Harmonically Weighted Processes*

Uwe Hassler† and Mehdi Hosseinkouchack

Goethe University Frankfurt‡

December 31, 2018

Abstract

We discuss a model for long memory and persistence 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. Theoretically, the harmonically weighted [HW] process displays less persistence and weaker memory than the classical competitor, fractional integration [FI] of order d. Still, we establish that a test rejects the null hypothesis of $d = 0$ if the process is HW. Similarly, a bias approximation shows that estimators of d will fail to distinguish between HW and FI given realistic sample sizes. The difficulties to disentangle HW and FI are illustrated experimentally and with U.S. inflation data.

MSC: 60F17 (functional limit theorems), 60G10 (stationary processes), 62M10 (time series)

Keywords Long memory; persistence; fractional integration; inflation

---

* **Acknowledgements:** An earlier version of this paper was presented at the Workshop in Time Series Econometrics in Zaragoza, March 30-31, 2017. The authors are grateful to Anindya Banerjee, Matei Demetrescu, Jesús Gonzalo, Javier Hualde, Marc Pohle, to the Co-Editor Morten Nielsen and to three anonymous referees for very helpful comments.

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

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

1 Introduction

It is a stylized fact in different fields of science that many time series display long memory in the sense that the sequence of their autocovariances dies out only slowly, see e.g. Beran, Feng, Ghosh, and Kulik (2013, Sect. 1.2) for empirical examples. The most widely used model to account for long memory is fractional integration [FI] of order d, where the autocovariances at lag h vanish at rate h^{2d-1} as $h \to \infty$, $0 < d < 1/2$. This is mirrored in the frequency domain by the spectrum diverging with λ^{-2d} as λ approaches the origin from the right. The estimation of d has been subject to intense research over the last decades, but there seems to be little consensus on how to proceed in practice. Unfortunately, different researchers therefore measure different degrees of memory from identical data. Therefore, we suggest and discuss in this paper a model of long memory that does not require parameter estimation, namely harmonically weighted processes.

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 Breitung and Hassler (2002, eq. (3)), and it was used to construct a Lagrange Multiplier test for fractional integration. The autocovariances of this process are not summable, see e.g. Palma (2007, Prob. 3.15). In the context of fractional integration testing, Demetrescu, Kuzin, and Hassler (2008) considered more generally the process $\sum_{j=1}^{t-1} j^{-1}x_{t-j}$ where the filtered input $\{x_t\}$ 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 such harmonically weighted processes are characterized by a sequence of square summable autocovariances. Except for these results, little seems to be known about harmonically weighted processes [HWP].

The present paper has two main contributions. First, we discuss the persistence and long memory properties of HWP that differ from the well known features under fractional integration. They are characterized by a singularity in the spectrum at the origin that is of order $\ln^2 \lambda$ for $\lambda \to 0$, and the autocovariances at lag h vanish at rate $\ln h/h$, see Proposition 1. Consequently, it follows in Proposition 2 that a functional central limit theorem [FCLT] requires normalization with $\sqrt{T}\ln T$ where T denotes the sample size. Second, we address the (im)possibility to discriminate between HW and FI. Specifically, Proposition 3 characterizes the behaviour of the Lagrange Multiplier [LM] test by Robinson (1991) under the true null hypothesis $d = d_0$ if a *local* HW component is present. If the HW component is downweighted by T^κ with $\kappa \leq 0.25$, then the test still has nonnegligible power against this weaker form of long memory. At the same time one has to be careful with a rejection of the null, which may not be hastily interpreted as evidence in favour of $d > d_0$. When it comes to estimating d, we use a bias approximation in order to show that disentangling HW and FI is out of reach even for samples of size $T = 10^4$. Finally, we illustrate with real U.S. inflation data that the model of harmonic weighting may in practice do as good a job in accounting for long memory as the model of fractional integration.

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. Further, it discusses the harmonic inverse transformation required for an autoregressive representation. The third section presents the asymptotic theory for partial sums of HWP, containing a nonstandard central limit theorem [CLT]. Section 4 addresses the discrimination between HWP and FI on theoretical grounds. Finite sample evidence from Monte Carlo experiments is provided in Section 5, while Section 6 contains an empirical example. Concluding remarks are offered in the final section. 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, $\stackrel{D}{\to}$ and $\stackrel{p}{\to}$ represent convergence in distribution and in probability, respectively, 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 $h(L)$ 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 of 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. We next 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 $c(L)$,

$$y_t = \mu + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}\,, \quad b_j = \sum_{k=0}^{j} \frac{c_k}{j+1-k}\,, \tag{3}$$

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

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

a) *for the moving average coefficients that*

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

b) *for the spectrum that*

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

c) *and for the autocovariances that*

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

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*

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

$$\gamma_y(h) = \sigma^2 \sum_{j=0}^{\infty} \frac{1}{(j+1)\,(j+1+h)} = \sigma^2 \frac{1}{h} \sum_{j=1}^{h} \frac{1}{j}\,, \quad h > 0\,, \tag{4}$$

see also Palma (2007, Prob. 3.15).

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 $\sum_{h=0}^{H} |\gamma_y(h)| \to \infty$ as $H \to \infty$.[1] For comparison with the traditional long memory model, we briefly recap the well known fractionally integrated [FI] process $\{z_t\}$ of order d, for short $z_t \sim I(d)$, which relies on the fractional integration operator with the usual binomial expansion: $(1-L)^{-d} = \sum_{j=0}^{\infty} \binom{-d}{j} (-L)^j$.

Assumption 2 *Let* $z_t = \mu + (1-L)^{-d} x_t$, $t \in \mathbb{Z}$, $0 \leq d < 1/2$, *where* $\{x_t\}$ *is from Assumption 1.*

This FI process $\{z_t\}$ is often called of type I since the work by Marinucci and Robinson (1999). The Wold decomposition provides $z_t = \mu + \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j}$ with $\beta_j \sim \frac{c(1)}{\Gamma(d)} j^{d-1}$, see Hassler (2019, Lemma 5.4). Note that $\Gamma(x) \sim x^{-1}$ at the origin. Hence, $j^{d-1}/\Gamma(d)$ converges to zero (for fixed j) as $d \to 0$, and HW is not a special case of FI. The spectrum of $\{z_t\}$ becomes $f_z(\lambda) = (2\sin\lambda/2)^{-2d} f_x(\lambda)$, see e.g. Giraitis, Koul, and Surgailis (2012, Prop. 3.2.2). Consequently, it holds for the autocovariances $\gamma_z(h)$ that $\gamma_z(h) \sim C_d\, h^{2d-1}$ for a constant C_d defined in Giraitis et al. (2012, Prop. 3.1.1) or Hassler (2019, Coro. 6.1); note that $C_d \to 0$ as $d \to 0$. Consequently, the persistence of the HWP and its degree of long memory are not as strong as under the assumption of fractional integration:

$$\lim_{j\to\infty} \frac{1/j}{j^{d-1}} = 0\,, \quad \lim_{\lambda\to 0} \frac{\ln^2(\lambda)}{\lambda^{-2d}} = 0 \text{ and } \lim_{h\to\infty} \frac{\ln h/h}{h^{2d-1}} = 0 \quad \text{for } 0 < d < 1/2\,. \tag{5}$$

Although the persistence and memory of FI and HW processes have different qualities asymptotically, matters may be different in finite samples. Assume a sample of size T. One typically estimates spectra at harmonic frequencies $\lambda_j = 2\pi j/T$. For that reason, we plot in Figure 1 spectra of HW noise and of FI noise ($d = 0.3$ and $d = 0.4$) for different T (with $\sigma^2 = 2\pi$), where we focus on frequencies only up to $\pi/4$. For $d = 0.3$, the HW spectrum turns out to be higher than the $I(0.3)$ spectrum even at λ_j close to the origin. For $d = 0.4$,

[1] Note that our definition of long memory follows e.g. Giraitis et al. (2012, eq. (3.1.1)), while Haldrup and Vera Valdés (2017, Def. (i)) define long memory by the hyperbolic decay h^{2d-1}, which is characteristic for fractional integration.

