Semiparametric Estimation of Fractional Integration: An Evaluation of Local Whittle Methods

JASON R. BLEVINS

The Ohio State University

December 16, 2025

Abstract. Fractionally integrated time series, exhibiting long memory with slowly decaying autocorrelations, are frequently encountered in economics, finance, and related fields. Since the seminal work of Robinson (1995a), a variety of semiparametric local Whittle estimators have been proposed for estimating the memory parameter d. However, applied researchers must determine which estimator to use, and under what conditions. This paper compares several local Whittle estimators, first replicating key findings from the literature and then extending these with new Monte Carlo experiments and in-depth empirical studies. We compare how each estimator performs in the presence of short-run dynamics, unknown means, time trends, and structural breaks, and discuss how to interpret potentially conflicting results with real datasets. Based on the findings, we offer guidance to practitioners on estimator choice, bandwidth selection, and the potential diagnostic value of disagreements between estimators.

Keywords: fractional integration, fractional differencing, nonstationarity, long memory, local Whittle estimation.

JEL Classification: C22, C87.

1. Introduction

A fundamental problem in econometrics concerns the estimation of the memory parameter d for fractionally integrated time series, a quantity of interest in macroeconomics, finance, and climate econometrics. Since the seminal work of Granger and Joyeux (1980) and Hosking (1981), who introduced fractionally integrated processes to model intermediate persistence between stationary and unit root behavior, an extensive literature on semiparametric estimation of d has emerged. The early contributions of Geweke and Porter-Hudak (1983) and Robinson (1995b) concerned log-periodogram regressions for frequency-domain estimation of d. Alternatively, by maximizing a quasi-likelihood function based on the Whittle likelihood, the local Whittle (LW) estimator of Künsch (1987) and Robinson (1995a) attains the conjectured[1] asymptotic variance lower bound of $\frac{1}{4}$ for stationary processes with $d \in (-\frac{1}{2}, \frac{1}{2})$.

[1] Robinson (1995a, p. 1640) conjectured that the semiparametric efficiency bound for this class of estimators is $\frac{1}{4}$. Although no estimator has achieved lower asymptotic variance, the semiparametric efficiency bound

Velasco (1999) and Hurvich and Chen (2000) developed tapered LW variants that extended the method to nonstationary processes, but at the cost of efficiency. Shimotsu and Phillips (2005) introduced the exact local Whittle (ELW) estimator, which restored efficiency through exact fractional differencing while extending the method to the entire parameter space. The two-step ELW (2ELW) procedure of Shimotsu (2010) further extended the approach to processes with unknown means and time trends. The modified local Whittle estimator of Hou and Perron (2014) (LWLFC), provides robustness to low-frequency contaminations such as level shifts.

Other recent work has further extended the scope of local Whittle methods. Abadir, Distaso, and Giraitis (2007) developed the fully extended local Whittle (FELW) estimator, which is valid for the range $d \in (-\frac{3}{2}, \infty)$, using a generalized discrete Fourier transform. Arteche (2020) developed ELW estimation for models with multiple poles, to jointly estimate memory at standard, seasonal, and other cyclical frequencies. Wingert, Leschinski, and Sibbertsen (2020) addressed contamination from seasonality by omitting affected frequencies. Bias-reducing methods introduced by Robinson and Henry (2003) and Andrews and Sun (2004) yield improvements in the presence of short-run dynamics. Frederiksen, Nielsen, and Nielsen (2012) extended the local polynomial approach to perturbed fractional processes. Extensions to fractional cointegration (Nielsen, 2007; Shimotsu, 2012), spatiotemporal data (Chen and Wang, 2017), functional time series (Li, Robinson, and Shang, 2021), and high-dimensional settings (Baek, Düker, and Pipiras, 2023) have adapted local Whittle methods to accommodate more complex processes.

Although the theoretical foundations are now well-established and multiple estimators are available, applied researchers face a practical difficulty: when methods yield different estimates, how should they proceed? This could be due to estimator limitations, features of the data, or statistical uncertainty about the memory parameter. Two examples from our empirical analysis illustrate this point. In the case of French inflation (log-differenced CPI), the six estimators we consider yield estimates $\hat{d}$ from 0.26 to 0.82. This range includes the nonstationarity boundary at 0.50, leading to quite different conclusions about persistence. Similarly, for U.S. industrial production (log monthly index), the ELW estimator yields $\hat{d} = 0.10$, suggesting behavior close to $I(0)$, while tapered estimators yield $\hat{d} \approx 1.3$, indicating strong nonstationarity. Although the underlying reasons for these differences are distinct, we show how to identify and correct both.

Previous authors have carried out finite-sample comparison studies that provide a foundation for our work. Nielsen and Frederiksen (2005) conducted an extensive Monte Carlo comparison of parametric, semiparametric, and wavelet estimators. They concluded that frequency-domain maximum likelihood was superior among parametric methods. Among

has not been formally established.

bias-reducing semiparametric estimators, they found that the local polynomial Whittle estimator of Andrews and Sun (2004) and the bias-reduced log-periodogram regression of Andrews and Guggenberger (2003) substantially reduced bias in the presence of persistent short-run dynamics, at the cost of increased variance. A related study by García-Enríquez and Hualde (2019) compared standard local Whittle estimation with the bias-reducing variants of Robinson and Henry (2003) and Andrews and Sun (2004). They found that the benefits depend on the nature of the underlying short-run dynamics. These methods are helpful for moderate autoregressive short-run dynamics, but they can underperform the simpler standard LW estimator when short-run dynamics are mild. Both of these previous studies, however, are based entirely on Monte Carlo experiments and they do not consider the more recent ELW or 2ELW estimators of Shimotsu and Phillips (2005) and Shimotsu (2010) or the break-robust LWLFC estimator of Hou and Perron (2014).

This paper addresses this gap in three ways. First, we carry out a detailed comparison of the performance of six local Whittle estimators across an expanded parameter range. We consider the standard LW and Velasco-tapered LW estimators as well as four methods not previously compared: the Hurvich-Chen-tapered LW, ELW, 2ELW, and LWLFC estimators. Second, we show that differences among methods have diagnostic value. For example, large gaps between ELW and tapered methods signal mean or trend contamination, estimates clustering at validity boundaries indicate true nonstationarity, and full-sample estimates that exceed subsample estimates by a wide margin suggest spurious memory from structural breaks. Third, we demonstrate through a series of detailed case studies how practitioners can use these findings in practice, showing that the French inflation disagreement reflects structural breaks with within-regime memory of $d \approx 0.3$–0.5, while the industrial production discrepancy is due to ELW's sensitivity to the unknown mean.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical development of local Whittle estimation methods and reports our replications of several main results from the literature for each method. Section 3 presents a new, cross-method Monte Carlo comparison with controlled contaminations. Section 4 presents our extended empirical analyses based on several macroeconomic, financial, and climate time series. Section 5 summarizes our practical guidance for practitioners for selecting and applying local Whittle methods to real data. Section 6 concludes. All replication code and datasets are available at `https://github.com/jrblevin/lws`. Our analysis was conducted using PyELW, a cross-platform library for local Whittle estimation in Python (Blevins, 2025).

2. Semiparametric Estimation of Fractional Integration

In this section, we briefly review six major local Whittle estimators. Consider the fractionally integrated process X_t defined as

$$(1) \qquad (1-L)^{d_0} X_t = u_t \, 1\{t \geq 1\}, \quad t = 0, \pm 1, \pm 2, \ldots,$$

where L is the lag operator, $d_0 \in \mathbb{R}$ denotes the true memory parameter, $1\{\cdot\}$ denotes the indicator function, and u_t is a stationary, mean-zero process with spectral density $f_u(\lambda)$. The indicator function imposes the initial condition $X_t = 0$ for $t \leq 0$.[2] The memory parameter d_0 determines the fundamental properties of X_t: $I(0)$ behavior when $d_0 = 0$, stationary long memory for $d_0 \in (0, \frac{1}{2})$, nonstationary but mean-reverting dynamics for $d_0 \in [\frac{1}{2}, 1)$, and the unit root case when $d_0 = 1$. Negative values of d_0 generate antipersistent processes, while $d_0 > 1$ yields explosive nonstationarity.

The estimators we consider are frequency-domain methods that exploit the distinctive spectral properties of long-memory processes without specifying a parametric functional form for $f_u(\lambda)$. The Whittle likelihood (Whittle, 1951) provides a computationally efficient approximation to the Gaussian likelihood by working in the frequency domain, where the discrete Fourier coefficients are asymptotically independent complex normal random variables (Brillinger, 1981, Theorem 4.4.1). Time-domain maximum likelihood estimation requires $O(n^3)$ operations due to the inversion of the $n \times n$ autocovariance matrix. By focusing on low-frequency behavior, where the memory parameter dominates the shape of the spectral density, local Whittle methods offer two advantages: they avoid both the computationally intensive matrix inversion and the need for a complete parametric specification.

2.1. Local Whittle Estimation

The local Whittle estimator of Robinson (1995a) exploits the power-law behavior of the spectral density of fractionally integrated processes near zero. The spectral density satisfies

$$(2) \qquad f_X(\lambda) \sim G\lambda^{-2d_0} \quad \text{as} \quad \lambda \to 0^+,$$

where $a \sim b$ means that $\frac{a}{b} \to 1$ and $G = f_u(0) > 0$.[3] This relationship captures the quintessential long-memory properties while remaining agnostic about the specific parametric form of the short-run dynamics.

The LW approach restricts attention to the first $m = m(n) \ll n$ Fourier frequencies

[2] The process defined in (1) is a "Type II" process, as opposed to a "Type I" process which applies the filter to a doubly-infinite sequence.

[3] For nonstationary processes with $|d_0| \geq \frac{1}{2}$, the classical spectral density is undefined due to infinite variance. In this case, $f_X(\lambda)$ is interpreted as a pseudo-spectral density.

where (2) holds most accurately, avoiding contamination from short-run dynamics at higher frequencies. We define the discrete Fourier transform of a_t at frequencies $\lambda_j = \frac{2\pi j}{n}$ as

$$w_a(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} a_t e^{it\lambda_j}.$$

Let $I_a(\lambda_j) = |w_a(\lambda_j)|^2$ denote the periodogram. The Whittle approximation to the (negative) Gaussian log-likelihood is

$$\sum_{j=1}^{m} \left[\log f_X(\lambda_j) + \frac{I_X(\lambda_j)}{f_X(\lambda_j)} \right].$$

Substituting the power-law approximation $f_X(\lambda_j) \approx G\lambda_j^{-2d}$ for the unknown spectral density and concentrating out the nuisance parameter G yields the local Whittle objective function:

$$R(d) = \log \left(\frac{1}{m} \sum_{j=1}^{m} \lambda_j^{2d} I_X(\lambda_j) \right) - \frac{2d}{m} \sum_{j=1}^{m} \log \lambda_j. \tag{3}$$

The local Whittle estimator is defined as:

$$\hat{d}_{\mathrm{LW}} = \underset{d \in [\Delta_1, \Delta_2]}{\arg\min}\ R(d),$$

where $[\Delta_1, \Delta_2]$ denotes the admissible parameter space. Since the periodogram $I_X(\lambda_j)$ does not depend on d, it can be precomputed from the data and so the objective function requires only $O(m)$ operations per evaluation. The objective function $R(d)$ is also convex in d (Baum, Hurn, and Lindsay, 2020, Appendix A), meaning that any local minimum reached by an optimization algorithm is also the global minimum. Robinson (1995a) established regularity conditions under which for $d_0 \in (-\frac{1}{2}, \frac{1}{2})$, the local Whittle estimator is $\sqrt{m}$-consistent:

$$\sqrt{m}(\hat{d}_{\mathrm{LW}} - d_0) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \tfrac{1}{4}\right). \tag{4}$$

The asymptotic variance of $\frac{1}{4}$ implies an asymptotic standard error of $\frac{1}{\sqrt{4m}}$ for $\hat{d}_{\mathrm{LW}}$.

