1

Introduction

1.1 Empirical Examples

Figure 1.1 displays 663 annual observations of minimal water levels of the Nile river. This historical data is from Beran (1994, Sect. 12.2) and ranges from the year 622 until 1284. The second panel contains the sample autocorrelations $\hat{\rho}(h)$ at lag $h \in \{1, 2, \dots, 30\}$. The maximum value, $\hat{\rho}(1) = 0.57$, is not particularly large, but the autocorrelogram dies out only very slowly with $\hat{\rho}(30) = 0.15$ still being significantly positive. Such a slowly declining autocorrelogram is characteristic of what we will define as long memory or strong persistence. It reflects that the series exhibits a very persistent behavior in that we observe very long cyclical movements or (reversing) trends. Note, e.g. that from the year 737 until 805, there are only three data points above the sample average (=11.48), i.e. there are seven decades of data below the average. Then the series moves above the average for a couple of years, only to swing down below the sample mean for another 20 years from the year 826 on. Similarly, there is a long upward trend from 1060 on until about 1125, followed again by a long-lasting decline. Such irregular cycles or trends due to long-range dependence, or persistence, have first been discovered and discussed by Hurst, a British engineer who worked as hydrologist on the Nile river; see in particular Hurst (1951). Mandelbrot and Wallis (1968) coined the term *Joseph effect* for such a feature; see also Mandelbrot (1969). This alludes to the biblical seven years of great abundance followed by seven years of famine, only that *cycles* in Figure 1.1 do not have a period of seven years, not even a constant period.

Long memory in the sense of strong temporal dependence as it is obvious in Figure 1.1 has been reported in many fields of science. Hipel and McLeod (1994, Section 11.5) detected long memory in hydrological or meteorological series like annual average rainfall, temperature, and again river flow data; see also Montanari (2003) for a survey. A further technical area beyond geophysics with long memory time series is the field of data network traffic in computing; see Willinger et al. (2003).

Time Series Analysis with Long Memory in View, First Edition. Uwe Hassler.

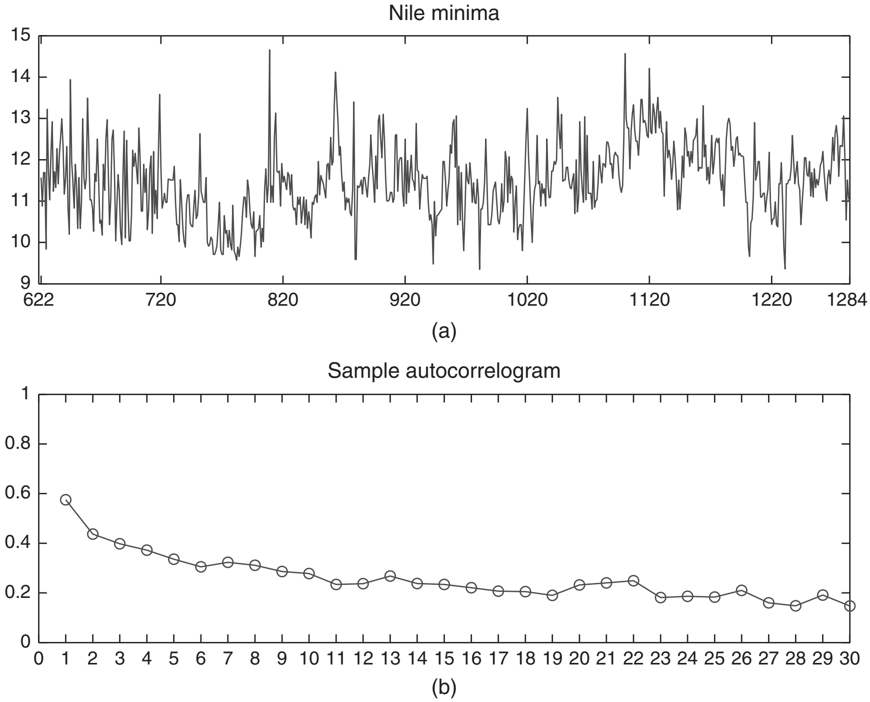

Figure 1.1 Annual minimal water levels of the Nile river.