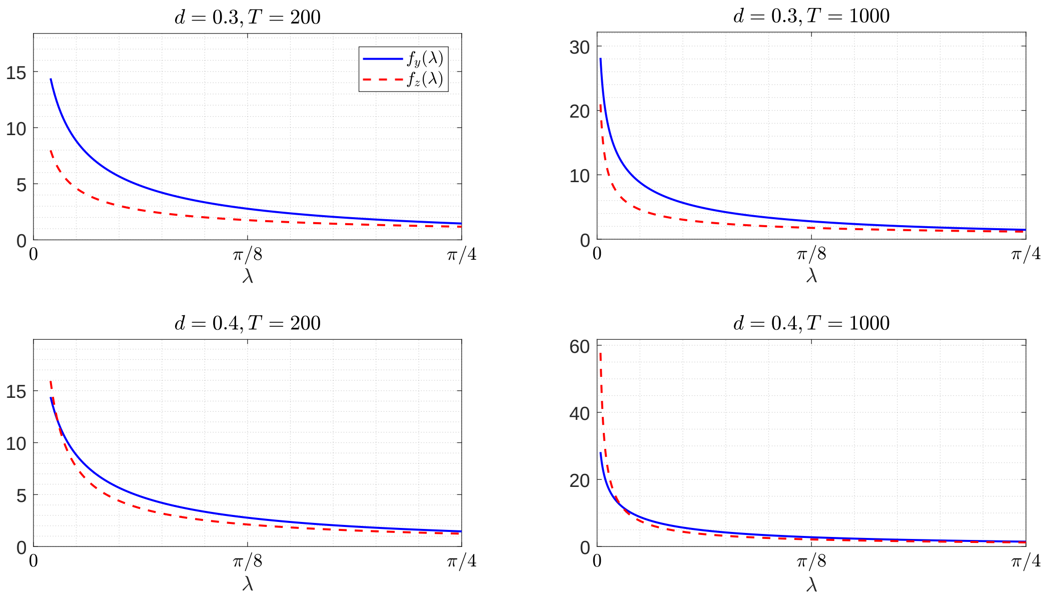

Figure 1: Spectra at $\lambda_j = 2\pi j/T$ for HW noise (solid) and FI noise (dashed) where $0 < \lambda_j \leq \pi/4$

the spectra of the $I(d)$ process (dashed line) and the HW process (solid line) 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 will return to the difficulties to disentangle the two models statistically after the next section.

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

$$g(L) = \frac{1}{h(L)} = -\frac{L}{\ln(1-L)} = 1 - \sum_{j=1}^{\infty} g_j L^j \,, \tag{6}$$

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, $g_j > 0$. It holds 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)):

$$g_j \sim \frac{1}{j \ln^2 j} \quad \text{as } j \to \infty \,. \tag{7}$$

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 with $g(L)$ not only removes the long memory but the mean at the same time. With the absolutely summable filter $g(L)$, it is straightforward to obtain an AR(∞) representation for HWP. Under the additional assumption that the spectrum f_x is strictly positive, we have from Brillinger (1975, pp. 76, 77) that $\{x_t\}$ from Assumption 1 has an AR representation building on

$$\varepsilon_t = \frac{x_t}{c(L)} = \sum_{j=0}^{\infty} a_j x_{t-j} \quad \text{with } \sum_{j=0}^{\infty} j|a_j| < \infty \,.$$

The convolution of $g(L)$ and $(c(L))^{-1} = \sum_{j=0}^{\infty} a_j L^j$ then results in an absolutely summable AR(∞) representation of $\{y_t\}$.

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

$$g_+(L)y_t := g(L)y_t \mathbf{1}_{(t>0)}(t) = y_t - \sum_{j=1}^{t-1} g_j y_{t-j} \,, \quad t = 1, \ldots, T \,. \tag{8}$$

Here, we employ the indicator function

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

Similarly, one may consider h_+ as a truncated version of $h(L)$ and define $\{y_t^+\}$ as follows:

$$y_t^+ := \mu + h_+(L)x_t := \mu + h(L)x_t \mathbf{1}_{(t>0)}(t) = \mu + \sum_{j=0}^{t-1} \frac{x_{t-j}}{j+1} \,, \quad t = 1, \ldots, T \,. \tag{9}$$

This process is only asymptotically stationary. However, it follows from the proof of Lemma 2 in Demetrescu et al. (2008) that

$$y_t - y_t^+ = O_p\left(\frac{1}{\sqrt{t}}\right) . \tag{10}$$

That's why we focus on $\{y_t\}$ from Assumption 1 for the rest of the exposition.

3 (Functional) Central limit theorem

We now turn to large sample properties of the sample mean of HWP. We obtain the behaviour 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 $E\left(\varepsilon_t^2\right) = \sigma^2$ and $E\left(|\varepsilon_t|^p\right) < \infty$ 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*

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

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 $\overline{y} = T^{-1}\sum_{t=1}^{T} y_t$,

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

Although $\text{Var}(\overline{y})$ 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. Still, it is remarkable that the limiting process in *b*) is standard: a Wiener process with independent increments. This contrasts again the case of FI, where the limiting process is a so-called fractional Brownian motion with dependent increments, see Abadir et al. (2014, Coro. 4.1). This reflects that the HWP displays a weaker form of long memory than FI.

To close this section, 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 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 2 we evaluate $\text{Var}(\widetilde{\mu})/\text{Var}(\overline{y})$ for T ranging from 50 up to 2000. 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.

4 HWP versus FI

Now, we turn to the discrimination between HWP and FI. Both features are embedded in the harmonically weighted fractionally integrated [HWFI] process $\{\xi_t\}$,

$$\xi_t = \mu + h(L)(1-L)^{-d} x_t\,, \quad t \in \mathbb{Z}\,,\ 0 \le d < \frac{1}{2}\,, \tag{11}$$

where $\{x_t\}$ is from Assumption 1. The HW process $\{y_t\}$ is a special case of (11) for $d = 0$. It follows for the spectrum of $\{\xi_t\}$ along the lines of the proof of Proposition 1 that

$$f_\xi(\lambda) = \left[\ln^2\left(2\sin\frac{\lambda}{2}\right) + \left(\frac{\pi-\lambda}{2}\right)^2\right]\left(4\sin^2\frac{\lambda}{2}\right)^{-d} f_x(\lambda)\,, \quad \lambda > 0\,,$$

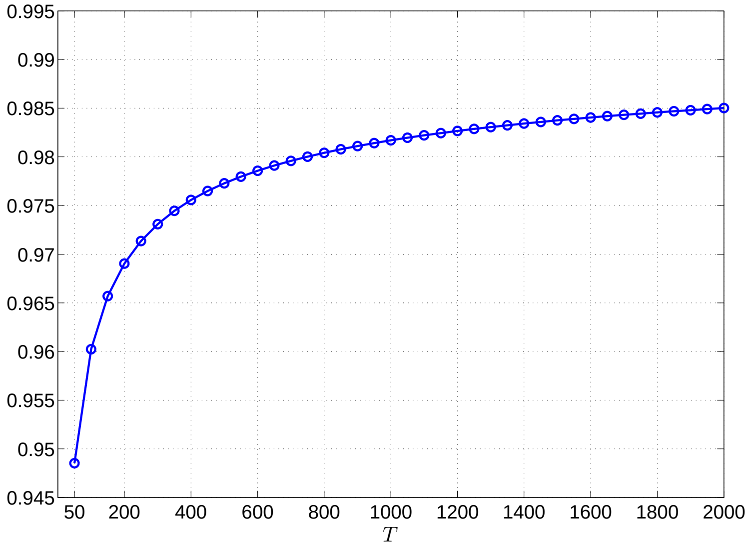

Figure 2: $\mathrm{Var}(\widetilde{\mu})/\mathrm{Var}(\overline{y})$ for $T = 50, \dots, 2000$

and $f_\xi(\lambda) \sim \ln^2(\lambda)\,\lambda^{-2d} f_x(0)$ as $\lambda \to 0$. Model (11) is a special case of the more general case considered in Robinson (2014, eq. (2)), in that $\ln^2\lambda$ is a particular parameterization of a slowly varying function. How do HW and FI interact statistically?

We first consider the LM test for d suggested by Robinson (1991), which is efficient against fractional alternatives, see Robinson (1994). Following Robinson (1994), we assume fractional integration of type II and difference the data under H_0: $d = d_0$, i.e. $\xi_{t,d} := \Delta_+^{d_0}\xi_t$. For $\xi_t \sim I(d_0+\theta)$ this means that $\xi_{t,d} \sim I(\theta)$. With the auxiliary variable $\xi^*_{t-1,d} := h_+(L)\xi_{t-1,d} = \sum_{j=1}^{t-1}\xi_{t-j,d}/j$ the test statistic of the LM test in the time domain becomes

$$t_{LM} := \frac{\sqrt{6}}{\pi\widehat{\sigma}^2}\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\xi_{t,d}\xi^*_{t-1,d}\,, \quad \widehat{\sigma}^2 = T^{-1}\sum_{t=2}^{T}\xi^2_{t,d},$$

see also Tanaka (1999). Under the null hypothesis one has $\xi_{t,d} \sim I(0)$, or H_0: $\theta = 0$. From Robinson (1994, Thm. 2) it follows that the test has power against local fractional alternatives in a Pitman sense. Although the HWP is not $I(\theta)$ with $\theta > 0$, it violates the null $I(0)$ since it displays long memory. Hence, (11) violates the null although the fractional alternative is not true. We now investigate the power under local harmonically weighted processes.

