Harmonically Weighted Processes: The case of U.S. inflation*

Uwe Hassler[†] and Mehdi Hosseinkouchack

Goethe University Frankfurt[‡]

January 3, 2018

Abstract

We suggest a model for long memory in time series that amounts to harmonically weighting short memory processes, $\sum_j x_{t-j}/(j+1)$. A nonstandard rate of convergence is required to establish a Gaussian functional central limit theorem. Further, we study the asymptotic least squares theory when harmonically weighted processes are regressed on each other. The regression estimators converge to Gaussian limits upon the conventional normalization with square root of the sample size, and standard testing procedures apply. Harmonically weighted processes do not allow - or require - to choose a memory parameter. Nevertheless, they may well be able to capture dynamics that have been modelled by fractional integration in the past, and the conceptual simplicity of the new model may turn out to be a worthwhile advantage in practice. The harmonic inverse transformation that removes this kind of long memory is also developed. We successfully apply the procedure to monthly U.S. inflation, and provide simulation evidence that fractional integration of order d is well captured by harmonic weighting over a relevant range of d in finite samples.

JEL classification C22 (time-series models);

Keywords Long memory; persistence; fractional integration

*[*] **Acknowledgements:** Earlier versions of this paper were presented at the Workshop in Time Series Econometrics in Zaragoza, March 30-31, 2017, and at seminars at Freie Universität Berlin and Universität zu Köln. The authors are grateful to Anindya Banerjee, Jesús Gonzalo, Javier Hualde and Marc Pohle for helpful comments.

[†]Corresponding author: hassler@wiwi.uni-frankfurt.de

[‡]Campus Westend, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany.

1 Introduction

Much of the economic and financial literature equates long memory in time series with the model of fractional integration [FI] popularized by Granger and Joyeux (1980) and Hosking (1981). Fractional integration of order $d < 1/2$ implies autocovariances $\gamma(h)$ converging to zero at rate h^{2d-1}; hence, they are not (absolutely) summable as long as $d > 0$. In the frequency domain this translates into a pole of order λ^{-2d} as the frequency λ approaches the origin. Such a feature has been described by Granger (1966) as "typical spectral shape of an economic variable". Using the fractionally integrated model, Geweke and Porter-Hudak (1983) revealed long memory in different U.S. price indices. Further independent work on long memory in inflation rates was by Delgado and Robinson (1994) for Spain, and by Hassler and Wolters (1995) and Baillie, Chung, and Tieslau (1996) for international evidence, followed by abundant evidence in e.g. Franses and Ooms (1997), Baum, Barkoulas, and Caglayan (1999), Franses, Ooms, and Bos (1999), Hsu (2005), Kumar and Okimoto (2007), and Martins and Rodrigues (2014). Similarly in the field of finance, long memory in realized volatility is sometimes considered to be a stylized fact since the papers by Andersen, Bollerslev, Diebold, and Ebens (2001), Andersen, Bollerslev, Diebold, and Labys (2003), see also Maasoumi and McAleer (2008) and the recent evidence by Hassler, Rodrigues, and Rubia (2016). Most of these papers assume explicitly or implicitly fractional integration to capture and detect long memory. With $d \in (0, 1/2)$, fractional integration offers an overwhelming flexibility in modelling the strength of memory under stationarity, and the extension to the nonstationarity region is straightforward. While flexibility is a virtue on the one hand, it is a burden at the same time: When regressing two fractionally integrated series on each other, we have to ensure that their orders of integration are equal in order to avoid unbalanced regressions, and the issue of fractional cointegration comes in, see Granger (1981). Even if both orders of integration are equal spurious regression (in the sense of diverging t-statistics) may occur, see Tsay and Chung (2000). This triggered a huge literature on the estimation and testing of d, see e.g. the recent books by Giraitis, Koul, and Surgailis (2012) and Beran, Feng, Ghosh, and Kulik (2013).

Despite ample evidence in favour of long memory, it has been argued that it does not necessarily have to result from fractional integration. This strand of literature has been labelled "spurious long memory" since Lobato and Savin (1998), and many authors have contributed, see e.g. Diebold and Inoue (2001), Granger and Hyung (2004), Ohanissian, Russell, and Tsay (2008), Perron and Qu (2010) and Qu (2011). With harmonic weighting we suggest yet another model for long memory, which is extremely simple and falls into the class of linear models just like fractional integration. Similarly, Corsi (2009) proposed a simple model for long memory in realized volatility. But we go one step beyond and discuss the regression of harmonically weighted processes, too. Notwithstanding the long memory, the asymptotic theory remains standard, asymptotic χ^2 tests apply. Further, we argue that harmonically weighted processes are in finite samples almost observationally equivalent to fractionally integrated ones over a wide range of relevant values of d. Hence, the applied researcher may allow long memory series of this persistence to enter his or her stationary regressions without having to worry about nonstandard inference.

Let $\{\varepsilon_t\}$ denote a sequence of white noise [WN] with $\mathrm{E}(\varepsilon_t) = 0$. Harmonically weighted noise, $\sum_{j=1}^{t-1} j^{-1}\varepsilon_{t-j}$, shows up in the derivative of the log-likelihood function of Gaussian fractionally integrated noise, see Tanaka (1999, eq. (40)), and it was used to construct a Lagrange Multiplier test for fractional integration. With the same purpose, Demetrescu, Kuzin, and Hassler (2008) considered more generally the processes

$$y_{t-1}^{+} = \sum_{j=1}^{t-1} \frac{x_{t-j}}{j} \quad \text{and} \quad y_{t-1} = \sum_{j=1}^{\infty} \frac{x_{t-j}}{j} \,,$$

where $\{y_{t-1}^{+}\}_{t=2,\ldots,T}$ is the finite sample counterpart of $\{y_{t-1}\}_{t\in\mathbb{Z}}$, where $\mathbb{Z}$ denotes the set of all integers. The filtered process $\{x_t\}$ is assumed to be a stationary regular process with absolutely summable moving average coefficients and positive spectrum. Demetrescu et al. (2008, Lemma 4) showed that $\{y_{t-1}\}$ possesses a sequence of square summable autocovariances. Without squaring the autocovariances are not summable, which was shown for the particular case of harmonically weighted noise by Pesaran (2015, p. 347). In fact, it is not hard to show that the autocovariance at lag h decays at rate $(\ln h)/h$ in case of harmonically weighted noise, see eq. (6) below.

Except for the above results, little seems to be known about harmonically weighted processes [HWP]. Here, we discuss their persistence and long memory properties that differ from the well known features under fractional integration. Their persistence and long memory are characterized by a pole in the spectrum at the origin that is of order $\ln^2 \lambda$ for $\lambda \to 0$, see Proposition 1. Consequently, it follows from Proposition 2 that the sample mean converges only with variance $(\ln^2 T)/T$, such that the true ensemble mean is harder to estimate than in the case of standard stationary processes (integrated of order zero). Further, we discuss the inversion of the filter with harmonic weights, called harmonic inverse transformation [HIT]. When applying the HIT to some data with a spectral pole of order $\ln^2 \lambda$, then this transformation removes the pole. At the same time a mean different from zero will be effectively removed without having to be estimated. Processes like $\{y_{t-1}\}$ defined above are not only of theoretical interest, showing up in the Lagrange Multiplier test for fractional integration. They are also interesting for modelling empirical series, and we demonstrate the usefulness and appropriateness of the HIT with monthly U.S. inflation data that have been modelled previously by means of fractional integration;

see Figure 2 below.

The rest of the paper is organized as follows. Section 2 becomes precise on the assumptions and contains the properties of HWP in the time and frequency domains. The third section presents the asymptotic theory for partial sums of HWP, with a nonstandard central limit theorem [CLT] as special case. Asymptotic least squares theory when harmonically weighted processes are regressed on each other is given in Section 4. The harmonic inverse transformation is introduced and discussed in Section 5. Section 6 compares fractional differencing with the harmonic inverse transformation for monthly U.S. inflation data. Section 7 compares systematically the properties of HWP with the more common long memory model, namely fractional integration, and discusses the possibility (and difficulty) to discriminate between the two of them. The final section offers some conclusions. Mathematical proofs are relegated to the Appendix.

A final word on notation: Throughout this paper, $\Rightarrow$ stands for weak convergence as the sample size T diverges, $\stackrel{D}{\to}$ represents convergence in distribution, and $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to $x \geq 0$, $x \in \mathbb{R}$. Further, (probabilistic) Landau symbols $O(\cdot)$ (and $O_p(\cdot)$) have their usual meaning, and $\sim$ denotes asymptotic equivalence of two sequences or functions.

2 Properties of HWP

In terms of the usual lag operator L we define the harmonically weighted filter

$$h(L) = 1 + \frac{L}{2} + \frac{L^2}{3} + \cdots$$

given by the formal expansion of $\ln(1-L)$:

$$h(L) := -\frac{\ln(1-L)}{L} = \sum_{j=0}^{\infty} \frac{L^j}{j+1} \,. \tag{1}$$

This defines a harmonically weighted process, HWP, as follows.

