University of Pretoria
Department of Economics Working Paper Series

Long-Memory Modeling and Forecasting: Evidence from the U.S. Historical Series of Inflation

Heni Boubaker
International University of Rabat
Giorgio Canarella
University of Nevada
Rangan Gupta
University of Pretoria
Stephen M. Miller
University of Nevada
Working Paper: 2018-69
November 2018

Department of Economics
University of Pretoria
0002, Pretoria
South Africa
Tel: +27 12 420 2413

Long-Memory Modeling and Forecasting: Evidence from the U.S. Historical Series of Inflation

Heni Boubaker
International University of Rabat, BEAR LAB,
Technopolis Rabat-Shore Rocade-Sale, Marocco

Giorgio Canarella
University of Nevada, Las Vegas, 4505 S. Maryland Parkway,
Las Vegas, Nevada, USA

Rangan Gupta
University of Pretoria, Pretoria, 0002, South Africa

Stephen M. Miller
University of Nevada, Las Vegas, 4505 S. Maryland Parkway,
Las Vegas, Nevada, USA

November 4, 2018

Keywords: long memory, wavelet analysis, time-varying persistence

JEL classification: C13, C22, C32, C54, E31

Abstract

We report the results of applying semi-parametric long-memory estimators to the historical monthly series of U.S. inflation, and analyze their empirical forecasting performance over 1, 6, 12, and 24 months using in-sample and out-of-sample procedures. For comparison purposes, we also apply two parametric estimators, the naive AR(1) and the ARFIMA(1, d, 1) models. We evaluate the forecasting accuracy of the competing methods using the mean square error (MSE) and mean absolute error (MAE) criteria. We evaluate the statistical significance of forecasting accuracy of competing forecasts using the Diebold-Mariano (1995) test. Overall, our results preforms slightly better than the Lahiani and Scaillet (2009) threshold estimator based on the MSE and MAE criteria. This improvement in performance does not prove significant enough to cause a rejection of the null hypothesis of equality of predictive accuracy. The Boubaker (2017) estimator, on the other hand, significantly outperforms the time-invariant estimators over longer horizons. Over shorter horizons, however, the Boubaker (2017) estimator does not exhibit a significantly better predictive performance than the time-invariant long-memory estimators with the exception of the naive AR(1) model.

1 Introduction

This paper provides the first attempt to conduct a comparative analysis of alternative long-memory estimators applied to the historical series of U.S. inflation. A large literature estimates the statistical properties of inflation, in particular, the property of persistence, generally defined as the speed at which an inflation shock dissipates. As frequently noted in the monetary literature, this notion plays an important role in the design of monetary policy, since it determines if the effect of a shock is transitory or permanent (Gerlach and Tillmann, 2012; Fuhrer, 1995; Fuhrer and Moore, 1995).

One stylized fact of inflation is that it is a long-memory process. Inflation conforms to a fractional differencing process of order d, that is, I(d) where d is a real number. A long-memory process experiences shocks that possess long-lasting effects, but for which the underlying process reverts to its mean. In other words, a long-memory process is not the exclusive property of non-stationary processes. Stationary processes may exhibit long memory as well. This most important feature distinguishes long-memory models from unit-root models, where non-stationarity does not admit mean-reversion (Granger and Joyeux, 1980).

Researchers document significant evidence of long memory in the inflation processes of many countries over the last twenty years of empirical research. For example, see Delgado and Robinson (1994), Hassler and Wolters (1995), Baillie (1996), Baillie et al. (1996), Baum et al. (1999), Bos et al (1999, 2002), Gadea and Mayoral (2005), Kumar and Okimoto (2007) Beechey and Osterholm (2009), Meller and Nautz (2012), Kouretas and Wohar (2012), Canarella and Miller (2016, 2017), and Boubaker et al. (2017), among many others.

The emergence of inflation targeting by nearly 70 central banks or governments around the world by 2018 makes the persistence of the inflation rate an important issue in these countries. [1] The more persistent the inflation rate is, the more difficulty the central bank or government authorities will experience in achieving the target inflation rate. Moreover, a more persistent inflation rate makes it more likely that the inflation rate will overshoot itself as the authorities try to push the inflation rate toward its target. That is, more difficulty will confront the authorities in easing their policy effort to move the inflation rate to their target.

The existing literature proposes many estimators of the long-memory parameter d. Most semi-parametric procedures rely on the frequency domain and, more recently, the wavelet domain. While the frequency estimation of long-memory processes is well-known (see, e.g., Hosking, 1981, 1984; Geweke and Porter-Hudak, 1983), the wavelet methodology is relatively new. One most useful property of wavelet analysis is it's the time-scale

[1] See Central Bank News, `http://www.centralbanknews.info/p/inflation-targets.html`.

decomposition (Gencay et al., 2003; Boubaker et al., 2017), which is the ability to decompose any signal into its time-scale components and to isolate short-lived phenomena from long-term trends in a signal. Wavelets can prove most useful when the signal shows a different behavior in different time periods or when the signal is localized in time as well as frequency. High-frequency components reflect short-term behavior, whereas low-frequency components capture long-term dynamics. As it enables a more flexible approach in time-series analysis, wavelet analysis is a refinement of Fourier analysis. [2] Boubaker et al. (2017) estimate inflation persistence in a time-varying framework, implementing a methodology based on the wavelet approach and the instantaneous least-squares estimator (ILSE).

Within this framework, we discuss and evaluate the performance of the following alternative estimators of inflation persistence using monthly United States data from January 1871 to April 2018: (1) Jensen (1999) wavelet version of OLS (WOLS1); (2) Veitch and Abry (1999) wavelet version of OLS (WOLS2); (3) Geweke and Porter-Hudak (1983) log periodogram estimator (GPH); (4) Lee (2005) wavelet version of the GPH estimator (WGPL); (5) Shimotsu and Phillips (2005) exact local Whittle (ELW) estimator; (6) Boubaker and Peguin-Feissolle (2013) wavelet version of the exact local Whittle (WELW). These estimators assume that the parameter d measuring inflation persistence is constant in the sample. In empirical applications, treating the long-memory parameter d as a constant implies that the long-range dependence structure of inflation persists over time. This assumption seems too restrictive when dealing with historical series such as ours, due to the potential presence of problems of structural breaks and policy shifts. Our data span the evolution of the modern monetary history of the United States and include significant monetary policy and volatility shifts, such as the classical gold standard era, the Bretton Woods system, and the post-Bretton Woods system, and, thus, provide a unique opportunity to document that U.S. inflation may exhibit different degrees of long memory over time and to appraise how inflation persistence may vary across different monetary regimes and institutions. In view of these considerations, we generalize the standard long-memory modeling incorporated by the nine estimators listed above by assuming that the long-memory parameter d is time-varying, and consider two additional long-memory estimators: (10) Boubaker (2017) instantaneous least squares (ILSE) estimator, which accounts for long-memory and smooth-transition regimes (STR); and, (8) The Lahiani and Scaillet (2009) estimator, which simultaneously accounts for long-memory and threshold effects.

[2] The well-known Fourier transform involves the projection of a series onto an orthonormal set of trigonometric components. In particular, Fourier series do not fade away (have infinite energy) and do not change over time (have finite power). In contrast, wavelets grow and decay in a limited time period (have finite energy and compact support).

The rest of paper is organized as follows. The next section describes the ARFIMA model and frequency and wavelet approaches to estimation of inflation persistence under the assumption that inflation persistence is constant over the sample, that is, estimators (1) through (6). Section 3 relaxes this assumption and presents a discussion of estimators (7) and (8). Section 4 presents and discusses the empirical results from the application of the eight alternative estimators. For comparison, we also report results from estimators (7) to (9), although we do not discuss the estimators in any detail. Section 5 outlines a comparative analysis of their respective predictive performance, and Section 6 concludes.

2 The time-invariant long-memory model

2.1 ARFIMA model

Let $X(t), t = 1, \ldots, T$ denote a time-series process. Following Granger and Joyeux (1980), the usual ARFIMA(p, d, q) process is expressed as

$$\Phi(B)(1 - B)^d (X(t) - u) = \Theta(B)\varepsilon(t), \tag{1}$$

where B is the backshift operator such that $B^i X(t) = X(t-i)$, $\Phi(B) = I + \phi_1 B + \cdots + \phi_p B^p$ and $\Theta(B) = I + \theta_1 B + \cdots + \theta_q B^q$ are polynomials in B involving autoregressive and moving average coefficients of orders p and q respectively with their roots strictly outside the unit circle and no common factors, d is the fractional integration parameter, u is the mean of the process, and $\varepsilon(t) \sim$ i.i.d. $\mathcal{N}(0, \sigma_\varepsilon^2)$.

Following Granger and Joyeux (1980) and Hosking (1981), the fractional differencing lag operator for non-integer values of d can be defined by the binomial expansion

$$(1 - L)^d = \sum_{i=0}^{+\infty} \begin{bmatrix} d \\ i \end{bmatrix} (-B)^i = 1 - dB - \frac{1}{2}d(1-d)B^2 - \frac{1}{6}d(1-d)(2-d)B^3 - \ldots, \tag{2}$$

we can rewrite this expression as

$$(I - B)^d = \sum_{i=0}^{+\infty} \frac{\Gamma(i-d)B^i}{\Gamma(-d)\Gamma(i+1)}, \tag{3}$$

where $\Gamma(.)$ denotes the gamma function.