More specifically, we consider the following local type II model:

$$\Delta_+^d \xi_t = \eta_t + T^{-\kappa} h_+(L)\varepsilon_t\,, \quad 0 < \kappa \leq 0.5\,. \tag{12}$$

Then we have the following result.

Proposition 3. *Let* $\{\eta_t\}$ *be an iid process with zero mean and variance* σ_η^2 *independent of the iid process* $\{\varepsilon_t\}$ *with variance* σ_ε^2 *and with finite fourth moments. It then holds under (12) that*

$$t_{LM} \begin{cases} \xrightarrow{D} \mathcal{N}(0,1) & \text{if } \kappa > 0.25 \\ \xrightarrow{D} \mathcal{N}\left(\frac{2\sqrt{6}}{\pi}\frac{\sigma_\varepsilon^2}{\sigma_\eta^2}\zeta(3), 1\right) & \text{if } \kappa = 0.25 \\ \to \infty & \text{if } \kappa < 0.25 \end{cases}$$

as $T \to \infty$, *where* $\zeta(\cdot)$ *is Riemann's zeta function.*

Proof. See Appendix.

Hence, if the HW component is downweighted by $T^{0.25}$, the test has still non-negligible power, and it even rejects with probability one if the local HW component is stronger. On the one hand, this is good news: The LM test designed against fractional alternatives has power against the weaker form of long memory of a HW process, too. On the other hand, this calls for attention: A (one-sided) rejection of $d = d_0$ or $\theta = 0$ is typically interpreted as $d > d_0$ or $\theta > 0$, which is a wrong conclusion under (12), i.e. a HW component will appear as FI(θ).

Second, we turn to the estimation of d when an FI process is perturbed by HW, where we maintain again model (11). Can one estimate d without systematic bias notwithstanding the additional long memory due to HW? The answer will be yes (asymptotically) and no (for samples even of size $T = 10^5$).

Robinson (2014) proved that the widely used log-periodogram regression [LPR] studied by Geweke and Porter-Hudak (1983), Robinson (1995) and Hurvich, Deo, and Brodsky (1998) provides a consistent estimator for d under (11). Using the notation by Robinson (2014), the estimator becomes

$$\widehat{d}_{LPR} = -\frac{\sum_{j=1}^m \nu_j \ln(I_\xi(\lambda_j))}{2\sum_{j=1}^m \nu_j^2}\,, \quad \nu_j := \ln(j) - \ln(m!)/m\,,$$

where I_ξ is the periodogram of ξ_t, $\lambda_j = 2\pi j/T$. With $U_j = \ln(I_\xi(\lambda_j)/f_\xi(\lambda_j))$ it follows that $\widehat{d}_{LPR}$ is made up by three terms:

$$\begin{aligned}\widehat{d}_{LPR} &= \frac{\sum\limits_{j=1}^{m} \nu_j \left(d \ln(4 \sin^2 \frac{\lambda_j}{2}) - \ln \left[\ln^2 \left(2 \sin \frac{\lambda_j}{2} \right) + \left(\frac{\pi - \lambda_j}{2} \right)^2 \right] - U_j \right)}{2 \sum_{j=1}^{m} \nu_j^2} \\ &\approx d + b(m,T) + o_p(1) \xrightarrow{p} d \quad \text{as } \frac{1}{m} + \frac{m}{T} \to 0 \,.\end{aligned}$$

The first terms equals approximately d since $\ln(4 \sin^2 \frac{\lambda_j}{2}) \sim 2(\ln(j) + \ln(2\pi/T))$; the third one depending on U_j is $o_p(1)$ by Robinson (2014, Ass. 1); and the middle term is the approximate bias,

$$b(m,T) := - \frac{\sum\limits_{j=1}^{m} \nu_j \ln \left[\ln^2 \left(2 \sin \frac{\lambda_j}{2} \right) + \left(\frac{\pi - \lambda_j}{2} \right)^2 \right]}{2 \sum_{j=1}^{m} \nu_j^2} \,, \tag{13}$$

which vanishes asymptotically by Robinson (2014, eq. (30)). In Figure 3 we evaluate this bias term for growing T with $m = \lfloor T^{0.65} \rfloor$. Even for $T = 10^4$, the value is above 0.25, and for $T = 10^5$ one has $b(\lfloor T^{0.65} \rfloor, T) = 0.222$. What is more, we will observe in the next section that $b(m,T)$ does a very good job in explaining the bias quantified by means of computer experiments (also for more efficient estimators). Hence, a reliable estimation of d seems to be out of reach for realistic sample sizes.

5 Monte Carlo results

All computer experiments below were executed with MATLAB. The results rely on 10^4 replications.

5.1 Central limit theorem

In this subsection we turn to Proposition 2. As true data generating process [DGP] we consider the case $\mu = 0$, i.e. $y_t = h(L)x_t$, $t = 1, 2, \ldots, T$. In order to simulate a sample from a stationary type I HW process of length T, we generated $T + 5000$ observations from the type II model (9) and discarded the

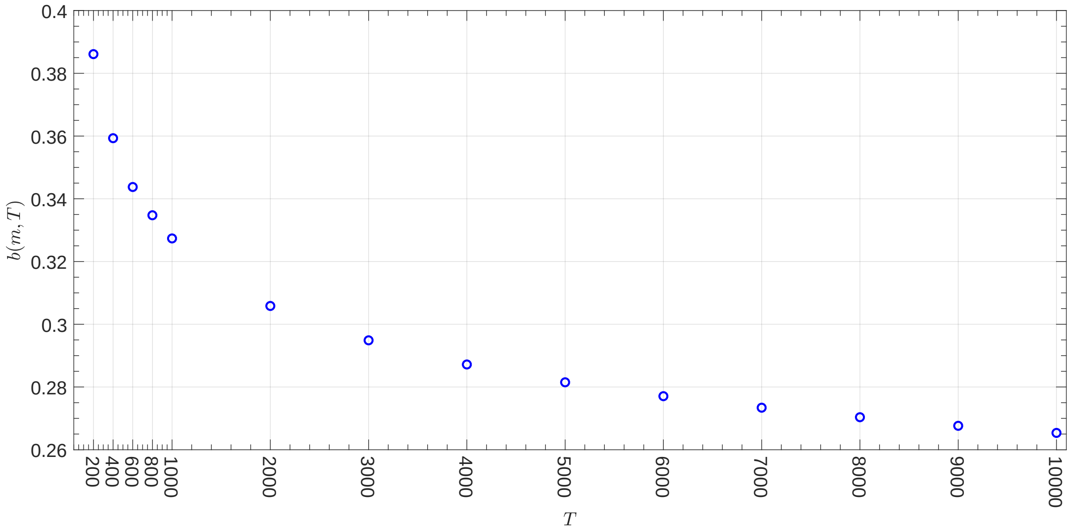

Figure 3: $b(m,T)$ for $T \in \{200, 400, \ldots, 1000, \ldots, 10000\}$ with $m = \lfloor T^{0.65} \rfloor$

first 5000 observations. With the so-called long-run variance $\omega_x^2 = 2\pi f_x(0)$, we define the infeasible test statistic

$$\mathcal{T}_0 := \frac{\sqrt{T}}{\ln T} \frac{\overline{y}}{\omega_x},$$

which is compared in absolute value with $z_{0.975} = 1.96$ for a two-sided test at nominal 5% level. $\mathcal{T}_1$ and $\mathcal{T}_2$ are the corresponding test statistics with ω_x^2 replaced by a consistent estimator. The estimator is computed from $g_+(L)y_t$. For $\mathcal{T}_1$, the estimation builds on a Bartlett kernel with data-driven bandwidth according to Andrews (1991, eq. (5.3)), while the estimation behind $\mathcal{T}_2$ relies on the quadratic spectral kernel advocated by Andrews (1991) with the deterministic bandwidth choice $\lfloor 4(T/100)^{1/4} \rfloor$.