The second data set that we look into is from political science. Let p_t denote the poll data on partisanship, i.e. the voting intention measured by monthly opinion polls in England. More precisely, p_t is the portion of people supporting the Labor Party. The sample ranges from September 1960 until October 1996 and has been analyzed by Byers et al. (1997).[1] Figure 1.2 contains the logit transformation of this poll data,

$$y_t = \ln\left(\frac{p_t}{1-p_t}\right),$$

such that $y_t = 0$ for $p_t = 50\%$; here $\ln(x)$ stands for the natural logarithm of x. We observe long-lasting upswings followed by downswings amounting to a pseudocyclical pattern or reversing trends. This is well reflected and quantified by the sample autocorrelations in the lower panel, decreasing from $\hat{\rho}(1) = 0.9$ quite slowly to $\hat{\rho}(24) \approx 0.2$. Independently of Byers et al. (1997), Box-Steffensmeier and Smith (1996) detected long memory in US opinion poll data on partisanship. Long memory in political popularity has been confirmed

1 We downloaded the data from James Davidson's homepage on May 5, 2016. The link is http://people.exeter.ac.uk/jehd201/bdpdata.txt.

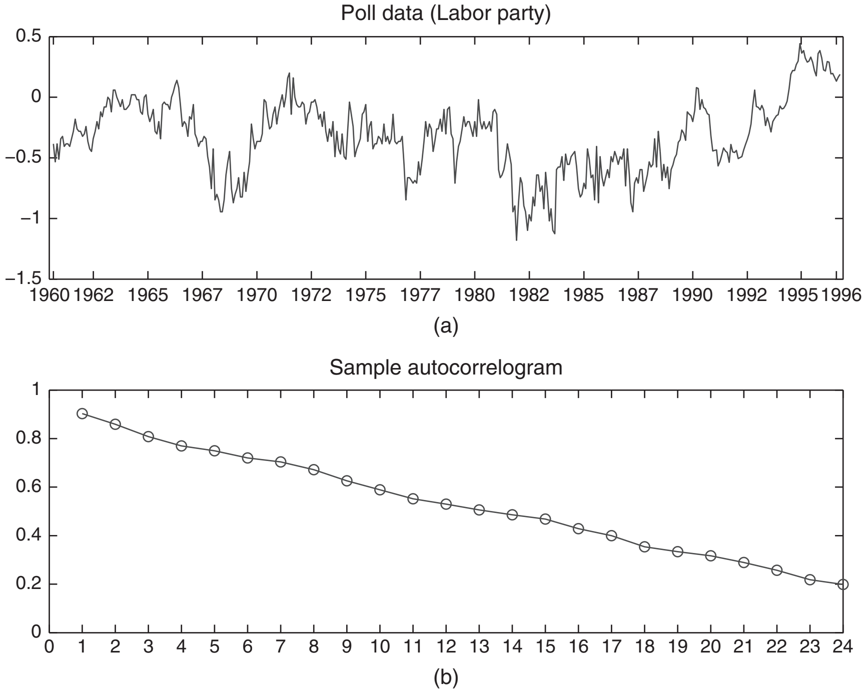

Figure 1.2 Monthly opinion poll in England, 1960–1996.

in a sequence of papers; see Byers et al. (2000, 2007), and Dolado et al. (2003); see also Byers et al. (2002) for theoretical underpinning of long memory in political popularity. Further evidence on long memory in political science has been presented by Box-Steffensmeier and Tomlinson (2000); see also the special issue of *Electoral Studies* edited by Lebo and Clarke (2000).

Since Granger and Joyeux (1980), the fractionally integrated autoregressive moving average (ARMA) model gained increasing popularity in economics. The empirical example in Granger and Joyeux (1980) was the monthly US index of consumer food prices. Granger (1980) had shown theoretically how the aggregation of a large number of individual series may result in an index that is fractionally integrated, which provided theoretical grounds for long memory as modeled by fractional integration in price indices. A more systematic analysis by Geweke and Porter-Hudak (1983) revealed long memory in different US price indices. These early papers triggered empirical research in long memory in inflation rates in independent work by Delgado and Robinson (1994) for Spain and by Hassler and Wolters (1995) and Baillie et al. (1996) for international evidence. Since then, there has been offered abundant evidence in favor of long memory in inflation rates; see, e.g. Franses and Ooms (1997),