The bandwidth parameter m controls the bias-variance trade-off. Increasing m includes more frequencies which reduces variance, but this increases bias since higher frequencies are influenced more by short-run dynamics. Henry and Robinson (1996) show that the theoretical optimal rate is $m \propto n^{\frac{4}{5}}$, which minimizes the mean squared error of the estimator. However, this rate can be sensitive to model misspecification and short-run dynamics, leading practitioners to often use more conservative choices $m = \lfloor n^{\alpha} \rfloor$ with $\alpha < \frac{4}{5}$. See Henry (2001), Baillie, Kapetanios, and Papailias (2014), and particularly the discussion in Arteche and

Table 1. LW Estimator: Replication of Right Panel of Table 1 of Shimotsu and Phillips (2005)

	Original			Replication		
d	Bias	S.D.	MSE	Bias	S.D.	MSE
-1.3	0.4109	0.2170	0.2160	0.4114	0.2182	0.2169
-0.7	0.0353	0.0885	0.0091	0.0373	0.0882	0.0092
-0.3	-0.0027	0.0781	0.0061	-0.0008	0.0780	0.0061
0.0	-0.0075	0.0781	0.0062	-0.0060	0.0775	0.0060
0.3	-0.0066	0.0785	0.0062	-0.0060	0.0778	0.0061
0.7	0.0099	0.0812	0.0067	0.0095	0.0812	0.0067
1.3	-0.2108	0.0982	0.0541	-0.2106	0.0986	0.0541

Notes: LW estimator, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = \lfloor n^{0.65} \rfloor = 56$ frequencies, 10,000 replications.

Orbe (2017) for further developments on bandwidth selection, which we revisit in Section 4.

In Table 1, we replicate the right panel of Table 1 from Shimotsu and Phillips (2005), which demonstrates the finite-sample performance of the LW estimator for a range of d values. Each sample consists of $n = 500$ simulated observations from an ARFIMA$(0, d, 0)$ process. The table shows the mean, standard deviation, and mean squared error of the estimates over 10,000 replications. The number of frequencies used was $m = \lfloor n^{0.65} \rfloor = 56$. The "Original" columns report the original results from Shimotsu and Phillips (2005) while the "Replication" columns show the independent results from our Python implementation. Our results correspond closely with the original results for all specifications.

The results confirm the theoretical properties of the local Whittle estimator. Within the valid range $d \in (-0.5, 0.5)$, the estimator has very little bias, standard deviations in the range of the theoretical asymptotic value of $\frac{1}{\sqrt{4m}} = 0.067$, and low mean squared errors. The estimator performs particularly well for $d \in [-0.3, 0.3]$. However, the theoretical limitations are apparent for the more extreme values $d = -1.3$ and $d = 1.3$, where validity breaks down and bias dominates the mean squared error.

2.2. Extending the Range via Tapering

Although the LW estimator achieves the asymptotic variance of $\frac{1}{4}$, it is only valid for $d_0 \in (-\frac{1}{2}, \frac{1}{2})$. This range excludes nonstationary processes, ruling out time series with unit root or near-unit root behavior. To address this limitation, Velasco (1999) developed a class of tapered LW estimators that trades efficiency for a wider validity range.

Formally, tapering involves weighting the original series X_t by a sequence h_t that is symmetric around $\lfloor \frac{n}{2} \rfloor$, with $\max_t h_t = 1$, and that decays smoothly to zero at the sample

boundaries. The tapered periodogram

$$I_h(\lambda_j) = |w_h(\lambda_j)|^2, \quad w_h(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} h_t X_t e^{it\lambda_j},$$

replaces the standard periodogram $I_X(\lambda_j)$ in the objective function. The underlying idea is that down-weighting observations near the sample boundaries reduces contamination from the nonstationary component of the process at low frequencies, which are used to identify the memory parameter. For Velasco tapers, the periodogram uses subsampled frequencies $\lambda_j = \frac{2\pi j}{n}$ with $j = p, 2p, 3p, \ldots, m$, where p is the taper order. As with the standard LW estimator, evaluating these tapered objective functions remains computationally inexpensive and the criterion functions are amenable to simple scalar minimization algorithms.

Using tapers of higher orders p extends the valid range of d_0 values and provides robustness to trends of order $p-1$. The triangular Bartlett window taper ($p = 2$) is valid for $d_0 \in (-\frac{1}{2}, \frac{3}{2})$ and robust to linear trends, while the cosine bell and Zhurbenko-Kolmogorov tapers ($p = 3$) are valid for $d_0 \in (-\frac{1}{2}, \frac{5}{2})$ and robust to linear and quadratic time trends. However, as mentioned, this robustness is at the expense of increased variance. The variance of the limiting distribution of the tapered LW estimator is $\frac{p\Phi_p}{4}$ where Φ_p is a constant that depends on the order p of the taper. For $p = 2$ with $\Phi_2 = 1.05000$, the asymptotic variance is $2 \times 1.05 = 2.1$ times the LW variance; for $p = 3$ with $\Phi_3 = 1.00354$, the variance is 3.01 times larger.

We replicate the right panel of Table 2 from Shimotsu and Phillips (2005) to illustrate the finite-sample behavior of the Bartlett-tapered LW estimator of Velasco (1999). The results are shown in Table 2. As before, our Python implementation yields results that are very close to the original paper across all memory parameter values. Within the range $d \in [-1.7, 1.7]$ (which even includes values far outside the method's valid range), the estimates have low bias and standard deviations of around 0.120, close to the asymptotic value of $\sqrt{\frac{2.1}{4m}} = 0.097$. The Bartlett taper is of order $p = 2$, implying a variance inflation factor of 2.1 relative to the LW estimator. As expected, at more extreme values of d outside the method's valid range, we see large biases and mean squared errors.

Since PyELW implements all three tapers considered by Velasco (1999), we compare their finite-sample performance in Table 3. MSE values larger than 0.05 are shaded in gray. (This threshold is arbitrary and was chosen as a simple way to highlight obvious failures.) The results reflect the fundamental trade-off between bias reduction and variance inflation due to tapering. Within the range $d \in [-0.7, 0.7]$, all three tapers perform well. The Bartlett taper ($p = 2$) achieves the lowest variance with standard deviations around 0.120.

We can see the advantages of higher-order tapering in the nonstationary range. For $d \in [1.3, 2.3]$, the cosine bell and Kolmogorov tapers ($p = 3$) show lower bias than the

Table 2. Velasco Tapered LW Estimator: Replication of Right Panel of Table 2 of Shimotsu and Phillips (2005)

	Original			Replication		
d	Bias	S.D.	MSE	Bias	S.D.	MSE
−3.5	1.6126	0.3380	2.7148	1.6231	0.3326	2.7450
−2.3	0.2155	0.1726	0.0762	0.2173	0.1734	0.0773
−1.7	0.0259	0.1235	0.0159	0.0279	0.1211	0.0154
−1.3	0.0081	0.1211	0.0147	0.0090	0.1203	0.0146
−0.7	−0.0068	0.1219	0.0149	−0.0048	0.1200	0.0144
−0.3	−0.0133	0.1224	0.0151	−0.0097	0.1201	0.0145
0.0	−0.0138	0.1224	0.0152	−0.0111	0.1201	0.0145
0.3	−0.0132	0.1235	0.0154	−0.0103	0.1201	0.0145
0.7	−0.0068	0.1227	0.0151	−0.0051	0.1203	0.0145
1.3	0.0140	0.1232	0.0154	0.0170	0.1222	0.0152
1.7	0.0456	0.1288	0.0187	0.0485	0.1258	0.0182
2.3	−0.1781	0.1419	0.0519	−0.1774	0.1410	0.0514
3.5	−1.4541	0.1338	2.1322	−1.4529	0.1378	2.1298

Notes: Velasco tapered LW estimator with Bartlett taper, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = \lfloor n^{0.65} \rfloor = 56$ frequencies, 10,000 replications.

Table 3. Comparison of Velasco (1999) Tapers

	Bartlett ($p = 2$)			Cosine ($p = 3$)			Kolmogorov ($p = 3$)		
d	Bias	S.D.	MSE	Bias	S.D.	MSE	Bias	S.D.	MSE
−3.5	1.6231	0.3326	2.745	0.1254	0.1644	0.043	0.2207	0.1796	0.081
−2.3	0.2173	0.1734	0.077	0.0698	0.1644	0.032	0.0757	0.1648	0.033
−1.7	0.0279	0.1211	0.015	0.0515	0.1643	0.030	0.0542	0.1649	0.030
−1.3	0.0090	0.1203	0.015	0.0419	0.1642	0.029	0.0431	0.1648	0.029
−0.7	−0.0048	0.1200	0.014	0.0315	0.1640	0.028	0.0313	0.1645	0.028
−0.3	−0.0097	0.1201	0.015	0.0274	0.1636	0.028	0.0268	0.1640	0.028
0.0	−0.0111	0.1201	0.015	0.0258	0.1632	0.027	0.0255	0.1635	0.027
0.3	−0.0103	0.1201	0.015	0.0257	0.1629	0.027	0.0259	0.1631	0.027
0.7	−0.0051	0.1203	0.014	0.0281	0.1625	0.027	0.0296	0.1626	0.027
1.3	0.0170	0.1222	0.015	0.0410	0.1628	0.028	0.0432	0.1622	0.028
1.7	0.0485	0.1258	0.018	0.0722	0.1654	0.033	0.0597	0.1620	0.030
2.3	−0.1774	0.1410	0.051	0.2316	0.2010	0.094	0.1020	0.1636	0.037
3.5	−1.4529	0.1378	2.130	−0.3994	0.1041	0.170	−0.4160	0.1835	0.207

Notes: Comparison of three Velasco tapers, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = \lfloor n^{0.65} \rfloor = 56$ frequencies, 10,000 replications. Shaded cells indicate MSE > 0.05.

Bartlett taper, though at the cost of increased variance with standard deviations around 0.16, somewhat higher than the theoretical value of $\sqrt{\frac{3.01}{4m}} = 0.116$. At more extreme values such as $d = 3.5$, the third-order tapers maintain lower bias than the second-order Bartlett taper, but all three eventually fail far outside their valid ranges.

2.3. Efficiency Gains via Complex-Valued Tapering

Hurvich and Chen (2000) (HC) developed an alternative tapered local Whittle estimator based on applying a complex-valued exponential taper $h_t = \exp(i\pi t/n)$ to first-differenced data $\Delta X_t = X_t - X_{t-1}$. The HC tapered LW estimator extends the valid range to $d_0 \in (-\frac{1}{2}, \frac{3}{2})$ while achieving lower variance than Velasco tapers. It is also robust to linear trends, due to first-differencing the data before applying the taper. After estimating the memory parameter of the differenced data, the researcher adds back one degree of integration to the resulting estimate. The HC method uses shifted frequencies $\lambda_{\tilde{j}} = 2\pi(j+0.5)/n$ to avoid the singularity at zero. The estimator is defined as

$$\hat{d}_{\mathrm{HC}} = 1 + \arg \min_{d \in [\Delta_1 - 1, \Delta_2 - 1]} R_{\mathrm{HC}}(d),$$

where $R_{\mathrm{HC}}(d)$ has the same form as $R(d)$, but uses the tapered periodogram of ΔX_t. The asymptotic variance of the HC estimator is $\frac{1.5}{4}$, which is higher than $\frac{1}{4}$ for the LW estimator, but lower than $\frac{2.1}{4}$ when using a second-order Velasco taper.