In 5 sets of experiments, the input sequence $\{x_t\} = \{\varepsilon_t\}$ is free of serial correlation. We consider the case of a standard normal distribution $\mathcal{N}(0,1)$ and of a t distribution $t(3)$ with 3 degrees of freedom. Further, $\mathcal{AL}_i$, $i = 1, 2$, represents two asymmetric Laplace distributions with the following distribution

Table 1: Empirical size testing for true $\mu = 0$ at $\alpha = 5\%$

	$\mathcal{N}(0,1)$	$t(3)$	$\mathcal{AL}_1$	$\mathcal{AL}_2$	$GARCH$	$AR(1)$	$MA(9)$
				$T = 250$			
$\mathcal{T}_0$	5.05	4.88	5.13	5.14	5.25	5.23	4.91
$\mathcal{T}_1$	5.64	5.75	6.04	6.05	5.83	8.60	10.39
$\mathcal{T}_2$	5.70	5.89	6.04	5.89	5.57	8.35	13.71
				$T = 500$			
$\mathcal{T}_0$	4.98	4.66	4.52	4.82	4.62	4.79	4.92
$\mathcal{T}_1$	5.35	4.87	5.22	5.40	5.08	6.87	8.80
$\mathcal{T}_2$	5.25	4.84	5.06	5.28	5.02	7.11	13.17
				$T = 1000$			
$\mathcal{T}_0$	4.64	4.67	4.32	4.79	4.76	4.40	4.50
$\mathcal{T}_1$	4.86	4.84	4.75	5.16	4.98	6.10	6.76
$\mathcal{T}_2$	4.82	4.73	4.76	5.23	4.89	5.91	9.01

Note: $\mathcal{T}_0$ is the infeasible statistic computed from $y_t = h(L)x_t$, $\mathcal{T}_1$ and $\mathcal{T}_2$ rely on estimates described in the text. x_t is either free of serial correlation or serially correlated; for a description of the columns see the text.

function

$$F\left(x; m, \lambda, \tau\right) = \begin{cases} \frac{\tau^2}{1+\tau^2} \exp\left(\frac{\lambda}{\tau}\left(x - m\right)\right) & \text{for } x \leq m \\ 1 - \frac{1}{1+\tau^2} \exp\left(-\lambda\tau\left(x - m\right)\right) & \text{for } x > m \end{cases} .$$

For $i = 1$ and $i = 2$, we chose $m_1 = \frac{3}{2}$, $\lambda_1 = 1$, $\tau_1 = 2$ and $m_2 = -\frac{3}{2}$, $\lambda_2 = 1$, $\tau_2 = \frac{1}{2}$, respectively. The expected values are 0 in both cases, and the skewness coefficients amount to $-\frac{126}{17\sqrt{17}}$ and $\frac{126}{17\sqrt{17}} \approx 1.8$, respectively. Further, we consider stationary GARCH(1,1) innovations with (using standard notation) $\alpha_0 = 0.05$, $\alpha_1 = 0.10$ and $\beta_1 = 0.85$ where the underlying shocks are iid and distributed according to a standard Gaussian law. The corresponding kurtosis is 3.77. Finally, we allow $\{x_t\}$ to be serially correlated; first as AR(1) with $x_t = 0.5x_{t-1} + \varepsilon_t$, ε_t is iid $\mathcal{N}(0,1)$, and second as MA(9): $x_t = \varepsilon_t + 0.9\varepsilon_{t-1} + \cdots + (1 - {}^{9}\!/_{10})\varepsilon_{t-9}$.

Table 1 contains the rejection rates for different sample sizes. Under all circumstances where $\{x_t\} = \{\varepsilon_t\}$, no notable size distortions show up. For the mildly persistent AR(1) case, mild distortions are observed when ω_x has to be estimated from small samples; in the slightly more persistent MA(9) case, the distortions are slightly stronger. All in all, we find that Proposition 2 provides a reliable guideline for finite sample inference under a variety of realistic distributional and dynamic assumptions.

5.2 Discrimination between I(d) and HW

In this subsection, we have the model (11) in mind and focus without loss of generality on $d = 0$. More specifically, we maintain (12) while fixing $\kappa = 0$: $y_t = h_+(L)\varepsilon_t$. Hence, the DGP is now of type II, which is conformable with the assumption in Robinson (1994).

To begin with, we turn to Proposition 3 and quantify the effect of the presence of a HW component on tests for FI in finite samples. To mimic a realistic practical strategy, we allow for additional short memory employing the augmented LM [ALM] test by Demetrescu, Kuzin, and Hassler (2008). This version of the test is executed by regressing $\xi_{t,d}$ on the auxiliary regressor $\xi^*_{t-1,d}$ and k endogenous lags, $\xi_{t-j,d}$, $j = 1, \ldots, k$. The (absolute) value of the t statistic testing for insignificance of $\xi^*_{t-1,d}$ is compared with the standard normal. Following the recommendation by Demetrescu et al. (2008), we choose $k = \lfloor 4(T/100)^{1/4} \rfloor$. Note that a data driven lag selection e.g. with information criteria cannot be advised due a devastating post model selection effect, see Demetrescu, Hassler, and Kuzin (2011) for a quantification. In Table 2 we report rejection frequencies at nominal 5% level for a selection of T between 200 and 10^4. Trivially, the one-sided test rejects more often than the two-sided version. The rejection rates grow with T. This can be read as increasing power when having the null hypothesis in mind, H_0: $y_t \sim I(0)$, which is violated under HW. At the same time a word of warning is due. A one-sided rejection of $d = d_0$, must not be hastily interpreted as evidence in favour $d > d_0$, since it may as well result from the presence of HW.

Table 2: ALM test at nominal size 5%

T	200	400	600	800	1000	2000	5000	10000
one-sided	20.77	31.26	37.36	47.71	49.02	72.09	93.43	99.22
two-sided	14.23	22.36	27.34	36.88	37.87	61.90	88.71	98.30

Note: Frequency of rejections testing for true $d = 0$ from HW noise, $y_t = h_+(L)\varepsilon_t$

Next, we relate to the bias approximation when estimating d, see (13). For Figure 4, we estimated the order of integration by means of the LPR and by

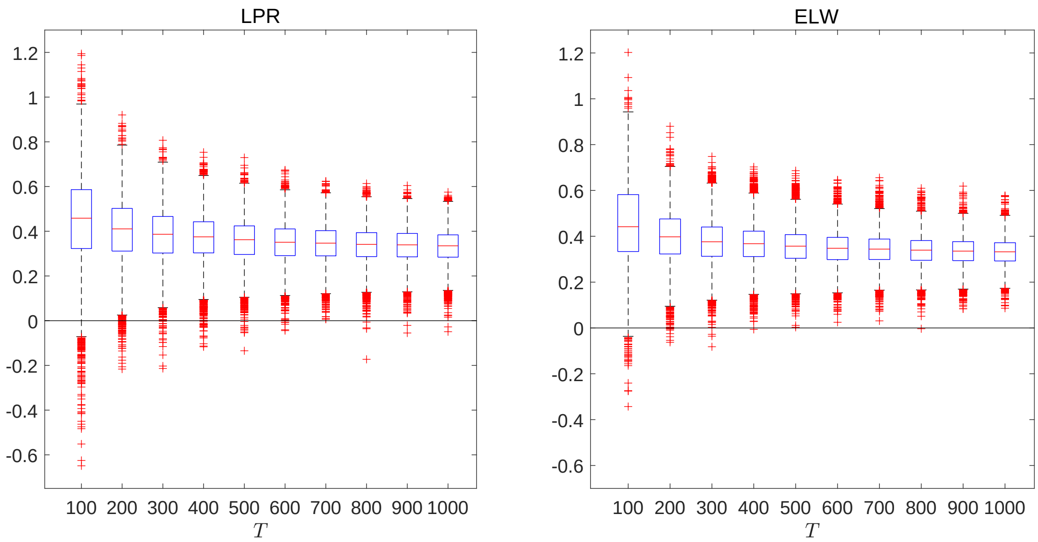

Figure 4: Box plots of estimates of $d = 0$; the true process is $y_t = h(L)\varepsilon_t$

means of the more efficient exact local Whittle [ELW] estimator $\widehat{d}_{ELW}$ proposed by Shimotsu and Phillips (2005) and Shimotsu (2010) with bandwidth $m = \lfloor T^{0.65} \rfloor$. The true DGP is now again of type I as in the previous subsection: $y_t = h(L)\varepsilon_t$. Asymptotically, the estimates should concentrate around the true $d = 0$. In finite samples, however, things are quite different. In Figure 4 we present Box plots of the estimates. For $T = 100$, the median is well above 0.4, for $T = 1000$, the median is roughly 0.33, which well corresponds to the approximating values in Figure 3. Figure 5 presents experimental evidence that also for $T = 10^4$ the mean and median closely relate to $b(m, 10^4)$. What is more, out of 10^4 experiments, *all* estimates behind Figure 5 are larger than the true value $d = 0$. This means for realistic sample sizes in practice that a fractional specification will be mislead for sure under HW.

