EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 10, No. 1, 2017, 82-103
ISSN 1307-5543 – www.ejpam.com
Published by New York Business Global

Sir Clive W.J. Granger Memorial Special Issue on Econometrics

Long Memory and Fractional Differencing: Revisiting Clive W. J. Granger's Contributions and Further Developments

Ryoko Ito

Nuffield College and the Department of Economics, Oxford University, Oxford, UK.

Abstract. In 1980, Sir Clive W. J. Granger discovered the fractional differencing operator and its fundamental properties in discrete-time mathematics, which sparked an enormous literature concerning the fractionally integrated autoregressive moving average models. Fractionally integrated models capture a type of long memory and have useful theoretical properties, although scientists can find them difficult to estimate or intuitively interpret. His introductory papers from 1980, one of which with Roselyne Joyeux, show his early and deep understanding of this subject by showing that familiar short memory processes can produce long memory effects under certain conditions. Moreover, fractional differencing advanced our understanding of cointegration and the properties of traditional Dickey-Fuller tests, and motivated the development of new unit-root tests against fractional alternatives. This article honors his significant contributions by identifying key areas of research he inspired and surveying recent developments in them.

2010 Mathematics Subject Classifications: 62M10, 62M15, 62M20, 62P20, 91B84, 37M10
Key Words and Phrases: Cointegration, long-range dependence, unit-root, frequency domain

1. Introduction to Fractional Differencing

A stationary process is said to have *long memory* (or *long-range dependence*) if its autocorrelation decays slower than the exponential rate. Hurst [76] noticed this phenomenon in the streamflow of the Nile, when he was studying the storage capacity of reservoirs along the river. His work resulted in a measure of long-range dependence called the Hurst exponent and the development of certain long memory processes in continuous-time mathematics by Mandelbrot and Van Ness [96] and Mandelbrot [95]. Also see Lawrence and Kottegoda [87] and Hipel and McLeod [73].

Email address: ryoko.ito@economics.ox.ac.uk (R. Ito)

Long memory processes have been documented in many scientific fields. Finance and economics are no exceptions: Ding et al. [46] and Bollerslev and Mikkelsen [19], for instance, illustrated that stock market volatility can be described by a long memory process. Chortareas et al. [27] made similar findings in the volatility of chosen exchange rates with the Euro. Fleming and Kirby [52] found that both return volatility and trading volume of a selected set of twenty equities display long memory. Degiannakis et al. [39], Caporale and Gil-Alana [22], and many others found more empirical evidence of long memory in finance and economics.

When scientists apply standard regression frameworks to nonstationary time series, it is common to difference the series first to remove a unit root. A popular differencing filter is the n-th difference filter, $(1 - L^n)$, where $n \in \mathbb{N}$ and L is the lag operator. Hylleberg and Granger [77] developed tests for roots of this type. This filter is one of the simplest methods of adjusting for either unit roots or seasonality, and is frequently viewed as a practical response to time series with persistence that is incompatible with the standard autoregressive moving average (ARMA) models. From the viewpoint of Granger and Joyeux [65], a major concern here was that this form of differencing can lead to over-differencing, in the sense that it can distort interesting short-run dynamics and lead to misleading conclusions, just like any method of de-trending or de-seasonalizing. The *fractional differencing* operator formally introduced in the ARMA context by Granger [60], Granger and Joyeux [65], and Hosking [74] deepened our understanding of the concept of differencing, and allowed scientists to model long memory processes using the language of discrete-time mathematics.

A process y_t is said to be integrated of order d (denoted by $I(d)$) if the differenced process, $(1 - L)^d y_t$, is stationary. $d = 0$ if y_t is an ARMA process and $d = 1$ if y_t has a unit-root. In fractional differencing, d can be a fraction. Then we may say that y_t is *fractionally integrated* (or, simply, *fractional*). One example is the autoregressive fractionally integrated moving average (ARFIMA) model of the following form:

$$\phi(L)(1 - L)^d y_t = \theta(L)\varepsilon_t, \tag{1}$$

where $t \in \mathbb{Z}$, ε_t is white noise, and $\phi(z)$ and $\theta(z)$ are finite-order polynomials. Granger and Joyeux [65] and Hosking [74] showed that $d \in (-1/2, 1/2)$ is required for this process to be stationary and invertible. If d in (1) is greater than one, d can be re-centered to be inside the unit-circle by applying $(1 - L)$ a sufficient number of times to y_t.

The autocorrelation functions of fractionally integrated processes decay at a hyperbolic rate for $d \in (0, 1/2)$. If we write $S_T = \sum_{t=1}^{T} y_t$, it can be shown that

$$\lim_{T\to\infty} T^{-(2d+1)}\mathrm{Var}(S_T) = \text{a positive constant}. \tag{2}$$

This means that $\ln(\mathrm{Var}(S_T))$ should be a linear function of $\ln(T)$ when T is large. As an example, Figure 1 shows the daily absolute logarithmic return series, denoted by $|r_t|$, of the S&P500 index,* where the logarithmic returns are defined by $r_t = 100(\ln(p_t) - \ln(p_{t-1}))$

*The data source is Yahoo! UK & Ireland Finance (uk.finance.yahoo.com).

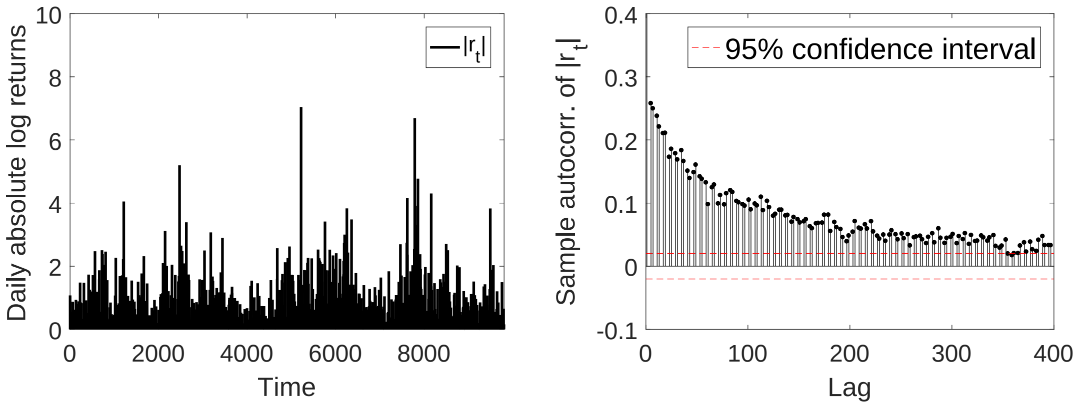


Figure 1: Left: $|r_t|$ for $t \in \mathbb{N}$, where r_t is the log-return series of S&P500. Right: the sample autocorrelation of $|r_t|$. The sample period is between 4 January 1978 and 18 November 2016. The 2476th observation (which corresponds to 19 October 1987) of $|r_t|$ was removed from the analysis since it was 23.0% and appeared to be an outlier.