The parameters found in $\Phi(B)$ and $\Theta(B)$ constitute the short-memory parameters and affect only the short-run dynamics of the process, whilst the fractional integration parameter d captures the long-memory behavior of the process. Various cases can occur: If $-0.5 < d < 0$, the process exhibits anti-persistence. If $0 < d < 0.5$, the process is stationary long-memory and possesses shocks that disappear hyperbolically.[3] If $0.5 \leq d < 1$, the process is non-stationary, but mean-reverting, with finite impulse response weights. When $d = 0$, the process reduces to the standard ARMA and when $d = 1$, the process becomes ARIMA and implies infinite persistence of the mean to a shock in the returns.[4]

[3] In this paper, we will concentrate on fractionally-integrated stationary processes.

[4] See Beran (1994) for more details on long-memory processes.

2.2 Wavelet Analysis

Wavelet theory requires the orthonormal bases obtained by dyadically dilating and translating a pair of specially constructed functions φ and ψ, which are called father wavelets and mother wavelets, respectively, such that

$$\int \varphi(t)\,dt = 1, \text{ and} \int \psi(t)\,dt = 0. \tag{4}$$

The smooth and the low-frequency parts of the series captures the father wavelet while the detail and the high-frequency components captures the mother wavelet. The obtained wavelet basis can be given, respectively, by the pair of functions

$$\varphi_{j,k}(t) = 2^{j/2}\varphi\left(2^j t - k\right), \text{ and } \psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right), \tag{5}$$

where $j = 1, \ldots, J$ indexes the scale and $k = 1, \ldots, 2^j$ indexes the translation. The parameter j dilates the waves' functions. This parameter j adjusts the support of $\psi_{j,k}(t)$ to locally capture the characteristics of high or low frequencies. The parameter k relocates the wavelets in the temporal scale. The maximum number of scales that can be considered in the analysis limits the number of observations ($T \geq 2^J$).

One special property of the wavelet expansion incorporates the localization property, where the coefficient of $\psi_{j,k}(t)$ reveals the information content of the function at approximate location $k2^{-j}$ and frequency 2^j. Using wavelets, any function in $L^2(\mathbb{R})$ can be expanded over the wavelet basis, uniquely, as a linear combination at arbitrary level $J_0 \in \mathbb{N}$ across different scales of the type

$$X(t) = \sum_k s_{J_0,k}\varphi_{J_0,k}(t) + \sum_{j>J_0}\sum_k d_{j,k}\psi_{j,k}(t), \tag{6}$$

where $\varphi_{J_0,k}$ is a scaling function with the corresponding coarse and fine scale coefficients $s_{J_0,k}$ and $d_{j,k}$, respectively, by

$$s_{J_0,k} = \int X(t)\,\varphi_{J_0,k}(t)\,dt, \text{ and } d_{j,k} = \int X(t)\,\psi_{j,k}(t)\,dt. \tag{7}$$

These coefficients measure the contribution of the corresponding wavelet to the function. The expression (6) represents the decomposition of $X(t)$ into orthogonal components at different resolutions and constitutes the wavelet multiresolution analysis (MRA).

In practical applications, we invariably deal with sequences of values indexed by integers rather than functions defined over the entire real axis. Instead of actual wavelets, we use short sequences of values referred to as wavelet filters. The number of values in the sequence equals the width of the wavelet filter. Thus, the wavelet analysis considered from a filtering perspective is then well-suited for time-series analysis. For the discrete wavelet transform, the wavelet coefficients can be calculated from the MRA scheme. The recursive MRA scheme[5], which is implemented by a two-channel filter bank (i.e, a high-pass wavelet filter $\{h_l,\ l = 0,\ldots,L-1\}$[6] and its associated low-pass scaling filter $\{g_l,\ l = 0,\ldots,L-1\}$[7] satisfying the quadrature mirror relationship given by $g_l = (-1)^{l+1} h_{L-1-l}$ for $l = 0,\ldots,L-1$, where $L \in \mathbb{N}$ is the length of the filter) representation of the wavelet transform, is divided into decomposition and reconstruction schemes according to the forward and inverse wavelet transform.

Daubechies (1992) defined a useful class of wavelet filters. Daubechies compactly supported wavelet filters of width L possessing the smallest support for a given number of vanishing moments[8] and distinguishes between two choices - the extremal phase filters $D(L)$ and the least asymmetric filters $LA(L)$. A modified version of the Discrete Wavelet Transform (DWT)[9] is the non-decimated or Maximal Overlap Discrete Wavelet Transform (MODWT) (Percival and Walden (2000)). The MODWT algorithm carries out the same filtering steps as the standard DWT, but does not subsample (decimate by 2); therefore, the number of scaling and wavelet coefficients at each level of the transform equals the number of sample observations (see Percival and Walden (2000) and Gençay et al. (2002) for more details).

2.3 Jensen (1999) estimator

The Wavelet Ordinary Least Square (WOLS) estimate of the fractional differencing parameter was introduced by Jensen (1999). He proved that the wavelet coefficients, $d_{j,k}$, associated with a mean zero ARFIMA$(0,d,0)$ model with $|d| < 0.5$ are distributed $\mathcal{N}\left(0, \sigma^2 2^{-2jd}\right)$, where σ^2 is a finite constant.[10] The wavelet coefficient's variance at a scale j is defined by

$$Var(d_{j,k}) = R(j) = \sigma^2 2^{-2jd}. \tag{8}$$

[5]A robust theoretical framework for critically sampled wavelet transformation is Mallat's Multiresolution Analysis (for more details, see Mallat (1989))

[6]The wavelet function (filter) of support L proceeds as a special filter possessing specific properties, such that (i) it integrates to zero, i.e, $\sum_{l=0}^{L-1} h_l = 0$, (ii) has unit energy, i.e, $\sum_{l=0}^{L-1} h_l^2 = 1$ and (iii) is orthogonal to its even shifts, i.e, $\sum_{l=0}^{L-1} h_l h_{l+2n} = \sum_{l=-\infty}^{\infty} h_l h_{l+2n} = 0,\ \forall\, n \in \mathbb{N}^*$.

[7]The scaling filter of support L satisfies the following properties, (i) $\sum_{l=0}^{L-1} g_l = \sqrt{2}$, (ii) $\sum_{l=0}^{L-1} g_l^2 = 1$ and (iii) $\sum_{l=0}^{L-1} g_l g_{l+2n} = \sum_{l=-\infty}^{\infty} g_l g_{l+2n} = 0,\ \forall\, n \in \mathbb{N}^*$.

[8]For Daubechies wavelets, the number of vanishing moments is half the filter length.

[9]In practice, the DWT is implemented via a pyramid algorithm (see Mallat (1989)), which is a design method underlying the conception of the DWT and the construction of the wavelet bases.

[10]See Jensen (1999) for more details about the estimation method.

By taking the logarithms on both sides of equation (8), we have

$$\log\{R(j)\} = \log\left\{\sigma^2\right\} - d\log\left\{2^{2j}\right\}. \tag{9}$$

The estimate of the fractional differencing parameter d comes from ordinary least-squares estimation of equation (9). Following Jensen (1999), the wavelet ordinary least-squares estimate of d is given by

$$\hat{d}_{WOLS_1} = \frac{\sum_{j=0}^{J-1} y_j \log\left\{\bar{R}(j)\right\}}{\sum_{j=0}^{J-1} y_j^2}, \tag{10}$$

where $y_j = \log\left\{2^{-2j}\right\} - \frac{1}{J}\sum_{j=0}^{J-1}\log(2^{-2j})$ and $\bar{R}(j) = \frac{1}{2^j}\sum_{k=0}^{2^j-1} d_{j,k}^2$. Consequently, if $d > -1/4$, $Var(\bar{R}(j)) \to 0$ as $j \to \infty$ and we get $\bar{R}(j)$ will tend in probability to $R(j)$ as $j \to \infty$.

Using the Taylor expansion of $\log\left\{\bar{R}(j)\right\}$ around $\log\{R(j)\}$, we find

$$\log\left\{\bar{R}(j)\right\} = \log\{R(j)\} + o_n(1), \tag{11}$$

Thus, we obtain

$$\log\left\{\bar{R}(j)\right\} = \log\left\{\sigma^2\right\} - d\log\left\{2^{2j}\right\} + o_n(1), \tag{12}$$

Jensen shows that the $\hat{d}_{WOLS_1}$ consistently estimates the fractional-integration parameter d when. $j \to \infty$.

On the other hand, to compute the variance of $\hat{d}_{WOLS_1}$ Jensen demonstrates that

$$\hat{d}_{WOLS_1} - d = \theta^{1/2}2^{-j/2}Z + o_n\left(2^{-j/2}\right), \tag{13}$$

where $\theta = \theta\left(1, 2^{-1}, \ldots, 2^{1-n}\right)$ is a constant and Z is a random variable with unit variance (see Jensen (1999) for a formal proof).