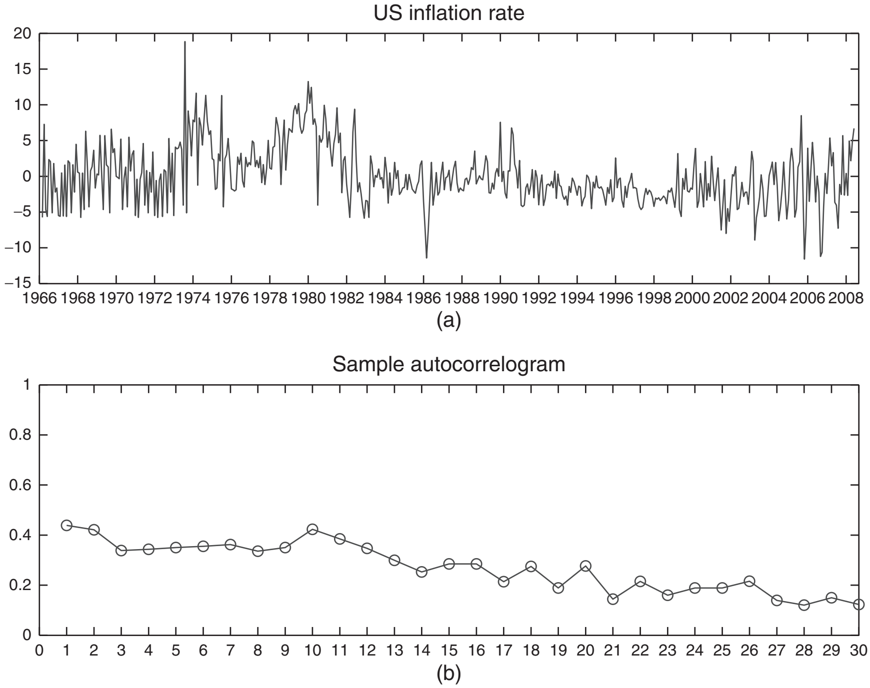

Figure 1.3 Monthly US inflation, 1966–2008.

Baum et al. (1999), Franses et al. (1999), Hsu (2005), Kumar and Okimoto (2007), Martins and Rodrigues (2014), and Hassler and Meller (2014), where the more recent research focused on breaks in persistence, i.e. in the order of fractional integration. For an early survey article on further applications in economics, see Baillie (1996).

Figure 1.3 gives a flavor of the memory in US inflation. The seasonally adjusted and demeaned data from January 1966 until June 2008 has been analyzed by Hassler and Meller (2014). The autocorrelations fall from $\hat{\rho}(1) = 0.44$ to a minimum of $0.12 = \min\{\hat{\rho}(h)\}$, $h = 1, 2, \ldots, 30$. Again, this slowly declining autocorrelogram mirrors the reversing trends in inflation, although Hassler and Meller (2014) suggested that the persistence may be superimposed by additional features like time-varying variance.

The fourth empirical example is from the field of finance. Figure 1.4 displays daily observations from January 4, 1993, until May 31, 2007. This sample of 3630 days consists of the logarithm of realized volatility of International Business Machines Corporation (IBM) returns computed from underlying five-minutes data; see Hassler et al. (2016) for details. Although the dynamics of the series