6 Empirical example: U.S. inflation

Granger (1980) argued that the aggregation of individual (price) series may result in an index that is fractionally integrated. Consequently, Granger and Joyeux (1980) studied as an empirical example for fractional integration the

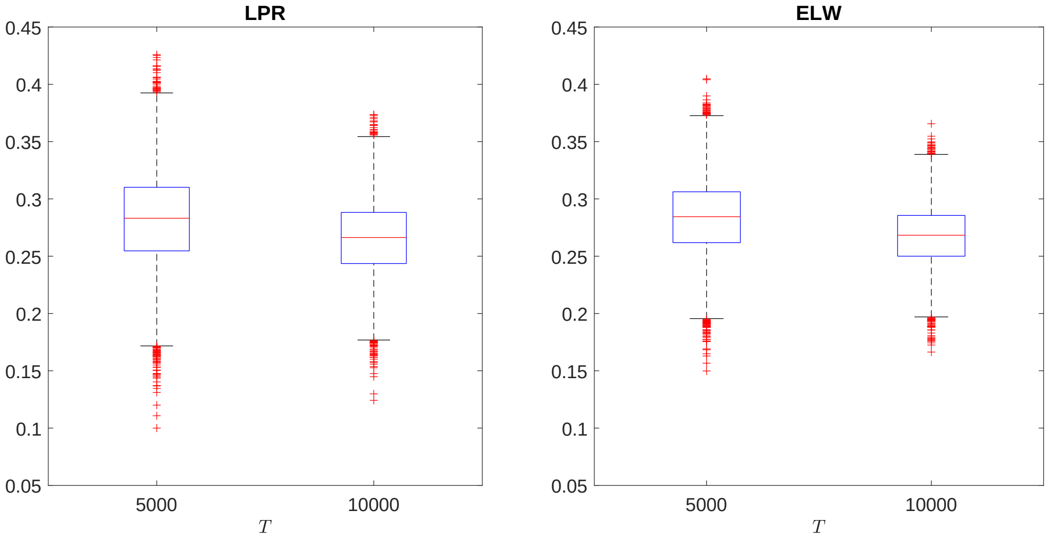

Figure 5: See Figure 4

monthly U.S. index of consumer food prices. More systematically, Geweke and Porter-Hudak (1983) applied fractional integration to different U.S. price indices. Their work triggered independent studies on long memory in inflation by Delgado and Robinson (1994), Hassler and Wolters (1995) and Baillie, Chung, and Tieslau (1996). Long memory in inflation is sometimes considered as a stylized fact supported by abundant evidence over the last decades.

Let P_t stand for the seasonally adjusted monthly consumer price index from December 1969 until August 2017, more precisely: Consumer Price Index for all urban consumers (all items), retrieved from the Federal Reserve Bank of St. Louis. The inflation series is computed as $\pi_t = 100\,(P_t - P_{t-1})/P_{t-1}$, $t = 1, \dots, T = 572$, see the northwestern graph in Figure 6. The sample autocorrelogram in the northeastern graph is indicative of long memory with $\widehat{\rho}_\pi(h) > 0.3$ up to $h = 20$. 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 differencing parameter is $\widehat{d} = 0.43$ when estimated by exact local Whittle [ELW] with bandwidth $\lfloor T^{0.65} \rfloor = 61$. 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 6 (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,

$$Q_{dif}(25) = T \sum_{h=1}^{25} \left(\widehat{\rho}_{dif}(h)\right)^2 = 69.74 \quad \text{and} \quad Q_{hit}(25) = T \sum_{h=1}^{25} \left(\widehat{\rho}_{hit}(h)\right)^2 = 70.11\,.$$

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 $Q_{dif}(25)$ and $Q_{hit}(25)$ are both approximately equal to 70, suggesting 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 model is radically more simple, it does not require to choose an estimator of d, and it does not require to pick a bandwidth m. Further, note that the semiparametric estimation of d is plagued by large variances. For ELW one obtains as approximate confidence interval at 95% level $[\widehat{d} \pm 1.96/15.62] = [0.3045, 0.5555]$.

7 Concluding remarks

From Proposition 1 we learn that HW processes are strongly persistent and display long memory in the sense that the moving average coefficients and the autocovariances are not summable. Still, by (5) the strength of persistence and the length of memory are of a different, weaker quality than with the traditional model of fractional integration. One might be tempted to introduce a new category of "weak long memory" or "intermediate memory" to characterize HW. We think there is no need to do so, since Proposition 2 helps to clarify what distinguishes HW from FI. Remember the concept of summability 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,

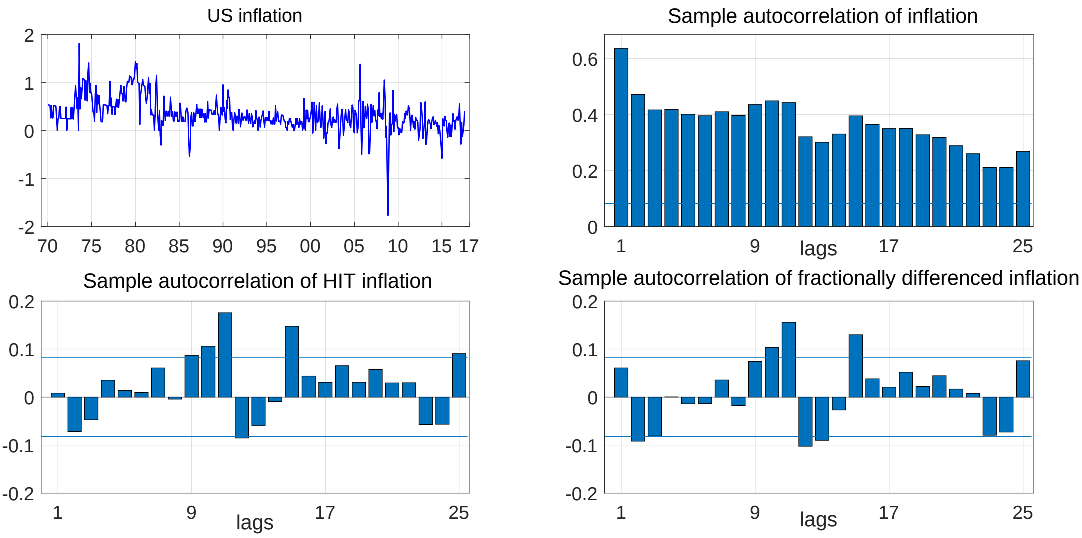

Figure 6: U.S. inflation

if δ is the minimum number such that

$$\frac{L(T)}{T^{\delta}\sqrt{T}} \sum_{t=1}^{T} (\xi_t - \mu) = O_p(1)\,. \tag{14}$$

Since $1/\ln T$ is slowly varying at infinity, Proposition 2 implies that the HWP $\{y_t\}$ is summable of order $\delta = 0$. Once more, this contrasts the long memory FI case: If $d > 0$, then the FI process is summable of order d, see Berenguer-Rico and Gonzalo (2014, Prop. 1). In that sense, the HWP fills a gap between short memory processes as characterized in Assumption 1 and FI processes from Assumption 2. The process $\{x_t\}$ from Assumption 1 has short memory and is summable of order 0, and the process $\{z_t\}$ from Assumption 2 with $d > 0$ has long memory and is summable of order d; the HWP $\{y_t\}$ is in-between: it has long memory but is summable of order 0.