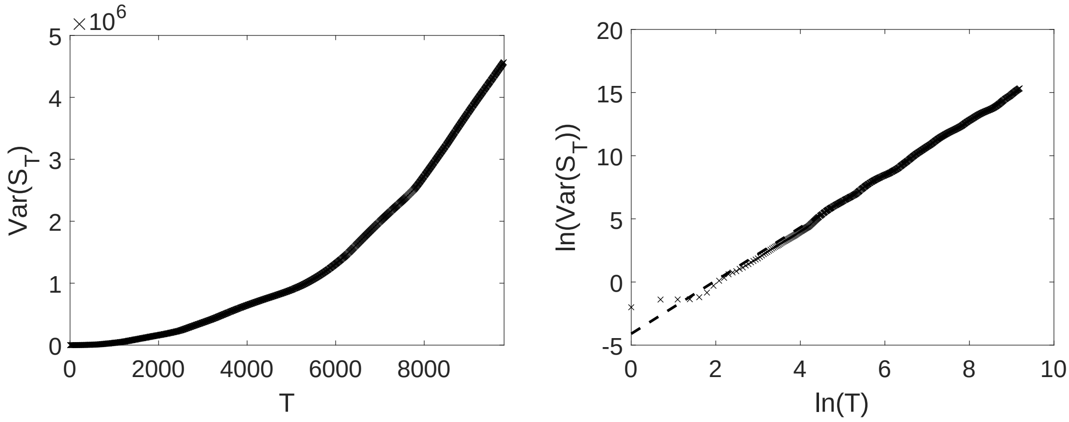


Figure 2: Left: $\mathrm{Var}(S_T)$ as T becomes large, where $S_T = \sum_{t=1}^{T} |r_t|$ and r_t is the log-return series of S&P500 displayed in Figure 1. Right: $\ln(T)$ against $\ln(\mathrm{Var}(S_T))$. The variance is computed using a common sample variance formula.

and p_t is the level of the index at time $t \in \mathbb{N}$. The sample autocorrelation of $|r_t|$ decays very slowly. Figure 2 gives the appearance of a linear relationship between $\ln(\mathrm{Var}(S_T))$ and $\ln(T)$ for large T. Hence one might speculate that the absolute logarithmic returns of S&P500 could have long memory. Areal and Taylor [9] gave a similar example. A graphical illustration of this kind, where both axes are in the logarithmic scale, should be interpreted with caution because the logarithmic transformation can obfuscate the behavior of the argument when it is large. Hence, although these pictures might tempt us to view (2) as a favorable description in this case, formal estimation and testing procedures are needed to conclude that there is long memory.

The fractional differencing operator, $(1 - L)^d$, introduces a type of long memory in a

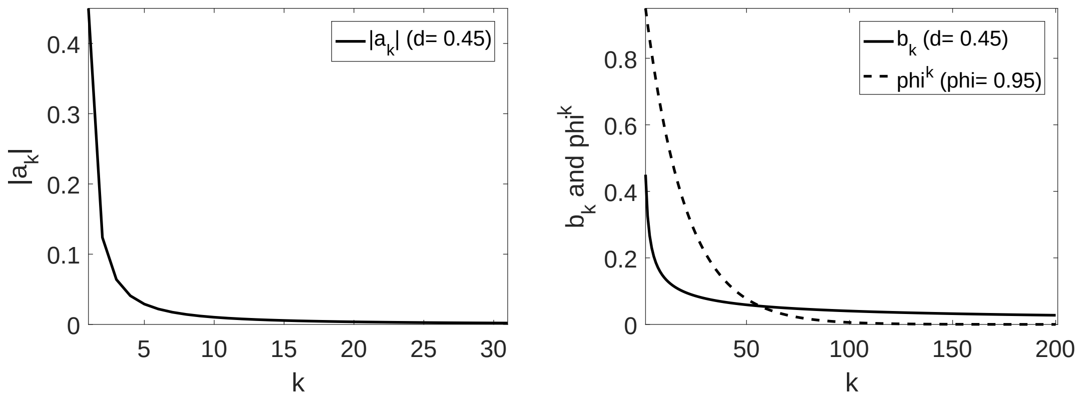


Figure 3: $|a_k|$ and b_k when $d = 0.45$, compared with ϕ^k when $\phi = 0.95$.

model because it translates to an infinite number of very slowly decaying coefficients on the lags of the variable to which it is applied. To see this, note that one version of the Binomial formula permits the following expansion:

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)} L^k = \sum_{k=0}^{\infty} a_k L^k, \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function, $a_k \equiv \prod_{j=0}^{k-1}(j-d)/(j+1)$, and $\prod_0^{-1} \cdot = 1$. In an infinite moving average representation, the following expansion may be used:

$$(1-L)^{-d} = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} L^k = \sum_{k=0}^{\infty} b_k L^k, \tag{4}$$

where $b_k \equiv \prod_{j=0}^{k-1}(j+d)/(j+1)$. Although $|a_k| \to 0$ and $|b_k| \to 0$ as $k \to \infty$, the speed of convergence can be very slow. $(b_k)_{k\in\mathbb{N}}$ are the weights on $(\varepsilon_{t-k})_{k\in\mathbb{N}}$ that can be very slow to decay, and they induce long-lasting effects of shocks on today's outcome. See Figure 3. These features illustrate how d governs the long-term behavior of the process. (Also see Hosking [75].) But these features also make fractionally integrated models a challenge to estimate, as we discuss in Section 2.

2. Estimating Fractionally Integrated Models

It is common to estimate fractionally integrated models in two stages by first estimating d to fractionally difference (or filter) the series, and then use the filtered series to estimate the short memory (ARMA) dynamics. See, for instance, Harvey [67] for discussions. A famous example we outline below was by Geweke and Porter-Hudak [56], who proposed to estimate d in the frequency domain. Consider the process,

$$(1-L)^d y_t = \varepsilon_t \tag{5}$$

for $t \in \mathbb{Z}$. Suppose we observe $y_1, \ldots, y_T$ for some $T \in \mathbb{N}$. Here, ε_t is a stationary and invertible process with the power spectrum, $f_\varepsilon(\lambda)$, where $\lambda \in [0, \pi]$ is frequency. λ in the negative range is not considered since the power spectrum is symmetric around zero. The power spectrum of y_t, denoted by $f_y(\lambda)$, is

$$f_y(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_\varepsilon(\lambda) = 2^{-d}(1 - \cos\lambda)^{-d} f_\varepsilon(\lambda) = 4^{-d}\left(\sin^2\frac{\lambda}{2}\right)^{-d} f_\varepsilon(\lambda).$$

Taking the logarithm of this equation and rearranging both sides using the sample spectral density,[†] $I(\lambda)$, gives

$$\log I(\lambda) = \log f_\varepsilon(0) - d \log\left[4\sin^2\left(\frac{\lambda}{2}\right)\right] + \log\left(\frac{f_\varepsilon(\lambda)}{f_\varepsilon(0)}\right) + \log\left[\frac{I(\lambda)}{f_y(\lambda)}\right].$$

Suppose the last term can be considered to be negligible. Then based on this equation, we have a regression model:

$$\log I(\lambda_j) = \beta_0 + \beta_1 \log\left[4\sin^2\left(\frac{\lambda_j}{2}\right)\right] + v_j, \tag{6}$$

where $v_j \equiv \log\left(f_\varepsilon(\lambda_j)/f_\varepsilon(0)\right)$ is assumed to be independently and identically distributed (i.i.d.) with zero mean and variance, $\pi^2/6$. We typically set $\lambda_j = 2\pi j/T$ for $j = 0, \ldots, m$ and $m \ll T$ such that $\lambda_j \in [0, \pi]$ for each j. We denote estimated quantities by $\widehat{\cdot}$. Regressing $\log I(\lambda_j)$ on $\log[4\sin^2(\lambda_j/2)]$ and taking minus $\widehat{\beta}_1$ yields $\widehat{d}$. For some appropriate choice of m, we have

$$(\widehat{d} - d)/\sqrt{\mathrm{Var}(\widehat{d}\,)} \xrightarrow{d} N(0, 1),$$

where $\mathrm{Var}(\widehat{d}\,)$ is obtained from the regression.