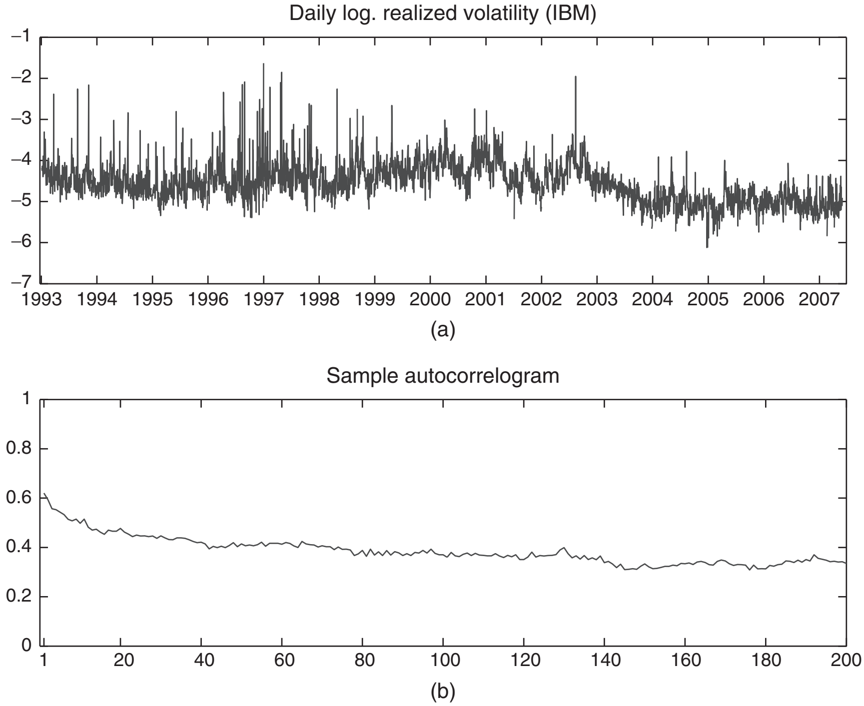

Figure 1.4 Daily realized volatility, 1993–2007.

is partly masked by extreme observations, one clearly may distinguish periods of weeks where the data tends to increase, followed by long time spans of decrease. The high degree of persistence becomes more obvious when looking at the sample autocorrelogram. Starting off with $\hat{\rho}(1) = 0.62$, the decline is extremely slow with $\hat{\rho}(400)$ still being well above 0.2. Long memory in realized volatility is sometimes considered to be a stylized fact since the papers by Andersen et al. (2001, 2003). Such a view is supported by the special issue in *Econometric Reviews* edited by Maasoumi and McAleer (2008).

Finally, with the last example we return to economics. Figure 1.5 shows 435 monthly observations from 1972 until 2008. The series is the logarithm of seasonally adjusted US unemployment rates (number of unemployed persons as a percentage of the civilian labor force); see Hassler and Wolters (2009) for details. The sample average of log-unemployment is 1.7926; compare the straight line in the upper panel of Figure 1.5. Here, the trending behavior is so strong that the sample average is crossed only eight times over the period of 35 years. The deviations from the average are very pronounced and very long relative to the sample size. In that sense the series from Figure 1.5 seems to be most persistent of all the five examples considered in this introduction.

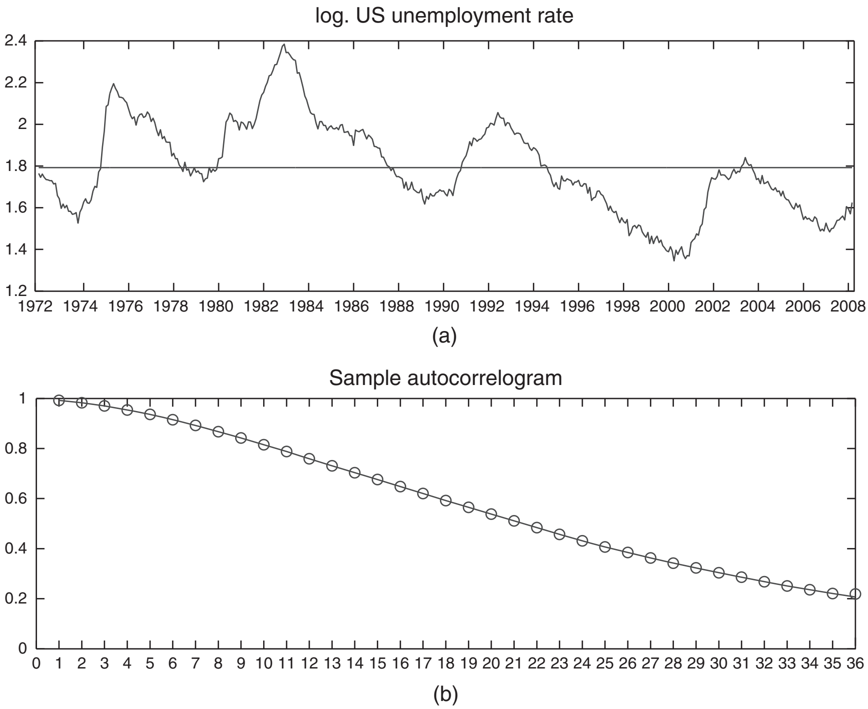

Figure 1.5 Monthly unemployment rate, 1972–2008.