Assumption 1 *Let*

$$y_t = \mu + h(L) x_t \,, \quad t \in \mathbb{Z} \,,$$

where $\{x_t\}$ *is a stationary process with mean zero and*

$$x_t = c(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \,, \quad \varepsilon_t \sim \text{ WN } (0, \sigma^2) \,, \text{ i.e. } \mathrm{E}(\varepsilon_t \, \varepsilon_{t+h}) = \begin{cases} \sigma^2 \,, & h = 0 \\ 0 \,, & h \neq 0 \end{cases} \,,$$

and with $(c_0 = 1)$

$$\sum_{j=0}^{\infty} j\,|c_j| < \infty \quad \text{and} \quad c(1) = \sum_{j=0}^{\infty} c_j \neq 0\,. \tag{2}$$

The process $\{x_t\}$ behind Assumption 1 is sometimes called integrated or order zero, $I(0)$. The restriction of one-summability, $\sum_{j=0}^{\infty} j\,|c_j| < \infty$, is a rather weak and widely used assumption since Phillips and Solo (1992). All stationary and invertible autoregressive moving average processes [ARMA] meet (2), since c_j is geometrically bounded in the ARMA case.

Obviously, $\{y_t\}$ is conformable with the definition from the introduction, except for the expectation μ. The finite analogue may be written by means of the indicator function,

$$\mathbf{1}_{(t>0)}(t) = \begin{cases} 1\,, & t > 0 \\ 0\,, & \text{else} \end{cases}\,.$$

We then have h_+ and $\{y_t^+\}$ defined as follows:

$$y_t^+ := \mu + h_+(L)x_t := \mu + h(L)x_t\mathbf{1}_{(t>0)}(t) = \mu + \sum_{j=0}^{t-1} \frac{x_{t-j}}{j+1}\,, \quad t = 1, \ldots, T\,. \tag{3}$$

For the rest of the exposition we focus on $\{y_t\}$ from Assumption 1, since Demetrescu et al. (2008) showed in the proof of their Lemma 2 that

$$y_t - y_t^+ = O_p\left(\frac{1}{\sqrt{t}}\right)\,. \tag{4}$$

We now give properties of $\{y_t\}$ in terms of $\{x_t\}$ with autocovariances γ_x and spectrum f_x:

$$\gamma_x(h) = \sigma^2 \sum_{j=0}^{\infty} c_j c_{j+h}\,,\ h = 0, 1, \ldots, \ \text{and} \quad f_x(\lambda) = \frac{\sigma^2}{2\pi}\left|\sum_{j=0}^{\infty} c_j e^{\mathrm{i}j\lambda}\right|^2\,,\ \mathrm{i}^2 = -1\,.$$

Correspondingly, f_y and γ_y stand for the spectrum and the autocovariances of $\{y_t\}$, respectively. The moving average representation of the process is given by convolution of $h(L)$ and $c(L)$,

$$y_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}\,, \quad b_j = \sum_{k=0}^{j} \frac{c_k}{j+1-k}\,, \tag{5}$$

where $\{\varepsilon_t\}$ is the white noise from Assumption 1.

Proposition 1. *The harmonically weighted process* $\{y_t\}$ *from Assumption 1 is stationary with mean* μ *and*

a) *moving average coefficients*

$$b_j \;\sim\; \frac{\sum_{k=0}^{\infty} c_k}{j} = \frac{c(1)}{j}\,, \quad j \to \infty\,;$$

b) *spectrum*

$$\begin{aligned} f_y(\lambda) &= \left[\ln^2\left(2\,\sin\frac{\lambda}{2}\right) + \left(\frac{\pi-\lambda}{2}\right)^2\right] f_x(\lambda)\,, \quad \lambda > 0\,, \\ &\sim \ln^2(\lambda)\; f_x(0)\,, \quad \lambda \to 0\,; \end{aligned}$$

c) *autocovariances*

$$\gamma_y(h) \;\sim\; 2\pi f_x(0)\frac{\ln h}{h}\,, \quad h \to \infty\,.$$

Proof. See Appendix.

REMARK 1 *Let us consider the special case of harmonically weighted noise, where* $x_t = \varepsilon_t$ *and* $2\pi f_x(0) = \sigma^2$. *It is straightforward to show in this case that*

$$\gamma_y(0) = \sigma^2 \sum_{j=0}^{\infty} (j+1)^{-2} = \sigma^2 \frac{\pi^2}{6}\,,$$

$$\gamma_y(h) = \sigma^2 \sum_{j=0}^{\infty} \frac{1}{(j+1)\,(j+1+h)} = \sigma^2 \frac{1}{h} \sum_{j=1}^{h} \frac{1}{j}\,. \tag{6}$$

For the general HW process, we have a spectral singularity of order $\ln^2(\lambda)$ at the origin. This reflects that the sum over the Wold coefficients diverges at logarithmic rate:

$$\lim_{J\to\infty} \frac{1}{\ln J} \sum_{j=0}^{J} b_j = c(1)\ .$$

In that sense, the HW process is strongly persistent. Further, it displays long memory since

$$\sum_{h=0}^{H} |\gamma_y(h)| \;\to\; \infty \quad \text{as } H \to \infty\,.$$

This persistence and this degree of long memory are, however, not as strong as under the assumption of fractional integration [FI]. To make this statement precise, we briefly recap the model of FI. It relies on the fractional integration operator with the usual binomial expansion:

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \binom{-d}{j} (-L)^j .$$

We now define fractionally integrated processes, for short $z_t \sim I(d)$, often called of type I since the work by Marinucci and Robinson (1999).

Assumption 2 *Let*

$$z_t = \mu + (1-L)^{-d} x_t \,, \quad t \in \mathbb{Z}\,,\ 0 < d < \frac{1}{2}\,,$$

where $\{x_t\}$ *is from Assumption 1.*

By convolution it holds that

$$z_t = \mu + \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j} \,,$$

where $\{\varepsilon_t\}$ is from Assumption 1. The impulse responses β_j vanish at rate j^{d-1}:

$$\beta_j \ \sim \ \frac{\sum_{k=0}^{\infty} c_k}{\Gamma(d)} j^{d-1} \,, \quad j \to \infty \,.$$

From this it further follows that

$$f_z(\lambda) \sim \lambda^{-2d} f_x(0) \,, \quad \lambda \to 0 \,,$$

$$\gamma_z(h) \ \sim \ C\, h^{2d-1} \,, \quad h \to \infty \,,$$

where $C = 2\pi f_x(0)\Gamma(1-2d)/(\Gamma(d)\Gamma(1-d))$. Consequently, we find that the HW process has theoretically less memory and persistence than any FI process with positive d:

$$\lim_{j\to\infty} \frac{b_j}{\beta_j} = 0 \,, \quad \lim_{\lambda\to 0} \frac{f_y(\lambda)}{f_z(\lambda)} = 0 \,, \quad \lim_{h\to\infty} \frac{\gamma_y(h)}{\gamma_z(h)} = 0 \,, \quad d > 0 \,. \tag{7}$$

In finite samples, however, matters may be different, see Section 7 below.

3 (Functional) Central Limit Theorem

We now turn to large sample properties of the sample mean of HWP. We obtain the behavior of the variance of cumulated HWP, which is used to establish a functional central limit theorem [FCLT].

Proposition 2. *Let us maintain Assumption 1, where $\{\varepsilon_t\}$ is a martingale difference sequence with $E(\varepsilon_t^2) = \sigma^2$ and $E(|\varepsilon_t|^p) < \infty$ for some $p > 2$. It is further assumed to be either strictly stationary and ergodic or to satisfy Abadir, Distaso, Giraitis, and Koul (2014, Ass. 2.1). It then holds as $T \to \infty$*

a) *that*

$$\frac{\mathrm{Var}\left(\sum_{t=1}^{T} y_t\right)}{T \, \ln^2 T} \;\to\; 2\pi f_x(0)\,,$$

b) *and that*

$$\frac{\sum_{t=1}^{\lfloor rT \rfloor}(y_t - \mu)}{\sqrt{T}\, \ln T} \Rightarrow \sqrt{2\pi f_x(0)}\, W(r)\,,$$

where W is a standard Wiener process, $0 \le r \le 1$.

Proof. See Appendix.

Our proof of Proposition 2 *b*) relies on Abadir et al. (2014). Hence, we maintain their assumptions. Note that Abadir et al. (2014, Ass. 2.1) allow for conditional heteroskedasticity meeting certain requirements with respect to conditional moments, see also the discussion in Abadir et al. (2014, Sect. 4.1). For $r = 1$, we have the following central limit theorem for $\overline{y} = T^{-1} \sum_t y_t$,

$$\sqrt{T}\, \frac{(\overline{y} - \mu)}{\ln T} = \frac{\sum_{t=1}^{T}(y_t - \mu)}{\sqrt{T}\, \ln T} \;\stackrel{D}{\to}\; \mathcal{N}\left(0,\, 2\pi f_x(0)\right)\,.$$

Although $\mathrm{Var}(\overline{y})$ converges to zero with T, it does so more slowly than in the standard case of absolutely summable processes like $\{x_t\}$ characterized in Assumption 1:

$$\sqrt{T}\,(\overline{x} - 0) = \frac{\sum_{t=1}^{T} x_t}{\sqrt{T}} \;\stackrel{D}{\to}\; \mathcal{N}\left(0,\, 2\pi f_x(0)\right)\,.$$

At the same time, $\overline{y}$ converges faster than in case of FI with long memory. Let $\overline{z}$ be the sample mean of a processes satisfying Assumption 2. For $I(d)$ processes we know from Abadir et al. (2014, Cor. 4.1) that limiting normality arises when normalizing with $T^{-1/2+d}$. Consequently, by Proposition 2:

$$\frac{\mathrm{Var}(\overline{y})}{\mathrm{Var}(\overline{z})} = O_p\left(\frac{\ln^2 T}{T^{2d}}\right) \;\to\; 0\,, \quad d > 0\,.$$