2.4 Veitch and Abry (1999) estimator

The Veitch and Abry (1999) is based on the DWT coefficients $d_{j,k}$ defined in equation (10) of $X(t)$, $t = 1, 2, \ldots, T$, where $X(t)$ is an ARFIMA$(0, d, 0)$. Following Veitch and Abry (1999), we have

$$\hat{\mu}_j = \frac{1}{\upsilon_j}\sum_{k=1}^{\upsilon_j} d_{j,k}^2, \tag{14}$$

where υ_j is the number of the wavelet coefficients at octave j available for computation. As shown by Veitch and Abry (1999),

$$\hat{\mu}_j \sim \frac{z_j}{\upsilon_j}\chi^2_{\upsilon_j}, \tag{15}$$

where $z_j = c2^{2dj}$, $c > 0$, and $\chi^2_{\upsilon_j}$ is Chi-squared random variable with υ_j degrees of freedom. By taking the logarithms on both sides of equation (15), we have

$$\log(\hat{\mu}_j) \sim 2dj + \log_2(c) + \frac{\log(\chi^2_{\upsilon_j})}{\log 2} - \log_2(\upsilon_j), \tag{16}$$

The expected value and the variance of the variable $\log(\chi^2_\upsilon)$ are given by

$$E\left\{\log(\chi^2_\upsilon)\right\} = \xi(\frac{\upsilon}{2}) + \log 2, \tag{17}$$

$$Var\left\{\log(\chi^2_\upsilon)\right\} = \zeta(2, \frac{\upsilon}{2}). \tag{18}$$

where $\xi(h) = \partial h/\partial h \log\{\Gamma(h)\}$, and $\zeta(2, \frac{\upsilon}{2})$ is the Hurwitz zeta function,[11] defined by the formula

$$\zeta(r, s) = \sum_{n=0}^{+\infty} \frac{1}{(n+s)^r}, \tag{19}$$

Equation (16) can be written as follows:

$$\vartheta_j = \alpha + \beta w_j + \varepsilon_j, \tag{20}$$

where $\vartheta_j = \log_2(\hat{\mu}_j) - g_j$, $\alpha = \log_2(c)$, $\beta = 2d$, $w_j = \log_2(2^j) \simeq j$, $\varepsilon_j = \log_2\left\{\log(\chi^2_{\upsilon_j})\right\} - \log_2(\upsilon_j) - g_j$, $g_j = \xi(\upsilon_j/2) - \log(\upsilon_j/2)$. ε_j satisfies

$$E(\varepsilon_j) \simeq 0, and \tag{21}$$

$$Var(\varepsilon_j) = \frac{\zeta(2, \frac{\upsilon_j}{2})}{[\log 2]^2} \simeq \left\{2\upsilon_j \log^2 2\right\}^{-1}. \tag{22}$$

The WOLS2 estimate of Veitch and Abry (1999) is given by

$$\hat{d}_{WOLS_2} = \frac{\hat{\beta}}{2}, \tag{23}$$

where $\hat{\beta}$ is the ordinary least-squares estimate obtained from equation (20). Veitch and Abry (1999) shows that under some regularity conditions, $\hat{d}_{WOLS_2}$ is efficient and consistent.

[11] The Hurwitz zeta function is a generalization of the Riemann zeta function, defined by $\zeta(r) = \frac{1}{\Gamma(r)}\int_0^{+\infty}\frac{u^{r-1}}{\exp(u)-1}du = \frac{1}{1-2^{1-r}}\sum_{n=1}^{+\infty}\frac{(-1)^{n-1}}{n^r}$.

2.5 Geweke and Porter-Hudak estimator (1983)

The Geweke and Porter-Hudak (1983) method is based on the behavior of the spectral density of an ARFIMA process when the frequencies tend towards zero. As previously, this method estimates only the long-memory parameter d. The GPH estimator presents a bias related to the periodogram estimator.[12] Let $I(\lambda_n, T)$ be the periodogram evaluated at the Fourier frequencies $\lambda_{n,T} = 2\pi n/T$, $n = 1, 2, \ldots, m$, where m is the number of frequencies used in the regression; we have for $j = 1, \ldots, m$

$$\log\{I(\lambda_n)\} = \log\{f_\varepsilon(0)\} - d\log\left[\left\{4\sin^2(\frac{\lambda_n}{2})\right\}\right] + \log\{f_\varepsilon(\lambda_n)/f_\epsilon(0)\} + \log\{I(\lambda_n)/f(\lambda_n)\}. \quad (24)$$

where $f(\lambda)$ is the spectral density of $X(t)$ and $f_\varepsilon(\lambda)$ is the spectral density of $\varepsilon_t = (1 - B)^d X_t$, assumed to be a finite and continuous function on the interval $[-\pi, \pi]$. Under these two assumptions: $\log\{f_\varepsilon(\lambda_n)/f_\varepsilon(0)\}$ is negligible for sufficiently low frequencies and the random variables $\log\{I(\lambda_n)/f(\lambda_n)\}_{n=1,2,\ldots,m}$, are asymptotically *i.i.d*. We can write the linear regression following GPH (1983) based on a fractional time series $X(t)$ with length T as

$$\log\{I_X(\lambda_n)\} = \alpha + d\log\left[\left\{2\sin(\frac{\lambda_n}{2})\right\}^{-2}\right] + \epsilon_n. \quad (25)$$

The estimated slope coefficient of equation (**??**) estimates the fractional differencing parameter d, denoted by $\hat{d}_{GPH}$. Here, $I_X(\lambda_n)$ is the periodogram that is normalized by 2π at the n-th Fourier frequency λ_n, where $\lambda_n = 2\pi n/T$ and $\epsilon \sim i.i.d.\mathcal{N}(-c, \frac{\pi^2}{6})$ with c as Euler's constant.[13]

2.6 Wavelet GPH estimator

Lee (2005) proposed the idea of the wavelet GPH estimate using the discrete wavelet transform of $X(t)$, as

$$d_{j,k} = \sum_k X(t)\psi_{j,k}(t). \quad (26)$$

The spectral density of the wavelet transform at the scale j around zero frequency for $d \in (0, 1.5)$ is as follows;

$$f_j(\lambda) = C_j\,|\lambda|^{-2d}\,|\Lambda(\lambda)|^2 \text{ as } \lambda \to 0 = C_j\,|\lambda|^{-2(d-1)}\,g^2(\lambda) \text{ as } \lambda \to 0, \quad (27)$$

[12] Hurvich and Beltrao (1993) and Robinson (1994a, 1995a) proposed a modified version of the GPH estimator by using a smoothed periodogram or by discarding the first frequencies to reduce this bias.

[13] See Geweke and Porter-Hudak (1983) for more details.

where $C_j = c_j/2\pi < \infty$ is a constant term, and $|\Lambda| = \lambda^\nu g(\lambda)$ for all integer ν, with $g(t\lambda)/g(\lambda) = 1$ for all t as $\lambda \to 0$ and $0 < g(0) < \infty$.

For a fixed-scale j, the periodogram of $f_j(\lambda)$ is

$$I^{(j)} = \frac{1}{2\pi T} \sum_{k=0}^{2^j-1} \left| d_{j,k} \exp(i\lambda_n k) \right|^2, \ n = 1, 2, \ldots, m, \tag{28}$$

where $\lambda_n = 2\pi n/T$, m is the number of frequencies that are restricted such that $m \to \infty$ and $m/T \to 0$ as $T \to \infty$.

The wavelet-based GPH estimate denoted as d_{WGPH}, comes from a log-transformation of equation (28). More precisely, we regress the transformed log-periodogram, $\log\left\{I_n^{(j)}\right\}$, on the regressors $-2\log(\lambda_n)$ for $n = 1, 2, \ldots, m$, then add one to the estimate.

For $d \in (0, 1.5)$, Lee (2005) shows that $\hat{d}_{WGPH}$ is consistent and asymptotically normal if $m = o(T^{4/5})$. That is,

$$\sqrt{m}\left(\hat{d}_{WGPH} - d\right) \to \mathcal{N}\left(0, \frac{\pi^2}{24}\right) \text{ as } T \to \infty, \tag{29}$$

where $m = T^{4/5}$ is the optimal rate for the number of frequencies in terms of the mean squared error (see Hurvich et al. (1998) and Andrews and Guggenberger (2003)).

2.7 Exact Local Whittle estimator

Let $X(t)$, $t = 1, \ldots, T$ be a time series generated by the following fractional model

$$(1 - L)^d X(t) = \epsilon(t)\mathbb{1}\{t \geq 1\}, \tag{30}$$

where $\epsilon(t)$ is stationary with zero mean and power transfer function $f_\epsilon(\lambda)$ satisfying $f_\epsilon(\lambda) \sim G$ as $\lambda \to 0$. We define the discrete Fourier transform and the periodogram of a generic time series evaluated at the fundamental frequencies as

$$\mathcal{F}_X(\lambda_n) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X(t) \exp(it\lambda_n), \ t = 0, 1, \ldots, T. \tag{31}$$

$$I_X(\lambda_n) = \left|\mathcal{F}_X(\lambda_n)\right|^2, \tag{32}$$

where $\lambda_n = 2\pi n/T$, and $n = 1, \ldots, T$.

Shimotsu and Phillips (2005) proposed to estimate (d, G) by minimizing the objective function

$$Q_m(d, G) = \frac{1}{m} \sum_{n=1}^{m} \left[\log\left(G\lambda_n^{-2d}\right) + \frac{1}{G} I_\epsilon(\lambda_n)\right], \tag{33}$$

where m is some integer less than T such as $m = T^{\alpha}$, $0 < \alpha < 1$.

Thus, we have

$$\left(\hat{d}, \hat{G}\right) = \underset{G \in (0,\infty),\, d \in [\triangle_1, \triangle_2]}{\arg\min} Q_m(d, G), \tag{34}$$

where $-\infty < \triangle_1 < \triangle_2 < +\infty$ are the lower and upper bounds of the admissible values of d and m is the bandwidth parameter determines the number of periodogram ordinates used in the estimation.

Concentrating $Q_m(d, G)$ with respect to G, Shimotsu and Phillips (2005) defined the Exact Local Whittle (ELW) estimate as

$$\hat{d}_{ELW} = \underset{d \in [\triangle_1, \triangle_2]}{\arg\min} R(d), \tag{35}$$

where $R(d) = \log\left\{\hat{G}(d)\right\} - 2d\frac{1}{m}\sum_{n=1}^{m}\log\lambda_n$ and $\hat{G}(d) = \frac{1}{m}\sum_{n=1}^{m} I_{\epsilon}(\lambda_n)$.