In Table 4, we replicate the HC tapered LW estimator results from Shimotsu and Phillips (2005, Table 2). Our results agree closely with the original results, confirming the HC estimator's good finite-sample performance in its valid range $d \in (-\frac{1}{2}, \frac{3}{2})$ and beyond, with degradation at more extreme values. The complex taper successfully reduces bias due to overdifferencing while improving efficiency relative to other tapered estimators. We defer a full cross-method comparison until Section 3.

2.4. Exact Local Whittle Estimation

The standard local Whittle estimator assumes that the power-law relationship between the periodograms of the observed and unobserved processes—$I_u(\lambda_j) \approx \lambda_j^{2d_0} I_X(\lambda_j)$—holds for frequencies near zero. This approximation is valid when $|d_0| < \frac{1}{2}$. The exact local Whittle (ELW) estimator of Shimotsu and Phillips (2005) instead works directly with fractionally differenced data, avoiding this approximation. From the model in (1), we see that the periodogram of the fractionally differenced series $\Delta^d X_t \equiv (1-L)^d X_t$ equals the periodogram of u_t precisely when $d = d_0$, regardless of the value of d_0. By fractionally differencing the series for each candidate d value during optimization, the ELW estimator removes the

Table 4. HC Tapered LW Estimator: Replication of Left Panel of Table 2 of Shimotsu and Phillips (2005)

	Original			Replication		
d	Bias	S.D.	MSE	Bias	S.D.	MSE
−3.5	2.5889	0.3037	6.7946	2.5920	0.2988	6.8075
−2.3	1.1100	0.2893	1.3157	1.1104	0.2882	1.3160
−1.7	0.4474	0.2154	0.2466	0.4493	0.2149	0.2481
−1.3	0.1551	0.1231	0.0392	0.1547	0.1239	0.0393
−0.7	0.0278	0.0957	0.0099	0.0294	0.0959	0.0101
−0.3	0.0100	0.0971	0.0095	0.0134	0.0972	0.0096
0.0	0.0034	0.0985	0.0097	0.0053	0.0978	0.0096
0.3	−0.0033	0.1004	0.0101	−0.0007	0.0983	0.0097
0.7	−0.0066	0.0994	0.0099	−0.0055	0.0983	0.0097
1.3	−0.0079	0.0987	0.0098	−0.0050	0.0974	0.0095
1.7	0.0008	0.0972	0.0095	0.0029	0.0957	0.0092
2.3	0.0528	0.0981	0.0124	0.0557	0.0981	0.0127
3.5	−0.4079	0.1142	0.1795	−0.4069	0.1144	0.1787

Notes: Hurvich-Chen tapered LW estimator, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = \lfloor n^{0.65} \rfloor = 56$ frequencies, 10,000 replications.

approximation error from the LW approach, extending consistency and asymptotic normality to the full parameter space.

The fractional differencing operator is defined by the binomial expansion

$$\Delta^d X_t = \sum_{k=0}^{t-1} \pi_k(d) X_{t-k},$$

where the coefficients $\pi_k(d)$ can be computed using the recurrence

$$\pi_0(d) = 1, \quad \pi_k(d) = \pi_{k-1}(d) \cdot \frac{k-1-d}{k}, \quad k \geq 1.$$

This operator transforms the data to remove d orders of integration, and when $d = d_0$, this operation recovers the short-memory process u_t exactly.

The ELW objective function has a similar functional form as (3) for LW, but is based on the periodogram $I_{\Delta^d X}(\lambda_j)$ of the fractionally differenced series:

$$R_{\mathrm{ELW}}(d) = \log\left(\frac{1}{m}\sum_{j=1}^{m} I_{\Delta^d X}(\lambda_j)\right) - \frac{2d}{m}\sum_{j=1}^{m} \log \lambda_j. \tag{5}$$

Unlike for the LW objective, the periodogram $I_{\Delta^d X}(\lambda_j)$ must be recomputed for each

candidate d. This introduces a nonlinear dependence on d and, as a result, the ELW objective is not guaranteed to be convex.

The primary advantage of the ELW estimator is that it remains valid no matter whether d_0 lies in the stationary region or far in the nonstationary region, provided that the width of the optimization interval is less than $\frac{9}{2}$.[4] Despite this, the ELW estimator retains the same asymptotic distribution as the LW estimator:

$$\sqrt{m}(\hat{d}_{\mathrm{ELW}} - d_0) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \tfrac{1}{4}\right).$$

The computational cost of ELW estimation is higher than for LW. The latter only required one precomputed periodogram calculation and no fractional differencing. For the ELW estimator, for each candidate value of d encountered during optimization, one needs to recompute the fractional difference $\Delta^d X_t$ and the corresponding periodogram $I_{\Delta^d X}$. While computing $\Delta^d X_t$ naïvely requires $O(n^2)$ operations, the fast fractional differencing algorithm of Jensen and Nielsen (2014), based on the fast Fourier transform, reduces this to $O(n \log n)$. The computational burden of ELW is therefore $O(n \log n)$ per objective function evaluation, as compared to $O(m)$ for LW. For this additional cost, we obtain a consistent estimator with asymptotic variance $\frac{1}{4}$ across the full parameter space.

Table 5 presents our replication of the left panel of Table 1 from Shimotsu and Phillips (2005), demonstrating the finite-sample performance of the ELW estimator over an extended range of d values. The table shows the mean, standard deviation, and mean squared error of the estimates over 10,000 replications each with $n = 500$ observations simulated from an ARFIMA$(0, d, 0)$ process spanning $d \in [-3.5, 3.5]$. Our results are very similar to the original results for all d values. The results highlight the primary advantage of the ELW estimator: under ideal conditions, it exhibits minimal bias, standard deviations near the theoretical asymptotic value of $\frac{1}{\sqrt{4m}} = 0.067$, and uniformly low mean squared errors for any value of d. This stands in contrast to the LW estimator's failure outside the stationary region $|d| \geq 0.5$.

2.5. Robustness to Unknown Mean and Trend

The ELW estimator assumes that the stochastic process X_t has zero mean and does not contain a trend. However, real economic time series may have nonzero means and time trends. This led Shimotsu (2010) to develop the two-step exact local Whittle (2ELW) estimator, which addresses both issues while preserving the $\mathrm{N}(0, \frac{1}{4})$ limiting distribution,

[4] The width restriction arises because the proof of consistency requires different techniques when $|d - d_0| > \frac{1}{2}$, which have been established only for $|d - d_0| \leq \frac{9}{2}$. See the discussion on pp. 1894–1895 of Shimotsu and Phillips (2005) for details.

Table 5. ELW Estimator: Replication of Left Panel of Table 1 of Shimotsu and Phillips (2005)

	Original			Replication		
d	Bias	S.D.	MSE	Bias	S.D.	MSE
−3.5	−0.0024	0.0787	0.0062	−0.0014	0.0777	0.0060
−2.3	−0.0020	0.0774	0.0060	−0.0015	0.0777	0.0060
−1.7	−0.0020	0.0776	0.0060	−0.0015	0.0778	0.0061
−1.3	−0.0014	0.0770	0.0059	−0.0015	0.0778	0.0061
−0.7	−0.0024	0.0787	0.0062	−0.0015	0.0778	0.0061
−0.3	−0.0033	0.0777	0.0060	−0.0015	0.0779	0.0061
0.0	−0.0029	0.0784	0.0061	−0.0016	0.0778	0.0061
0.3	−0.0020	0.0782	0.0061	−0.0015	0.0779	0.0061
0.7	−0.0017	0.0777	0.0060	−0.0015	0.0778	0.0061
1.3	−0.0014	0.0781	0.0061	−0.0014	0.0779	0.0061
1.7	−0.0025	0.0780	0.0061	−0.0014	0.0778	0.0061
2.3	−0.0026	0.0772	0.0060	−0.0012	0.0779	0.0061
3.5	−0.0016	0.0770	0.0059	−0.0013	0.0778	0.0061

Notes: Exact local Whittle estimator, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = \lfloor n^{0.65} \rfloor = 56$ frequencies, 10,000 replications.

although with a limited parameter range. The valid range depends on whether a trend is included or not, as we will discuss below.

First, consider the model with an unknown mean:

$$X_t = \mu_0 + X_t^0, \quad X_t^0 = (1-L)^{-d_0} u_t \, 1\{t \geq 1\},$$

where μ_0 is an unknown constant. The problem this introduces is subtle. The best way to handle the mean depends on the unknown value of d_0 itself. For stationary processes ($|d_0| < \frac{1}{2}$), the sample average $\hat{\mu} = \bar{X}$ is a reasonable estimator for μ_0. For nonstationary, highly persistent processes ($d_0 \geq \frac{1}{2}$), the first observation $\hat{\mu} = X_1$ is a better estimator. To address this indeterminacy, Shimotsu (2010) proposed an adaptive mean estimator $\tilde{\mu}(d)$ and a modified ELW objective function based on fractionally differencing the mean-corrected data $X_t - \tilde{\mu}(d)$. The estimator $\tilde{\mu}(d)$ is based on a weight function that varies smoothly with the value of d:

$$\tilde{\mu}(d) = w(d)\bar{X} + (1 - w(d))X_1,$$

where $w(d)$ is the weight function

$$(6) \qquad w(d) = \begin{cases} 1 & \text{if } d \leq \frac{1}{2}, \\ \frac{1}{2}[1 + \cos(4\pi d)] & \text{if } \frac{1}{2} < d < \frac{3}{4}, \\ 0 & \text{if } d \geq \frac{3}{4}. \end{cases}$$

The intuition behind this weighting is as follows: $w(d) = 1$ for stationary processes, so $\bar{X}$ receives full weight. ELW with $\hat{\mu} = \bar{X}$ remains consistent for $d_0 < 1$, but it loses asymptotic normality for $d_0 \geq \frac{3}{4}$. Using X_1 instead extends asymptotic normality to $d_0 < 2$, so $w(d) = 0$ for $d \geq \frac{3}{4}$, placing full weight on X_1. The smooth cosine transition balances these two extremes.

Estimation of the mean in this way significantly complicates the asymptotic theory, so Shimotsu (2010) instead proposed a two-step approach based on a $\sqrt{m}$-consistent first-step estimator. He suggested applying one of the tapered LW estimators discussed above to obtain an initial estimate $\hat{d}_T$. The second step applies a single Newton-Raphson step to the modified ELW objective function starting from $\hat{d}_T$. Shimotsu (2010) notes that the Newton-Raphson approach can become numerically unstable when the Hessian $R''(\hat{d}_T)$ takes very small values, resulting in extremely large updates. To address this instability, he recommended using $\max\{R''(\hat{d}_T), 2\}$ instead of the actual Hessian in the Newton-Raphson step. The lower bound avoids extremely large steps leading to extreme estimates $\hat{d}_{2\text{ELW}}$.

For data believed to contain a polynomial trend of order k, the procedure first removes the trend via an OLS regression of X_t on $(1, t, t^2, \ldots, t^k)$, and then applies the above two-step procedure to the residuals. Shimotsu (2010) shows that the 2ELW estimator is consistent with detrending, but the valid parameter range is restricted to $d_0 \in (-\frac{1}{2}, 1.75)$, compared to $d_0 \in (-\frac{1}{2}, 2)$ for the unknown mean case. The procedure achieves the same $\mathrm{N}(0, \frac{1}{4})$ limiting distribution as LW within these ranges while maintaining robustness to both unknown means and polynomial trends.