Now, we briefly turn to the issue of finite sample efficiency of $\overline{y}$. Let $\widetilde{\mu}$ denote the generalized least squares [GLS] estimator of μ under Assumption 1, i.e. the best linear

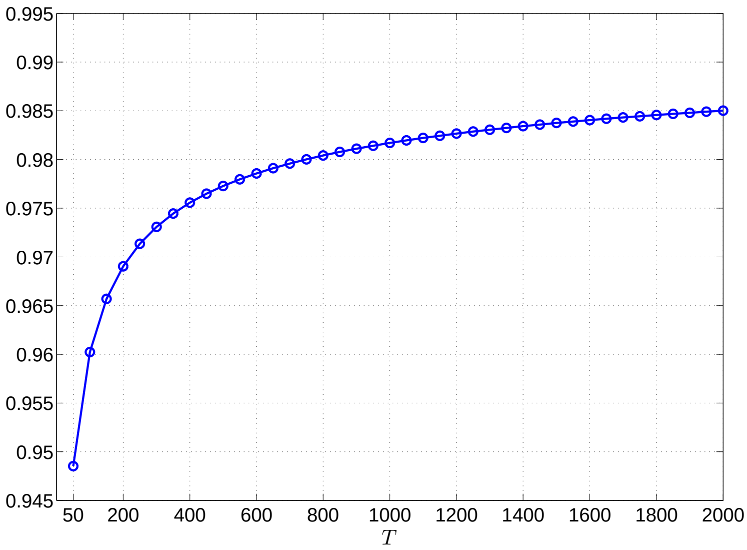

Figure 1: $\text{Var}(\widetilde{\mu})/\text{Var}(\overline{y})$

unbiased estimator. We now consider an example to quantify potential efficiency gains beyond $\overline{y}$. Assume $x_t = \varepsilon_t$ with known σ^2, such that $\{y_t\}$ is harmonically weighted noise. With $\mathbf{1}$ denoting a T vector of ones, we have

$$\frac{\text{Var}(\widetilde{\mu})}{\text{Var}(\overline{y})} = \frac{T^2}{\mathbf{1}'\Omega\mathbf{1} \cdot \mathbf{1}'\Omega^{-1}\mathbf{1}},$$

where Ω contains $\omega_{i,i+h} = \gamma_y(h)/\sigma^2$ with $\gamma_y(h)$ being from Remark 1. In Figure 1 we evaluate $\text{Var}(\widetilde{\mu})/\text{Var}(\overline{y})$ for T ranging from 50 up to 2,000. It is obvious that the efficiency gains of $\widetilde{\mu}$ relative to $\overline{y}$ are very small in larger samples. The estimation of μ is inevitably plagued by the strong persistence or long memory of HWP resulting in the slow rate of convergence observed in Proposition 2.

From Proposition 2 it follows that HW processes fall into a class that has been characterized recently by Berenguer-Rico and Gonzalo (2014). Let $L(x)$ be slowly varying at infinity in Karamata's sense, $L(cx)/L(x) \to 1$ as $x \to \infty$ for all $c > 0$. Then, according to Berenguer-Rico and Gonzalo (2014), a process $\{\xi_t\}$ is summable of order δ, if δ is the minimum number such that

$$\frac{L(T)}{T^\delta\sqrt{T}} \sum_{t=1}^{T} (\xi_t - \mu) = O_p(1).$$

Since $1/\ln T$ is slowly varying at infinity, Proposition 2 implies that the HWP $\{y_t\}$ is summable of order $\delta = 0$ in this sense. At the same time it is worth repeating that the HWP is not integrated of order 0 (by Proposition 1).

4 Regression Results

Notwithstanding the nonstandard rate of convergence found in Proposition 2, we will be able to establish standard regression results for harmonically weighted series under appropriate error assumptions. Let the vector of regressors, $\mathbf{r}_t$, contain a constant intercept and K harmonically weighted regressors,

$$\mathbf{r}_t = \begin{pmatrix} 1 \\ r_{1,t} \\ \vdots \\ r_{K,t} \end{pmatrix} = \begin{pmatrix} 1 \\ \mu_1 \\ \vdots \\ \mu_K \end{pmatrix} + h(L) \begin{pmatrix} 0 \\ x_{1,t} \\ \vdots \\ x_{K,t} \end{pmatrix} , \tag{8}$$

where the stochastic component is $\mathbf{x}_t' = (x_{1,t}, \dots, x_{K,t})$. The maintained single equation regression model becomes ($t = 1, \dots, T$)

$$y_t = \beta_0 + \sum_{k=1}^{K} \beta_k r_{k,t} = \beta' \mathbf{r}_t + \varepsilon_t \, , \quad \beta' = (\beta_0, \beta_1, \dots, \beta_K) \, . \tag{9}$$

The stochastic assumptions are as follows.

Assumption 3 *Let* $\mathrm{E}(\mathbf{x}_t) = 0$ *and let further hold*

a) that $\{(\varepsilon_t, \mathbf{x}_t')'\}$ *is a strictly stationary and ergodic vector;*

b) that

$$\mathrm{E}(\varepsilon_t | \mathbf{r}_t, \varepsilon_{t-1}, \mathbf{r}_{t-1}, \varepsilon_{t-2}, \mathbf{r}_{t-2}, \dots) = 0 \, , \quad 0 < \mathrm{E}(\varepsilon_t^2 | \mathbf{r}_t) = \sigma^2 < \infty \, ;$$

c) that the matrix $\Sigma_r = \mathrm{E}(\mathbf{r}_t \mathbf{r}_t')$ *is finite and positive definite.*

Now, consider the ordinary least squares [OLS] regression,

$$y_t = \widehat{\beta}' \mathbf{r}_t + \widehat{\varepsilon}_t \, , \quad t = 1, \dots, T \, .$$

Note that Assumption 3 is standard when maintaining a stationary regression model. It guarantees that $\{\varepsilon_t\}$ and $\{\mathbf{r}_t \varepsilon_t\}$ are both strictly stationary and ergodic martingale difference sequences [mds] with $\mathrm{Var}(\varepsilon_t) = \sigma^2$ and $\mathrm{Cov}(\mathbf{r}_t \varepsilon_t) = \Sigma_r \sigma^2$. Hence, we are able to prove the following result.

Proposition 3. *Let model (8) with (9) hold true. Under Assumption 3 it follows that*

$$\sqrt{T}(\widehat{\beta} - \beta) \overset{D}{\to} \mathcal{N}\left(0, \, \Sigma_r^{-1} \sigma^2\right) ,$$

$$\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 \xrightarrow{p} \sigma^2 \,, \quad T\left(\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}_t'\right)^{-1} \xrightarrow{p} \Sigma_r^{-1} \,,$$

as $T \to \infty$.

Proof. See Appendix.

REMARK 2 *It follows that standard inference applies: the usual t statistics and Wald statistics result in limiting standard normal and χ^2 distributions under the respective null hypotheses. Further, the Durbin-Watson statistic converges to 2, while the coefficient of determination obviously tends to* $1 - \sigma^2/\operatorname{Var}(y_t)$.

REMARK 3 *It is straightforward to extend Proposition 3 to allow for stationary regressors that are not harmonically weighted. Assume instead of (8) that we have H harmonically weighted regressors and $K - H$ further regressors without long memory* ($0 \le H \le K$)

$$\mathbf{r}_t = \begin{pmatrix} 1 \\ \mu_1 \\ \vdots \\ \mu_H \\ \mu_{H+1} \\ \vdots \\ \mu_K \end{pmatrix} + \begin{pmatrix} 0 \\ h(L)x_{1,t} \\ \vdots \\ h(L)x_{H,t} \\ x_{H+1,t} \\ \vdots \\ x_{K,t} \end{pmatrix} . \tag{10}$$

Then Proposition 3 continues to hold. Further, the regressors may contain lagged endogenous variables as long as the regression model (9) remains stable.

Next, we present computer experimental evidence on tests relying on Proposition 3. We report the frequency of rejections of a true null hypothesis at the 5% level from 10,000 replications. We consider two data generating processes [DGPs]. The respective regressors are

$$\text{DGP1: } \mathbf{r}_t = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ h(L)x_{1,t} \\ h(L)x_{2,t} \end{pmatrix} , \quad \text{DGP2: } \mathbf{r}_t = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ h(L)x_{1,t} \\ x_{2,t} \end{pmatrix} , \tag{11}$$

with

$$y_t = \beta'\mathbf{r}_t + \varepsilon_t \,, \quad \beta' = (1,1,1) \,, \quad \begin{pmatrix} \varepsilon_t \\ x_{1,t} \\ x_{2,t} \end{pmatrix} \sim \mathcal{N}(0, I_3) \,. \tag{12}$$

Table 1: Experimental size for DGP1 at 5%

	$h(L)$			$h_+(L)$		
T	t_1	t_2	$2\,F_{1,2}$	t_1	t_2	$2\,F$
100	5.18	5.14	5.18	5.20	5.36	5.59
250	5.07	4.79	4.78	5.47	5.44	5.43
500	5.01	4.66	5.02	4.73	5.42	5.00
1000	5.11	5.03	5.17	5.20	4.72	4.85
5000	4.90	5.08	4.87	4.93	5.05	5.13

Table 2: Experimental size for DGP2 at 5%

	$h(L)$			$h_+(L)$		
T	t_1	t_2	$2\,F_{1,2}$	t_1	t_2	$2\,F$
100	5.27	5.69	5.52	5.07	5.25	5.58
250	5.31	4.99	5.49	4.89	5.16	5.14
500	4.96	4.50	4.95	5.04	5.03	5.17
1000	5.11	5.23	5.19	5.46	4.59	4.95
5000	4.83	5.14	5.08	5.01	4.78	5.04

In the columns labelled by $h(L)$ in Table 1 and 2, we truncated the infinite filter:

$$h(L)x_{k,t} \approx \sum_{j=0}^{1{,}000+t-1} \frac{1}{j+1} x_{k,t-j}\,.$$

In the last three columns labelled by $h_+(L)$, we worked with the finite filter:

$$h_+(L)x_{k,t} = \sum_{j=0}^{t-1} \frac{1}{j+1} x_{k,t-j}\,.$$

The test statistics are the standard t statistics t_1 and t_2 when testing individually for $\beta_1 = 1$ and $\beta_2 = 1$, respectively, and $2\,F$ is twice the standard F statistic testing for $\beta_1 = \beta_2 = 1$ jointly, where the latter is compared with quantiles from $\chi^2(2)$.