Shimotsu and Phillips (2005) found that the ELW is consistent[14] for $d \in (\triangle_1, \triangle_2)$, and asymptotically normal

$$\sqrt{m}\left(\hat{d}_{ELW} - d\right) \to \mathcal{N}\left(0, \frac{1}{4}\right) \text{ as } T \to \infty, \tag{36}$$

provided that $\Delta_2 - \Delta_1 \leq \frac{9}{2}$.

2.8 Wavelet Exact Local Whittle estimator

The wavelet Exact Local Whittle estimate ($WELW$), as defined by Boubaker and Péguin-Feissolle (2013), is based on the discrete wavelet transform of $X(t)$ in equation (27). This method minimizes the objective function $Q_m\left(G_j, d_j\right)$ representing the likelihood function of Whittle for each scale j.

We define the objective function, at fixed-scale j, by

$$Q_m\left(d_j, G_j\right) = \frac{1}{m}\sum_{n=1}^{m}\left[\log\left(G_j\lambda_n^{-2d_j}\right) + \frac{1}{G_j} I_{\epsilon j}^{(j)}(\lambda_n)\right]. \tag{37}$$

where $I_{\epsilon j}^{(j)}$ is the periodogram of the wavelet at scale j of $\Delta^{d_j} X_t = (1-L)^{d_j} X_t$ as defined by equation (29) for $n = 1, 2, \ldots, m$, $j = 1, \ldots, J$ and $T = 2^J$.

The estimated values $\left(\hat{d}_j, \hat{G}_j\right)$ are given by

$$\left(\hat{d}_j, \hat{G}_j\right) = \underset{G_j \in (0,\infty),\, d_j \in \left[\Delta_{1j}, \Delta_{2j}\right]}{\arg\min} Q_m\left(d_j, G_j\right), \tag{38}$$

[14] See Shimotsu and Phillips (2005) for more details about the consistency of the estimate.

where Δ_{1j} and Δ_{2j} are the lower and upper bounds of the admissible values of d_j at fixed-scale j such that $-\infty < d_{1j} < d_{2j} < +\infty$.

Concentrating $Q_m\left(d_j, G_j\right)$ with respect to G_j, we define the $WELW$ estimator at fixed-scale j as

$$\widehat{d}_{j(WELW)} = \underset{d_j \in \left[\Delta_{1j}, \Delta_{2j}\right]}{\arg\min} R_{WELW}\left(d_j\right), \tag{39}$$

with $R_{WELW}\left(d_j\right) = \log\left\{\hat{G}_j(d_j)\right\} - 2d_j\frac{1}{m}\sum_{n=1}^{m}\log\lambda_n$ and $\hat{G}(d) = \frac{1}{m}\sum_{n=1}^{m} I_{\epsilon j}^{(j)}(\lambda_n)$.

Boubaker and Péguin-Feissolle (2013) demonstrate that the $WELW$ estimator is consistent and asymptotically normal for any $d_j \in \left(\Delta_{1j}, \Delta_{2j}\right)$ if $\Delta_{2j} - \Delta_{1j} \leq \frac{9}{2}$ and under fairly mild assumptions on m and the stationary component $\epsilon(t)$. They conclude that the statistical properties of the $WELW$ estimator are the same those of the ELW estimator.

3 The time-varying long-memory model

It seems too restrictive to assume that the fractional-integration parameter d is constant over time, which implies that the long-range dependence structure of the underlying phenomenon persists over time with a constant degree. That is, structural changes in the fractional-integration parameter, will have long-range dependence to evolve in time. Thus, the long-memory parameter is likely a time-varying d_t. That is, the degree of persistence to shocks varies over time and some type of non-stationarity exists.

3.1 Time-varying ARFIMA model

If we suppose that d varies over time, i.e., d_t, we obtain the time-varying ARFIMA(p, d, q) model or TV-ARFIMA(p, d_t, q). Let $X(t)$, $t = 0, \ldots, T-1$ be a stochastic process defined by

$$\Phi(B)(X(t) - u) = \Theta(B)(1 - B)^{-d_t}\varepsilon(t), \tag{40}$$

where $d_t < 0.5$ is the time-varying fractional-integration parameter, $\Phi(B)$ and $\Theta(B)$ are stable polynomials with roots strictly outside the unit circle, u is the mean of the process, and $\varepsilon(t)$ is a white noise process with zero mean and variance σ_ε^2. For simplicity, we suppose that we observe d_t on a finer grid, i.e., making d_t rescaled on $[0, 1]$ so that we can denote it by $d(t/T)$. We can also model the short-memory parameters found in $\Phi(B)$ and $\Theta(B)$ as functions of t, i.e., $\Phi_t(B)$ and $\Theta_t(B)$. In the following and for purpose of simplicity, we consider that the short-memory parameters are constant in time by setting $\Phi_t(B) = \Phi(B)$ and $\Theta_t(B) = \Theta(B)$ for all t.

This long-memory model (**??**) is a member of non-stationary class of processes known as locally-stationary processes in the sense of Dahlhaus (1996) and Whitcher and Jensen (2000), where the spectral representation of stochastic process $X(t)$, with realizations of length T, is given by

$$X(t) = \int_{-\pi}^{\pi} \exp(it\lambda) A_t^0(\lambda) \, d\xi(\lambda), \tag{41}$$

where $\xi(\lambda)$ is a stochastic process on $[-\pi, \pi]$ as defined in Dahlhaus (1996) with $\overline{\xi(\lambda)} = \xi(-\lambda)$ and $A_t^0(\lambda)$ is the transfer function given by

$$A_t^0(\lambda) \equiv \frac{\sigma_\varepsilon}{2\pi} \frac{\Theta(\exp(-i\lambda))}{\Phi(\exp(-i\lambda))} (1 - \exp(-i\lambda))^{-d_t}. \tag{42}$$

Let $X_0, \ldots, X_{T-1}$ be realizations of a locally-stationary process with transfer function $A_t^0(\lambda) \equiv A(n, \lambda)$, $n = t/T$ will represent a time point in the rescaled time domain $[0, 1]$, where $A(n, \lambda)$ is an even and 2π-periodic function that is uniformly Lipschitz continuous in $n \in [0, 1]$ and λ. The time-varying spectral density function is defined as follows

$$sdf(n, \lambda) \equiv |A(n, \lambda)|^2. \tag{43}$$

For a locally-stationary long-memory process $X(t)$, the time-varying spectral density function can be represented as

$$sdf(n, \lambda) \sim \lambda^{-2d(n)} \text{ as } \lambda \to 0^+. \tag{44}$$

Thus, if $d(n) > 0$, $sdf(n, \lambda)$ is smooth for frequencies close to zero, but is unbounded when $\lambda = 0$. In other words, the energy of $X(t)$ is concentrated over those frequencies associated with long-term cycles. If $d(n) < 0$, then $sdf(n, \lambda) = 0$ and $X(t)$ is a locally-stationary series that is anti-persistent. As a result of the time-varying long-memory parameter, X_t will be smoother with less variation in its amplitude during time periods when $d(n) > 0$, and will have large fluctuations in its values when $d(n) < 0$.

Denote by $cov_X(n, g-h)$ the local auto-covariance function for a locally-stationary long-memory process $X(t)$ at time n, we have

$$cov_X(n, g-h) = \int_{-\pi}^{\pi} \exp(i(g-h)\lambda) \, sdf(n, \lambda) \, d\lambda. \tag{45}$$

By substituting the spectral density function of a locally stationary long-memory model into equation (**??**) and using the property of gamma function, the local auto-covariance function can be simplified as

$$cov_X(n, g-h) \sim |g-h|^{2d(n)-1} \text{ as } |g-h| \to \infty. \tag{46}$$

As indicated, the slow hyperbolic decay of $cov_X(n, g-h)$ is the feature most often noted when discussing the dynamics of a long-memory process.

Nevertheless, to specify the evolution of d , many studies assume that the time-varying fractional-integration parameter d_t follows several regime-switching models.

3.2 Boubaker (2017) estimator

Following Boutahar et al. (2008), Aloy et al. (2013) and Boubaker (2017), we can assume that d_t evolves according to STR model advanced by Teräsvirta (1994, 1998):

$$d_t = d_1 [1 - F(s_t; \gamma, c)] + d_2 F(s_t; \gamma, c). \tag{47}$$

where d_1 and d_2 are the values of the fractional-integration parameter in the first and second regime, respectively. $F(s_t; \gamma, c)$ is the continuous and bounded transition function between 0 to 1 with s_t denoting the transition variable which can be one of the lagged endogenous variable $s_t = d_{t-i}, \forall t > i, t = 0, \ldots, T-1$ or an exogenous variable $s_t = X(t-i), \forall t > i, t = 0, \ldots, T-1$ or the trend $s_t = t, t = 0, \ldots, T-1$.[15] The slope parameter γ measures the speed of the transition between the two extreme regimes (associated with the extreme values 0 and 1 of the transition function) that can be either positive or negative depending upon whether the logistic curve is increasing or not. The parameter c represents the threshold for the transition variable s_t, which defines the underlying regimes: $s_t \leq c$ ($s_t > c$) means that the underlying regime is the first (the second) one. We find in the literature two types of transition function, the logistic function STR model and the exponential function STR model, respectively, expressed as follows:

$$F(s_t; \gamma, c) = (1 + \exp(-\gamma(s_t - c)))^{-1}, and \tag{48}$$

$$F(s_t; \gamma, c) = 1 - \exp\left(-\gamma(s_t - c)^2\right). \tag{49}$$

To simplify the exposition, we assume that $\Phi(B) = \Theta(B) = 1$, i.e., the simpler model with no short-term components. Let $(\Omega, F, \mathcal{P})$ be the probability triple, and assume that $X(t)$ and d_t are defined on the set of all possible outcomes Ω and given respectively by (**??**) and (**??**). The stochastic fractional filter $(1-B)^{-d_t}$ can be defined as follows

$$\forall w \in \Omega, \quad (1-B)^{-d(n,w)} = \sum_{i=0}^{\infty} \frac{\Gamma(i + d(n,w))}{\Gamma(i+1)\Gamma(d(n,w))} B^i, \tag{50}$$

[15]For more details about STR models and transition variables, see van Dijk et al. (2002).

where $d(n,w) := d_t(w)$ and Γ is the gamma function.