Table 6 presents our replication of the 2ELW Monte Carlo results from Table 2 of Shimotsu (2010). Our results show close agreement with the original results across the parameter range. The 2ELW estimator maintains good finite-sample properties with minimal bias and low variance for $m = 57$, where the asymptotic variance is $\frac{1}{4m} = 0.0044$. Performance remains stable across the stationary ($d = 0.0, 0.4$) and nonstationary ($d = 0.8, 1.2$) regions. Thus, the two-step procedure successfully combines consistency and robustness of the first-stage tapered estimator with the efficiency of the second-stage ELW approach.

This experiment verifies the basic theoretical properties. We only consider specifications from Table 2 of Shimotsu (2010) for which there are no short-run dynamics ($\rho = 0$). Later in Section 3, we compare all estimators across a more thorough set of benchmarks including

Table 6. 2ELW Estimator: Replication of Table 2 of Shimotsu (2010)

	Original		Replication	
d	Bias	Variance	Bias	Variance
0.0	−0.0022	0.0058	−0.0019	0.0057
0.4	0.0001	0.0058	0.0053	0.0066
0.8	−0.0003	0.0058	−0.0012	0.0058
1.2	−0.0006	0.0057	−0.0014	0.0060

Notes: Two-step ELW estimator, $n = 512$ observations from ARFIMA$(0, d, 0)$, $m = \lfloor n^{0.65} \rfloor = 57$ frequencies, 10,000 replications.

cases with AR(1) short-term dynamics with $\rho > 0$, unknown means, and time trends. In these simulations, we will further investigate 2ELW's robustness to such contaminations.

2.6. Robustness to Low-Frequency Contamination

In practice, time series may be contaminated by level shifts that distort the spectrum at low frequencies. Diebold and Inoue (2001) and Granger and Hyung (2004) showed that such contaminations can manifest as spurious long memory. To address this, Hou and Perron (2014) developed a modified local Whittle estimator designed to be robust to low-frequency contamination (LWLFC) such as random level shifts, deterministic level shifts, and certain types of normalized trends. However, LWLFC is only valid for stationary processes with $d_0 \in [0, \frac{1}{2})$.

Specifically, let Y_t denote the observed time series, which is assumed to follow

$$Y_t = \mu_0 + X_t + L_t,$$

where μ_0 is a constant, X_t is a fractionally integrated process with memory parameter d_0 as in (1), and L_t captures low-frequency disturbances.

The estimator introduces an additional parameter $\theta \geq 0$, representing a contamination ratio—the variance ratio of the low-frequency contamination L_t to the long-memory component X_t. The modified objective function is

$$J(d, \theta) = \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{I_Y(\lambda_j)}{g_j}\right) + \frac{1}{m}\sum_{j=1}^{m}\log(g_j),$$

where $g_j = \lambda_j^{-2d} + \theta\lambda_j^{-2}/n$. The term $\theta\lambda_j^{-2}/n$ corrects the spectral density for contribution

of the low-frequency contamination. The LWLFC estimator is defined as

$$(\hat{d}_{\text{LWLFC}}, \hat{\theta}_{\text{LWLFC}}) = \underset{(d,\theta)\in[0,\frac{1}{2})\times[0,\infty)}{\arg\min} J(d,\theta).$$

Hou and Perron (2014) state regularity conditions for which the LWLFC estimator is consistent and asymptotically normal with

$$\sqrt{m}(\hat{d}_{\text{LWLFC}} - d_0) \xrightarrow{\text{d}} \text{N}\left(0, \frac{1}{4}\right).$$

The asymptotic variance is the same as that of the standard LW estimator, meaning that accounting for low-frequency contamination does not reduce efficiency. If there is no contamination, then $\hat{\theta}_{\text{LWLFC}} \xrightarrow{\text{p}} 0$ and the LWLFC estimator is asymptotically equivalent to the standard LW estimator.

3. Monte Carlo Comparisons Across Methods

In this section, we compare the finite-sample performance of the six local Whittle estimators across several data-generating processes. We consider the impact of short-run dynamics, unknown means, and time trends, each under varying degrees of fractional integration.

Our Monte Carlo specifications follow conventions established in the literature. We use a sample size of $n = 500$ with Gaussian innovations and bandwidth $m = \lfloor n^{0.65} \rfloor$. These choices are consistent with the baseline specifications in Shimotsu and Phillips (2005), Shimotsu (2010), and Nielsen and Frederiksen (2005). Hurvich and Chen (2000) also used $n = 500$ and Gaussian innovations, but with bandwidth $m = \lfloor 0.25n^{0.80} \rfloor$. To focus on the performance of estimators without bandwidth variability, for now we use a fixed bandwidth rule. We will evaluate data-driven bandwidth choices in empirical applications later in Section 4. Our empirical case studies also compare estimator performance with both shorter and longer time series.

3.1. Comprehensive Estimator Comparison with Short-Run Dynamics

To establish a baseline for cross-estimator comparison, we revisit the Monte Carlo design from Table I of Hurvich and Chen (2000). Their experiment used 500 replications of simulated ARFIMA$(1, d, 0)$ processes with AR(1) short-run dynamics (autocorrelation parameter ρ). We use bandwidth $m = 56$ following the standard rule $m = \lfloor n^{0.65} \rfloor$ mentioned above.

We extend their original study by comparing all six estimators: LW, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW, 2ELW, and LWLFC. Furthermore, we evaluate

these estimators over a wider range of 33 parameter combinations

$$(d, \rho) \in \{-2.2, -1.8, -1.2, -0.6, -0.3, 0.0, 0.3, 0.6, 1.2, 1.8, 2.2\} \times \{0.0, 0.5, 0.8\}.$$

This allows for a more thorough evaluation across ranges of fractional integration from strongly persistent to antipersistent. The results, reported in Table 7, reveal several patterns regarding the relative performance of estimators under short-run AR(1) dynamics.

Table 7. Comprehensive Estimator Comparison

		Bias						MSE					
d	ρ	LW	V	HC	ELW	2ELW	LWLFC	LW	V	HC	ELW	2ELW	LWLFC
-2.2	0.0	1.505	0.072	0.992	-0.002	1.198	1.259	2.359	0.032	1.064	0.006	1.505	1.605
-1.8	0.0	0.991	0.057	0.548	-0.002	0.753	0.824	1.067	0.030	0.356	0.006	0.617	0.685
-1.2	0.0	0.318	0.039	0.111	-0.003	0.271	0.229	0.138	0.028	0.024	0.006	0.091	0.057
-0.6	0.0	0.020	0.031	0.025	-0.002	0.003	-0.007	0.007	0.028	0.010	0.006	0.006	0.008
-0.3	0.0	-0.002	0.024	0.011	-0.002	-0.004	-0.024	0.006	0.027	0.010	0.006	0.006	0.009
0.0	0.0	-0.007	0.026	0.005	-0.002	-0.002	-0.033	0.006	0.027	0.010	0.006	0.006	0.010
0.3	0.0	-0.006	0.024	-0.002	-0.002	-0.001	-0.046	0.006	0.027	0.010	0.006	0.006	0.014
0.6	0.0	0.003	0.029	-0.007	-0.003	0.012	-0.082	0.006	0.027	0.010	0.006	0.006	0.045
1.2	0.0	-0.125	0.036	-0.008	-0.002	-0.002	-0.125	0.022	0.028	0.010	0.006	0.006	0.024
1.8	0.0	-0.738	0.063	0.004	-0.003	-0.003	-0.733	0.560	0.031	0.009	0.006	0.006	0.557
2.2	0.0	-1.160	0.095	0.038	-0.002	-0.002	-1.156	1.358	0.036	0.011	0.006	0.006	1.353
-2.2	0.5	1.318	0.191	0.881	0.100	1.088	1.206	1.825	0.063	0.847	0.016	1.248	1.456
-1.8	0.5	0.843	0.178	0.487	0.100	0.693	0.803	0.778	0.058	0.275	0.016	0.515	0.645
-1.2	0.5	0.275	0.161	0.181	0.101	0.275	0.237	0.096	0.052	0.043	0.016	0.089	0.060
-0.6	0.5	0.110	0.150	0.132	0.100	0.100	0.099	0.019	0.050	0.027	0.016	0.016	0.017
-0.3	0.5	0.098	0.148	0.122	0.100	0.100	0.089	0.016	0.049	0.025	0.016	0.016	0.015
0.0	0.5	0.094	0.145	0.114	0.100	0.100	0.084	0.015	0.048	0.023	0.016	0.016	0.014
0.3	0.5	0.093	0.145	0.108	0.099	0.101	0.079	0.015	0.048	0.022	0.016	0.017	0.015
0.6	0.5	0.102	0.147	0.104	0.100	0.105	0.066	0.017	0.048	0.021	0.016	0.016	0.032
1.2	0.5	-0.096	0.158	0.102	0.099	0.099	-0.094	0.020	0.052	0.021	0.016	0.016	0.020
1.8	0.5	-0.735	0.183	0.114	0.100	0.100	-0.732	0.558	0.060	0.023	0.016	0.016	0.556
2.2	0.5	-1.160	0.207	0.144	0.100	0.100	-1.157	1.358	0.070	0.030	0.016	0.016	1.354
-2.2	0.8	1.250	0.536	0.925	0.414	1.111	1.205	1.638	0.316	0.905	0.179	1.278	1.454
-1.8	0.8	0.824	0.526	0.632	0.417	0.792	0.813	0.723	0.306	0.418	0.181	0.649	0.663
-1.2	0.8	0.475	0.513	0.483	0.417	0.457	0.465	0.234	0.294	0.243	0.181	0.215	0.225
-0.6	0.8	0.414	0.502	0.457	0.415	0.415	0.413	0.179	0.282	0.220	0.180	0.179	0.178
-0.3	0.8	0.409	0.499	0.450	0.416	0.416	0.408	0.174	0.279	0.214	0.180	0.180	0.174
0.0	0.8	0.406	0.503	0.446	0.416	0.419	0.406	0.172	0.282	0.211	0.180	0.184	0.172
0.3	0.8	0.406	0.499	0.441	0.417	0.419	0.405	0.172	0.279	0.206	0.181	0.183	0.172
0.6	0.8	0.396	0.501	0.437	0.417	0.417	0.390	0.164	0.280	0.202	0.181	0.181	0.171
1.2	0.8	-0.035	0.511	0.435	0.417	0.417	-0.034	0.032	0.289	0.201	0.181	0.181	0.032
1.8	0.8	-0.731	0.526	0.441	0.416	0.416	-0.728	0.558	0.304	0.205	0.181	0.180	0.556
2.2	0.8	-1.158	0.551	0.462	0.416	0.416	-1.156	1.355	0.331	0.224	0.181	0.180	1.352

Notes: Monte Carlo results for ARFIMA$(1, d, 0)$ processes with $n = 500$, $m = 56$, 10,000 replications. Shaded cells indicate MSE > 0.05.
LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW, LWLFC = Local Whittle with Low Frequency Contamination.

Since we deliberately evaluated some estimators outside their valid parameter ranges (e.g., 2ELW for $d < -0.5$) this allows us to characterize failure modes, which can be informative for practitioners who do not know d a priori. In the absence of short-run dynamics ($\rho = 0$),

the ELW estimator performs best overall, with minimal bias (-0.003) and MSE values (0.006). The 2ELW estimator has similar behavior, but fails with large bias for antipersistent processes with $d < -0.6$. This is consistent with its theoretical restriction to $d > -0.5$. The standard LW estimator also fails outside its valid range $|d| < 0.5$, with bias of -1.160 and MSE of 1.358 at $d = 2.2$. Similarly, LWLFC shows good performance in its valid range $0 \leq d < 0.5$, extending somewhat beyond, but fails outside that range. The Velasco (V) estimator has positive bias larger than other estimators in the valid range, while the Hurvich-Chen (HC) estimator shows intermediate accuracy with generally smaller bias than V but larger MSE than the exact methods.