First, we observe that the experimental size of the asymptotic tests is close to the nominal one already for $T = 100$. Second, we find that the asymptotic theory established for $h(L)$ in Proposition 3 works equally well for $h_+(L)$. Third, Table 2 supports Remark 2: The size properties under DGP1 (both regressors are harmonically weighted) are very similar to the case of DGP2, where only one regressor is harmonically weighted. For robustness checks we considered further setups, replacing $\beta' = (1, 1, 1)$ by $\beta' = (1, 1, 0)$ and allowing for correlation between the regressors; this leaves the figures essentially unchanged, details are not reported here but available upon request.

It is worth noting that Proposition 3 holds for $\beta = 0$, which amounts to the regression of white noise ε_t on $\mathbf{r}_t$. The case where independent HW processes are regressed on each other is covered in the next proposition For simplicity, we restrict the presentation to the case of a simple regression.

Proposition 4. *Assume two independent processes, $y_{k,t} = h(L)x_{k,t}$, $k = 1, 2$, with autocorrelations $\rho_k(h)$ and variances $\gamma_k(0)$, where $\{x_{k,t}\}$ are from Assumption 1, and $\{\varepsilon_{k,t}\}$ from Assumption 1 are strictly stationary, ergodic martingale difference sequences. Consider the OLS regression*

$$y_{1,t} = \widehat{\alpha} + \widehat{\beta} y_{2,t} + \widehat{e}_t\,, \quad t = 1, \ldots, T\,,$$

with the usual t statistic $t_{\beta=0}$ testing for $\beta = 0$. It then holds that

$$\sqrt{T}\widehat{\beta} \overset{D}{\to} \mathcal{N}\left(0,\, \frac{\gamma_1(0)}{\gamma_2(0)} \sum_{h=-\infty}^{\infty} \rho_1(h)\rho_2(h)\right) \quad \text{and} \quad t_{\beta=0} \overset{D}{\to} \mathcal{N}\left(0,\, 1 + 2\sum_{h=1}^{\infty} \rho_1(h)\rho_2(h)\right)$$

as $T \to \infty$.

Proof. See Appendix.

Regardless of the long memory in both processes, no spurious regression arises under independence. First, $\widehat{\beta}$ from Proposition 4 converges to the true value at the standard rate. Second, the t statistic does not diverge, although its limiting normal distribution has of course a variance different from one due to the serial correlation in the residuals.

REMARK 4 *As an example, consider two independent harmonically weighted noise series, $y_{k,t} = h(L)\varepsilon_{k,t}$ where by Remark 1 $\rho_k(h) = \frac{6}{\pi^2}\frac{1}{h}\sum_{j=1}^{h}\frac{1}{j}$. We hence have*

$$2\sum_{h=1}^{\infty} \rho_1(h)\rho_2(h) = 2\sum_{h=1}^{\infty} \left(\rho_1(h)\right)^2 = \frac{17}{5}\,,$$

where the latter equality is from Borwein and Borwein (1995, eq. (3)). If one erroneously compares $t_{\beta=0}$ with quantiles from the standard normal distribution, then the probability to reject for a two-sided 5% level test becomes asymptotically

$$\lim_{T\to\infty} Pr\left(|t_{\beta=0}| > z_{0.975}\right) = 2\,\Phi\left(-\frac{\sqrt{5}}{\sqrt{22}}1.96\right) \approx 0.35\,,$$

where Proposition 4 was used with the standard normal distribution function $\Phi(\cdot)$.

In the previous section we learnt that the sample means of HW as well as FI processes display nonstandard rates of convergence, however the rate is slower for FI due to the stronger long memory. We close the present section with a comparison of respective regression results. When it comes to regressions of FI processes, it is crucial that the order of integration of the left-hand side equals the maximum order on the right-hand side; otherwise the equation is unbalanced, and the regressors cannot possibly explain the regressand. But even in the case of balanced regressions, FI may cause troubles. Consider in analogy to Proposition 4 that two independent, stationary FI processes are regressed on each other, where the order of integration is between 1/4 and 1/2. Then spurious regressions arise in that $t_{\beta=0}$ diverges in absolute value with increasing sample size, see Tsay and Chung (2000). Such pitfalls cannot happen with HW processes.

5 Harmonic Inverse Transformation

Next, we turn to the transformation of the data that removes the pole in the spectrum observed from Prop. 1 *a*). Thus the harmonic filter $h(L)$ is inverted to define

$$g(L) = \frac{1}{h(L)} = -\frac{L}{\ln(1-L)} = 1 - \sum_{j=1}^{\infty} g_j L^j \,, \tag{13}$$

where $\{g_j\}$ are the coefficients of the Taylor expansion, and $h(L)g(L) = 1$ yields the recursive relation

$$g_j = \frac{1}{j+1} - \sum_{i=1}^{j-1} \frac{g_i}{j-i+1} \,, \quad j \geq 1 \,, \; g_0 = 1 \,.$$

These coefficients are sometimes called Gregory coefficients, see e.g. Blagouchine (2016), and they are known to be positive, $g_j > 0$. An evaluation yields for the first terms of the sequence

$$g_0 = 1 \,, \; g_1 = \frac{1}{2} \,, \; g_2 = \frac{1}{12} \,, \; g_3 = \frac{1}{24} \,, \; g_4 = \frac{19}{720} \,, \; g_5 = \frac{3}{160} \,.$$

It is obvious that

$$g(1) = \lim_{z \to 1} g(z) = 0 \,,$$

see also Blagouchine (2016, eq. (20)). Hence, we have that $\sum_{j=1}^{\infty} g_j = 1$, such that the filter $g(L)$ is (absolutely) summable, and one even knows the rate at which the coefficients vanish, see Blagouchine (2016, eq. (18)):

$$g_j \sim \frac{1}{j \ln^2 j} \quad \text{as } j \to \infty \,. \tag{14}$$

Since the filter coefficients sum up to zero, it follows for HW processes from Assumption 1 that

$$g(L)y_t = g(1)\mu + x_t = x_t \, .$$

Hence, filtering the data not only removes the long memory but the mean at the same time, which is convenient, since we saw in Section 3 that the mean is hard to estimate. Similarly, we may allow for mean shifts that are removed by harmonic inverse transformation (HIT). In the simplest case, let one break occur at time $\lfloor \tau T \rfloor$ $(0 < \tau < 1)$:

$$m_t = \begin{cases} \mu_0 \, , & t \leq \lfloor \tau T \rfloor \\ \mu_1 \, , & t > \lfloor \tau T \rfloor \end{cases} \, .$$

Here, the removal is not exact for $t > \tau T$. We rather have that

$$g(L)m_t = \begin{cases} g(1)\mu_0 = 0 \, , & \text{if } \; t \leq \lfloor \tau T \rfloor \\ \left(1 - \sum_{j=1}^{t-\lfloor \tau T \rfloor - 1} g_j \right) \mu_1 - \mu_0 \sum_{j=t-\lfloor \tau T \rfloor}^{\infty} g_j \, , & \text{if } \; t > \lfloor \tau T \rfloor \end{cases} \, .$$

For some fixed $\epsilon > 0$ and $t = \lfloor \tau T \rfloor + \lfloor \epsilon T \rfloor$, the term $\sum_{j=t-\lfloor \tau T \rfloor}^{\infty} g_j$ converges to zero as $T \to \infty$, such that

$$\lim_{T \to \infty} g(L) m_{\lfloor \tau T \rfloor + \lfloor \epsilon T \rfloor} = g(1) \, \mu_1 - \mu_0 \, 0 = 0 \, .$$

In practice, given only a finite past, the HIT has to be truncated:

$$g_+(L)y_t := g(L)y_t \mathbf{1}_{(t>0)}(t) = \sum_{j=0}^{t-1} g_j y_{t-j} \, , \quad t = 1, \ldots, T \, . \tag{15}$$

Now, we are ready to study the dynamic properties of U.S. inflation data.

6 U.S. Inflation

Let P_t stand for the seasonally adjusted monthly consumer price index from December 1969 until August 2017.[1] The inflation series is computed as $\pi_t = 100 \, (P_t - P_{t-1})/P_{t-1}$, $t = 1, \ldots, T = 572$, see the northwestern graph in Figure 2. The sample autocorrelogram in the northeastern graph is indicative of long memory with $\widehat{\rho}_\pi(h) > 0.3$ for $1 \leq h \leq 20$. At the same time, $\widehat{\rho}_\pi(1)$ is clearly less than 1, so that we can rule out a unit root $(d = 1)$. The estimated integration parameter is very close to the region of stationarity, $\widehat{d} = 0.52$, when estimated by exact local Whittle [ELW] according to Shimotsu and Phillips (2005)

[1] Consumer Price Index for All Urban Consumers (all items), retrieved from FRED, Federal Reserve Bank of St. Louis, October 04, 2017.