This is also expressed by the sample autocorrelogram virtually beginning at one ($\hat{\rho}(1) = 0.992$) and $\hat{\rho}(h) > 0.2$ for $h \in \{1, 2, \dots, 36\}$. What is more, the autocorrelations decline almost linearly in h, which is indicative of an $I(1)$ process or an $I(d)$ process with even $d > 1$; see Hassler (1997, Corollary 3) and Section 7.5. Hence, the log-unemployment data seems to be most persistent, or most strongly trending, among our empirical examples.

1.2 Overview

There are two natural approaches to long memory modeling by fractional integration. The first one takes the nonstationary $I(1)$ model as starting point, i.e. processes integrated of order 1. Such processes are often labeled as unit root processes in econometrics, where they play a major role within the cointegration framework; see, for instance, Johansen (1995), Lütkepohl (2005), or Pesaran (2015). The extension from the $I(1)$ model to the more general $I(d)$ model might be considered as a nearby approach from an econometric point of view. The second approach starts off with the classical stationary time series model, where the moving average coefficients from the Wold decomposition are assumed to be absolutely summable and to sum to a value different from 0. For this model, which may be called integrated of order 0, $I(0)$ (see Chapter 6), it holds true that the scaled sample average converges with the square root of the sample size to a nondegenerate normal distribution. This $I(0)$ model underlying the major body of time series books from Anderson (1971) over Brockwell and Davis (1991) and Hamilton (1994) to Fuller (1996) may be generalized to the stationary $I(d)$ process for $d < 1/2$. The latter can be further extended to the region of nonstationarity ($d \geq 1/2$). Here, we follow this second route starting with the $I(0)$ case. More precisely, the outline of the book is as follows.

A definition of stationarity of stochastic processes is given in the next chapter. Moreover, Chapter 2 contains a discussion of ergodicity that corrects expositions found in some books (see Example 2.2). Next, we show that a familiar sufficient condition for ergodicity in the mean (defined in Definition 2.3) is also necessary; see Proposition 2.2. Then we distinguish between (short and long) memory (Definition 2.4) and different degrees of persistence on statistical grounds: Short memory is separated from long memory to characterize under what circumstances the variance of the sample average is of order $1/T$, where T denotes the sample size; see Proposition 2.3. Persistence is defined (Definition 2.5) to characterize the absence or presence and strength of a trend component in a process; see also Eq. (4.22).

Chapter 3 focuses on moving average processes of infinite order, sometimes called linear processes. This is motivated by Wold's theorem in Section 3.2. We thus have a unified framework to embed the classical process of moderate persistence as well as processes with antipersistence or strong persistence, which may or may not display long memory at the same time. The discussion of memory vs. persistence is picked up again in Section 3.3. The discussion of Examples 3.2 through 3.5 shows that the series from Figures 1.1 to 1.5 display both long memory and strong persistence, which motivates the model of fractional integration in Chapter 6. Before leaving Chapter 3, we provide some interesting results on the summability of the classical ARMA process (Proposition 3.5) established with a sequence of technical lemmata.

Chapter 4 introduces to the frequency domain where much of the long memory analysis is settled. The frequency domain is not only useful for data analysis, but it also allows for a deeper theoretical study. For instance, the classical concept of invertibility can be recast following Bloomfield (1985) and Bondon and Palma (2007) in a way (Proposition 4.6) that extends the region of invertibility of fractionally integrated processes; see Proposition 6.2. Next, we introduce the so-called exponential model formulated in the frequency domain. This exponential model is typically not treated in time series books, although it is particularly convenient in the context of long memory as modeled by fractional integration. Similarly, time series books typically do not deal with so-called

Whittle estimation, which is a frequency domain approximation to maximum likelihood that we present in Section 4.6, thus laying the foundation for memory estimation in Chapters 8 and 9.