Substituting equation (**??**) into (**??**), equation (**??**) can be written as an infinite moving average process in terms of $\varepsilon(t-i)$

$$X(t) = (1-B)^{-d_t} \frac{\Theta(B)}{\Phi(B)} \varepsilon(t) \equiv \sum_{i=0}^{\infty} \frac{\Gamma(i+d_t)}{\Gamma(i+1)\Gamma(d_t)} \pi_i \varepsilon(t-i). \tag{51}$$

The expression (**??**) can be re-expressed as

$$X(t) \equiv \sum_{i=0}^{\infty} a_{it}(s_t) \pi_i \varepsilon \varepsilon(t-i), \tag{52}$$

$$\text{with} \begin{cases} a_{0t}(\cdot) = 1 \\ a_{1t}(\cdot) = D(\cdot) \\ a_{2t}(\cdot) = a_{1t-1} + \frac{D(\cdot)[1-D(\cdot)]}{2!} \\ a_{3t}(\cdot) = D(\cdot) a_{2t-1} + \frac{D(\cdot)[1-D(\cdot)]}{2!} a_{1t-2} + \frac{D(\cdot)[1-D(\cdot)][2-D(\cdot)]}{3!} \\ a_{4t}(\cdot) = D(\cdot) a_{3t-1} + \frac{D(\cdot)[1-D(\cdot)]}{2!} a_{2t-2} + \frac{D(\cdot)[1-D(\cdot)][2-D(\cdot)]}{3!} a_{1t-3} + \frac{D(\cdot)[1-D(\cdot)][2-D(\cdot)][3-D(\cdot)]}{4!} \\ \vdots \end{cases}$$ [16],

where π_i is the solution to $\frac{\Theta(z)}{\Phi(z)} = \sum_i \pi_i z^i$, and $D(\cdot) = d_1[1-F(\cdot;\gamma,c)] + d_2 F(\cdot;\gamma,c)$.[17]

The instantaneous least-squares estimator uses a single wavelet coefficient from each scale of resolution. That is, we only use d_{j,t_j}^2 to estimate $V_X^2(\tau_j)$, where t_j is the time index of the j^{th} level MODWT coefficient associated with time t in $X(t)$, $t = 0, \ldots, T-1$. The time index t_j can be meaningfully determined only if we use a linear phase wavelet filter. In particular, let $X_0, \ldots, X_{T-1}$ be a time series of interest, which is the realization of a stationary process with variance σ_X^2. If $V_X^2(\tau_j)$ designates the wavelet variance for scale $\tau_j \equiv 2^{j-1}$, then we have

$$\sigma_X^2 \equiv \sum_{j=1}^{J} V_X^2(\tau_j). \tag{53}$$

The wavelet variance is estimated using the MODWT coefficients through

$$V_X^2(\tau_j) \equiv \frac{1}{T-L_j+1} \sum_{t=L_j-1}^{T-1} d_{j,t}^2, \tag{54}$$

[16] We note that a time-varying d generates an infinite dimensional memory parameter. For that reason, we use a classes of time-varying linear filters (see Surgailis (2008) and Philippe et al. (2008) for a robust theoretical framework).

[17] $D(\cdot) = d_1 + (d_2-d_1)[1+\exp(-\gamma(\cdot - c))]^{-1}$ if the transition function is logistic and $D(\cdot) = d_1 + (d_2-d_1)\left[1-\exp\left(-\gamma(\cdot - c)^2\right)\right]$ if the transition function is exponential

where $d_{j,t}$ is the MODWT wavelet coefficient of X_t at scale τ_j and $L_j \equiv \left(2^j - 1\right)(L-1)+1$ is the width of the scale τ_j wavelet filter.

Formally, let the vector of dimension containing the wavelet coefficients be obtained by the MODWT transform. The instantaneous least squares estimator $\hat{d}_{ILSE,t}$ is given by

$$\hat{d}_{ILSE,t} = \frac{\Delta_J \sum_{j=J_0}^{J} \ln\left(\tau_j\right) Y_t\left(\tau_j\right) - \sum_{j=J_0}^{J} \ln\left(\tau_j\right) \sum_{j=J_0}^{J} Y_t\left(\tau_j\right)}{2\left(\Delta_J \sum_{j=J_0}^{J} \ln^2\left(\tau_j\right) - \left(\sum_{j=J_0}^{J} \ln\left(\tau_j\right)\right)^2\right)} + \frac{1}{2}, \tag{55}$$

where $\Delta_J = J - J_0 + 1$ and $Y_t\left(\tau_j\right) \equiv \ln\left(d^2_{j,t_j}\right) - psi\,(1/2) - \ln(2)$, with psi as the digamma function (see, also, a MODWT-based weighted least-squares estimator developed by Percival and Walden (2000)). To decrease the variability of the estimates, Δ_J should ideally be set as large as feasible. This estimator, however, is independent of the entire time series and utilizes only certain coefficients that are co-located in time.

In this paper, we use an iterative procedure by performing multi-step estimation given by Boubaker (2017) to efficiently estimate a time-varying long-memory model defined, respectively, by equations (**??**) and (**??**).

3.3 Lahiani and Scaillet (2009) estimator

This method introduces a new class of threshold ARFIMA models to account simultaneously for long-memory and regime switching. The threshold effect appears in the autoregressive and/or the fractional-integration parameters which we can test for using LM tests. Let $X(t)$, $t = 0, \ldots, T-1$ be a stochastic process defined by equation (**??**) the long-memory threshold ARFIMA(p, d, q) model. The ARFIMA(p, d^-, d^+, q) can be defined as follows:

$$\Phi(B)(1-B)^{d^-} X(t)^- + \Phi(B)(1-B)^{d^+} X(t)^+ = \Theta(B)\,\varepsilon(t), \tag{56}$$

where $\left(X(t)^-, X(t)^+\right) = \left((X(t) - u^-)\mathbb{1}\{z_t \leq \delta\}, (X(t) - u^+)\mathbb{1}\{z_t > \delta\}\right)$, the threshold parameter is δ, and $z_{t-1} = z\left(\varepsilon(t-1), \cdots, \varepsilon(t-p)\right)$ is a known function of the noise terms. The fractional integration parameters are d^- when $z_{t-1} \leq \delta$ and d^+ when $z_{t-1} > \delta$. The errors $\varepsilon(t)$, however, are assumed to be i.i.d. $\mathcal{N}(0, \sigma^2_\varepsilon)$. Lahiani and Scaillet (2009) introduce a more general model for jointly modeling the threshold affects the long and short memory of the ARFIMA model to capture the two properties of long memory and nonlinearities in parameters of long and short memory. To test for a threshold effect in the parameters of the ARFIMA models, we can apply an LM test under the null hypothesis of no threshold effects. A threshold ARFIMA model reduces to a symmetric ARFIMA model. The parameter vector β decomposes into $\beta = \left(\beta_1', \beta_2'\right)$, where $\beta_1 = (d^-, u^-)$ and $\beta_2 = (d^+, u^+)$. The null hypothesis is $H_0 : \beta_1 = \beta_2$, while the alternative hypothesis is $H_0 : \beta_1 \neq \beta_2$. Under the null hypothesis, the parameter is not identified, and has to be treated as a nuisance parameter.

In this case, we use the reparameterization $\beta = \beta_1 - \beta_2$, which amounts to a test for $\beta = 0$. Then, we can use the SupLM-statistic or the ExpLM-statistic under a local-to-null approach.

4 Empirical results

This section contains an empirical application of ARFIMA(p, d, q) and TV-ARFIMA$(0, d_t, 0)$ models to modeling the dynamics of CPI time series. First, we describe the data, then we present the empirical results.

4.1 Data description

The data consist of monthly observations on the seasonally adjusted U.S. Consumer Price Index (CPI) and are obtained from the data segment on the R.J. Shiller website (http://www.econ.yale.edu/ shiller/data.htm), covering 1871:01 to 2018:04 and corresponding to 1768 observations. We define the inflation rate as the annualized monthly change in the log of CPI.

Table 1 reports the key descriptive statistics of the inflation series for the full sample and four subsamples, corresponding to the classical gold standard period (1870-1914), the interwar period (1915-1944), the Bretton Woods period (1945-1971) and the post-Bretton Woods period (1972 to present). Over the entire sample, inflation averages, on an annual basis, approximately 2 percent. But, compared over the different monetary regimes, average inflation shows some interesting characteristics. For example, during the period of the classical gold standard, average inflation, on an annual basis, is -0.4788 percent, although the period was characterized by two decades of secular deflation, followed by two decades of secular inflation (Bordo and Redish, 2003), with a maximum of 81 percent and a minimum of -81 percent. This is, however, the only period when inflation, on average, is negative. In the interwar period, inflation averages 1.88 percent, while during the Bretton Woods era, inflation averages about 3 percent. The post-Bretton Woods period, on the other hand, witnesses an average inflation of about 4 percent, with a maximum and minimum of 21 and -23 percent, respectively.