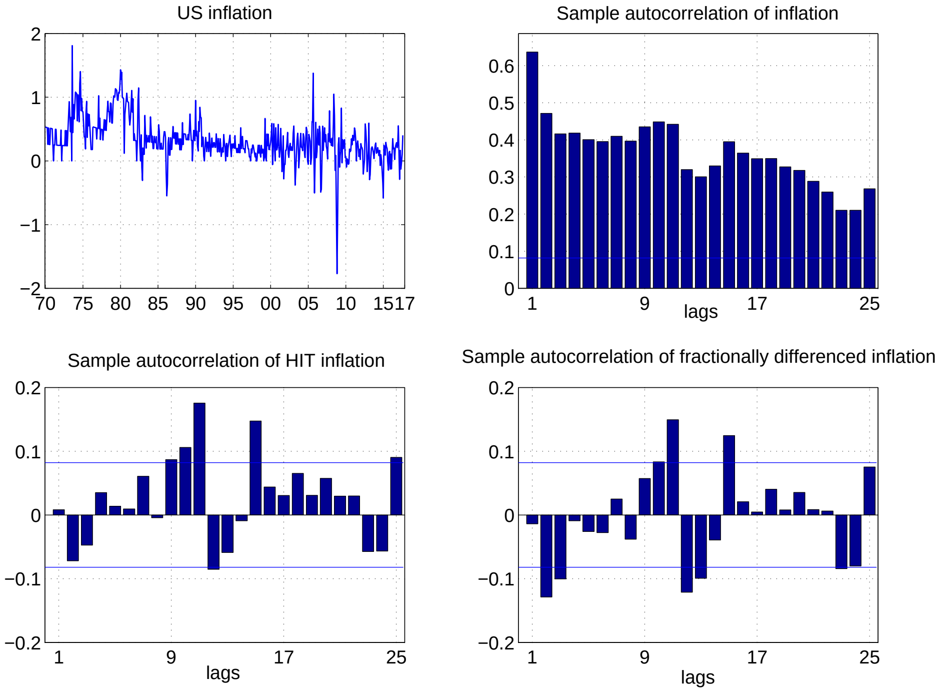

Figure 2: U.S. inflation

and Shimotsu (2010) with bandwidth $T^{0.60}$. This value was used to fractionally difference the series, and alternatively we use the harmonic inverse transformation, HIT:

$$dif_t = (1-L)^{\widehat{d}}\pi_t \quad \text{and} \quad hit_t = g(L)\pi_t .$$

Next, the sample autocorrelations of dif_t and hit_t are computed; they are plotted in the lower graphs of Figure 2 (right and left, respectively). The resulting sample autocorrelograms appear very similar by visual inspection. This suggests that the harmonically weighted model captures the long-range dependence of U.S. inflation just as well as fractional integration. To support this claim we compute the Box-Pierce statistics,

$$Q_{dif}(25) = T\sum_{h=1}^{25}(\widehat{\rho}_{dif}(h))^2 = 72.81 \quad \text{and} \quad Q_{hit}(25) = T\sum_{h=1}^{25}(\widehat{\rho}_{hit}(h))^2 = 70.11\,.$$

Clearly, these values are significantly different from zero at any reasonable level: We do not claim that fractional differencing or harmonic inverse transformation turn U.S. inflation into white noise. But the difference between Q_{dif} and Q_{hit} is small, supporting our claim that the model of harmonic weighting does as good a job in capturing the inflation persistence as the more popular model of fractional integration. At the same time, the HW is radically more simple, it does not require to choose an estimator $\widehat{d}$, and it does not require to pick a bandwidth m.

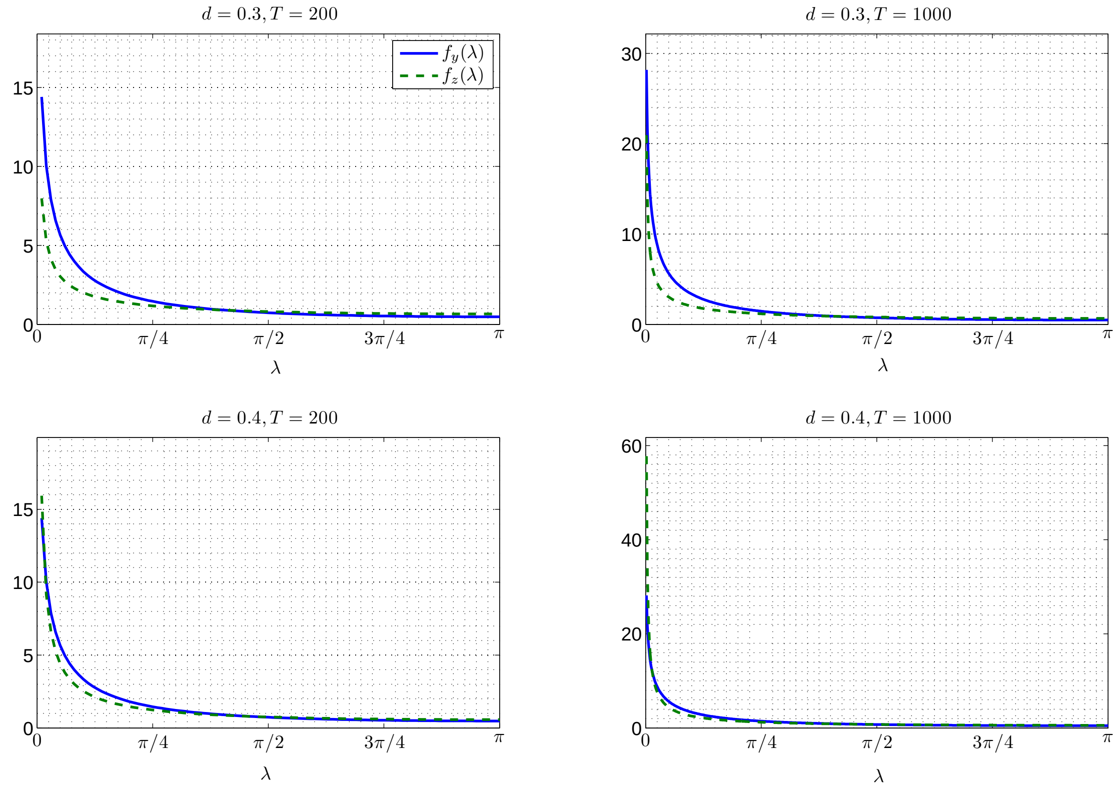

Figure 3: Spectra at harmonic frequencies $\lambda_j = 2\pi j/T$, $j = 1, \ldots, {}^T\!/_2 - 1$

Of course, we do not know whether the true DGP behind U.S. inflation is fractional integration. Therefore, the next section turns to a more systematic comparison between this traditional long memory model and the model of harmonic weighting.

7 HWP versus FI

Although we observed in (7) that the persistence and memory of FI and HW processes have different qualities in theory, matters may be different in finite samples. Given a sample size T, one typically estimates spectra at harmonic frequencies $\lambda_j = 2\pi j/T$, $j = 1, \ldots, {}^T\!/_2 - 1$. For that reason, we plot in Figure 3 HWP spectra and fractionally integrated spectra (for $d = 0.3$ and $d = 0.4$) for different T (with $\sigma^2 = 2\pi$). For $d = 0.3$, the HWP spectrum turns out to be higher than the $I(0.3)$ spectrum at λ_j close to the origin; for $d = 0.4$, the spectra of the $I(d)$ and HW processes are even closer and hard to distinguish by eyesight, and this will of course be all the more true when spectra are estimated in practice. We take this as preliminary evidence that the HWP may be an adequate way to capture memory in data that is otherwise modelled by fractional integration.

To gain further insights into the relation between FI and HWP, we conducted a Monte Carlo experiment with 10,000 replications. We simulated fractionally integrated noise (of type II according to Jensen and Nielsen (2014)), $z_t = \Delta_+^{-d}\varepsilon_t \sim I(d)$, where ε_t is standard normal. Then we HIT the data, $hit_t = g_+(L)z_t$, and estimate the order of integration of the filtered sequence hit_t by means of the ELW estimator mentioned previously. Theoretically, the order of integration should not be affected by HITing, i.e. $hit_t \sim I(d)$. In other words: $\widehat{d}$ computed from hit_t should vary around d. In finite samples, however, things are quite different. In Figure 4 we present Box plots of ELW estimates $\widehat{d}$ computed as described in the previous section. For $T = 100$, the median of $\widehat{d}$ is roughly $d - 0.4$; still for $T = 1000$ the difference between the median of $\widehat{d}$ and d is roughly 0.3. For $d = 0.3$ and $d = 0.4$ the zero line falls almost always between the lower and upper quartile of $\widehat{d}$ for all T, meaning the majority of these cases resembles upon HITing $I(0)$ rather than $I(d)$.