It is common to use a rule-of-thumb that sets m as a simple function of the sample size, T. See Geweke and Porter-Hudak [56] and Robinson [109] for consistency and asymptotic normality of $\widehat{d}$ and the conditions m needs to satisfy to get these results. The choice of m also affects the bias-variance trade-off. Bias in $\widehat{d}$ can be large if m is too large. A popular choice is $m = T^\delta$, where δ is between 0.5 and 0.8. Diebold and Rudebusch [44]

[†] The sample spectral density is given by

$$I(\lambda) = \frac{1}{2\pi}\left[c(0) + 2\sum_{\tau=1}^{T-1} c(\tau)\cos(\lambda\tau)\right],$$

where $c(\tau)$ is the sample autocovariance formula given by

$$c(\tau) = \frac{1}{T}\sum_{t=\tau+1}^{T}(y_t - \bar{y})(y_{t-\tau} - \bar{y}), \qquad \tau = 1, 2, \ldots,$$

with $\bar{y} = T^{-1}\sum_{t=1}^{T} y_t$. The sample spectrum computes $I(\lambda)$ at chosen frequencies, denoted by λ_j for $j = 1, \ldots, m$. Although λ_j can be any number in $[0, \pi]$, it is typically computed at $\lambda_j = 2\pi j/T$ for $j = 1, \ldots, \lfloor T/2 \rfloor$.

used $m = \lfloor T^{1/2} \rfloor$. Whether $\widehat{d}$ is a good estimator also depends on the serial dependence of the short memory process, ε_t. If it is very persistent, (6) may hold only for λ_j near zero. Alternative estimators of d can be also sensitive to the degree of autocorrelation in the fractionally differenced series, $(1-L)^d y_t$. See Agiakloglou et al. [3].

2.1. How to Treat Initial Values

The above estimation procedure handles the fractional difference operator in the frequency domain. In fact, it is common in this literature to transform a problem defined in the time domain to a problem in the frequency domain. For instance, the harmonized representations defined in the frequency domain of the ARFIMA models are often used instead. Davidson and Hashimzade [36] would describe (5) a *one-sided* (or *causal*) model that depends on lags, and its harmonized representation a *two-sided* model that depends on both lags and leads. These authors cautioned against viewing the two versions as being equivalent because their limiting properties in the time domain and the frequency domain can be different.

The need to convert the domain from the time domain to the frequency domain stems from the long lag structure of fractionally integrated models we noted at the end of Section 1. Although the following representation of (5) is permitted:

$$y_t = (1-L)^{-d}\varepsilon_t = \sum_{k=0}^{\infty} b_k \varepsilon_{t-k}, \tag{7}$$

for $d \in (0, 1/2)$, computational limitations might mean that we choose to simulate from

$$y_t^* = (1-L)^{-d}\varepsilon_t \mathbf{1}_{\{t>0\}} = \sum_{k=0}^{t-1} b_k \varepsilon_{t-k}, \tag{8}$$

where the finite sum on the right hand side is well-defined for any $d > 0$. (7) is a fractional model of *type I*, which is a function of an infinite number of lags. (8) is a *type II* model, which truncates lags at some point. Since fractional models of type I tend to be computationally demanding to simulate from and estimate, type II models are often used in simulation and estimation exercises. An alternative form of truncation may keep the length of the summation window fixed and roll it with t. If we assume that the true model is (7) but make inferences based on (8), the effects of the truncation may not be inconsequential, in the sense that the critical values of test statistics that correspond to type II models are different from the ones that correspond to type I models even in large sample. Davidson and Hashimzade [37] studied the distortions that arise from truncations and suggested a simulation method in the time domain to adjust for them. They noted that the computational techniques they investigated can be useful for estimating the ARFIMA models of type I.

The relation (8) assumes that the value of the process before time $t = 1$ is zero. It is a convenient assumption if the model is estimated by the least squares (or the conditional maximum likelihood) estimation procedure, which is based on the Gaussian likelihood. Since it minimizes the sum of squared residuals, this procedure is also called the conditional sum-of-squares (CSS). In the time domain, the CSS is more popular and computationally easier to implement than the maximum likelihood procedure that is based on the unconditional Gaussian likelihood function, because the latter involves repeatedly constructing and inverting the covariance matrix. (See, for instance, Harvey [67, p.149].) Doornik and Ooms [48] studied the computational properties of the maximum likelihood based on the unconditional Gaussian likelihood and the CSS.

It is common to assume that the initial values are zero in the CSS procedure. Johansen and Nielsen [80, 81] considered the use of non-zero initial values when fractional processes are nonstationary. The initial values are assumed to be a deterministic bounded sequence. More recently, Johansen and Nielsen [83] studied the role of initial values in the CSS estimator when the true fractional process is nonstationary. They considered a version of (5) in which $d > 1/2$, y_t is re-centered around $\mu \in \mathbb{R}$, and ε_t is i.i.d. with zero mean and variance σ^2. They assumed that data do not exist before time $t = -N_0$ for some $N_0 \in \mathbb{N}$, that unobservable data exist between $t = -N_0 + 1$ and $t = 0$, and that observable data exist from $t = 1$. They set aside the first $N \in \mathbb{N}$ observed data as the initial values and investigated their effects. They showed that, if the CSS is used to estimate (d, μ, σ^2), the estimator of d is consistent and asymptotically normal, but the estimator of μ is not consistent. The bias correction procedure they proposed eliminates the second-order bias completely if $N_0 = 0$, but only partly if $N_0 > 0$ because the second-order bias term can be only diminished by increasing N.

3. Why Long Memory

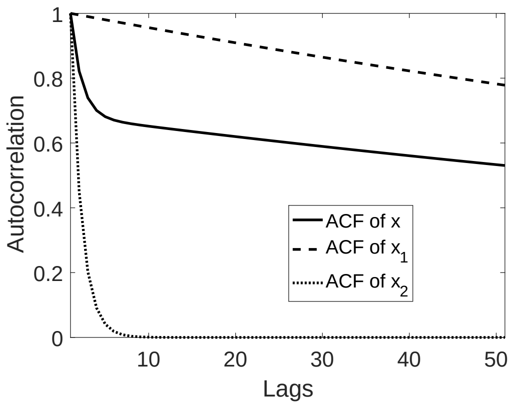
The above discussions highlight that fractional models are theoretically important, but computationally challenging mainly because of the long lag structure and the need for many data points to estimate them. It can be also difficult to interpret fractional differencing intuitively, or to extend it to multivariate cases. See, for instance, Corsi [31] and references therein. Researchers have sought to explain and replicate long memory behavior using models that are simpler, more intuitive, and easier to estimate.