The minimum and maximum values indicate that the inflation series varied considerably over the four sub-periods. Inflation exhibits greater variability during the classical gold standard. The remaining statistics provide evidence of fat tails and non-normality. Inflation displays a strong asymmetric distribution characterized by positive skewness, mainly during the interwar and the Bretton Woods periods. The kurtosis statistics indicate that inflation in all sub-periods and the entire period exhibits fat tails. The Jarque-Bera test statistic rejects normality for the entire sample and the four sab-samples. This result is not surprising and is frequently found in the empirical literature on inflation persistence. [18]

Table 1: Summary statistics of the U.S. monthly inflation series (in percent)

	1870-2018	**1870-1914**	**1915-1944**	**1945-1971**	**1972-present**
Mean	2.0352	-0.4788	1.8888	3.0984	3.8952
Median	1.6632	0.0000	0.0000	3.1452	3.5424
Maximum	81.6636	81.6636	58.548	68.5896	21.4764
Minimum	-81.6636	-81.6636	-38.5056	-10.1712	-23.2056
Std. Dev	12.5304	19.4304	11.5884	6.462	4.5312
Skewness	-0.1401	0.0932	0.4298	3.8601	-0.0928
Kurtosis	9.6554	4.8247	4.9526	36.0543	6.0015
Jarque-Bera	3266.963	73.8764	68.2789	15554.59	209.52
p-value	0.0000	0.0000	0.0000	0.0000	0.0000
Obs.	1767	527	360	324	556

4.2 Time-invariant long-memory result

First, we estimate the fractional-integration parameter (d) over the whole sample period to verify the presence of long-range dependence. For that, we consider several estimators in both the time and wavelet domains. In the time domain, we consider the GPH (1983) estimator that is theoretically valid for $0 < d < 0.5$. If the estimate of the memory parameter is on the verge of stationarity, we need to consider an estimator that is consistent for $d > 0.5$ as well as $0 < d < 0.5$. Further, if the number of frequencies m included in the regression is restricted such that $m = O\left(T^{4/5}\right)$, then we obtain the asymptotic normality (see Hurvich et al. (1998)). In addition, we use the Exact Local Whittle estimator developed by Shimotsu and Phillips (2005). It is a semi-parametric estimator, generally giving a good estimation method for the memory parameter in terms of consistency and limit distribution. This estimator is consistent and has an $N\left(0, \frac{1}{4}\right)$ limit distribution for all values of d if the optimization covers an interval of width less than $\frac{9}{2}$ and the mean of the process is known. Whereas, in the wavelet domain, we apply some ordinary least-squares estimators

[18]In addition, we apply three tests to assess the stationarity of the inflation rate series. The ADF and PP test statistics (with and without a constant) overwhelmingly reject the null hypothesis of unit root at the 1-percent level. The KPSS test (with and without a constant), on the other hand, confirms these results by failing to reject the null hypothesis of stationarity at the 1-percent level. Detailed results are available on request.

of the long-memory parameter from a fractionally-integrated process. The first estimator is a semi-parametric wavelet based estimator for the Hurst parameter as proposed by Abry and Veitch (1998). Under the general conditions and the Gaussian assumptions, this estimator is unbiased and efficient. The second estimator is developed by Jensen (1999b) based on a log linear relationship between the variance of the wavelet coefficients from the long-memory process and its scale equal to the long-memory parameter. This log-linear relationship yields a consistent ordinary least-squares estimator. The third estimator is the wavelet GPH estimate as proposed by Lee (2005) also known as the wavelet log-periodogram regression method. The fourth estimation method is the Wavelet Exact Local Whittle (WELW) estimator that constitutes an extension of the ELW estimator of Shimotsu and Phillips (2005) defined by Boubaker and Péguin-Feissolle (2013). They use the local Whittle method, but their likelihood function is based on the wavelet coefficients and their estimator is deduced from the asymptotic consistency of these coefficients.

Table 2 summarizes the results of the six estimators. In addition, it reports the average of the Boubaker (2017) time-varying estimates, i.e., $\hat{d}_{ILSE}$

Table 2: Estimation results of constant fractional integration parameter

r_t	$\hat{d}_{GPH}$	$\hat{d}_{ELW}$	$\hat{d}_{WOLS_1}$	$\hat{d}_{WOLS_2}$	$\hat{d}_{WGPH}$	$\hat{d}_{WELW}$	$\hat{d}_{ILSE}$
$\hat{d}$	0.2845***	0.2975***	0.2548***	0.2627***	0.2853***	0.2977***	0.2981***

Note: $\hat{d}_{GPH}$ is the GPH estimator of Geweke and Porter-Hudak (1983), $\hat{d}_{ELW}$ is the Exact Local Whittle estimator of Shimotsu and Phillips (2005), $\hat{d}_{WOLS_1}$ is the estimator of Abry and Veitch (1998), $\hat{d}_{WOLS_2}$ is the estimator of Jensen (1999b), $\hat{d}_{WGPH}$ is the wavelet GPH estimator of Lee (2005), $\hat{d}_{WELW}$ is the wavelet ELW estimator of Boubaker and Péguin-Feissolle (2013), and $\hat{d}_{ILSE}$ is the mean of instantaneous least-squares estimator given by $\hat{d}_{ILSE} = \frac{1}{T}\sum_{t=1}^{T}\hat{d}_{ILSE,t}$, where $\hat{d}_{ILSE,t}$ is the instantaneous least-squares estimator and T is the number of observations. ***indicate significance at the 1% significance.

We see that, for all estimation methods, strong empirical evidence exists of long-memory behavior in the whole period since the fractional-integration parameter $\hat{d}$ is positive and less than 0.5.

For comparison purposes, we present in Table 3 the estimation results of ARFIMA (1,d,1) model

Table 3: Estimation results of $ARFIMA(1, d, 1)$ model

	$\hat{\phi}_1$	$\hat{\theta}_1$	$\hat{d}$	$\hat{u}$
r_t	0.6268***	-0.7083**	0.2782***	0.0015*

Note: ***indicate significances at the 1% significance,**indicate significances at the 5% significance and *indicate significances at the 10% significance.

The ARFIMA results, which allow for short-term dynamics, provide further empirical evidence that inflation is reactionary and persistent. Finally, we note that the estimate on the lagged inflation in the AR(1) model is 0.2827, significant at the 1% level. One limitation of these estimators is their assumption that the fractional-integration parameter is constant and that the effect of dependence persists over time. In practice, however, the fractional-integration parameter and the persistence may vary over time. As announced previously, we estimate first the time-varying fractional-integration parameter using both approaches of Boubaker (2017) and Lahiani and Scaillet (2009).

Tables 4 and 5 report the empirical results. In Table 4, we report the time-varying estimates of the fractional-integration parameters, using the Boubaker (2017) approach. The estimates in the two regimes, $\hat{d}_1$ and $\hat{d}_2$, are positive and less than 0.5, which suggests that inflation is a stationary, long memory, mean-reverting process in both regimes. The estimate of the slope parameter $\hat{\gamma}$ is positive and significant, indicating that the logistic function is increasing. The estimate of the threshold parameter $\hat{c}$ defines the transition point from the first to the second regime. The likelihood ratio (LR) test statistic of 124.359 rejects the null hypothesis of equality of $\hat{d}_1$ and $\hat{d}_2$. We compute the test statistic as $LM = 2(L_A - L_0)$, where L_A is the log-likelihood under H_A (that the true model is a time-varying ARFIMA), and L_0 is the log-likelihood under H_0 (that the true model is a time invariant ARFIMA). The test statistic is asymptotically chi-squared distributed with one degree of freedom, equal to the number of restrictions in the test.

In Table 5, we report the time-varying estimates of the fractional-integration parameters using the Lahiani and Scaillet (2006) estimation method. The estimates in the two regimes, $\hat{d}^{+}$ and $\hat{d}^{-}$, and positive and less than 0.5, also suggesting that inflation is a stationary, long memory, and mean reversing process in both regimes. The fractional integration parameter in the second regime, d^{+}, is significantly less than the fractional-integration parameter in the first regime, d^{-}.

The LR test (84.572) rejects the null hypothesis of equality of d^{+} and d^{-}. In comparison, the 1% critical value of the chi-squared distribution with 1 degree of freedom is 6.643.

Table 4: Estimation results of time-varying fractional integration parameter Boubaker (2017)

r_t	$\hat{d}_1$	$\hat{d}_2$	$\hat{\gamma}$	$\hat{c}$
$\hat{d}_{ILSE,t}$	0.2589^{***}	0.2641^{***}	0.1978^{***}	0.2324^{***}

Note: $\hat{d}_{ILSE,t}$ is the instantaneous least squares estimator of Boubaker (2017). ***indicate significances at the 1% significance.

Table 5: Estimation results of time-varying fractional-integration parameter Lahiani and Scaillet (2009)

r_t	$\hat{d}^-$	$\hat{d}^+$
$\hat{d}_S$	0.2252***	0.1463***

Note: $\hat{d}_S$ is the threshold estimator of Lahiani and Scaillet (2009). ***indicate significances at the 1% significance.

We observe, however, that we cannot compare the two models since they are not nested. The only way to assess the superiority of one method over the other is in terms of predictive performance. We undertake this task in the next section.