The addition of moderate AR(1) persistence ($\rho = 0.5$) changes the relative performance landscape. All estimators show upward bias, with the ELW and 2ELW methods maintaining their advantage, although with higher bias (≈ 0.100). The performance of the standard LW estimator remains good in the valid range ($|d| < 0.5$), and actually improves somewhat for highly antipersistent cases, but still fails for extreme values of d, with bias of -0.735 at $d = 1.8$ and -1.160 at $d = 2.2$. The Velasco estimator maintains its stability but with higher mean bias for all d values. The HC estimator again shows intermediate performance, with bias increasing with $|d|$ as before.

With stronger AR(1) persistence ($\rho = 0.8$), the performance of all estimators suffers. Bias for central $|d|$ values exceeds 0.40 for most methods and MSE values increase by an order of magnitude. The ELW estimator has the smallest uniform bias (≈ 0.416), while the Velasco estimator exhibits bias around 0.500 for most d values. The LW estimator's behavior is somewhat erratic, with the lowest bias of all estimators at $d = 1.2$ but the highest bias of all at $d = 2.2$ and $d = -2.2$. As before, the performance of 2ELW is similar to ELW for central values of d, but it fails more severely for large negative d.

In terms of overall robustness to AR(1) dynamics, the exact methods (ELW and 2ELW) show the most stability. Their performance degrades smoothly as autocorrelation increases. Although the behavior of the Velasco estimator is also very uniform across d values, the bias and MSE is generally higher than that of ELW and 2ELW, while the HC estimator shows intermediate robustness and more sensitivity under antipersistence. The standard LW and LWLFC estimators are most vulnerable to autocorrelation when $|d|$ is large, noting that such values are beyond their valid range. Among these estimators, we therefore recommend using exact local Whittle methods when contamination from AR(1) dynamics is a concern, with the exception of 2ELW when strong antipersistence is expected. Researchers may also consider the local polynomial Whittle estimator of Andrews and Sun (2004), noting that this method is also designed for stationary processes with $|d| < 0.5$. Using a more conservative or data-driven bandwidth is another simple step towards robustness.

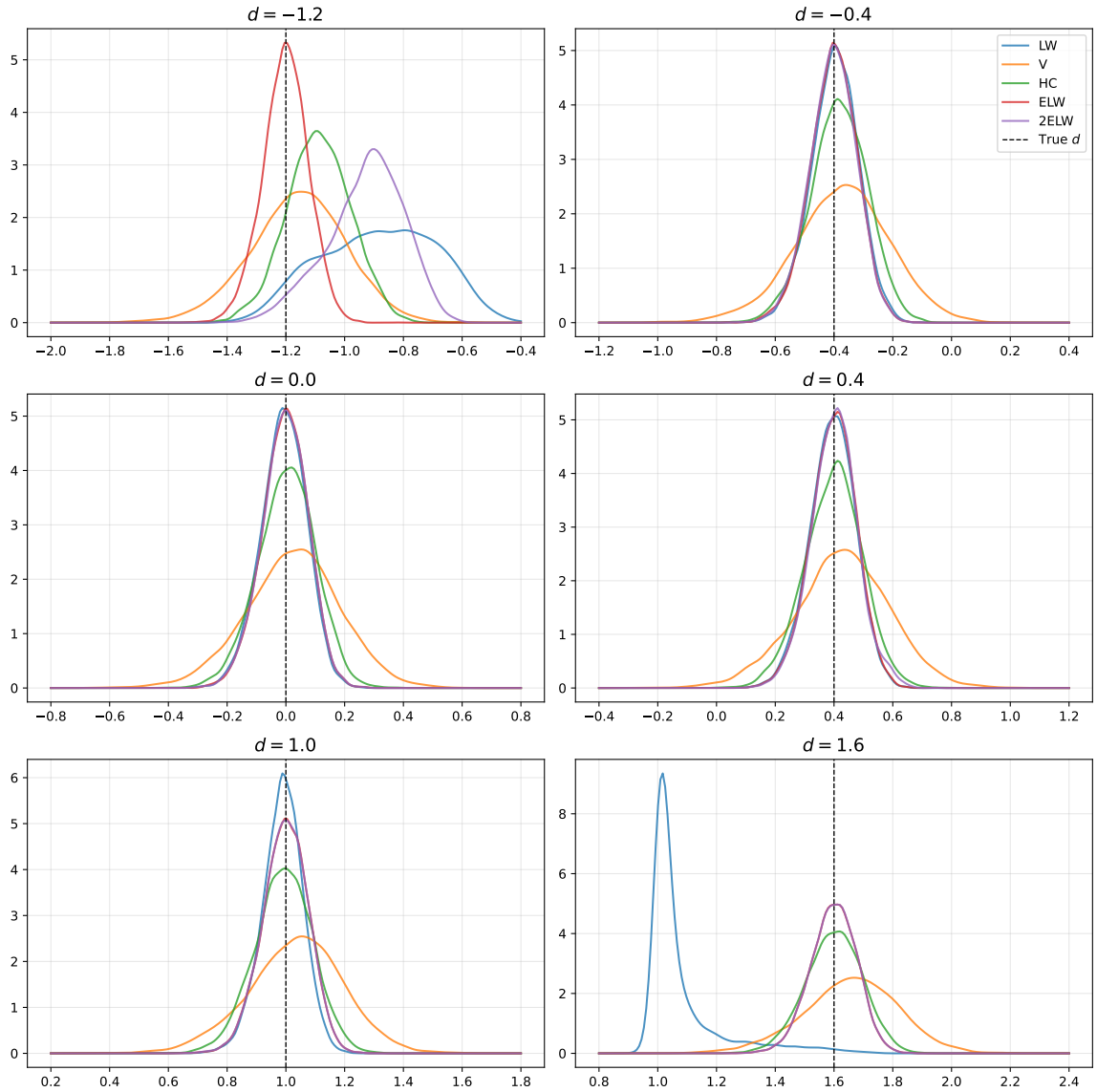

FIGURE 1. Sampling distributions of local Whittle estimators with $n = 500$ over 10,000 replications. V denotes the Velasco estimator with Kolmogorov taper.

3.2. Sampling Distribution Comparison

To further illustrate the finite-sample behavior of local Whittle estimators across a wide range of memory parameters, we plot the sampling distributions of the estimators in different scenarios. Figure 1 presents kernel density estimates of the sampling distributions for five estimators—LW, V, HC, ELW, and 2ELW—based on 10,000 replications of ARFIMA$(0, d, 0)$ processes with $n = 500$ observations and bandwidth $m = \lfloor n^{0.65} \rfloor$. The analysis spans six representative values of the memory parameter: $d \in \{-1.2, -0.4, 0.0, 0.4, 1.0, 1.6\}$, covering

antipersistent, stationary long memory, and strongly persistent processes.

The distributions provide verification of the theoretical properties discussed in Section 2 and offer insights about estimator performance in different persistence regimes. For the stationary models ($|d| < 0.5$), in accordance with theory, all estimators have approximately symmetric distributions centered around the true parameter value. The less efficient tapered estimators (V and HC) exhibit increased dispersion relative to the more efficient LW, ELW, and 2ELW estimators. In the case of nonstationary, explosive processes ($d = 1.6$), the LW estimator is inconsistent (Phillips and Shimotsu, 2004, Theorem 3.2). Its distribution concentrates around $d = 1.0$, while the exact methods (ELW and 2ELW) remain centered around the true values. For the highly antipersistent case ($d = -1.2$), in accordance with the theoretical properties, only the ELW estimator remains unbiased and concentrated around the true value. LW and 2ELW are biased upward, with means around -0.9. The V and HC estimators also have positive bias, but to a lesser degree.

3.3. Robustness to Unknown Mean

Next, we consider the sensitivity of the estimators to unknown population means. We compare performance across three approaches to mean correction: no correction ($\hat{\mu} = 0$), sample mean correction ($\hat{\mu} = \bar{X}$), and first-observation correction ($\hat{\mu} = X_1$). Table 8 shows baseline MSE values (with $\mu = 0$) and MSE ratios (with $\mu = 5$ relative to $\mu = 0$) for simulated ARFIMA$(0, d, 0)$ processes with $d \in [-2.2, 2.2]$. Ratios near unity indicate robustness to nonzero means; larger values indicate performance degradation. For each correction, we transform the data as $\tilde{X}_t = X_t - \hat{\mu}$ before applying the estimator. We apply this transformation in both cases $\mu = 0$ and $\mu = 5$. The only exception is 2ELW, which receives the original data and applies its own internal adaptive mean correction. (As a result, the MSE values for 2ELW are identical across the three mean corrections.)

In the top panel, with no mean correction, the presence of a nonzero mean reveals differences in robustness. The Velasco estimator shows both good baseline performance (MSE ≈ 0.03) and robustness (MSE ratios near unity) for all d values, providing reliable behavior in the presence of mean contamination. The LW, HC, 2ELW, and LWLFC estimators also exhibit MSE ratios near unity. This indicates insensitivity to the mean, but note that their baseline MSE values are elevated at extreme d values (e.g., LW has baseline MSE of 2.359 at $d = -2.2$). In those cases, a MSE ratio near 1.0 indicates consistently poor performance rather than robustness. In contrast, ELW has extremely low baseline MSE values (≈ 0.006 across d), but exhibits catastrophic sensitivity to mean contamination for $d \leq 0.6$, especially for antipersistent processes (MSE ratio of 5.3 at $d = 0.0$, increasing to 1410 at $d = -2.2$), with robustness only for $d \geq 1.2$.

In the middle panel, we apply a sample mean correction ($\hat{\mu} = \bar{X}$). This has the effect of