The question whether it is possible to discriminate between the different rates of memory of HW and FI processes comes up naturally. From Proposition 3 we learn the following for a process $\{\xi_t\}$. Assume that the hypothetical order of integration d_0 equals the true one, and is removed from the data: $\Delta^{d_0}\xi_t$. If these differences are tested for short memory, I(0), then the widely used LM test will reject with high probability in the presence of HW, which is supported by Monte Carlo evidence. Upon rejection, one would thus like to unveil the nature of the remaining memory, i.e. to know whether $\{\xi_t\}$ is FI with $d = d_0 + \theta$, $\theta > 0$, or whether it is a harmonically weighted FI(d_0) process, see (11). When applying the log-periodogram regression to $\Delta^{d_0}\xi_t$, we know from Robinson (2014) that the estimator of the order of integration will converge to the true value of zero, asymptotically, notwithstanding the fact that $\Delta^{d_0}\xi_t$ displays long memory under HW. However, the bias approximation (13) shows that this convergences is incredibly slow. Even with $T = 10^5$ observations the estimation will be misleading. This is supported by Monte Carlo evidence, and for the so-called exact local Whittle estimator, too. Hence, in practice we see little chances to disentangle HW and FI.

FI processes offer an overwhelming flexibility in modelling persistence and long memory. This is a virtue and a burden at the same time: on the one hand, there is a continuity of long memory depending on the order of integration d, but on the other hand the estimation of d is notoriously difficult and troubled by large variances of slowly converging semiparametric estimators. With U.S. inflation data we illustrate that a HW process may be just as able to capture persistence as the more involved FI model. The admitted simplicity and rigidity of the HW model, which does not allow - or require - to choose a memory parameter, may turn out to be a practical advantage in applied work, and empirical researchers may prefer the HW model without having to choose an estimator of d, which typically is plagued by the need of further decisions like picking a bandwidth. Of course, we need more empirical evidence to learn whether and for what values of d and in which fields of application HW may be a serious competitor to FI.

There are further open issues. First, one may wish to step beyond the univariate model and consider a 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:*

$$If \lim_{n\to\infty} \frac{s_{n+1} - s_n}{\sigma_{n+1} - \sigma_n} = \ell \,, \quad then \quad \lim_{n\to\infty} \frac{s_{n+1}}{\sigma_{n+1}} = \ell \,. \tag{15}$$

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):

$$\sum_{h=1}^{T} f(h) = \int_1^T f(x)\,\mathrm{d}x + \frac{1}{2}\left(f(T) + f(1)\right) + \frac{1}{12}\left(f^{(1)}(T) - f^{(1)}(1)\right) + R\,, \tag{16}$$

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

$$\int_1^T \left|f^{(3)}(x)\right|\,\mathrm{d}x \le 1 + \frac{13}{3}T - \frac{16}{3T^2} - 2\frac{\ln T}{T^2}\,.$$

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

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

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

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

Consequently, $j \sum_{k\leq j/2} c_k/(j+1-k) \to \sum_{k=0}^{\infty} c_k$ for $j \to \infty$, 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 (15) 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 (4). 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 (15) it hence follows that

$$\frac{s_{T-1} - s_{T-2}}{\sigma_{T-1} - \sigma_{T-2}} \;\to\; 2\pi f_x\,(0) \;.$$

Again by (15), 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 $\sigma_T^2 = \mathrm{Var}(S_T(1))$. Then we first establish the convergence of the finite dimensional distributions of $\sigma_T^{-1}S_T(r)$ for $0 \leq r \leq 1$. To do so we first observe that

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

For brevity define $a_{t-1} = \sum_{m=0}^{t-1} b_m$ with b_m from (3). With $S_j = \sum_{t=1}^{j} y_t = \sum_{t=1}^{j} a_{t-1}\varepsilon_t$ we easily see for $j \geq k$ that $\mathrm{Cov}(S_j, S_k) = \mathrm{Var}(S_k)$, since

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

Therefore, following the same steps as in the proof of Proposition 3.1 in Abadir et al. (2014), we may conclude that

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

where $\overset{fdd}{\to}$ 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 $\mathrm{E}(|\varepsilon_t|^p) < \infty$ for some $p > 2$. Note that with some positive constant c

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

$$\leq \quad c\left|\frac{\lfloor rT\rfloor}{T}-\frac{\lfloor sT\rfloor}{T}\right|^{\frac{p}{2}},$$

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

Our proof of Proposition 3 builds on a technical lemma that we establish first.

Lemma B. *Let $y_t = h_+(L)\varepsilon_t$ and $y^*_{t-1} := h_+(L)y_{t-1}$. For $\{\varepsilon_t\}$ being iid$(0, \sigma^2_\varepsilon)$ with finite fourth moments, it holds that*

$$S_T := \frac{1}{T}\sum_{t=2}^{T} y_t y^*_{t-1} \ \xrightarrow{p}\ 2\sigma^2_\varepsilon \zeta(3)$$

as $T \to \infty$.

Proof. First, we determine an expression for the moving average coefficients of y^*_{t-1}. To that end, rewrite

$$y_j = \sum_{k=1}^{j}\frac{\varepsilon_k}{j-k+1} \text{ and } y^*_{t-1} = \sum_{j=1}^{t-1}\frac{y_j}{t-j},$$

such that

$$\begin{aligned} y^*_{t-1} &= \sum_{j=1}^{t-1}\sum_{k=1}^{j}\frac{1}{(t-j)(j-k+1)}\varepsilon_k \\ &= \sum_{k=1}^{t-1}\sum_{j=k}^{t-1}\frac{1}{(t-j)(j-k+1)}\varepsilon_k. \end{aligned}$$

Hence, the coefficients are

$$\sum_{j=k}^{t-1}\frac{1}{(t-j)(j-k+1)} \quad = \quad \sum_{j=k}^{t-1}\frac{(t-j)+(j-k+1)}{(t-j)(j-k+1)}\frac{1}{t-k+1}$$

$$
\begin{aligned}
&= \frac{1}{t-k+1}\left(\sum_{j=k}^{t-1}\frac{1}{j-k+1}+\sum_{j=k}^{t-1}\frac{1}{t-j}\right) \\
&= \frac{2}{t-k+1}\sum_{j=1}^{t-k}\frac{1}{j}=\frac{2H(t-k)}{t-k+1},
\end{aligned}
$$

where $H(j)=\sum_{k=1}^{j}\frac{1}{k}$ denotes the j^{th} harmonic number.

Second, we derive the limit of $\mathrm{E}(S_T)$. To that end, consider

$$
\mathrm{E}\left(y_t y_{t-1}^*\right)=2\sigma_\varepsilon^2\sum_{k=1}^{t-1}\frac{H(t-k)}{(t-k+1)^2}=2\sigma_\varepsilon^2\sum_{k=1}^{t-1}\frac{H(k)}{(k+1)^2}\,.
$$

Next, note that

$$
\begin{aligned}
\sum_{k=1}^{t-1}\frac{H(k)}{(k+1)^2} &= \sum_{k=1}^{t-1}\frac{H(k+1)}{(k+1)^2}-\sum_{k=1}^{t-1}\frac{1}{(k+1)^3} \\
&= \sum_{k=1}^{t}\frac{H(k)}{k^2}-\sum_{k=1}^{t}\frac{1}{k^3} \\
&\to 2\zeta(3)-\zeta(3)\,,\quad \text{as } t\to\infty\,,
\end{aligned}
$$

where the second limit follows by definition, and the first limit is taken from Borwein and Borwein (1995, p. 1195), who attribute this result (and a generalization thereof) to Euler: $\sum_{j=1}^{\infty}\frac{H(j)}{j^n}=\left(1+\frac{n}{2}\right)\zeta(n+1)-\frac{1}{2}\sum_{k=1}^{n-2}\zeta(k+1)\zeta(n-k)$ for $n=2,3,\ldots$. We thus have $\mathrm{E}\left(y_t y_{t-1}^*\right)\to 2\sigma_\varepsilon^2\zeta(3)$, which implies that $\mathrm{E}(S_T)\to 2\sigma_\varepsilon^2\zeta(3)$.