5 Predictive performance

The empirical results presented in the previous section raise an important issue, which any empirical investigation cannot ignore. That is, does an estimator exist, among the ones discussed in this paper, that exhibits a statistical superiority with respect to the others? In this section, we present a comparative analysis of the forecast performance of our alternative estimators. Specifically, we evaluate the in-sample and out-of-sample forecasting performance for different estimation techniques. Concerning the out-of-sample forecasting performance, we estimate the models starting with the in-sample period from t_0 to $t_1 - 1$ and forecast inflation for the periods t_1 through t_s. We then re-estimate the model from t_0 to t_1 and forecast inflation for the period through $t_1 + 1$ to $t_s + 1$. This procedure continues until all the data are exhausted. To evaluate the accuracy of the forecasts, we apply two evaluation criteria, namely the mean squared error (MSE) and the mean absolute error (MAE), given respectively by

$$MSE = \frac{1}{T - t_1} \sum_{t=t_1}^{N} \left(r_{t+s} - \hat{r}_{t,t+s}\right)^2 , \tag{57}$$

$$MAE = \frac{1}{T - t_1} \sum_{t=t_1}^{N} \left|\left(r_{t+s} - \hat{r}_{t,t+s}\right)\right| . \tag{58}$$

where T is the number of observations, $T - t_1$ is the number of observations for predictive performance, r_{t+s} is the inflation series through period $t + s$, and $\hat{r}_{t+s}$ is the predicted inflation at horizon s at time t.

Additionally, we employ the Diebold-Mariano (1995) test to compare the predictive accuracy of two competing forecasts. The Diebold-Mariano test uses a loss function associated with the forecast error of each forecast and tests the null that the expected differential loss is zero, that is, $E(d_t) = 0$, where the loss differential $d_t = h(e_{1t}) - h(e_{2t})$. The two loss functions are computed as follows:

$$h(e_{1t}) = h(\hat{y}_{1t} - y_t), \tag{59}$$

and

$$h(e_{2t}) = h(\hat{y}_{2t} - y_t), \tag{60}$$

where y_t is the actual value of the series and $\hat{y}_{1t}$ and $\hat{y}_{2t}$ are two predictions for y_t, $t = 1, 2, ...T$. In most cases, the loss function is a square-error loss function or an absolute-error loss function. The hypotheses of interest are as follows:

$$H_0 : E(h(e_{1,t+h}) - h(e_{2,t+h})) = 0, \tag{61}$$

and

$$H_A : E(h(e_{1,t+h}) - h(e_{2,t+h})) \neq 0, \tag{62}$$

where $h \geq 1$ is the forecast horizon. The relevant statistic of the test, denoted by DM, is given by

$$S_1 = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}} \rightarrow N(0, 1), \tag{63}$$

$$\bar{d} = \frac{\sum_{t=1}^{n} d_t}{n}, \tag{64}$$

$$\hat{V}(\bar{d}) = \frac{\hat{\gamma}_0 + 2 \sum_{k=1}^{n-1} \hat{\gamma}_k}{n}, and \tag{65}$$

$$\hat{\gamma}_k = \frac{\sum_{t=k+1}^{n} (d_t - \bar{d})(d_{t-k} - \bar{d})}{n}. \tag{66}$$

The DM test has a standard normal limiting distribution under the null hypothesis.

We evaluate the predictive performance of the alternative long memory models of inflation estimated in section 4 (Geweke and Porter-Hudak 1983; Shimotsu and Phillips 2005; Jensen 1999; Veitch and Abry 1999; Lee 2005; Boubaker and Peguin-Feissolle 2013; Boubaker 2017) with the addition of the naïve AR(1) model and the ARFIMA (1, d, 0). We take into consideration four time horizons, 1-month, 6-months, 12-months, and 24-months ahead forecasting horizons.

The DM test uses the ILSE,t (Boubaker 2017) estimator as a benchmark and tests whether the superiority of the ILSE,t forecast is statistically significant or is simply due to in-sample variability. The results of the in sample predictive performance for horizons s = 1, 6, 12, and 24 months, where the model parameters are estimated once using the entire sample and are reported in Table 6. The results of out-of-sample forecasting for the same horizons are tabulated in Table 7. In this case, the first estimation sample period covers approximately the first half of the total sample (i.e., from 1870:2 to 1945:12), and the remaining 856 observations are used for forecast evaluation, where the models are estimated recursively and forecasts are generated starting from 1946:1 up to 2017:4. We generate, therefore, a sequence of 856 1-step ahead forecasts, 850 6-step ahead forecasts, 844 12-step ahead forecasts, and 832 24-step ahead forecasts.

Several comments about the findings are warranted. First, the results of the in-sample as well as the out-of-sample forecasts suggest that all models perform in the short run as well as the time-varying (Boubaker 2017) ILSE,t estimator. The only exception is the AR(1) model, which is outperformed by the ILSE,t model at all four horizons. The forecasts of the ILSE,t model, however, are preferred to the forecasts of the time invariant models over longer horizons. The time varying ILSE,t outperforms the non-wavelet-based estimators at s = 6, 12, and 24 in the case of the GPH, but only at s = 24 in the case of the ELW model, and outperforms the wavelet-based estimators of Jensen (1999) and Veitch and Abry (1999) at s = 6, 12, 24. On the other hand, the wavelet-based versions of the GPH (Lee 2005) and ELW (Boubaker and Peguin-Feissolle 2013) do not exhibit a significant improvement over their corresponding original estimators and are outperformed by the ILSE,t estimator at the same forecasting horizons as their non-wavelet-based counterparts.

Second, the ILSE,t (Boubaker 2017) forecasting performance is as accurate as the time-varying model of Lahiani and Scaillet (2009). The MSE and MAE values of the Boubaker (2017) model are smaller than those of the Lahiani and Scaillet (2009) model, but this reduction is not sufficient enough to reject the DM null hypothesis of equality of predictive accuracy.

Third, for the in-sample and out-of-sample forecasting performance, predictability does increase as we attempt to forecast further into the future. As the forecasting horizon increases, long-memory models become more and more useful. This result may reflect the mean-reversion characteristics of inflation over longer time horizons, where a steady rise of the forces of mean reversion is detectable in all models (Ca' Zorzi, Kolasa and Rubaszek 2016). A statistical explanation is that persistence in the ARFIMA model dissipates in the long run (Bondon 2005).

Table 6: In-sample forecasts

	Criteria	$s=1$	$s=6$	$s=12$	$s=24$
$\hat{d}_{GPH}$	MSE	0.0231	0.0143	0.0112	0.0106
	MAE	0.1471	0.1228	0.0953	0.0921
	DM	1.6784	1.9856	2.4384	2.6872
$\hat{d}_{ELW}$	MSE	0.0229	0.0141	0.0108	0.0101
	MAE	0.1398	0.1188	0.0912	0.0852
	DM	1.1169	1.5364	1.7782	2.1582
$\hat{d}_{WOLS_1}$	MSE	0.0247	0.0176	0.0132	0.0111
	MAE	0.1643	0.1345	0.1032	0.0978
	DM	1.6686	1.9852	2.4241	2.6472
$\hat{d}_{WOLS_2}$	MSE	0.0252	0.0186	0.0146	0.0112
	MAE	0.1637	0.1412	0.1213	0.1123
	DM	1.7236	2.0232	2.5421	2.7381
$\hat{d}_{WGPH}$	MSE	0.0228	0.0140	0.0111	0.0105
	MAE	0.1465	0.1224	0.0950	0.0920
	DM	1.7216	1.9581	2.6748	2.8130
$\hat{d}_{WELW}$	MSE	0.0227	0.0137	0.0105	0.0100
	MAE	0.1396	0.1185	0.0911	0.0851
	DM	1.1165	1.5361	1.7778	2.1575
$\hat{d}_{ILSE}$	MSE	0.0222	0.0134	0.0096	0.0089
	MAE	0.1376	0.1182	0.0891	0.0805
	DM	1.5148	1.7121	1.8654	2.1832
$\hat{d}_{S}$	MSE	0.0221	0.0134	0.0095	0.0088
	MAE	0.1375	0.1178	0.0912	0.0789
	DM	1.0126	1.2571	1.3244	1.5772
$ARFIMA(1,d,0)$	MSE	0.0234	0.0145	0.0102	0.0097
	MAE	0.1415	0.1254	0.1137	0.0108
	DM	1.6543	1.8213	1.9869	2.1872
$AR(1)$	MSE	0.0258	0.0185	0.0162	0.0135
	MAE	0.1639	0.1446	0.1310	0.1274
	DM	2.2783	2.8564	3.2576	3.7983
$\hat{d}_{ILSE,t}$	MSE	0.0218	0.0131	0.0092	0.0084
	MAE	0.1371	0.1176	0.0889	0.0786