Table 8. Robustness to Unknown Mean

	Baseline MSE ($\mu = 0$)						MSE Ratio ($\mu = 5$ / $\mu = 0$)					
d	LW	V	HC	ELW	2ELW	LWLFC	LW	V	HC	ELW	2ELW	LWLFC
Mean correction: None												
-2.2	2.359	0.032	1.064	0.006	1.505	1.605	0.993	0.987	0.999	1410.208	0.994	1.001
-1.8	1.077	0.029	0.355	0.006	0.618	0.684	0.993	1.024	0.988	1024.138	0.993	0.999
-1.2	0.139	0.028	0.024	0.006	0.091	0.057	0.970	1.013	0.997	537.797	0.983	0.988
-0.6	0.007	0.028	0.010	0.006	0.006	0.008	0.966	0.976	0.981	137.055	0.964	0.980
-0.3	0.006	0.027	0.010	0.006	0.006	0.009	1.010	0.993	1.004	32.781	1.004	0.985
0.0	0.006	0.027	0.010	0.006	0.006	0.010	0.978	1.006	0.977	5.359	0.979	0.970
0.3	0.006	0.027	0.010	0.006	0.006	0.014	1.016	0.985	1.020	14.515	1.016	0.997
0.6	0.007	0.028	0.010	0.006	0.006	0.047	0.973	0.963	0.981	2.369	0.974	0.936
1.2	0.022	0.027	0.010	0.006	0.006	0.025	1.007	1.009	0.997	1.019	1.006	1.014
1.8	0.557	0.031	0.009	0.006	0.006	0.554	1.006	1.003	1.011	1.032	1.027	1.006
2.2	1.359	0.036	0.011	0.006	0.006	1.355	1.001	0.993	0.990	1.015	1.010	1.001
Mean correction: $\hat{\mu} = \bar{X}$												
-2.2	2.359	0.032	1.064	2.099	1.505	1.605	0.993	0.987	0.999	0.995	0.994	1.001
-1.8	1.077	0.029	0.355	0.980	0.618	0.684	0.993	1.024	0.988	0.996	0.993	0.999
-1.2	0.139	0.028	0.024	0.121	0.091	0.057	0.970	1.013	0.997	0.977	0.983	0.987
-0.6	0.007	0.028	0.010	0.006	0.006	0.008	0.966	0.976	0.981	0.963	0.964	0.980
-0.3	0.006	0.027	0.010	0.006	0.006	0.009	1.010	0.993	1.004	1.004	1.004	0.985
0.0	0.006	0.027	0.010	0.006	0.006	0.010	0.978	1.006	0.977	0.979	0.979	0.970
0.3	0.006	0.027	0.010	0.006	0.006	0.014	1.016	0.985	1.020	1.017	1.016	0.997
0.6	0.007	0.028	0.010	0.006	0.006	0.047	0.973	0.963	0.981	0.968	0.974	0.933
1.2	0.022	0.027	0.010	0.012	0.006	0.025	1.007	1.009	0.997	1.009	1.006	1.014
1.8	0.557	0.031	0.009	0.425	0.006	0.554	1.006	1.003	1.011	1.008	1.027	1.006
2.2	1.359	0.036	0.011	1.120	0.006	1.355	1.001	0.993	0.990	1.002	1.010	1.001
Mean correction: $\hat{\mu} = X_1$												
-2.2	2.359	0.032	1.064	4.836	1.505	1.605	0.993	0.987	0.999	1.001	0.994	1.001
-1.8	1.077	0.029	0.355	3.128	0.618	0.684	0.993	1.024	0.988	1.003	0.993	0.999
-1.2	0.139	0.028	0.024	1.296	0.091	0.057	0.970	1.013	0.997	0.999	0.983	0.987
-0.6	0.007	0.028	0.010	0.278	0.006	0.008	0.966	0.976	0.981	1.003	0.964	0.980
-0.3	0.006	0.027	0.010	0.058	0.006	0.009	1.010	0.993	1.004	0.998	1.004	0.985
0.0	0.006	0.027	0.010	0.003	0.006	0.010	0.978	1.006	0.977	0.994	0.979	0.970
0.3	0.006	0.027	0.010	0.009	0.006	0.014	1.016	0.985	1.020	1.000	1.016	0.997
0.6	0.007	0.028	0.010	0.006	0.006	0.047	0.973	0.963	0.981	0.965	0.974	0.932
1.2	0.022	0.027	0.010	0.006	0.006	0.025	1.007	1.009	0.997	1.006	1.006	1.015
1.8	0.557	0.031	0.009	0.006	0.006	0.554	1.006	1.003	1.011	1.026	1.027	1.005
2.2	1.359	0.036	0.011	0.006	0.006	1.355	1.001	0.993	0.990	1.010	1.010	1.001

Notes: MSE results for ARFIMA$(0, d, 0)$ with $n = 500$, $m = 56$, 10,000 replications. Shaded cells indicate MSE > 0.05 or MSE Ratio > 2.0.
LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW with adaptive mean estimation applied to original series, LWLFC = Local Whittle with Low Frequency Contamination.

equalizing the relative performance of estimators for different d values. All estimators have MSE ratios near unity, meaning that their performance is similar whether the true mean is zero or nonzero. However, note that the performance of ELW at baseline is poor, with MSE of 2.099 at $d = -2.2$ and 1.120 at $d = 2.2$ for ELW. This indicates that sample mean correction itself—improperly demeaning data with $\mu = 0$ when $|d|$ is large—significantly affects the performance of ELW. The other estimators maintain their baseline performance levels with this correction. For example, although LW also has high baseline MSE values, by comparing the top left and middle left panels we can see that those are not affected by demeaning.

In the bottom panel, we consider the effects of the first-observation correction ($\hat{\mu} = X_1$). Similar to the case of the sample mean correction, all estimators have MSE ratios near unity. However, for ELW the baseline MSE effects are asymmetric: baseline MSE is elevated for $d < 0$ (4.836 at $d = -2.2$), but remains at the low level of 0.006 for $d \geq 0$. Therefore, in line with the 2ELW approach of Shimotsu (2010), this correction avoids the performance degradation for ELW that we saw with $\bar{X}$ for highly persistent processes, though at the cost of elevated baseline MSE for antipersistent processes.

Overall, these results demonstrate that the Velasco tapered LW estimator provides remarkable robustness to unknown means—corrected or not—over the full range of d values. LW, HC, and 2ELW are largely unaffected, but have ranges of d values where they perform much worse than Velasco. For ELW, the best mean correction critically depends on the expected range of d, which is precisely the intuition behind the 2ELW estimator's adaptive mean correction.

3.4. Robustness to Linear Time Trends

The Monte Carlo results in Table 9 examine the sensitivity of the estimators to linear deterministic time trends across the parameter range. The experiment involves simulated ARFIMA$(0, d, 0)$ processes under two scenarios: no trend ($\beta = 0.0$) and linear trend ($\beta = 0.05$). For both specifications, we report the performance both with and without detrending. As in the previous section, this table displays baseline MSE values (for $\beta = 0.0$) and MSE ratios ($\beta = 0.05$ relative to $\beta = 0.0$). For LW, V, HC, and ELW, in the "linear OLS detrending" panel we apply the estimators to pre-processed data that has been detrended (i.e., residuals from an OLS regression on a constant and t.) In contrast, we always apply 2ELW to the original unprocessed data and allow it to apply its own internal linear detrending. As a result, the 2ELW results are identical in the top and bottom panels.

With no trend correction (top panel), the Velasco and Hurvich-Chen tapered LW estimators demonstrate both good baseline MSE and robustness to a linear trend (i.e., MSE ratios are near unity). In contrast, LW, ELW, and LWLFC are highly sensitive to the trend

Table 9. Robustness to Time Trend

	Baseline MSE ($\beta = 0.0$)						MSE Ratio ($\beta = 0.05$ / $\beta = 0.0$)					
d	LW	V	HC	ELW	2ELW	LWLFC	LW	V	HC	ELW	2ELW	LWLFC
Trend correction: None												
-2.2	2.359	0.032	1.064	0.006	1.530	1.605	4.311	1.019	0.999	1487.482	0.995	6.333
-1.8	1.077	0.029	0.355	0.006	0.626	0.684	7.213	1.027	0.988	1099.768	0.993	11.358
-1.2	0.139	0.028	0.024	0.006	0.102	0.057	34.315	1.013	0.997	641.153	0.987	83.832
-0.6	0.007	0.028	0.010	0.006	0.007	0.008	336.676	0.976	0.981	287.097	0.973	293.749
-0.3	0.006	0.027	0.010	0.006	0.006	0.009	259.778	0.993	1.004	170.877	0.995	181.603
0.0	0.006	0.027	0.010	0.006	0.007	0.010	147.260	1.006	0.977	80.003	0.974	87.692
0.3	0.006	0.027	0.010	0.006	0.007	0.014	68.177	0.985	1.020	25.295	1.012	29.449
0.6	0.007	0.028	0.010	0.006	0.007	0.047	16.721	0.963	0.981	2.257	0.975	3.524
1.2	0.022	0.027	0.010	0.006	0.006	0.025	1.044	1.009	0.997	1.010	1.003	1.072
1.8	0.557	0.031	0.009	0.006	0.006	0.554	1.006	1.003	1.011	1.026	0.997	1.006
2.2	1.359	0.036	0.011	0.006	0.011	1.355	1.001	0.993	0.990	1.010	0.998	1.001
Trend correction: linear OLS detrending												
-2.2	2.551	0.032	1.064	2.443	1.530	1.585	0.994	0.987	0.999	0.994	0.995	1.001
-1.8	1.211	0.029	0.355	1.174	0.626	0.677	0.994	1.024	0.988	0.994	0.993	1.000
-1.2	0.180	0.028	0.024	0.177	0.102	0.054	0.973	1.013	0.997	0.974	0.987	0.988
-0.6	0.008	0.028	0.010	0.007	0.007	0.009	0.977	0.976	0.981	0.973	0.973	0.990
-0.3	0.006	0.027	0.010	0.006	0.006	0.009	1.001	0.993	1.004	0.995	0.995	0.985
0.0	0.007	0.027	0.010	0.007	0.007	0.010	0.974	1.006	0.977	0.974	0.974	0.969
0.3	0.007	0.027	0.010	0.007	0.007	0.013	1.012	0.985	1.020	1.012	1.012	0.987
0.6	0.008	0.028	0.010	0.007	0.007	0.029	0.974	0.963	0.981	0.978	0.975	0.952
1.2	0.015	0.027	0.010	0.009	0.006	0.020	0.989	1.009	0.997	0.980	1.003	1.043
1.8	0.317	0.031	0.009	0.317	0.006	0.308	1.012	1.003	1.011	1.013	0.997	1.013
2.2	0.870	0.036	0.011	1.029	0.011	0.837	1.011	0.993	0.990	1.006	0.998	1.009

Notes: MSE results for ARFIMA$(0, d, 0)$, n=500, m=56, 10,000 replications. Shaded cells indicate MSE > 0.05 or MSE Ratio > 2.0.
LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW with linear detrending and adaptive mean estimation applied to original series, LWLFC = Local Whittle with Low Frequency Contamination.

for $d \leq 0.6$. MSE ratios for LW exceed 100 for $-0.6 \leq d \leq 0.0$ (despite already poor baseline MSE at extreme d). MSE ratios for ELW are even larger, surpassing 170 at $d = -0.3$ and 1400 at $d = -2.2$ (despite excellent baseline MSE of 0.006). All three estimators are less sensitive to the trend as d increases (MSE ratios near unity for $d \geq 1.2$). The baseline MSE of 2ELW is elevated for $d \leq -1.2$, yet good robustness to the trend with MSE ratios near unity for all d values.

Detrending by OLS (bottom panel) equalizes performance between trend scenarios, with all estimators exhibiting MSE ratios near unity. However, detrending when no trend is actually present can itself be detrimental to baseline performance for ELW in particular. ELW baseline MSE increases from 0.006 to 2.443 at $d = -2.2$ and 1.029 at $d = 2.2$. The LW, Velasco, HC, and 2ELW estimators largely maintain their baseline performance with detrending.

These results illustrate an important practical point: the Velasco and Hurvich-Chen tapered LW estimators provide robustness to linear deterministic trends without any preprocessing. The internal trend correction of 2ELW is also effective. On the other hand, the LW, ELW, and LWLFC estimators fail unless the researcher detrends the data, and may still fail in spite of the correction for more extreme values of d.

4. Empirical Applications

We now turn from controlled Monte Carlo experiments to real economic data, where the memory parameter is unknown and the data may violate ideal assumptions. This section replicates and extends the empirical analysis of Hurvich and Chen (2000). Using their original datasets, but now comparing across multiple estimators, we examine anomalous estimates, bandwidth selection, and structural breaks.

4.1. Replication and Extension of Table III

We begin by replicating and extending the empirical analysis in Table III of Hurvich and Chen (2000), who applied their complex tapered LW estimator to seven real-world datasets. Motivated by their work, in Table 10 we compare five LW estimators using the same datasets.[5] These data include economic, financial, and environmental time series.

The global temperature series contains seasonally adjusted monthly temperatures for the northern hemisphere from 1854–1989, computed as deviations from monthly averages relative to 1950–1979 (Beran, 1994). The S&P 500 series consists of the natural logarithms of the daily stock index levels from July 1962 to December 1995. The remaining series were drawn from the International Monetary Fund's International Financial Statistics database and consist of monthly economic indicators from January 1957 to December 1997. These include CPI-based inflation rates for the U.S., U.K., and France (differences in logarithms of CPI), the logarithm of U.S. real manufacturing wages,[6] and the logarithm of U.S. industrial production. All of these economic series were seasonally adjusted.