We complement the experiment by testing the null hypothesis that the data upon harmonic inverse transformation, $hit_t = g_+(L)z_t$, is $I(0)$. Theoretically, this null is wrong, because $z_t \sim I(d)$. Still, we want to see how well a test discriminates at a 5% level. To that end we compute the (lag-)augmented LM [ALM] test by Demetrescu, Kuzin, and Hassler (2008). The test is executed by regressing the filtered data hit_t on the auxiliary regressor r_{t-1}^* and q endogenous lags, hit_{t-j}, $j = 1, \ldots, q$, where

$$r_{t-1}^* = \sum_{j=1}^{t-1} \frac{hit_{t-j}}{j} \, .$$

The absolute value of the t statistic testing for insignificance of r_{t-1}^* is compared with the standard normal. Following the recommendation by Demetrescu et al. (2008), we choose $q = \lfloor 4(T/100)^{1/4} \rfloor$. In Figure 5 we report rejection frequencies for $d \in \{0, 0.1, \ldots, 0.6\}$ and $T \in \{100, 200, \ldots, 1000\}$. For small samples, $T = 100$, the null of $I(0)$ is not rejected more often than in 5% of the cases when $d \in [0.2, 0.4]$. Even with $T = 1000$ and $d = 0.3$, the rejection frequency is only 10%, and only with $T = 1000$ and $d = 0$ or $d = 0.6$ we get somewhere close to 100% rejection. This points again at, loosely speaking, near-observational equivalence of harmonic weighting and fractional integration in finite samples over a wide range of combinations of d and T, meaning: If we HIT $I(d)$ processes, the persistence is effectively removed in finite samples for many combinations of d and T.

There is yet another aspect to Figure 5. Note that the highest frequency of rejection occurs for $d = 0$ for all values of T. This means that the ALM test, which is designed against fractional alternatives, has considerable power to detect long memory even if it is caused by harmonic weighting and not by fractional integration.

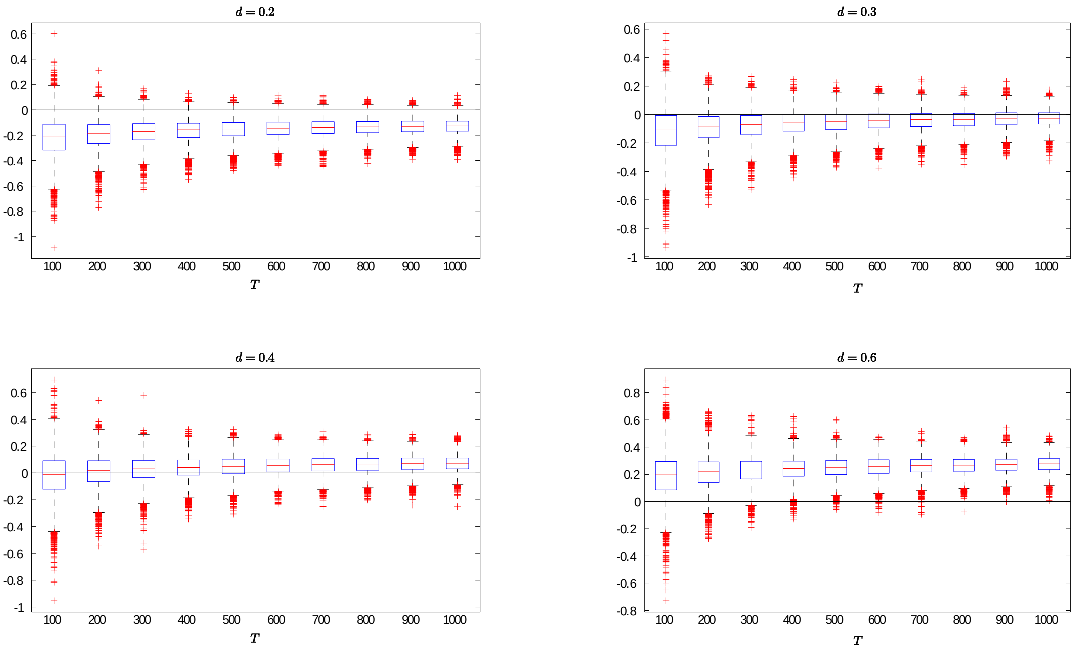

Figure 4: ELW estimates from $hit_t = g_+(L)z_t$, $z_t \sim I(d)$

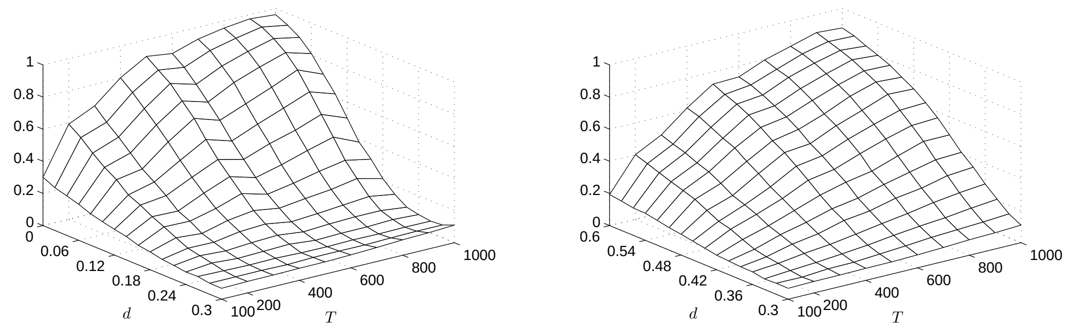

Figure 5: Rejection of ALM test when applied to $hit_t = g_+(L)z_t$, $z_t \sim I(d)$

8 Concluding Remarks

The standard model to account for long memory in finance and economics is fractional integration of order d, where this memory parameter may take on any positive value. Fractional integration thus offers an overwhelming flexibility in modelling persistence. This is a virtue and a burden at the same time: on the one hand, there is a high degree of flexibility in modelling long memory, but on the other hand the estimation of d is notoriously difficult and troubled by large variance of slowly converging semiparametric estimators. Here, we suggest an alternative model for long memory, which amounts to harmonically weighting short memory processes. With monthly U.S. inflation data we illustrate that this model may well be able to capture dynamics that have been modelled by fractional integration in the past. A large Monte Carlo experiment indeed shows that harmonic weighting may effectively capture the long memory usually modelled by fractional integration for a relevant range of d.

We also study the asymptotic least squares theory when harmonically weighted processes are regressed on each other. While limiting normality of the sample average requires the nonstandard normalization with $\sqrt{T}/\ln T$, the regression estimators converge to Gaussian limits upon the standard normalization with $\sqrt{T}$ allowing for standard inference. Computer experiments support the finite sample relevance of this limiting distribution theory. Consequently, the empirical economists may allow long memory series to enter their stationary regressions without having to worry about nonstandard inference – as long as the long memory may be considered as being caused or captured by harmonic weighting. The admitted simplicity and rigidity of the harmonically weighted model, which does not allow - or require - to choose a memory parameter, may therefore turn out to be a practical advantage in applied work.

There are open issues. First, one may wish to step beyond the single equation regression and allow for a truely multivariate framework where harmonically weighted vector autoregressive processes are allowed for. Second, one may account for nonstationarity and allow for processes where integer differencing is required to obtain harmonically weighted processes. Third, the harmonically weighted model may serve as a general forecasting device under long memory when the true data generating process is not known and might be fractionally integrated or spurious long memory. These issues are currently under investigation but beyond the scope of the present paper.

Appendix

Preliminary Results

Our proofs of Proposition 1 and 2 rely on what is sometimes called the Stolz-Cesàro Theorem. For the ease of reference, we give the result here, adopting the version by Mureşan (2009, Thm. 1.22).

Stolz-Cesâro Theorem *Let* $\{s_n\}$ *and* $\{\sigma_n\}$ *be real sequences,* $n \in \mathbb{N}$*, where* $\{\sigma_n\}$ *is strictly monotone and divergent. If* $(s_{n+1} - s_n)/(\sigma_{n+1} - \sigma_n)$ *converges, then* s_{n+1}/σ_{n+1} *converges, too, and has the same limit:*

$$If \lim_{n\to\infty} \frac{s_{n+1} - s_n}{\sigma_{n+1} - \sigma_n} = \ell\,, \quad then \lim_{n\to\infty} \frac{s_{n+1}}{\sigma_{n+1}} = \ell\,. \tag{16}$$

The proof by Mureşan (2009) also covers the case $\ell = \pm\infty$. For a historical exposition on this result we also recommend Knopp (1951, pp. 76, 77).

The proof of Proposition 2 requires a technical lemma that we provide next.