3.1. Sum of Short-Run Components

A useful insight into long memory came from Granger [60] and Granger and Joyeux [65]. They showed that long memory can be a consequence of aggregating short memory processes. Formally, it was shown that the autocorrelation function of the aggregated process,

$$x_t = \sum_{i=1}^{K} x_{i,t}, \qquad x_{i,t} = \phi_i x_{i,t-1} + \varepsilon_{i,t}, \tag{9}$$

where $(\varepsilon_{i,t})_{t\in\mathbb{N}}$ are i.i.d. white noise for each i, converges to that of a long memory process as $K \to \infty$ if $(\phi_i)_{i\in\mathbb{N}}$ are i.i.d. Beta random variables. The degree of persistence depends



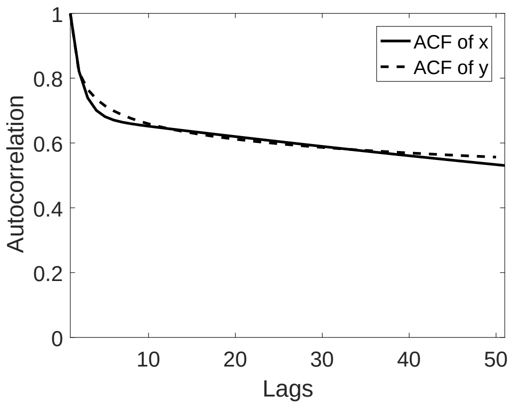


Figure 4: Left: the autocorrelation functions (ACFs) of $x_{1,t}$, $x_{2,t}$, and x_t with $K = 2$ when $\phi_1 = 0.995$, $\phi_2 = 0.45$, $\mathrm{Var}(\varepsilon_{1,t}) = 0.81$, and $\mathrm{Var}(\varepsilon_{2,t}) = 30.25$. Right: the autocorrelation functions of x_t and y_t defined by $(1 - L)^d y_t = \varepsilon_t$, where $d = 0.45$ and ε_t is white noise.

on the shape of the Beta distribution near unity. Linden [91] considered the special case of the standard uniformly distributed autoregressive coefficient. Zaffaroni [128] relaxed the Beta distribution assumption to a flexible semi-parametric probability distribution, and showed that the aggregated process is more persistent if the density of the autoregressive parameter is more concentrated around unity. Robinson [107] considered the properties of time series with a random autoregressive coefficient before Granger [60].

What Granger [60] had in mind in the above theoretical derivation was perhaps the fact that macroeconomic variables such as employment and gross domestic product are frequently the result of aggregating quantities that are measured for microunits such as firms and families. Granger [63] formally investigated the implication of aggregating microunit dynamics. The problem is still considered almost three decades later by, for instance, Pesaran and Chudik [101], who studied the effect of aggregating microunits in a dynamic panel setting.

The effect of aggregation can be translated to attractive stories of long memory. For instance, it is plausible that markets are driven by a large number of short-term fluctuations that lead to long memory phenomena in economic and financial time series. See Andersen and Bollerslev [5] and Taylor [123, p.340] for related discussions and financial applications. Striking empirical results reinforcing the usefulness of this insight are that, in practice, we don't need K to be large for x_t in (9) to mimic long memory; K can be two to four. Gallant et al. [54] empirically demonstrated that the sum of only two short memory components are able to mimic certain long memory dynamics. Figure 4 illustrates this possibility. Since component models consisting of short memory processes tend to be easier to estimate and interpret than fractionally integrated models, the former became a very popular mode of long memory models. See, for instance, Engle and Lee [50], Harvey [68, 69], Alizadeh et al. [4], Barndorff-Nielsen and Shephard [16, 17], LeBaron [88], and Pong et al. [104], and Andersen et al. [6], among many others.

3.2. Structural Breaks

About twenty years after Granger's [60] seminal work, Diebold and Inoue [43] presented another useful insight into long memory. They found that short memory processes with structural breaks can be easily mistaken as possessing long memory. This insight sparked studies on discriminating short memory with structural breaks from long memory. Granger and Hyung [64] contributed to this discussion by showing that occasional structural breaks cause the appearance of slowly decaying autocorrelations comparable to fractionally integrated processes.

Structural breaks can exist in different distribution parameters. Studies on this subject have mainly focused on the implications of breaks in the mean parameter or volatility. In terms of breaks in the fractional differencing parameter, d, Granger [60] showed that time series with a break in d can be represented by another long memory process with memory parameter, d^*, that is a weighted average of the pre- and the post-break memory parameters.

Many economic and financial time series are subject to structural breaks (see, for instance, Stock and Watson [121]). If breaks are at least partly responsible for long memory effects, one may want to control for them in order to mitigate bias, forecast failures, and misleading inference. For the importance of break corrections and procedures, see, for instance, Campos et al. [21], Hendry [72], Clements and Hendry [30], Pesaran and Timmerman [102, 103], Perron [100], Castle et al. [23], Castle et al. [24], and references therein.

But it is often a challenge to accurately estimate the timing or duration and the magnitude of breaks, and to decisively conclude that long memory in individual time series is purely due to structural breaks. Extending the results of Nunes et al. [98], Kuan and Hsu [86] showed that standard break tests applied to highly persistent stationary processes with no break can misleadingly suggest a significant break. Break detection in long memory time series is an active area of research today. In finance, for instance, Choi et al. [26] examined the possibility of structural breaks in a measure of volatility, called the daily realized volatility (RV), of selected currency exchange rates, and found that breaks in the mean are partly accountable for the persistence of the RV. Garvey and Gallagher [55] suggested that long memory in the volatility of sixteen chosen FTSE100 stocks between 1997 and 2003 is not due to breaks. Also see, for instance, Perron and Qu [99], Wang and Vasilakis [125], and Shi and Ho [116] for more empirical investigations and some proposed testing procedures.

4. Fractional Cointegration

Fractional differencing naturally led to the concept of *fractional cointegration* (see Granger [62]). It is a generalization of cointegration in the $I(0)/I(1)$ paradigm formally introduced by Engle and Granger [49], which is discussed by Jennifer L. Castle and David F. Hendry in this volume. Two $I(d)$ series, x_t and y_t, are said to be fractionally cointegrated (or *cofractional*) if there exists a linear combination (say, $y_t - \alpha x_t$ with $\alpha \in \mathbb{R}$) of x_t

and y_t that is $I(d-b)$ with $b > 0$. We commonly denote this relationship by $\mathrm{CI}(d,b)$. Since cointegration is usually about a stationary relationship between (nonstationary) variables, we usually concern ourselves with $d - b < 1/2$.