Table 10 compares the original Hurvich and Chen (2000) results (for HC only) with our independent results comparing all estimators produced using our Python implementation. Following the original paper, we report estimates up to two decimal places and standard errors to three decimal places. For all estimators, we use the same number of frequencies m as in the original study, where bandwidths were selected based on visual inspection of

[5]The reported number of observations is increased by one because Hurvich and Chen (2000) report the number of observations after first differencing.

[6]We note one important exception in our replication. Since the original IMF data on real manufacturing wages for the U.S. no longer appear to be available, we instead use data from FRED on average hourly earnings of production and nonsupervisory employees in manufacturing (CES3000000008) divided by CPI.

Table 10. Hurvich and Chen (2000) Datasets: LW Estimator Comparison

			Original	Replication				
Series	n	m	HC	LW	V	HC	ELW	2ELW
Global temp.	1632	130	0.45	0.50	0.44	0.45	0.50	0.47
			(0.060)	(0.046)	(0.076)	(0.060)	(0.045)	(0.044)
S&P 500	8432	1383	0.99	0.98	0.98	0.97	0.99	0.98
			(0.018)	(0.014)	(0.023)	(0.017)	(0.013)	(0.013)
Inflation, US	491	40	0.57	0.63	0.66	0.56	0.64	0.64
			(0.123)	(0.080)	(0.137)	(0.121)	(0.082)	(0.079)
Inflation, UK	491	40	0.33	0.46	0.49	0.34	0.47	0.47
			(0.123)	(0.072)	(0.137)	(0.121)	(0.074)	(0.079)
Inflation, FR	491	40	0.67	0.46	0.82	0.68	0.50	0.45
			(0.123)	(0.083)	(0.137)	(0.121)	(0.087)	(0.079)
Real wages, US	492	35	1.43	1.06	1.56	1.43	0.06	1.29
			(0.121)	(0.096)	(0.147)	(0.132)	(0.007)	(0.085)
Ind. prod., US	492	100	1.34	1.00	1.31	1.36	0.10	1.31
			(0.075)	(0.054)	(0.087)	(0.071)	(0.003)	(0.050)

Notes: Estimates of d with standard errors in parentheses. Original column shows results from Hurvich and Chen (2000) Table III for HC. Replication columns report results from multiple estimators: LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW.

log-log periodogram plots. For the HC tapered LW estimator, the agreement between our results and the original estimates is close for all datasets, with differences typically at the second decimal place or smaller.

In terms of the economic implications of these estimates, there is broad agreement across estimators. Global temperatures, with $\hat{d} \in [0.44, 0.50]$, appear to be stationary with moderate long memory. The S&P 500 index appears to have near-unit root behavior, with $\hat{d} \in [0.97, 0.99]$. The inflation series reveal heterogeneous persistence across countries: U.S. inflation has moderate persistence, with $\hat{d} \in [0.56, 0.66]$, while U.K. inflation has somewhat weaker dependence, with $\hat{d} \in [0.34, 0.49]$. The five estimators yield widely varying estimates of inflation persistence in France, with $\hat{d} \in [0.45, 0.82]$. This represents a very interesting case study that we examine in more detail in Section 4.4.

U.S. real manufacturing wages appear to be highly nonstationary, with $\hat{d} \in [1.06, 1.56]$ (with the notable exception of ELW), suggesting that wage shocks have permanent effects. The lower estimate from LW ($\hat{d} = 1.06$) is consistent with the inconsistency of LW for $d > 1$ (as we saw in our Monte Carlo results). ELW's extremely low estimate ($\hat{d} = 0.06$) reflects sensitivity to the unknown mean relative to 2ELW and other methods based on tapering. We will return to this point in the following section. Similarly, industrial production shows evidence of nonstationarity, with estimates $\hat{d} \in [1.00, 1.36]$ (again excluding ELW, and with LW being apparently bounded near 1.0), indicating that productivity shocks have persistent effects on growth. These results add nuance to the traditional $I(0)$ or $I(1)$ classifications of time series, underscoring the empirical relevance of fractional integration.

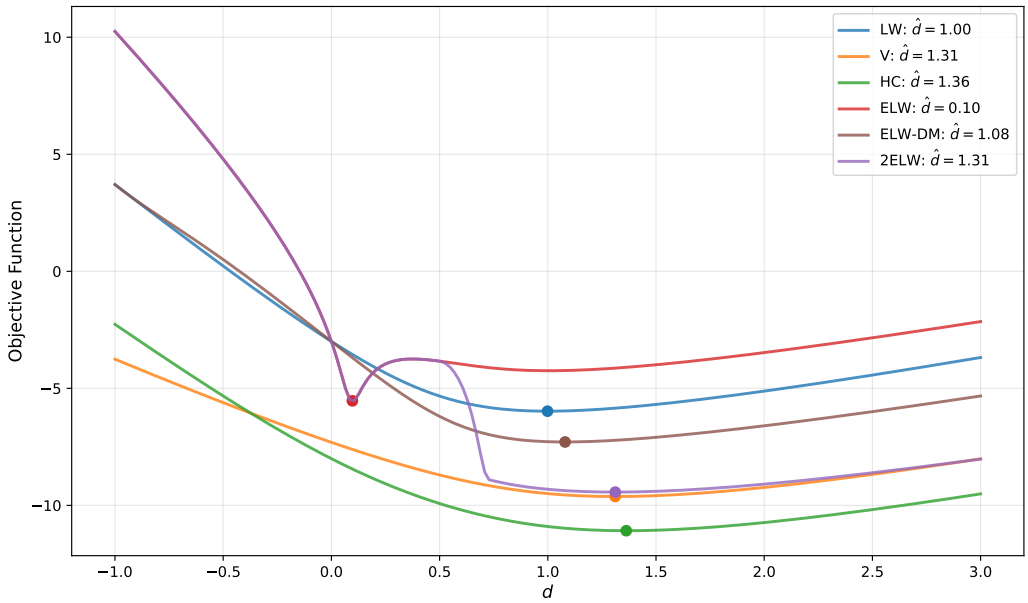

FIGURE 2. Local Whittle objective functions for U.S. industrial production (log monthly index, 1957–1997). The LW-based estimators (LW, V, HC) have convex objective functions, while ELW and 2ELW exhibit non-convex objectives with multiple local minima. ELW is an outlier with minimum around $d \approx 0.10$.

4.2. Anomalous Memory Estimates for U.S. Industrial Production

We now return to the U.S. industrial production time series, where ELW yields a spuriously low estimate ($\hat{d} = 0.10$) relative to tapered estimators as well as 2ELW ($\hat{d} \in [1.31, 1.36]$). Even the inconsistent untapered LW estimator ($\hat{d} = 1.00$) yields a more reasonable estimate than ELW. To understand this difference, we plot the objective functions for each estimator in Figure 2. The LW objective functions (LW, V, HC) are all convex with a single global minimum. The ELW and 2ELW objective functions, on the other hand, are non-convex and have multiple local minima. Due to the functional form of the piecewise adaptive weight function defined in (6), the objective functions for ELW and 2ELW coincide exactly up to $d = 0.5$. After $d = 0.5$, the weight function begins to apply positive weight to $\hat{\mu} = X_1$, causing the 2ELW objective function to diverge from that of ELW. The weight fully shifts away from $\hat{\mu} = \bar{X}$ to $\hat{\mu} = X_1$ at $d = 0.75$, where the 2ELW objective function has a kink. Here, the ELW estimate is anomalously low, $\hat{d} = 0.10$, while the first-step HC estimator provides a consistent starting value that helps 2ELW locate the global minimum at $\hat{d} = 1.31$. When we apply the ELW estimator to demeaned data, an estimator discussed in Section 3 of Shimotsu (2010)—denoted ELW-DM in the figure—we obtain an estimate $\hat{d} = 1.08$. This suggests that the low estimate produced by ELW is due to the fact that it assumes the

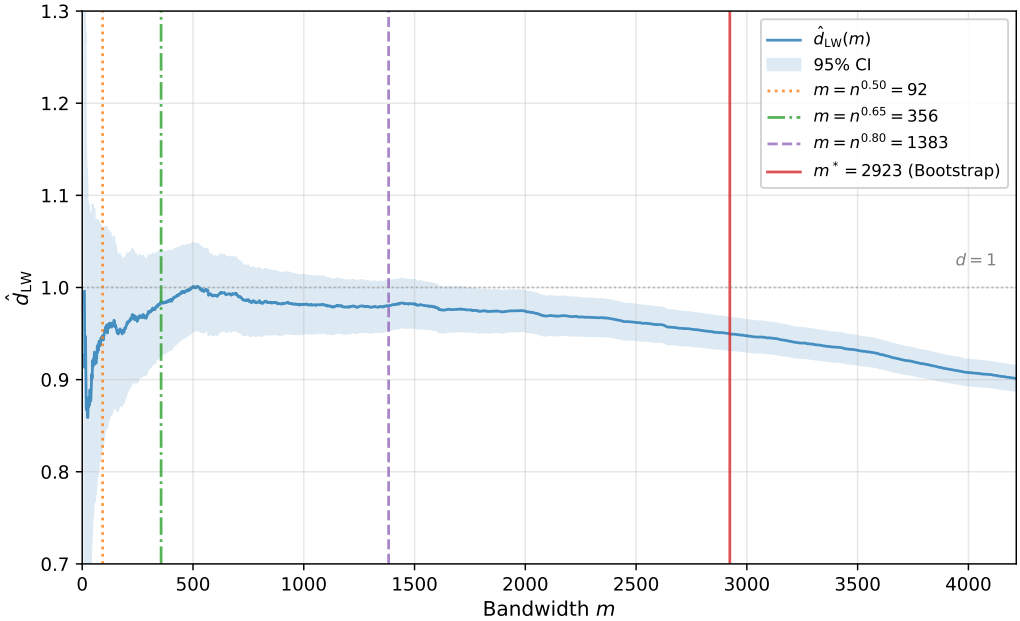

FIGURE 3. Local Whittle estimates by bandwidth using S&P 500 data. The plotted line shows estimates $\hat{d}_{\mathrm{LW}}(m)$ with 95% confidence bands. Vertical lines indicate bandwidth choices: power rules $m = n^{\alpha}$ (dashed) and the Arteche and Orbe (2017) optimal m^* (solid).

mean is known and equal to zero.

4.3. Bandwidth Selection for Daily Values of the S&P 500 Index

We now consider the effects of bandwidth choice m. In applied work it is common to choose $m = \lfloor n^{\alpha} \rfloor$ with $\alpha \in \{0.5, 0.65, 0.7, 0.8\}$. Monte Carlo and empirical work in the literature often uses the rule $m = \lfloor n^{0.65} \rfloor$ as a baseline. This rate was used by Shimotsu and Phillips (2005) and Shimotsu (2010) in their simulations and implemented as the default in both Stata's `whittle` command (Baum et al., 2020) and PyELW (Blevins, 2025). In empirical examples, Shimotsu (2010) used the slightly less conservative $m = \lfloor n^{0.7} \rfloor$.

Baillie et al. (2014) proposed cross-validation approaches focused on forecasting performance, while Arteche and Orbe (2017) developed a data-driven bandwidth selection strategy based on minimizing a bootstrap approximation of the mean squared error. Specifically, for each candidate bandwidth m, bootstrap samples of the locally standardized periodogram are generated, LW estimates are computed for each sample, and the bootstrap MSE is calculated. The optimal bandwidth m^* minimizes this estimated MSE.