Lemma A. *It holds that*

$$\sum_{h=1}^{T} \frac{(T-h)\,\ln h}{h} = \frac{T}{2}\,\ln^2 T - T\,\ln T + O(T)\,.$$

Proof. We define the function $f\,(x) = \frac{(T-x)\ln x}{x}$ with kth derivative $f^{(k)}$. In order to evaluate $\sum_{h=1}^{T} f(h)$, we use Euler's summation formula taken from Knopp (1951, p. 524):

$$\sum_{h=1}^{T} f(h) = \int_1^T f(x)\,\mathrm{d}x + \frac{1}{2}\,(f(T) + f(1)) + \frac{1}{12}\left(f^{(1)}(T) - f^{(1)}(1)\right) + R\,, \tag{17}$$

where

$$|R| \le \frac{1}{2\pi^3} \int_1^T \left|f^{(3)}\,(x)\right|\,\mathrm{d}x\,.$$

For the third derivative we obtain in absolute value that

$$\left|f^{(3)}\,(x)\right| = \left|\frac{11T}{x^4} - \frac{6T\ln x}{x^4} - \frac{2}{x^3}\right| \le \frac{11T}{x^4} + \frac{6T\ln x}{x^4} + \frac{2}{x^3}\,.$$

It is elementary to verify that

$$\int_1^T f(x)\,\mathrm{d}x = \frac{1}{2}T\ln^2 T - T\ln T + T - 1\,,$$

$f(1) = f(T) = 0$, $f^{(1)}(T) - f^{(1)}(1) = {}^{-\ln T}/_{T} - (T-1)$, and that

$$|R| \leq \frac{1}{2\pi^3}\left(O(T) + O(T) + O(1)\right) = O(T)\,.$$

Hence,

$$\sum_{h=1}^{T} f(h) = \frac{1}{2} T \ln^2 T - T \ln T + O(T)\,,$$

which proves the result. ∎

Proof of Proposition 1

The stationarity and the expectation follow from Fuller (1996, Thm. 2.2.3) since $b_j = \sum_{k=0}^{j} c_k/(j+1-k)$ is given by convolution of an absolutely summable and a square summable filter.

a) Let us decompose $jb_j = j\sum_{k\leq j/2} c_k/(j+1-k) + j\sum_{k>j/2} c_k/(j+1-k)$. We consider the second sum first:

$$j\left|\sum_{k>j/2} \frac{c_k}{j+1-k}\right| \leq \sum_{k>j/2} 2\,k\,\frac{|c_k|}{j+1-k} \leq \sum_{k>j/2} 2\,k\,|c_k| \;\rightarrow\; 0\,.$$

Second, we study the difference of the first sum and $\sum_{k\leq j/2} c_k$:

$$\begin{aligned}
\left|\sum_{k\leq j/2} c_k - j\sum_{k\leq j/2} \frac{c_k}{j+1-k}\right| \;&\leq\; \sum_{k\leq j/2} |c_k|\,\frac{|1-k|}{j+1-k} = \frac{|c_0|}{j+1} + \sum_{k=2}^{j/2} |c_k|\,\frac{k-1}{j+1-k} \\
&\leq\; \frac{1}{j+1} + \sum_{k=2}^{j/2} |c_k|\,\frac{k}{j+1-j/2} \;\rightarrow\; 0\,.
\end{aligned}$$

Consequently, $j\sum_{k\leq j/2} c_k/(j+1-k) \rightarrow \sum_{k=0}^{\infty} c_k$ for $j \rightarrow \infty$, as required.

b) For $\lambda > 0$ we have

$$f_y(\lambda) = \left|h(e^{\mathrm{i}\lambda})\right|^2 f_x(\lambda)\,, \quad h(e^{\mathrm{i}\lambda}) = -\frac{\ln(1-e^{\mathrm{i}\lambda})}{e^{\mathrm{i}\lambda}}\,,$$

where $\left|h(e^{\mathrm{i}\lambda})\right|^2 = \ln(1-e^{\mathrm{i}\lambda})\ln(1-e^{-\mathrm{i}\lambda})$. Note that

$$\ln(1-e^{\mathrm{i}\lambda}) = \ln(r(\lambda)\,e^{\mathrm{i}\theta(\lambda)}) = \ln(r(\lambda)) + \mathrm{i}\theta(\lambda)$$

with

$$r(\lambda) = \sqrt{(1-\cos\lambda)^2 + \sin^2\lambda} = \sqrt{4\,\sin^2\frac{\lambda}{2}}\,,$$

and

$$\theta(\lambda) = \arctan \frac{-\sin\lambda}{1-\cos\lambda}\,, \quad \lambda > 0\,.$$

With $\ln(1-e^{-\mathrm{i}\lambda}) = \ln(r(\lambda)) - \mathrm{i}\theta(\lambda)$ we obtain

$$\left|h(e^{\mathrm{i}\lambda})\right|^2 = \ln^2(r(\lambda)) + \theta^2(\lambda) = \ln^2\left(2\,\sin\frac{\lambda}{2}\right) + \arctan^2 \frac{\sin\lambda}{1-\cos\lambda}\,.$$

Further, focusing on the principal value,

$$\arctan \frac{\sin\lambda}{1-\cos\lambda} = \arctan\cot\frac{\lambda}{2} = \frac{\pi}{2} - \frac{\lambda}{2}\,,$$

where we used the usual double-angle formulae and $\tan(\pi/2 - x) = \cot x$ for the last two equations, respectively. Hence, we have at the origin that

$$\frac{\left|h(e^{\mathrm{i}\lambda})\right|^2}{\ln^2\lambda} \;\to\; 1 \quad \text{as } \lambda \to 0\,.$$

This implies the spectral results as required.

c) We write b_j as

$$b_j = \frac{1}{j+1}\sum_{k=0}^{j} \frac{c_k}{1-\frac{k}{j+1}} = \frac{1}{j+1}B_j\,,$$

where B_j was defined implicitly. From part *a*) we have that $B_j \to c(1)$. Now, define $s_j - s_{j-1} = b_j b_{j+h}$ and $\sigma_j - \sigma_{j-1} = \frac{1}{j+1}\frac{1}{j+h+1}$. It holds by part *a*) that $(s_j - s_{j-1})/(\sigma_j - \sigma_{j-1}) = B_j B_{j+h} \to (c(1))^2$. Therefore, by (16) we have

$$\frac{\sum_{j=0}^{\infty} b_j b_{j+h}}{\sum_{j=0}^{\infty}\frac{1}{j+1}\frac{1}{j+h+1}} = \frac{\gamma_y(h)/\sigma^2}{\frac{1}{h}\sum_{j=1}^{h}\frac{1}{j}} = (c(1))^2\,,$$

where the first equality is by (6). This means that

$$\gamma_y(h) \sim 2\pi f_x\,(0)\,\frac{\ln h}{h}\,.$$

Hence, the proof is complete.

Proof of Proposition 2

a) Define $s_{T-1} = \sum_{h=1}^{T-1}(T-h)\,\gamma_y\,(h)$ and $\sigma_{T-1} = \sum_{h=1}^{T-1}(T-h)\,\frac{\ln h}{h}$ with

$$\frac{s_{T-1} - s_{T-2}}{\sigma_{T-1} - \sigma_{T-2}} = \frac{\sum_{h=1}^{T-1}\gamma_y\,(h)}{\sum_{h=1}^{T-1}\frac{\ln h}{h}}\,.$$

By Proposition 1 c), we have $\gamma_y(T-1) \sim 2\pi f_x(0)\frac{\ln(T-1)}{T-1}$. Using (16) it hence follows that

$$\frac{s_{T-1}-s_{T-2}}{\sigma_{T-1}-\sigma_{T-2}} = \frac{\sum_{h=1}^{T-1}\gamma_y(h)}{\sum_{h=1}^{T-1}\frac{\ln h}{h}} \to 2\pi f_x(0) \ .$$

Again by (16), this time applied to s_{T-1} and σ_{T-1}, we conclude that

$$\frac{\sum_{h=1}^{T-1}(T-h)\gamma_y(h)}{\sum_{h=1}^{T-1}(T-h)\frac{\ln h}{h}} \to 2\pi f_x(0) \ .$$

We may expand the left-hand side,

$$\frac{\sum_{h=1}^{T-1}(T-h)\gamma_y(h)}{\sum_{h=1}^{T-1}(T-h)\frac{\ln h}{h}} = \frac{\sum_{h=1}^{T-1}(T-h)\gamma_y(h)}{\frac{1}{2}T\ln^2 T}\frac{\frac{1}{2}T\ln^2 T}{\sum_{h=1}^{T-1}(T-h)\frac{\ln h}{h}} ,$$

where the second factor on the right-hand side converges to 1 by Lemma A, such that

$$\frac{\sum_{h=1}^{T-1}(T-h)\gamma_y(h)}{\frac{1}{2}T\ln^2 T} \to 2\pi f_x(0) \ .$$

Consequently,

$$\frac{Var\left(\sum_{t=1}^{T} y_t\right)}{T\ln^2 T} = \frac{\gamma_y(0)}{\ln^2 T} + \frac{2\sum_{h=1}^{T-1}(T-h)\gamma_y(h)}{T\ln^2 T} \to 2\pi f_x(0) \ ,$$

as required.

b) Define $S_T(r) = \sum_{t=1}^{\lfloor rT \rfloor}(y_t-\mu)$ and $\sigma_T^2 = \operatorname{Var}(S_T(1))$. Then we first establish the convergence of the finite dimensional distributions of $\sigma_T^{-1}S_T(r)$ for $0 \le r \le 1$. To do so we first observe that

$$\frac{\operatorname{Var}\left(\sum_{t=1}^{\lfloor \tau T \rfloor}(y_t-\mu)\right)}{\operatorname{Var}\left(\sum_{t=1}^{T}(y_t-\mu)\right)} = \frac{\tau T\ln^2(\tau T)(1+o(1))}{T\ln^2(T)(1+o(1))} \to \tau.$$

For brevity define $a_{t-1} = \sum_{m=0}^{t-1} b_m$. With $S_j = \sum_{t=1}^{j} y_t = \sum_{t=1}^{j} a_{t-1}\varepsilon_t$ we easily see for $j \ge k$ that $\operatorname{Cov}(S_j, S_k) = \operatorname{Var}(S_k)$, since

$$\operatorname{Var}(S_j - S_k) \ = \ \operatorname{Var}\left(\sum_{t=k+1}^{j} a_{t-1}\varepsilon_t\right) = \sigma_\varepsilon^2\sum_{t=k+1}^{j} a_{t-1}^2 = \sigma_\varepsilon^2\sum_{t=1}^{j} a_{t-1}^2 - \sigma_\varepsilon^2\sum_{t=1}^{k} a_{t-1}^2.$$