Chapter 5 opens the route to fractional integration. It is a short chapter on the fractional difference and integration operator, respectively. We provide four technical lemmata that will be used repeatedly in subsequent chapters. Chapter 6 defines the stationary fractionally integrated process (of type I), building on a precise definition of $I(0)$ processes; see Assumption 6.2. Conditions for (different degrees of) persistence follow under minimal restrictions from Lemma 5.4, while Proposition 6.1 translates this into the frequency domain. Corollary 6.1 and Proposition 6.3 reflect the persistence as (short or long) memory in the time domain. After a discussion of parametric fractionally integrated models in Section 6.2, two different types of nonstationarity are discussed in Section 6.3: First, type II fractionally integrated processes are only asymptotically stationary if $d < 1/2$. Second, the case $d \geq 1/2$ covers nonstationarity for both type I and type II processes. Proposition 6.6 shows that classical parametric models imply frequency domain assumptions often entertained in the literature. For the rest of the book, we assume the fractionally integrated models as introduced in Chapter 6.

Chapter 7 sets off with what seems to be the most general central limit theorem currently available for moving average processes. It is applied to the sample average of fractionally integrated processes, closing in particular the gap at $d = -1/2$ in the literature; see Corollary 7.1. Section 7.3 extends the central limit theorem to a functional central limit theory, where fractional Brownian motions show up in the limit. Two seemingly different representations of type II fractional Brownian motion are shown to be identical in Lemma 7.2. Finally, this chapter contains in Section 7.5 an exposition on the behavior of the sample autocorrelations under fractional integration.

The eighth chapter is dedicated to the estimation of all other parameters except for the mean, assuming a fully parametric model of fractional integration. Theorem 8.1 gives the general structure of the limiting covariance matrix of the asymptotic normal distribution for (different approximations to) maximum likelihood, while Corollary 8.1 focuses in particular on the integration parameter d. Approximations to maximum likelihood may be settled in the time domain (Proposition 8.2) or in the frequency domain (Whittle estimation, Proposition 8.3). In particular, we find that the nonstationarity-extended Whittle estimator (Proposition 8.4) overcomes all pitfalls of exact maximum likelihood, except for being parametric of course. Section 8.5 paves the way to semiparametric estimation in that it studies the log-periodogram regression in the presence of a so-called exponential model for the short memory component. While consistency is established in Proposition 8.5, we learn that the estimator is less efficient than corresponding estimators rooted in the maximum likelihood principle.

Chapter 9 begins with the already familiar log-periodogram regression, however, now in the presence of short memory, which is not parametrically modeled. The whole chapter is dedicated to procedures that are semiparametric in the sense that they are robust with respect to short memory. This comes at the price of reduced efficiency. Indeed, we obtain a slower rate of convergence compared with parametric estimators. Within the class of semiparametric estimators, there exist differences in efficiency, too, and the local Whittle estimator (Proposition 9.4 or 9.5), respectively its versions allowing for nonstationarity (Proposition 9.6 or 9.7), turn out to be superior.

Since semiparametric estimators are burdened with large variances, it is interesting to have powerful tests that allow to discriminate statistically, e.g. between short memory and long memory or between stationarity and nonstationarity. This issue is addressed in Chapter 10. The first test builds on a classical rescaled range analysis that can be traced back to Hurst (1951). It has been improved by the rescaled variance test that is designed to provide a better balance of power and size in finite samples. A different approach is adopted in Section 10.4 dedicated to Lagrange multiplier (LM) tests. In Section 10.6, the original LM test is recast in a convenient lag-augmented regression framework (Proposition 10.8), and it has the nice property of robustness against conditional and even unconditional heteroskedasticity (Proposition 10.9). At the same time it is asymptotically most powerful against local alternatives.

Long memory is a still rapidly growing field of applied and theoretical research. Therefore, we close the book with a collection of further topics in the final chapter.

All chapters contain a final section called "Technical Appendix: Proofs" (except for the last chapter "Further Topics"). There, we give the mathematical proofs of results provided and discussed in the main text. Some proofs just accomplish or spell out simple steps to adapt proofs from the literature to our context. Other proofs are truly original in that they establish new results that cannot be drawn from the literature. By separating the proofs from the propositions in the main text, we hope to improve the readability of the book. Finally, it should be stressed that the book is not fully self-contained. While in some propositions we spell out all required assumptions, there are many cases where we refer to the literature. For brevity and convenience one finds in the latter case formulations like "…satisfying Assumption 6.3 […] and some further restrictions by Robinson (1995b);" such that the reader is expected to read up details from the provided reference, namely, Robinson (1995b), for this example from Proposition 9.4.