Third, we are left with the second moment: $\mathrm{E}(S_T^2)=T^{-2}\sum_{t=2}^{T}\sum_{s=2}^{T}\mathrm{E}\left(y_t y_{t-1}^* y_s y_{s-1}^*\right)$. We proceed by analyzing

$$
\begin{aligned}
\mathrm{E}\left(y_t y_{t-1}^* y_s y_{s-1}^*\right) &= \sum_{k=1}^{t}\sum_{j=1}^{t-1}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}\left(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m\right) \\
&= \sum_{k=1}^{s}\sum_{j=1}^{s}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}\left(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m\right) \\
&\quad +\sum_{k=s+1}^{t}\sum_{j=s+1}^{t-1}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}\left(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{k=1}^{s}\sum_{j=s+1}^{t-1}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m)\\
&+\sum_{k=s+1}^{t}\sum_{j=1}^{s}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m)\\
=\;&\sum_{k=1}^{s}\sum_{j=1}^{s}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m)\\
&+\sum_{k=s+1}^{t}\sum_{j=s+1}^{t-1}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m)\\
=\;&S_1(t,s)+S_2(t,s)\ ,\quad \text{say}\,.
\end{aligned}
$$

To simplify the summations we consider first the case when $t>s$. For $S_1(t,s)$ we have

$$
\begin{aligned}
S_1(t,s)\;=\;&\sum_{k=1}^{s}\sum_{j=1}^{s}\sum_{l=1}^{s}\sum_{m=1}^{s-1}\frac{4H(t-j)H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)}\mathrm{E}(\varepsilon_k\varepsilon_j\varepsilon_l\varepsilon_m)\\
=\;&\sum_{k=1}^{s-1}\frac{4H(t-k)H(s-k)}{(t-k+1)^2(s-k+1)^2}E\left(\varepsilon_k^4\right)\\
&+\sum_{m=1}^{s-1}\frac{4H(t-s)H(s-m)}{(t-s+1)(t-s+1)(s-m+1)(s-m+1)}\mathrm{E}(\varepsilon_s\varepsilon_s\varepsilon_m\varepsilon_m)\\
&+\sum_{m=1}^{s-1}\frac{4H(t-m)H(s-m)}{(t-m+1)(t-s+1)(s-s+1)(s-m+1)}\mathrm{E}(\varepsilon_s\varepsilon_m\varepsilon_s\varepsilon_m)\\
&+\sum_{m=1}^{s-1}\frac{4H(t-s)H(s-m)}{(t-s+1)(t-m+1)(s-s+1)(s-m+1)}\mathrm{E}(\varepsilon_m\varepsilon_s\varepsilon_s\varepsilon_m)\,,\\
&+\sum_{k=1}^{s-1}\sum_{j=1,j\neq k}^{s-1}\frac{4H(t-j)H(s-k)}{(t-j+1)(t-k+1)(s-j+1)(s-k+1)}\mathrm{E}\left(\varepsilon_k^2\right)\mathrm{E}\left(\varepsilon_j^2\right)\\
&+\sum_{k=1}^{s}\sum_{j=1,j\neq k}^{s-1}\frac{4H(t-j)H(s-j)}{(t-j+1)(t-k+1)(s-k+1)(s-j+1)}\mathrm{E}\left(\varepsilon_j^2\right)\mathrm{E}\left(\varepsilon_k^2\right)\\
&+\sum_{k=1}^{s-1}\sum_{l=1,l\neq k}^{s-1}\frac{4H(t-k)H(s-l)}{(t-k+1)^2(s-l+1)^2}\mathrm{E}\left(\varepsilon_k^2\right)\mathrm{E}\left(\varepsilon_l^2\right)\\
=\;&S_{1,1}(t,s)+S_{1,2}(t,s)+S_{1,3}(t,s)+S_{1,4}(t,s)+S_{1,5}(t,s)+S_{1,6}(t,s)\\
&+S_{1,7}(t,s)\,,
\end{aligned}
$$

where the first counts $k = j = l = m$, the second term counts $k = j = s \neq l = m$, the third term counts $k = l = s \neq j = m$, the fourth term counts $j = l = s \neq k = m$, the fifth term counts $k = m \neq j = l$, the sixth term counts $k = l \neq j = m$ and the last term counts $k = j \neq l = m$. We begin with $S_{1,1}(t,s)$:

$$\begin{aligned}
\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,1}(t,s) &= \frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\frac{4H(t-k)\,H(s-k)}{(t-k+1)^2\,(s-k+1)^2}\,\mathrm{E}\left(\varepsilon_k^4\right) \\
&\leq \frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\frac{4H(t-k)\,H(s-k)}{(t-k+1)\,(s-k+1)^2}\,\mathrm{E}\left(\varepsilon_k^4\right) \\
&= \frac{4\mu_4}{T^2}\sum_{s=2}^{T}\sum_{k=1}^{s-1}\sum_{t=s+1}^{T}\frac{H(t-k)}{(t-k+1)}\frac{H(s-k)}{(s-k+1)^2} \\
&\leq \frac{4\mu_4}{T^2}\sum_{s=2}^{T}\sum_{k=1}^{s-1}\sum_{t=1}^{T}\frac{H(t)}{t}\frac{H(s-k)}{(s-k+1)^2} \\
&= \frac{4\mu_4}{T^2}\sum_{t=1}^{T}\frac{H(t)}{t}\sum_{s=2}^{T}\sum_{k=1}^{s-1}\frac{H(s-k)}{(s-k+1)^2},
\end{aligned}$$

where $\mu_4 = \mathrm{E}\left(\varepsilon_t^4\right)$. Now, as $\sum_{t=1}^{T}\frac{H(t)}{t} = O\left(\log^2 T\right)$ and as $\sum_{s=2}^{T}\sum_{k=1}^{s-1}\frac{H(s-k)}{(s-k+1)^2} = O(T)$ we have that

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,1}(t,s) = O\left(\frac{\ln^2 T}{T}\right). \tag{17}$$

Next, we look at $S_{1,2}(t,s)$:

$$\begin{aligned}
\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,2}(t,s) &= \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{m=1}^{s-1}\frac{H(t-s)\,H(s-m)}{(t-s+1)^2\,(s-m+1)^2} \\
&= \frac{\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=1}^{T-s}\sum_{m=1}^{s-1}\frac{H(t)\,H(m)}{(t+1)^2\,(m+1)^2} \\
&= \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=1}^{T-s}\frac{H(t)}{(t+1)^2}\sum_{m=1}^{s-1}\frac{H(m)}{(m+1)^2} \\
&\leq \frac{4\sigma_\varepsilon^4\zeta(3)^2}{T}.
\end{aligned}$$

Hence,

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,2}(t,s) = O\left(\frac{1}{T}\right). \tag{18}$$

For $S_{1,3}$ we have

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,3}(t,s) = \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{m=1}^{s-1}\frac{H(t-m)\,H(s-m)}{(t-m+1)(t-s+1)(s-m+1)}.$$

Here noting that

$$\begin{aligned}\sum_{m=1}^{s-1}\frac{1}{(t-m+1)(s-m+1)} &= \frac{1}{t-s}\left(\sum_{m=1}^{s-1}\frac{1}{s-m+1}-\sum_{m=1}^{s-1}\frac{1}{t-m+1}\right)\\ &\leq \frac{1}{t-s}\sum_{m=1}^{s}\frac{1}{m},\end{aligned}$$

and as $H(s-m) < H(t-m) \leq C\ln T$, we obtain

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,3}(t,s) \leq C\frac{\ln^3 T}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\frac{1}{(t-s)^2}.$$

Therefore,

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,3}(t,s) = O\left(\frac{\ln^3 T}{T}\right). \tag{19}$$

For $S_{1,4}(t,s)$ we have

$$\begin{aligned}\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,4}(t,s) &= \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{m=1}^{s-1}\frac{H(t-s)\,H(s-m)}{(t-s+1)(t-m+1)(s-m+1)}\\ &\leq \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{m=1}^{s-1}\frac{H(t-s)\,H(s-m)}{(t-s+1)(s-m+1)^2}\\ &= \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=1}^{T-s}\sum_{m=1}^{s-1}\frac{H(t)\,H(m)}{(t+1)(m+1)^2}.\end{aligned}$$

Using $\sum_{m=1}^{s-1}\frac{H(m)}{(m+1)^2}\to\zeta(3)$ and as $\frac{1}{\ln^2 T}\sum_{t=1}^{T}\frac{H(t)}{t+1} = O(1)$, we have that

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T} S_{1,4}(t,s) = O\left(\frac{\ln^2 T}{T}\right). \tag{20}$$

We now turn to $S_{1,5}(t,s)$:

$$\begin{aligned}\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}S_{1,5}(t,s) &= \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\sum_{j=1,j\neq k}^{s-1}\frac{H(t-j)\,H(s-k)}{(t-j+1)(t-k+1)(s-j+1)(s-k+1)}\\ &\leq \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\sum_{j=1}^{s-1}\frac{H(t-j)\,H(s-k)}{(t-j+1)(t-k+1)(s-j+1)(s-k+1)}\\ &= \frac{4\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\frac{H(s-k)}{(t-k+1)(s-k+1)}\sum_{j=1}^{s-1}\frac{H(t-j)}{(t-j+1)(s-j+1)}.\end{aligned}$$

