1,$$

where the factor $f(h; d)$ is positive for $d > 0$,

$$f(h; d) = \frac{h-1+d}{h-d} > 0 \quad \text{with} \quad \frac{\partial f(h; d)}{\partial d} > 0,$$

such that $\rho(h; d) > 0$ since $\rho(0; d) = 1$. Hence, we have

$$\frac{\partial \rho(h; d)}{\partial d} = \frac{\partial f(h; d)}{\partial d} \rho(h-1; d) + f(h; d) \frac{\partial \rho(h-1; d)}{\partial d},$$

which is positive by induction since

$$\frac{\partial \rho(1; d)}{\partial d} = \frac{\partial f(1; d)}{\partial d} > 0.$$

Hence, $\rho(h; d)$ is growing with d as required.

5.8 The spectral density of the MA(1) component $e_t = \varepsilon_t + b\varepsilon_{t-1}$ is known from Example 4.3:

$$f_e(0) = (1 + b)^2 \sigma^2 / 2\pi.$$

We now express x_t in terms of a fractional noise called $y_t = \Delta^{-d} \varepsilon_t$:

$$\begin{aligned} x_t &= \Delta^{-d} B(L) \varepsilon_t = B(L) \Delta^{-d} \varepsilon_t \\ &= (1 + bL) y_t. \end{aligned}$$

Let $\gamma_x(h)$ and $\gamma_y(h)$ denote the autocovariances of $\{x_t\}$ and $\{y_t\}$, respectively. It holds that

$$\begin{aligned} \gamma_x(h) &= E[(y_t + by_{t-1})(y_{t+h} + by_{t+h-1})] \\ &= (1 + b^2)\gamma_y(h) + b(\gamma_y(h-1) + \gamma_y(h+1)). \end{aligned}$$

With the behavior of $\gamma_y(h)$ from Proposition 5.1 it follows

$$\begin{aligned} \frac{\gamma_x(h)}{h^{2d-1}} &\rightarrow (1 + b^2) \gamma_d \sigma^2 + 2b \gamma_d \sigma^2 \\ &= \gamma_d f_e(0) 2\pi, \end{aligned}$$

as required.

References

- Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73, 5–59.
- Bondon, P., & Palma, W. (2007). A class of antipersistent processes. *Journal of Time Series Analysis*, 28, 261–273.
- Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods* (2nd ed.). New York: Springer.
- Demetrescu, M., Kuzin, V., & Hassler, U. (2008). Long memory testing in the time domain. *Econometric Theory*, 24, 176–215.
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74, 427–431.
- Giraitis, L., Koul, H. L., & Surgailis, D. (2012). *Large sample inference for long memory processes*. London: Imperial College Press.

- Gradshteyn, I. S., & Ryzhik, I. M. (2000). *Table of integrals, series, and products* (6th ed.). London/San Diego: Academic Press.
- Granger, C. W. J. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, *16*, 121–130.
- Granger, C. W. J., & Joyeux, R. (1980). An Introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*, 15–29.
- Hassler, U. (2012). Impulse responses of antipersistent processes. *Economics Letters*, *116*, 454–456.
- Hassler, U. (2014). Persistence under temporal aggregation and differencing. *Economics Letters*, *124*, 318–322.
- Hassler, U., & Kokoszka (2010). Impulse responses of fractionally integrated processes with long memory. *Econometric Theory*, *26*, 1855–1861.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, *68*, 165–176.
- Maasoumi, E., & McAleer, M. (2008). Realized volatility and long memory: An overview. *Econometric Reviews*, *27*, 1–9.
- Rudin, W. (1976). *Principles of mathematical analysis* (3rd ed.). New York: McGraw-Hill.
- Sydsæter, K., Strøm, A., & Berck, P. (1999). *Economists' mathematical manual* (3rd ed.). Berlin/New York: Springer.
- Trench, W. F. (2013). *Introduction to real analysis*. Free Hyperlinked Edition 2.04 December 2013. Downloaded on 10th May 2014 from <http://digitalcommons.trinity.edu/mono/7>.