For modeling the cofractional relationships of order (d,b), Granger [62] proposed the following vector autoregressive model:

$$\Phi^*(L)(1-L)^d \boldsymbol{y}_t = -(1-(1-L)^b)(1-L)^{d-b}\boldsymbol{\gamma}\boldsymbol{\alpha}^\top \boldsymbol{y}_{t-1} + \theta(L)\boldsymbol{\varepsilon}_t, \tag{10}$$

where the dimension of $\boldsymbol{y}_t$ is $D \in \mathbb{N}$, $\Phi^*(0)$ is the $D \times D$ identity matrix, $\Phi^*(1)$ is of full-rank, and $\boldsymbol{\varepsilon}_t$ is a serially independent zero mean process with a joint distribution. $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are $D \times r$ matrices of rank $r \leq D$. Johansen [79] extended (10) by replacing $\Phi^*(L)$ with another polynomial, $\Phi((1-(1-L)^b)$. When cointegration models defined in the $I(0)/I(1)$ paradigm are generalized to encompass fractional dynamics in this fashion, it is important to know the conditions under which the solutions of new dynamic equations are well-defined and are allowed to be fractional, so that the equations serve as valid platforms for making inference on cofractional relationships and fractional orders. See Granger [62] and Johansen [79] for a detailed account on this issue.

The maximum likelihood method based on the conditional Gaussian likelihood can be used to estimate cofractional models in the time domain. Johansen and Nielsen [81] considered this for the above vector autoregressive model with $\Phi^*(L)$ replaced by a polynomial in $(1-L)^b$. Their analysis assumed that errors are i.i.d. with suitable moment conditions and that initial values are a deterministic bounded sequence. They showed that the maximum likelihood estimator is consistent, and that the likelihood ratio test for cointegration rank has standard limiting distributions (Gaussian or chi-squared) when $b < 1/2$. Johansen and Nielsen [80] derived analogous results in the context of univariate nonstationary fractional processes. The choice of the treatment of initial values can have a non-negligible impact on the inference as we discussed in Section 2.1. Alternatively, Davidson [33, 34] implemented bootstrap procedures to test cofractional relationship. Their procedures can be useful if one suspects that the assumptions of standard test statistics might be violated. Fractional cointegration can be estimated also in the frequency domain using the narrow-band least squares estimator proposed by Robinson [111]. The properties of this estimator were studied by, for instance, Lobato [93], Robinson and Marinucci [113, 114], and Christensen and Nielsen [28].

In finance and economics, fractional cointegration shed light on several empirical puzzles that have attracted attention. For instance, spot exchange rates that are individually found to be nonstationary often appear to be tied to each other in the long-run. This relationship could not be sufficiently explained by cointegration models in the $I(0)/I(1)$ paradigm. Baillie and Bollerslev [11] found that allowing for long-range dependence in individual series can reveal fractionally cointegrated relationships between them.

Another puzzle concerns the relationship between implied volatility (IV) and realized volatility (RV), which we touched on earlier.[‡] If traders are rational and markets are

---

[‡]IV is the value of the volatility parameter in Black and Scholes's [18] option pricing formula that is consistent with observed option prices. RV is the sum of squared returns of the underlying asset measured

efficient, financial theories suggest that IV is an unbiased predictor of RV. However, IV tended to be biased empirically. Using either selected foreign exchange rates or stock prices, studies showed that allowing for long-range dependence in both IV and RV can unveil a fractionally cointegrated relationship between them with the associated $\widehat{\alpha}$ suggesting the unbiasedness of IV. See Kellard et al. [84], Nielsen [97], Bandi and Perron [15], and Christensen and Nielsen [28].

Another example of fractional cointegration is between spot commodity prices and the present price of the corresponding futures contract. See Baillie and Bollerslev [10], Figuerola-Ferretti and Gonzalo [51], Westerlund and Narayan [126], and Cavaliere et al. [25] for discussions.

5. Unit-Root and Fractional Alternatives

Around the same time as Granger's [60, 61, 62] introduction to fractional differencing and (fractional) cointegration, Dickey and Fuller [41, 42] developed unit-root tests, which have been routinely applied to discriminate between stationary and nonstationary time series. Clive Granger's work also influenced the way we understand unit-root tests, and compelled researchers to develop unit-root tests against fractional alternatives.

Sowell [120] formally showed that misspecifying the order of fractional integration can lead to the use of different and wrong limiting distributions in tests for nonstationarity. Sowell [120] derived asymptotic distributions for the ordinary least squares (OLS) estimate of the first-order autoregression when the series are $I(1+d')$ with $d' \in (-1/2, 1/2)$, and showed that the support of the limiting distribution (termed the *fractional unit-root* distribution) is non-positive if $d' \in (-1/2, 0)$, nonnegative if $d' \in (0, 1/2)$, and the entire real line if $d' = 0$. Moreover, Sowell [120] showed that the t-statistic that results from this OLS estimation converges only when $d' = 0$.

The Dickey-Fuller (DF) unit-root tests cannot be used to distinguish between unit-root processes and fractional processes by construction because the series being tested are $I(0)$ under the alternative hypothesis. See Diebold and Rudebusch [45] and Hassler and Wolters [70] for discussions. Robinson [108, 110] introduced new testing procedures based on the Lagrange multiplier (LM) statistics that can be used to test unit-roots against fractional alternatives. The proposed testing procedure is attractive since the test statistic asymptotically has a familiar (chi-squared) distribution, and is locally most powerful under Gaussianity. Agiakloglou and Newbold [2] extended the setting of the LM test to general ARFIMA processes. Breitung and Hassler [20] proposed an augmented LM test for unit roots against fractional alternatives that can be generalized to multivariate cointegration tests for determining the cointegration rank of fractionally integrated processes. This work follows Johansen's [78] development of likelihood ratio (LR) tests for cointegration rank, which can be interpreted as a generalization of the DF test in the I(0)/I(1) paradigm. Also see Johansen and Nielsen's [80, 81] LR tests in the univariate and multivariate fractional settings, which we mentioned in Section 2.1.

---

in high frequency.

The development of testing procedures in this context have typically relied on the assumption of homoscedastic errors. Cavaliere et al. [25] cautioned against the conventional use of likelihood-based tests in the presence of heteroscedastic errors, as the tests may not be asymptotically correctly sized under the null and can lead to misleading conclusions about the order of integration. For related discussions, also see Baillie et al. [12], Ling and Li [92], Demetrescu et al. [40], Hassler et al. [71], and Kew and Harris [85]. The development of heteroscedasticity-robust tests in this setting is an active field of research.

6. Other Developments and Frontiers

In Section 2.1, we noted the work by Davidson and Hashimzade [38] on the distortions that arise from truncating the lags of fractional models. Their discussions use the limiting distribution derived by Davidson and Hashimzade [38] of the sample covariance between a nonstationary fractionally integrated process and the stationary increments of itself, or of another process. Their theoretical results can be used in the theory of cofractional models.

Li and McLeod [89] and Robinson [110] introduced the use of the CSS to estimate fractionally integrated models. Chung and Baillie [29] considered the small sample properties of the CSS estimator when it is used to estimate the ARFIMA models. Sowell [119] discussed computational procedures to evaluate the likelihood function in the unconditional Gaussian maximum likelihood. Lieberman and Phillips [90] derived asymptotic expansions of the distribution of the estimator for d that results in this procedure when the fractional processes are stationary. A minimum distance estimator, similar to the generalized method of moment estimator, has been considered by Tieslau et al. [124]. It minimizes the difference between sample and population autocorrelations.