Table 7: Out-of-sample forecasts

	Criteria	$s = 1$	$s = 6$	$s = 12$	$s = 24$
$\hat{d}_{GPH}$	MSE	0.0242	0.0166	0.0124	0.0106
	MAE	0.1563	0.1257	0.0974	0.0943
	DM	1.7352	2.2145	2.6735	2.8462
$\hat{d}_{ELW}$	MSE	0.0237	0.0158	0.0120	0.0101
	MAE	0.1476	0.1195	0.0927	0.0874
	DM	1.2157	1.5893	1.9145	2.2675
$\hat{d}_{WOLS_1}$	MSE	0.0256	0.0182	0.0137	0.0113
	MAE	0.1668	0.1378	0.1047	0.0098
	DM	1.8351	2.1245	2.7782	2.8589
$\hat{d}_{WOLS_2}$	MSE	0.0258	0.0177	0.0135	0.0110
	MAE	0.1665	0.1374	0.1049	0.0964
	DM	1.8762	2.2574	2.8574	2.9862
$\hat{d}_{WGPH}$	MSE	0.0240	0.0163	0.0121	0.0108
	MAE	0.1558	0.1221	0.0952	0.0924
	DM	1.8271	1.9893	2.7352	2.8544
$\hat{d}_{WELW}$	MSE	0.0236	0.0155	0.0118	0.0099
	MAE	0.1472	0.1191	0.0914	0.0821
	DM	1.2183	1.5985	1.8769	2.2542
$\hat{d}_{ILSE}$	MSE	0.0231	0.0151	0.0098	0.0092
	MAE	0.1437	0.1186	0.0895	0.0812
	DM	1.5851	1.7874	1.8982	2.2145
$\hat{d}_{S}$	MSE	0.0239	0.0161	0.0119	0.0109
	MAE	0.1481	0.1226	0.0918	0.0921
	DM	1.0134	1.3852	1.4986	1.6980
$ARFIMA(1, d, 0)$	MSE	0.0245	0.0167	0.0119	0.0098
	MAE	0.1534	0.1275	0.0092	0.0092
	DM	1.4765	1.6572	1.8531	2.1437
$AR(1)$	MSE	0.0267	0.0187	0.0165	0.0137
	MAE	0.1642	0.1452	0.1315	0.1276
	DM	2.2953	2.9845	3.3741	3.9856
$\hat{d}_{ILSE,t}$	MSE	0.0227	0.0148	0.0097	0.0090
	MAE	0.1415	0.1182	0.0893	0.0803

6 Conclusion

The last few decades of macroeconomic research have resulted in a vast array of important contributions in the area of long-memory modeling and forecasting of inflation. We obtain estimates of inflation persistence in the United States using monthly data from 1870:02 to 2018:04 from several different estimators. The estimators include the following: (1) the Jensen (1999) wavelet version of OLS (WOLS1); (2) the Veitch and Abry (1999) wavelet version of OLS (WOLS2); (3) the Geweke and Porter-Hudak (1983) log periodogram estimator (GPH); (4) the Lee (2005) wavelet version of the GPH estimator (WGPH); (5) the Shimotsu and Phillips (2005) exact local Whittle (ELW) estimator; (6) the Boubaker and Peguin-Feissolle (2013) wavelet version of the exact local Whittle (WELW) estimator; (7) the average estimator of Boubaker (2017); (8) the ARFIMA(1,d,1). One drawback of these methods is their assumption that the parameter d measures inflation persistence as a constant in the sample. In empirical applications, especially when dealing with historical series like ours, this assumption may be too restrictive, since we cannot exclude the potential presence of problems of structural breaks and policy shifts. Consequently, we introduce two long memory estimators that assume the long memory parameter d is time-varying: (10) the Boubaker (2017) instantaneous least squares (ILSE) estimator, which accounts for long memory and smooth transition regimes (STR); and, the Lahiani and Scaillet (2009) estimator that simultaneously accounts for long-memory and threshold effects. For all estimation methods, we find strong evidence that United States inflation is a stationary, but persistent, process, defined by a fractional process with integration parameter positive and less than 0.5. To evaluate the forecasting performance of the alternative estimators, we generate in-sample and out-of-sample forecasts that indicate that, for longer horizons, the time-varying estimators proposed by Boubaker (2017) and Lahiani and Scaillet (2009) provide an appreciable improvement in predictive performance relative to non-time-varying long-memory forecasts. For shorter horizons, however, the forecasting improvements are not statistically significant with the exception of the AR(1) estimator. This evidence is encouraging and suggestive of the potential usefulness of nonlinear fractional modeling and forecasting of inflation.

References

[1] Andrews, D., and Guggenberger, P. (2003). A bias-reduced log periodogram regression estimator for the long-memory parameter. Econometrica, 71, 675-712.

[2] Baillie, R. T., 1996. Long-memory processes and fractional integration in econometrics. Journal of Econometrics, 73, 5–59.

[3] Baillie, R. T., Chung, C. F., and Tieslau, M. A., 1996. Analyzing inflation by the fractionally integrated ARFIMA-GARCH model. Journal of Applied Econometrics, 11, 23-40.

[4] Beechey, M., and Österholm, P., 2009. Time-varying inflation persistence in the Euro area. Economic Modelling, 26, 532-535.

[5] Beran, J. (1994). Statistics for long memory processes. New York: (Chapman & Hall).

[6] Bordo, M.D., and Redish, A., 2003. Is Deflation depressing? Evidence from the Classical Gold Standard, NBER Working Papers 9520, National Bureau of Economic Research, Inc.

[7] Bos, C., Franses, P.H., and Ooms, M., 1999. Long Memory and Level Shifts: Reanalyzing Inflation Rates. Empirical Economics, 24, 427- 449.

[8] Bos, C.S., Franses, P.H., and Ooms, M., 2002. Inflation, Forecast Intervals and Long Memory Regression Models. International Journal of Forecasting, 18, 243-264.

[9] Boubaker, H., and Péguin-Feissolle, A. (2013). Estimating the Long-Memory Parameter in Nonstationary Processes using Wavelets. Computational Economics, 42, 291-306.

[10] Boubaker, H. (2017). A Generalized ARFIMA Model with Smooth Transition Fractional Integration Parameter. Journal of Times Series Econometrics, 10, 1, - .

[11] Boubaker, H., Canarella, G., Gupta, R., and Miller, S.M., 2017. Time-varying persistence of inflation: evidence from a Wavelet-based approach. Studies in Nonlinear Dynamics and Econometrics, 21, 1-38.

[12] Canarella, G., and Miller, S. M., 2017. Inflation Persistence Before and After Inflation Targeting: A Fractional Integration Approach. Eastern Economic Journal, 43, 78-103.

[13] Ca' Zorzi, M., Kolasa, M., and Rubaszek, M. 2016. Exchange rate forecasting with DSGE models. ECB Working Paper Series No. 1905.

[14] Delgado, M. A., and Robinson, P. M., 1996, Optimal spectral bandwidth for long memory. Statistica Sinica, 6, 97-112.

[15] Daubechies, I. (1992). Ten Lectures on Wavelets. Philadelphia, SIAM.

[16] Diebold, F.X., and Mariano, R. S. (1995). Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13, 253-263.

[17] Doukhan, P., Oppenheim, G., and Taqqu, M. S. (2003). Theory and applications of long-range dependence. Boston: Birkhauser.

[18] Fuhrer, J., 1995. The persistence of inflation and the cost of disinflation. New England Economic Review January, 3-16.

[19] Fuhrer, J. C., and Moore, G., 1995. Inflation persistence. Quarterly Journal of Economics, 110, 127-160.

[20] Gadea, M. D., and Mayoral, L., 2006. The persistence of inflation in OECD countries: A fractionally integrated approach. International Journal of Central Banking, 4, 51-104.

[21] Gençay, M., Selçuk, F., and Whitcher, B. (2002). An introduction to wavelets and other filtering methods in finance and economics. Academic Press.

[22] Gencay, R., Selcuk, F., and Whitcher, B., 2003. Systematic risk and time scales. Quantitative Finance, 3, 108-116.

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

[24] Granger, C. W. J., and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis, 1, 15-29.

[25] Hassler, U., and Wolters, J., 1995. Long memory in inflation rates: International evidence. Journal of Business and Economic Statistics, 13, 37–45.

[26] Hosking, J. (1981). Fractional differencing. Biometrika, 68, 165-176.

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

[28] Hurvich, C., Deo, R., and Brodsky, J. (1998). The mean squared error of geweke and porter-hudak's estimator of the long memory parameter of a long-memory time series. Journal of Time Series Analysis, 19, 1095-1112.

[29] Jensen, M. (1999). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. Journal of Forecasting, 18, 17-32.

[30] Kouretas, G., and Wohar, M. E., 2012. The dynamics of inflation: A study of a large number of countries. Applied Economics, 44, 2001-2026.

[31] Kumar, M. S., and Okimoto, T., 2007. Dynamics of persistence in international inflation rates. Journal of Money, Credit and Banking, 39, 1457-1479.

[32] Lahiani, A., and Scaillet, O. (2009). Testing for threshold effect in ARFIMA models: Application to US unemployment rate data. International Journal of Forecasting, 25, 418-428.

[33] Lee, J. (2005). Estimating memory parameter in the us inflation rate. Economics Letters, 87, 207-210.

[34] Mallat, S. (1989). A theory for multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 674-693.

[35] Meller, B., and Nautz, D., 2012. Inflation persistence in the Euro area before and after the European Monetary Union. Economic Modelling, 29, 1170-1176.

[36] Meyer, I. (1993). Wavelets: Algorithms and Applications. Philadelphia: SIAM.

[37] Percival, D. B., and Walden, A. T. (2000). Wavelet Methods for Time Series Analysis. Cambridge University Press, Cambridge.

[38] Shimotsu, K., and Phillips, P. (2005). Exact local whittle estimation of fractional integration. Annals of statistics, 20, 87-127.

[39] Veitch, D., and Abry, P. (1999). A wavelet-based joint estimator of the parameters of long-range dependence. IEEE Transactions on Information Theory, 45, 878-897.

[40] Whittle, P. (1951). Hypothesis Testing in Time Series Analysis, New York.

[41] Willinger, W., Taqqu, M. S., and Teverovsky, V. (1999). Stock market prices and long-range dependence. Finance and Stochastics, 3, 1-3.