Here, note that $H(s-k)\leq C\ln T$ and $H(t-j)\leq C\ln T$ and

$$\sum_{k=1}^{s-1}\frac{1}{(t-k+1)(s-k+1)} = \frac{1}{t-s}\left(\sum_{k=1}^{s-1}\frac{1}{s-k+1}-\sum_{k=1}^{s-1}\frac{1}{t-k+1}\right)\leq\frac{C\ln T}{t-s}$$

for a generic constant C, such that

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}S_{1,5}(t,s) \leq C\frac{\ln^2 T}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\frac{1}{(t-s)^2}=O\left(\frac{\ln^2 T}{T}\right). \quad (21)$$

$S_{1,6}(t,s)$ is the same as $S_{1,5}(t,s)$:

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}S_{1,6}(t,s)=O\left(\frac{\ln^2 T}{T}\right). \quad (22)$$

Further, we have

$$\begin{aligned}\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}S_{1,7}(t,s) &\leq \frac{\sigma_\varepsilon^4}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\sum_{l=1}^{s-1}\frac{4H(t-k)\,H(s-l)}{(t-k+1)^2(s-l+1)^2}\\ &\leq \frac{C\ln^2 T}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\sum_{l=1}^{s-1}\frac{1}{(t-k+1)^2(s-l+1)^2}\\ &= \frac{C\ln^2 T}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\frac{1}{(t-k+1)^2}\sum_{l=1}^{s-1}\frac{1}{(s-l+1)^2}\\ &\leq \frac{C\ln^2 T}{T^2}\sum_{s=2}^{T}\sum_{t=s+1}^{T}\sum_{k=1}^{s-1}\frac{1}{(t-k+1)^2},\end{aligned}$$

as $\sum_{l=1}^{s-1} \frac{1}{(s-l+1)^2} \leq \frac{\pi^2}{6}$. Noting that

$$\int_{s=2}^{T} \int_{t=s+1}^{T} \int_{k=1}^{s-1} \frac{\mathrm{d}s \, \mathrm{d}t \, \mathrm{d}k}{(t-k+1)^2} = 4 - 2T + 2\ln T + \ln \frac{T+1}{12} + T \ln \frac{T+1}{3},$$

we obtain

$$\frac{1}{T^2} \sum_{s=2}^{T} \sum_{t=s+1}^{T} S_{1,7}(t,s) = O\left(\frac{\ln^3 T}{T}\right). \tag{23}$$

Using (17)-(23) and noting that all the terms with $s \geq t$ behave similarly to those with $s < t$ lead to

$$\frac{1}{T^2} \sum_{s=2}^{T} \sum_{t=2}^{T} S_1(t,s) = O\left(\frac{\ln^3 T}{T}\right). \tag{24}$$

Now, we turn to $\frac{1}{T^2} \sum_{s=2}^{T} \sum_{t=s+1}^{T} S_2(t,s)$ and note that

$$\begin{aligned} S_2(t,s) &= \sum_{k=s+1}^{t} \sum_{j=s+1}^{t-1} \sum_{l=1}^{s} \sum_{m=1}^{s-1} \frac{4H(t-j)\,H(s-m)}{(t-j+1)(t-k+1)(s-l+1)(s-m+1)} \mathrm{E}(\varepsilon_k \varepsilon_j \varepsilon_l \varepsilon_m) \\ &= \sigma_\varepsilon^4 \sum_{j=s+1}^{t-1} \sum_{m=1}^{s-1} \frac{4H(t-j)\,H(s-m)}{(t-j+1)^2 (s-m+1)^2}. \end{aligned}$$

Hence, we need to analyze

$$\begin{aligned} \frac{1}{T^2} \sum_{s=2}^{T} \sum_{t=s+1}^{T} S_2(t,s) &= \frac{4\sigma_\varepsilon^4}{T^2} \sum_{s=2}^{T} \sum_{t=s+1}^{T} \sum_{j=s+1}^{t-1} \sum_{m=1}^{s-1} \frac{H(t-j)\,H(s-m)}{(t-j+1)^2 (s-m+1)^2} \\ &= \frac{4\sigma_\varepsilon^4}{T^2} \sum_{s=2}^{T} \sum_{t=s+1}^{T} \sum_{j=1}^{t-s} \sum_{m=1}^{s-1} \frac{H(j)\,H(m)}{(j+1)^2 (m+1)^2} \\ &= \frac{4\sigma_\varepsilon^4}{T^2} \sum_{s=2}^{T} \sum_{t=3}^{T-s} \sum_{j=1}^{t} \sum_{m=1}^{s-1} \frac{H(j)\,H(m)}{(j+1)^2 (m+1)^2} \\ &\to \frac{1}{2} \left(2\sigma_\varepsilon^2 \zeta(3)\right)^2. \end{aligned}$$

Because of $\sum_{m=1}^{s} \frac{H(m)}{(m+1)^2} - \zeta(3) = O\left(\frac{\ln T}{T}\right)$, we have that

$$\frac{1}{T^2} \sum_{s=2}^{T} \sum_{t=s+1}^{T} S_2(t,s) = \frac{1}{2} \left(2\sigma_\varepsilon^2 \zeta(3)\right)^2 + O\left(\frac{\ln T}{T}\right),$$

which in turn implies that

$$\frac{1}{T^2}\sum_{s=2}^{T}\sum_{t=2}^{T} S_2(t,s) = \left(2\sigma_\varepsilon^2 \zeta(3)\right)^2 + O\left(\frac{\ln T}{T}\right). \tag{25}$$

Using (24) and (25) we hence have that $\mathrm{Var}(S_T) = O\left(\frac{\ln^3 T}{T}\right)$, and the proof of Lemma B is complete. ■

Equipped with Lemma B, the proof of Proposition 3 is straightforward. With $\eta_{t-1}^* = h_+(L)\eta_{t-1}$ consider

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T} \xi_{t,d}\xi_{t-1,d}^* = \frac{1}{\sqrt{T}}\left[\sum_{t=2}^{T}\eta_t\eta_{t-1}^* + \sum_{t=2}^{T}\eta_t\frac{y_{t-1}^*}{T^\kappa} + \sum_{t=2}^{T}\frac{y_t}{T^\kappa}\eta_{t-1}^* + \sum_{t=2}^{T}\frac{y_t y_{t-1}^*}{T^{2\kappa}}\right]. \tag{26}$$

The first term on the right-hand side converges to $\mathcal{N}(0, \pi^2\sigma_\eta^4/6)$ since $\eta_t\eta_{t-1}^*$ forms a martingale difference sequence [mds], see Robinson (1991); the second term converges to zero as long as $\kappa > 0$, since $\eta_t y_{t-1}^*$ forms a mds, too. Along the lines of proof of Lemma B it follows that $\mathrm{Var}\left(\sum_{t=2}^{T} y_t\eta_{t-1}^*\right) = O(T\ \ln^3 T)$; in fact, one can establish as sharp rate that $\mathrm{Var}\left(\sum_{t=2}^{T} y_t\eta_{t-1}^*\right) = O(T)$. In any case, it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=2}^{T}\frac{y_t}{T^\kappa}\eta_{t-1}^* \xrightarrow{p} 0$$

as long as $\kappa > 0$. The behaviour of the fourth term on the right-hand side of (26) is obvious from Lemma B. Since $\widehat{\sigma}^2 \xrightarrow{p} \sigma_\eta^2$, the proof of Proposition 3 is complete.

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.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*, 817–858.

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.

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.

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

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston.

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., U. Hassler, and V. Kuzin (2011). Pitfalls of post-model-selection testing: experimental quantification. *Empirical Economics 40*(2), 359–372.

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

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. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*, 227–238.

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.

Haldrup, N. and J. E. Vera Valdés (2017). Long memory, fractional integration, and cross-sectional aggregation. *Journal of Econometrics 199*, 1–11.

Hassler, U. (2019). *Time Series Analysis with Long Memory in View.* Wiley.

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

Hurvich, C. M., R. Deo, and J. Brodsky (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis 19*, 19–46.

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

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

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

Palma, W. (2007). *Long-Memory Time Series: Theory and Methods.* Wiley.

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

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics 47*(1), 67–84.

Robinson, P. M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association 89*, 1420–1437.

Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Annals of Statistics 23*, 1048–1072.

Robinson, P. M. (2014). The estimation of misspecified long memory models. *Journal of Econometrics 178*, 225–230.

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.