Fox and Taqqu [53] and Giraitis and Surgailis [58] considered the use of the quadratic term in the Gaussian likelihood to estimate long memory parameter. Robinson [112] considered maximizing an approximate form of frequency domain Gaussian likelihood. Whittle's [127] work on likelihood approximation in the frequency domain has been widely applied in this field. Several versions of the Whittle estimator (e.g. local, exact, and extended versions) have been developed for long memory processes. See Giraitis and Robinson [57], Dalla et al. [32], Shimotsu and Phillips [117, 118], Haldrup and Nielsen [66], Abadir et al. [1], Shao [115], and references therein.

Baillie et al. [13] studied the implications of the choice of estimator to the predictive performance of the ARFIMA models. Semiparametric local-Whittle estimators and the maximum likelihood estimators are compared. Studies have also been dedicated to examining bootstrap-based estimation procedures and inference. See, for instance, Andrews and Lieberman [7], Andrews et al. [8], Poskitt [105], and Poskitt et al. [106].

In Section 2, we discussed the importance of the choice of m, which determines the number of frequencies used in a frequency domain estimation. This choice of the parameter, m, also appears in other frequency-domain procedures such as the local-Whittle estimator. But m is still usually a simple function of the sample size. Baillie et al. [14] recently proposed a cross-validation method as an alternative for choosing m; this method is based on the forecasting ability of the model to improve both the estimation of d and model's predictive performance.

As regard unit-root tests, Dolado et al. [47] extended the DF test in the $I(0)/I(1)$ paradigm to the fractional case and introduced the so-called fractional Dickey-Fuller (F-DF) test in the time domain. Their proofs use the convergence properties of the partial sums of the truncated and non-truncated processes derived by Gourieroux et al. [59] and Davidson and de Jong [35]. Note that Johansen and Nielsen [82] investigated the necessary moment conditions under which the fractional functional central limit theorem for partial sums of fractional processes hold, and pointed out corrections to Davidson and de Jong's [35] asymptotic results. Lobato and Velasco [94] studied the efficiency property of the F-DF test, and introduced a simple two-step OLS estimation procedure that leads to a t-test, which can be interpreted as a Wald test and is asymptotically equivalent to Robinson's [108, 110] LM statistics. Tanaka [122] considered nonstationary fractional unit-root tests in the time domain.

7. Concluding Remarks

Sir Clive Granger's work on long memory has had a major impact on time series analysis. He developed invaluable mathematical tools for more accurate data analysis through his lifelong dedication to scientific discoveries and knowledge sharing. Scientists continue to learn from him beyond his lifetime, as discussed above and by other contributors of this volume.

Acknowledgements

I would like to thank Jennifer L. Castle, David F. Hendry, and Søren Johansen for their thoughtful comments on this article. I would like to thank Nuffield College, the Institute of New Economic Thinking (INET), and the Department of Economics at Oxford University for funding my research.

References

[1] K. M. Abadir, W. Distaso, and L. Giraitis. Nonstationarity extended local whittle estimation. *Journal of Econometrics*, 141:1353–1384, 2007.

[2] C. Agiakloglou and P. Newbold. Lagrange multiplier tests for fractional difference. *Journal of Time Series Analysis*, 15:253–262, 1994.

[3] C. Agiakloglou, P. Newbold, and M. Wohar. Bias in an estimator of the fractional difference parameter. *Journal of Time Series Analysis*, 14:235–246, 1992.

[4] S. Alizadeh, M. W. Brandt, and F. X. Diebold. Range-based estimation of stochastic volatility models. *Journal of Finance*, 57(3):1047–1091, 2002.

[5] T. G. Andersen and T. Bollerslev. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance*, 52:975–1005, 1997.

[6] T. G. Andersen, T. Bollerslev, P. F. Christoffersen, and F. X. Diebold. Volatility and correlation forecasting. In G. Elliott, C. W. J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, pages 778–878. Elsevier, 2006.

[7] D. W. Andrews and O. Lieberman. Valid edgeworth expansions for the whittle maximum likelihood estimator for stationary long-memory gaussian time series. *Econometric Theory*, 21:710–734, 2005.

[8] D. W. Andrews, O. Lieberman, and V. Marmer. Higher-order improvements of the parametric bootstrap for long-memory gaussian processes. *Journal of Econometrics*, 133:673–702, 2006.

[9] N. M. P. C. Areal and S. J. Taylor. The realized volatility of ftse-100 futures prices. *Journal of Futures Markets*, 22:627–648, 2002.

[10] R. T. Baillie and T. Bollerslev. The long memory of the forward premium. *Journal of International Money and Finance*, 13:565–571, 1994.

[11] R. T. Baillie and T. Bollerslev. Cointegration, fractional cointegration, and exchange rate dynamics. *Journal of Finance*, 49:737–745, 1994.

[12] R. T. Baillie, C. F. Chung, and M. A. Tieslau. Analyzing inflation by the fractionally integrated arfima-garch model. *Journal of Applied Econometrics*, 11:23–40, 1996.

[13] Richard T. Baillie, Chaleampong Kongcharoen, and George Kapetanios. Prediction from arfima models: Comparisons between mle and semiparametric estimation procedures. *International Journal of Forecasting*, 28:46–53, 2012.

[14] Richard T. Baillie, George Kapetanios, and Fotis Papailias. Bandwidth selection by cross-validation for forecasting long memory financial time series. *Journal of Empirical Finance*, 29:129–143, 2014.

[15] F. M. Bandi and B. Perron. Long memory and the relation between implied and realized volatility. *Journal of Financial Econometrics*, 4:636–670, 2006.

[16] O. E. Barndorff-Nielsen and N. Shephard. Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 63:167–207, 2001.

[17] O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B*, 64:253–280, 2002.

[18] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.

[19] Tim Bollerslev and Hans Ole Mikkelsen. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics*, 73(1):151–184, July 1996.

[20] J. Breitung and U. Hassler. Inference on the cointegration rank in fractionally integrated processes. *Journal of Econometrics*, 110:167–185, 2002.

[21] J. Campos, N. R. Ericsson, and D. F. Hendry. Cointegration tests in the presence of structural breaks. *Journal of Econometrics*, 70:187–220, 1996.

[22] G. M. Caporale and L. A. Gil-Alana. Long memory and fractional integration in high frequency data on the us dollar/british pound spot exchange rate. *International Review of Financial Analysis*, 29:1–9, 2013.

[23] J. L. Castle, N. W. P. Fawcett, and D. F. Hendry. Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics*, 158:25–36, 2010.

[24] J. L. Castle, J. A. Doornik, D. F. Hendry, and F. Pretis. Detecting location shifts during model selection by step-indicator saturation. *Econometrics*, 3:240–264, 2015.

[25] Giuseppe Cavaliere, Morten Ørregaard Nielsen, and A. M. R. Taylor. Bootstrap score test for fractional integration in heteroskedastic arfima models, with an application to price dynamics in commodity spot and futures markets. *Journal of Econometrics*, 187:557–579, 2015.

[26] K. Choi, W. C. Yu, and E. Zivot. Long memory versus structural breaks in modeling and forecasting realized volatility. *Journal of International Money and Finance*, 29: 857–875, 2010.