Therefore, using Abadir et al. (2014, Thm. 2.1), we may conclude that

$$\frac{S_T(r)}{\sigma_T} \overset{fdd}{\to} W(\tau) \ ,$$

where $\stackrel{fdd}{\to}$ denotes the finite dimensional convergence of distributions. To complete the proof we need to show that $\frac{S_T(r)}{\sigma_T}$ is tight with respect to the uniform metric, where we require $\mathrm{E}\left(|\varepsilon_t|^p\right) < \infty$ for some $p > 2$. Note that with some positive constant c

$$
\begin{aligned}
\mathrm{E}\left|\frac{S_T(r)}{\sigma_T} - \frac{S_T(s)}{\sigma_T}\right|^p &\leq c\left[\mathrm{E}\left(\frac{S_T(r)}{\sigma_T} - \frac{S_T(s)}{\sigma_T}\right)^2\right]^{\frac{p}{2}} \\
&= c\left[\mathrm{E}\left(\sigma_T^{-1}\sum_{t=1}^{\lfloor rT\rfloor - \lfloor sT\rfloor}(y_t - \mu)\right)^2\right]^{\frac{p}{2}} \\
&= c\left[\frac{(\lfloor rT\rfloor - \lfloor sT\rfloor)\ln^2(\lfloor rT\rfloor - \lfloor sT\rfloor)}{T\ln^2(T)}\frac{1+o(1)}{1+o(1)}\right]^{\frac{p}{2}} \\
&\leq c\left|\frac{\lfloor rT\rfloor}{T} - \frac{\lfloor sT\rfloor}{T}\right|^{\frac{p}{2}},
\end{aligned}
$$

where the first inequality follows from Abadir et al. (2014, Lemma 3.1). By Billingsley (1968, Thm. 15.5), the last inequality shows that $\frac{S_T(r)}{\sigma_T}$ is tight with respect to the uniform metric. Hence, the proof is complete.

Proof of Proposition 3

Under our standard assumptions the proof is straightforward. By Assumption 3 *b*), the series $\{\varepsilon_t\}$ and $\{\mathbf{r}_t\varepsilon_t\}$ are both strictly stationary and ergodic mds with $\mathrm{Var}(\varepsilon_t) = \sigma^2$ and $\mathrm{Cov}(\mathbf{r}_t\varepsilon_t) = \Sigma_r\sigma^2$. By a mds CLT,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_t\varepsilon_t \stackrel{D}{\to} \mathcal{N}\left(0, \Sigma_x\sigma^2\right).$$

By ergodicity,

$$\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \stackrel{p}{\to} \Sigma_x,$$

where the limit is invertible by Assumption 3 *c*). Hence, the limit of $\sqrt{T}(\widehat{\beta}-\beta)$ is obvious. Finally, with $\widehat{\varepsilon}_t = \varepsilon_t - (\widehat{\beta}-\beta)' x_t$ it holds that

$$\sum_{t=1}^{T}\widehat{\varepsilon}_t^2 = \sum_{t=1}^{T}\varepsilon_t^2 + O_p(1),$$

such that $\widehat{\sigma}^2 \stackrel{p}{\to} \sigma^2$, as required to complete the proof.

Proof of Proposition 4

We observe

$$\sqrt{T}\,\widehat{\beta} = \frac{\frac{1}{\sqrt{T}}\sum_{t=1}^{T} y_{1,t}y_{2,t} - O_p\left(\frac{\ln^2 T}{\sqrt{T}}\right)}{\frac{1}{T}\sum_{t=1}^{T}(y_{2,t}-\overline{y}_2)^2}\,,$$

where the O_p term follows from Proposition 2. Define the stationary sequence $\{w_t\}$,

$$w_t = y_{1,t}y_{2,t} \quad \text{with } \mathrm{E}(w_t) = 0\,,\ \gamma_w(h) = \mathrm{E}(w_t w_{t+h}) = \gamma_1(h)\gamma_2(h)\,,$$

where $\gamma_k(h)$ are the autocovariances of $y_{k,t}$. By Proposition 1 a), the sequence $\{\gamma_w(h)\}$ dies out at rate $\frac{\ln^2 h}{h^2}$, and is hence summable. Thus we may define

$$\omega_w^2 = \sum_{h=-\infty}^{\infty} \gamma_w(h) < \infty\,.$$

By a CLT, see e.g. Abadir et al. (2014, Thm. 2.1), one has

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t \stackrel{D}{\to} \mathcal{N}(0,\,\omega_w^2)\,.$$

By ergodicity, the limiting distribution of $\sqrt{T}\widehat{\beta}$ is implied. The rest of the proof is straightforward.

References

Abadir, K. M., W. Distaso, L. Giraitis, and H. L. Koul (2014). Asymptotic Normality For Weighted Sums Of Linear Processes. *Econometric Theory 30*, 252–284.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica 71*, 579–625.

Baillie, R. T., C.-F. Chung, and M. A. Tieslau (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model. *Journal of Applied Econometrics 11*, 23–40.

Baum, C. F., J. T. Barkoulas, and M. Caglayan (1999). Persistence in International Inflation Rates. *Southern Economic Journal 65*(4), 900–913.

Beran, J., Y. Feng, S. Ghosh, and R. Kulik (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods.* Springer.

Berenguer-Rico, V. and J. Gonzalo (2014). Summability of stochastic processes: A generalization of integration for non-linear processes. *Journal of Econometrics 178*, 331–341.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

Blagouchine, I. V. (2016). Two series expansions for the logarithm of the gamma function involving Stirling numbers and containing only rational coefficients for certain arguments related to π^{-1}. *Journal of Mathematical Analysis and Applications 442*, 404–434.

Borwein, D. and J. M. Borwein (1995). On an intriguing integral and some series related to $\zeta(4)$. *Proceedings of the American Mathematical Society 123*, 1191–1198.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*, 174–196.

Delgado, M. A. and P. M. Robinson (1994). New methods for the analysis of long-memory time-series: Application to spanish inflation. *Journal of Forecasting 13*(2), 97–107.

Demetrescu, M., V. Kuzin, and U. Hassler (2008). Long Memory Testing In The Time Domain. *Econometric Theory 24*(01), 176–215.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics 105*, 131 – 159.

Franses, P. H. and M. Ooms (1997). A periodic long-memory model for quarterly UK inflation. *International Journal of Forecasting 13*(1), 117–126.

Franses, P. H., M. Ooms, and C. S. Bos (1999). Long memory and level shifts: Reanalyzing inflation rates. *Empirical Economics 24*(3), 427–449.

Fuller, W. A. (1996). *Introduction to Statistical Time Series* (2nd ed.). Wiley.

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

Giraitis, L., H. L. Koul, and D. Surgailis (2012). *Large Sample Inference for Long Memory Processes.* Imperial College Press.

Granger, C. W. and N. Hyung (2004). Occasional structural breaks and long memory with an application to the s & p 500 absolute stock returns. *Journal of Empirical Finance 11*(3), 399 – 421.

Granger, C. W. J. (1966). The typical spectral shape of an economic variable. *Econometrica 34*, 150–161.

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. and R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis 1*, 15–29.

Hassler, U., P. M. M. Rodrigues, and A. Rubia (2016). Quantile regression for long memory testing: A case of realized volatility. *Journal of Financial Econometrics 14*, 693–724.

Hassler, U. and J. Wolters (1995). Long memory in inflation rates: International evidence. *Journal of Business & Economic Statistics 13*, 37–45.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika 68*, 165–176.

Hsu, C.-C. (2005). Long memory or structural changes: An empirical examination on inflation rates. *Economics Letters 88*, 289–294.

Jensen, A. N. and M. Ø. Nielsen (2014). A fast fractional difference algorithm. *Journal of Time Series Analysis 35*(5), 428–436.

Knopp, K. (1951). *Theory and Application of Infinite Series* (second ed.). Blackie & Son.

Kumar, M. S. and T. Okimoto (2007). Dynamics of persistence in international inflation rates. *Journal of Money, Credit and Banking 39*, 1457–1479.

Lobato, I. N. and N. E. Savin (1998). Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics 16*, 261–268.

Maasoumi, E. and . M. McAleer (2008). Realized volatility and long memory: An overview. *Econometric Reviews 27*, 1–9.

Marinucci, D. and P. M. Robinson (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference 80*, 111–122.

Martins, L. F. and P. M. M. Rodrigues (2014). Testing for persistence change in fractionally integrated models: An application to world inflation rates. *Computational Statistics & Data Analysis 76*, 502–522.

Mureşan, M. (2009). *A Concrete Approach to Classical Analysis*. Springer.

Ohanissian, A., J. R. Russell, and R. S. Tsay (2008). True or spurious long memory? a new test. *Journal of Business & Economic Statistics 26*, 161–175.

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

Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics.* Oxford University Press.

Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics 20*, 971–1001.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics 29*, 423–438.

Shimotsu, K. (2010). Exact local Whittle estimation of fractional integration with unknown mean and trend. *Econometric Theory 26*, 501–540.

Shimotsu, K. and P. C. B. Phillips (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics 33*, 1890–1933.

Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory 15*, 549–582.

Tsay, W.-J. and C.-F. Chung (2000). The spurious regression of fractionally integrated processes. *Journal of Econometrics 96*(1), 155 – 182.