[27] G. Chortareas, Y. Jiang, and J. C. Nankervis. Forecasting exchange rate volatility using high-frequency data: Is euro different? *International Journal of Forecasting*, 27:1089–1107, 2011.

[28] B. Christensen and M. Nielsen. Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, 133:343–371, 2006.

[29] C. F. Chung and R. T. Baillie. Small sample bias in conditional sum of squared estimators of fractionally integrated arma models. *Empirical Economics*, 18:791–806, 1993.

[30] M. P. Clements and D. F. Hendry. *Forecasting Economic Time Series*. Cambridge University Press, 1998.

[31] F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7:174–196, 2009.

[32] V. Dalla, L. Giraitis, and J. Hidalgo. Consistent estimation of the long memory parameter for nonlinear time series. *Journal of Time Series Analysis*, 27:211–251, 2006.

[33] J. Davidson. A model for fractional cointegration, and tests for cointegration using the bootstrap. *Journal of Econometrics*, 110:187–212, 2002.

[34] J. Davidson. Alternative bootstrap procedures for testing cointegration in fractionally integrated processes. *Journal of Econometrics*, 133:741–777, 2006.

[35] J. Davidson and R. M. de Jong. The functional central limit theorem and weak convergence to stochastic integrals ii: Fractionally integrated processes. *Econometric Theory*, 2000.

[36] J. Davidson and N. Hashimzade. Alternative frequency and time domain versions of fractional brownian motion. *Econometric Theory*, 24:256–293, 2008.

[37] J. Davidson and N. Hashimzade. Type i and type ii fractional brownian motions: A reconsideration. *Computational Statistics and Data Analysis*, 53:2089–2106, 2009.

[38] J. Davidson and N. Hashimzade. Representation and weak convergence of stochastic integrals with fractional integrator processes. *Econometric Theory*, 25:1589–1624, 2009.

[39] S. Degiannakis, C. Floros, and P. Dent. Forecasting value-at-risk and expected shortfall using fractionally integrated models of conditional volatility: International evidence. *International Review of Financial Analysis*, 27:21–33, 2013.

[40] Matei Demetrescu, Vladimir Kuzin, and Uwe Hassler. Long memory testing in the time domain. *Econometric Theory*, 24:176–215, 2008.

[41] D. A. Dickey and W. A. Fuller. Distribution of estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.

[42] D. A. Dickey and W. A. Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49:1057–1072, 1981.

[43] F. X. Diebold and A. Inoue. Long memory and regime switching. *Journal of Econometrics*, 105:131–159, 2001.

[44] F. X. Diebold and G. D. Rudebusch. Long memory and persistence in aggregate output. *Journal of Monetary Economics*, 24(2):189–209, 1989.

[45] F. X. Diebold and G. D. Rudebusch. On the power of dickey-fuller tests against fractional alternatives. *Economic Letters*, 1991.

[46] Z. Ding, C. W. J. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106, 1993.

[47] J. J. Dolado, J. Gonzalo, and L. Mayoral. A fractional dickey-fuller test for unit roots. *Econometrica*, 70:1963–2006, 2002.

[48] J. A. Doornik and M. Ooms. Computational aspects of maximum likelihood estimation of autoregressive fractionally integrated moving average models. *Computational Statistics and Data Analysis*, 42:333–348, 2003.

[49] R. F. Engle and C. W. J. Granger. Cointegration and error correction: Representation, estimation and testing. *Econometrica*, 55:251–276, 1987.

[50] R. F. Engle and G. G. J. Lee. A permanent and transitory component model of stock return volatility. In R. F. Engle and H. White, editors, *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger*, pages 475–497. Oxford University Press, 1999.

[51] I. Figuerola-Ferretti and J. Gonzalo. Modelling and measuring price discovery in commodity markets. *Journal of Econometrics*, 158:95–107, 2010.

[52] J. Fleming and C. Kirby. Long memory in volatility and trading volume. *Journal of Banking and Finance*, 35:1714–1726, 2011.

[53] R. Fox and M. S. Taqqu. Large sample properties of parameter estimates for strong dependent stationary gaussian time series. *Annals of Statistics*, 14:517–532, 1986.

[54] A. R. Gallant, C. T. Hsu, and G. E. Tauchen. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *Review of Economics and Statistics*, 81:617–631, 1999.

[55] J. F. Garvey and L. A. Gallagher. The realized-implied volatility relationship: Recent empirical evidence from ftse-100 stock. *Journal of Forecasting*, 31:639–660, 2012.

[56] J. Geweke and S. Porter-Hudak. The estimation and application of long memory time series. *Journal of Time Series Analysis*, 4(4):221–238, 1983.

[57] L. Giraitis and P. M. Robinson. Edgeworth expansions for semiparametric whittle estimation of long memory. *Annals of Statistics*, 31:1325–1375, 2003.

[58] L. Giraitis and D. Surgailis. A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of whittle's estimate. *Probability Theory and Related Fields*, 86:87–104, 1990.

[59] C. Gourieroux, F. Maurel, and A. Monfort. Least squares and fractionally integrated regressors. Technical report, Document de Travail No. 8913, INSEE, 1989.

[60] C. W. J. Granger. Long memory relations and the aggregation of dynamic models. *Journal of Econometrics*, 14:227–238, 1980.

[61] C. W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, 1981.

[62] C. W. J. Granger. Developments in the study of cointegrated economic variables. *Oxford Bulletin of Economics and Statistics*, 48:213–228, 1986.

[63] C. W. J. Granger. Implications of aggregation with common factors. *Econometric Theory*, 3:208–222, 1987.

[64] C. W. J. Granger and N. Hyung. Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11:399–421, 2004.

[65] C. W. J. Granger and Roselyne Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1): 15–29, 1980. ISSN 1467-9892.

[66] N. Haldrup and M. O. Nielsen. Estimation of fractional integration in the presence of data noise. *Computational Statistics and Data Analysis*, 51:3100–3114, 2007.

[67] A. C. Harvey. *Time Series Models*. Harvester: Wheatsheaf, 2nd edition, 1993.

[68] A. C. Harvey. Long memory in stochastic volatility. In T. Knight and S. Satchell, editors, *Forecasting Volatility in Financial Markets*, pages 307–320. Butterworth-Heinemann, Oxford, 1998.

[69] A. C. Harvey. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Society Monograph. Cambridge University Press, 2013.

[70] U. Hassler and J. Wolters. On the power of unit root tests against fractionally integrated alternatives. *Economic Letters*, 45:1–5, 1994.

[71] U. Hassler, P. M. M. Rodrigues, and A. Rubia. Testing for general fractional integration in the time domain. *Econometric Theory*, 25:1793–1828, 2009.

[72] D. F. Hendry. The econometrics of macroeconomic forecasting. *The Economic Journal*, 107:1330–1357, 1997.

[73] W. H. Hipel and A. I. McLeod. Preservation of the rescaled adjusted range. *Parts 1-3, Water Resources Research*, 14:491–518, 1978.

[74] J. R. M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

[75] J. R. M. Hosking. Modeling persistence in hydrological time series using fractional differencing. *Water Resources Research*, 20:1898–1908, 1984.

[76] H. E. Hurst. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116:770–799, 1951.

[77] S. Hylleberg and C. W. J. Granger. Seasonal integration and cointegration. *Journal of Econometrics*, 44:215–238, 1990.

[78] S. Johansen. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12:231–254, 1988.

[79] S. Johansen. A representation theory for a class of vector autoregressive models for fractional processes. *Econometric Theory*, 24:651–676, 2008.

[80] S. Johansen and M. Ø. Nielsen. Likelihood inference for a nonstationary fractional autoregressive model. *Journal of Econometrics*, 158:51–65, 2010.

[81] S. Johansen and M. Ø Nielsen. Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, 80:2667–2732, 2012.

[82] S. Johansen and M. Ø Nielsen. A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory*, 28:671–679, 2012.

[83] S. Johansen and M. Ø. Nielsen. The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models. *Econometric Theory*, 32:1095–1139, 2016.

[84] N. M. Kellard, C. Dunis, and N. Sarantis. Foreign exchange, fractional cointegration and the implied-realized volatility relation. *Journal of Banking and Finance*, 34:882–891, 2010.

[85] H. Kew and D. Harris. Heteroskedasticity-robust testing for a fractional unit root. *Econometric Theory*, 25:1734–1753, 2009.

[86] C. M. Kuan and C. C. Hsu. Change-point estimation of fractionally integrated processes. *Journal of Time Series Analysis*, 19:693–708, 1998.

[87] A. J. Lawrence and N. T. Kottegoda. Stochastic modelling of river-flow time series. *Journal of the Royal Statistical Society: Series A*, 140:1–47, 1977.

[88] B. LeBaron. Stochastic volatility as a simple generator of financial power laws and long memory. *Quantitative Finance*, 1:621–631, 2001.

[89] W. K. Li and A. I. McLeod. Fractional time series modeling. *Biometrika*, 73:217–221, 1986.

[90] O. Lieberman and P. C. B. Phillips. Expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. *Econometric Theory*, 20: 464–484, 2004.

[91] M. Linden. Time series properties of aggregated ar(1) processes with uniformly distributed coefficients. *Economic Letters*, 64:31–36, 1999.

[92] S. Ling and W. K. Li. On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of American Statistical Association*, 92:1184–1194, 1997.

[93] I. N. Lobato. Consistency of the averaged cross-periodogram in long-memory series. *Journal of Time Series Analysis*, 18:137–155, 1997.

[94] I. N. Lobato and C. Velasco. Efficient wald tests for fractional unit roots. *Econometrica*, 75:575–589, 2007.

[95] B. B. Mandelbrot. Long-run linearity, locally gaussian process, h-spectra and infinite variances. *International Economic Review*, 10:82–111, 1969.

[96] B. B. Mandelbrot and J. W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10:422–437, 1968.

[97] M. Ø. Nielsen. Local whittle analysis of stationary fractional cointegration and the implied-realized volatility relation. *Journal of Business and Economic Statistics*, 25: 427–446, 2007.

[98] L. C. Nunes, C. M. Kuan, and P. Newbold. Spurious break. *Econometric Theory*, 11:736–749, 1995.

[99] P. Perron and Z. Qu. Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics*, 28:275–290, 2010.

[100] Pierre Perron. Dealing with structural breaks. In *Palgrave Handbook of Econometrics*, pages 278–352, 2006.

[101] M. H. Pesaran and A. Chudik. Aggregation in large dynamic panels. *Journal of Econometrics*, 178:273–285, 2014.

[102] M. H. Pesaran and A. Timmerman. Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129:183–217, 2005.

[103] M. H. Pesaran and A. Timmerman. Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137:134–161, 2007.

[104] S. Pong, S. J. Shackleton, S. J. Taylor, and X. Xu. Forecasting currency volatility: A comparison of implied volatilities and ar(fi)ma models. *Journal of Banking and Finance*, 28:2541–2563, 2004.

[105] D. S. Poskitt. Properties of the sieve bootstrap for fractionally integrated and non-invertible processes. *Journal of Time Series Analysis*, 29:224–250, 2008.

[106] D. S. Poskitt, Simone D. Grose, and Gael M. Martin. Higher-order improvements of the sieve bootstrap for fractionally integrated processes. *Journal of Econometrics*, 188:94–110, 2015.

[107] P. M. Robinson. Statistical inference for a random coefficient autoregressive model. *Scandinavian Journal of Statistics*, 5:163–168, 1978.

[108] P. M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47:67–84, 1991.

[109] P. M. Robinson. *Time Series with Strong Dependence, Advances in Econometrics: 6th World Congress*. Cambridge University Press, Cambridge, 1994.

[110] P. M. Robinson. Efficient tests of nonstationary hypotheses. *Journal of American Statistical Association*, 89:1420–1437, 1994.

[111] P. M. Robinson. Semiparametric analysis of long-memory time series. *Annals of Statistics*, 22:515–539, 1994.

[112] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23:1630–1661, 1995.

[113] P. M. Robinson and D. Marinucci. Narrow-band analysis of nonstationary processes. *Annals of Statistics*, 29:947–986, 2001.

[114] P. M. Robinson and D. Marinucci. Semiparametric frequency domain analysis of fractional cointegration. In P. M. Robinson, editor, *Time Series with Long Memory*, pages 334–373. Oxford University Press, 2003.

[115] X. Shao. Nonstationarity-extended whittle estimation. *Econometric Theory*, 26: 1060–1087, 2010.

[116] Yanlin Shi and Kin-Yip Ho. Long memory and regime switching: A simulation study on the markov regime switching arfima model. *Journal of Banking and Finance*, 61: 5189–5204, 2015.

[117] K. Shimotsu and P. C. B. Phillips. Exact local whittle estimation of fractional integration. *Annals of Statistics*, 33:1890–1933, 2005.

[118] K. Shimotsu and P. C. B. Phillips. Local whittle estimation of fractional integration and some of its variants. *Journal of Econometrics*, 130:209–233, 2006.

[119] F. Sowell. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165–188, 1992.

[120] F. B. Sowell. The fractional unit root distribution. *Econometrica*, 58:495–505, 1990.

[121] J. H. Stock and M. W. Watson. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, 14:11–30, 1996.

[122] K. Tanaka. The nonstationary fractional unit root. *Econometric Theory*, 15:549–582, 1999.

[123] S. J. Taylor. *Asset Price Dynamics, Volatility, and Prediction.* Princeton University Press, Princeton, N.J., 2005.

[124] M. A. Tieslau, P. Schmidt, and R. T. Baillie. A minimum-distance estimator for long memory processes. *Journal of Econometrics*, 71:249–264, 1996.

[125] S. H. Wang and C. Vasilakis. Recursive predictive tests for structural change of long memory arfima processes with unknown break-points. *Economics Letters*, 118: 389–392, 2012.

[126] J. Westerlund and P. Narayan. Testing the efficient market hypothesis in conditionally heteroskedastic futures markets. *Journal of Futures Markets*, 33:1024–1045, 2013.

[127] P. Whittle. *Hypothesis Testing in Time Series Analysis.* PhD Dissertation, University of Uppsala, 1951.

[128] P. Zaffaroni. Contemporaneous aggregation of linear dynamic models in large economies. *journal of Econometrics*, 120:75–102, 2004.