We now illustrate the effects of bandwidth selection using the daily S&P 500 index series from Hurvich and Chen (2000) ($n = 8{,}432$ observations, logarithms of daily index values) as an example. Figure 3 plots the LW estimates $\hat{d}_{\mathrm{LW}}(m)$ as a function of bandwidth m, from

Table 11. S&P 500: Bandwidth Selection Comparison

Method	Rule	m	$\hat{d}_{\text{LW}}$	SE	95% CI
Power Rule	$m = n^{0.50}$	92	0.948	0.059	$(0.831, 1.064)$
	$m = n^{0.65}$	356	0.984	0.029	$(0.928, 1.041)$
	$m = n^{0.80}$	1,383	0.980	0.014	$(0.953, 1.007)$
Arteche and Orbe (2017)	Min. Bootstrap MSE	2,923	0.950	0.009	$(0.932, 0.967)$

Notes: Logarithm of S&P 500 daily stock index value, $n = 8{,}432$. Hurvich and Chen (2000) used $m = n^{0.80}$. Bootstrap MSE bandwidth selected by minimizing bootstrap MSE with $B = 200$ replications and resampling width $k_n = 255$ following Arteche and Orbe (2017).

$m = 10$ to $m = \frac{n}{2}$. The shaded region corresponds to the 95% confidence intervals at each value of m. Estimates range from approximately 0.86 at extreme values of m to just over 1.0 at $m = 500$. The confidence bands narrow as m increases, reflecting the variance reduction from using additional frequencies.

The vertical lines in Figure 3 indicate different bandwidth choices. Power rules ($m = n^{0.5}$ through $m = n^{0.8}$) result in bandwidth choices from $m = 92$ to $m = 1{,}383$,[7] which yields estimates $\hat{d}_{\text{LW}} \in [0.948, 0.984]$. We also apply the bootstrap MSE minimization procedure of Arteche and Orbe (2017), which selects $m^* = 2{,}923$ and yields $\hat{d}_{\text{LW}} = 0.950$. We note that this bandwidth is much larger than that of any of the power rules. This occurs due to the relatively weak short-run dynamics: the bandwidth selection procedure first removes the power-law component in (2) implied by the memory parameter. The resulting locally normalized spectrum is relatively flat, resulting in a large bandwidth choice m^*.

Table 11 compares the estimates for different bandwidth selection methods. The 95% confidence intervals from the power rule methods each contain $d = 1.0$, consistent with a unit-root process. However, the slightly lower estimate and tighter confidence interval associated with the bootstrap MSE bandwidth m^* excludes unity.

Finally, in Figure 4, we plot the bootstrap MSE profile as a function of bandwidth. The bias-variance tradeoff can be seen clearly: MSE is high at small m due to high variance, then decreases to a minimum at $m^* = 2{,}923$ before rising again as m approaches $\frac{n}{2}$, where the bias begins to dominate. The power rules all yield smaller bandwidths, leading to conservative choices for this relatively large sample.

4.4. When Methods Disagree: French Inflation and Structural Breaks

The wide range of French inflation estimates in Table 10, where $\hat{d} \in [0.45, 0.82]$, raises a natural question: which estimator should practitioners trust? Disagreement of this

[7] To match Hurvich and Chen (2000), we use $m = \text{round}(n^{\alpha})$ here rather than $m = \lfloor n^{\alpha} \rfloor$, which is used elsewhere in this paper.

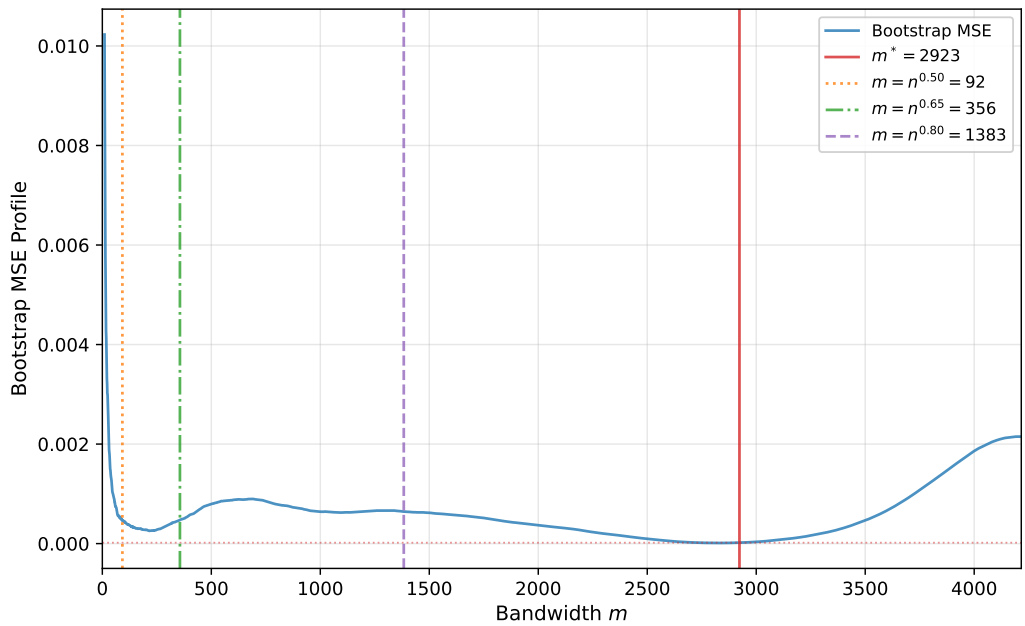

FIGURE 4. Bootstrap MSE profile for S&P 500 data. The blue line represents the estimated MSE for each bandwidth, with the minimum at $m^* = 2{,}923$ indicated by the red vertical line. Power rule bandwidths are indicated by dashed lines.

magnitude—with estimates on both sides of the nonstationary boundary $d = 0.5$—has implications for inference, forecasting, and policy. One concern is that structural breaks may inflate long-memory estimates: Diebold and Inoue (2001) and Granger and Hyung (2004) showed that regime-switching and occasional mean shifts generate autocorrelations with the same hyperbolic decay pattern as fractional integration. Indeed, Caporale, Gil-Alana, and Poza (2022) documented multiple structural breaks in G7 inflation series over 1973–2020. If structural breaks are indeed inflating full-sample estimates, then $\hat{d}$ should fall when estimated within stable regimes. We follow this logic to investigate the French inflation series.

We apply the structural break detection procedure of Bai and Perron (1998, 2003). Although minimizing BIC would select additional breaks, we focus on the two most prominent: March 1973 (before the first oil shock) and October 1984 (coinciding with the mid-1980s disinflation). This parsimonious specification preserves within-regime sample sizes for more reliable estimation, and the alignment with major events provides clear justification.

Figure 5 displays the inflation rates along with the detected break dates and regime-specific means. In the pre-oil-shocks period (1957–1973), there was moderate inflation of around 5.0% annually with high volatility (s.d. 1.7%). During the oil shocks era (1973–1984), France saw much higher inflation averaging 10.3%, but with lower volatility (s.d. 1.0%). During the disinflation period (1984–1997), inflation rates fell dramatically, to 2.5% annually,

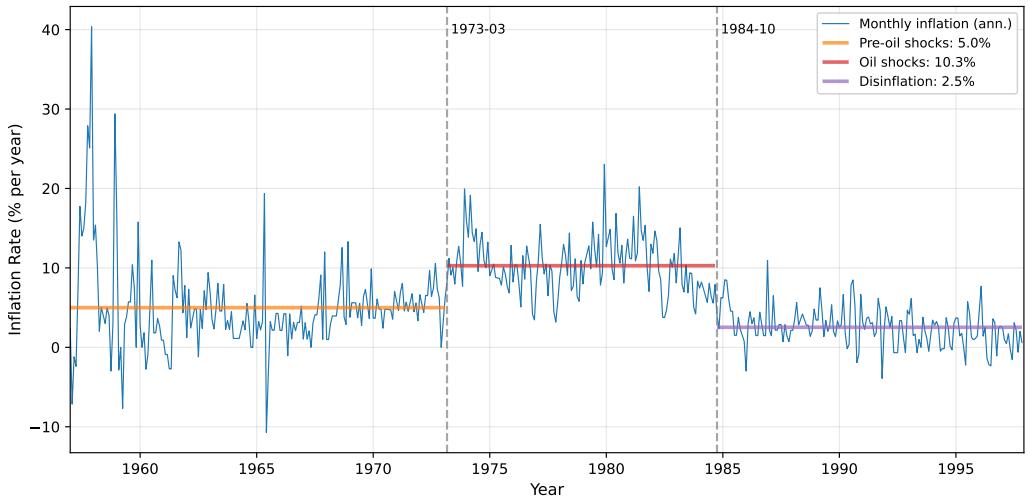

FIGURE 5. French inflation (annualized differences in log monthly CPI) with structural breaks. Horizontal lines indicate regime means: 5.0%/yr (pre-1973), 10.3%/yr (1973–1984), and 2.5%/yr (post-1984). Break dates detected using Bai and Perron (2003) methodology.

with even lower volatility (s.d. 0.7%).

In Table 12, we report the memory parameter estimates for French inflation using six local Whittle estimators. The top panel reproduces the previous results from Table 10 and adds the LWLFC estimator, which was not reported for this series previously. This break-robust estimator yields the lowest estimate among all estimators, with $\hat{d} = 0.263$, and the corresponding 95% confidence interval contains zero. This is consistent with the hypothesis that the full-sample estimates are inflated by structural breaks.

To investigate further, in Panels A and B of Table 12, we report the memory parameter estimates for each of these regimes with two bandwidth selection approaches. Panel A applies the same power rule as the full sample ($m = n^{0.60}$), yielding small subsample bandwidths of $m \in \{23, 19, 20\}$. Panel B uses the bootstrap-MSE-optimal selection procedure of Arteche and Orbe (2017), which results in larger subsample bandwidths of $m \in \{91, 32, 58\}$. We note that in both panels of this table, for 2ELW we specify a trend of order zero (i.e., demeaning only).

In Table 12, the full-sample LW and tapered LW estimates consistently exceed the subsample estimates. For example, in Panel A the Velasco-tapered LW estimate is $\hat{d} = 0.819$ for the full sample but only 0.248 for the pre-oil-shocks regime. Similarly, in Panel B the HC tapered LW estimate is $\hat{d} = 0.682$ for the full sample compared to 0.433 for the oil shocks regime. This pattern suggests that regime shifts inflate the full-sample estimates of persistence. However, the subsample estimates do not collapse entirely to zero, as a purely spurious long-memory explanation would predict. Using 2ELW as a benchmark

Table 12. Inflation in France: Subsample Analysis with Detected Structural Breaks

Period	n	m	LW	V	HC	ELW	2ELW	LWLFC
Full sample (1957–1997)	491	40	0.462	0.819	0.682	0.499	0.450	0.263
			(0.083)	(0.137)	(0.121)	(0.087)	(0.079)	(0.205)
Panel A: Power rule bandwidth ($m = n^{0.60}$)								
Pre-oil shocks	194	23	0.265	0.248	0.125	0.321	0.275	0.265
			(0.147)	(0.181)	(0.176)	(0.205)	(0.104)	(0.147)
Oil shocks	139	19	0.424	0.174	0.418	−0.001	0.453	0.424
			(0.152)	(0.199)	(0.202)	(0.050)	(0.115)	(0.152)
Disinflation	158	20	0.181	−0.096	0.012	−0.205	0.251	0.080
			(0.140)	(0.194)	(0.195)	(0.041)	(0.112)	(0.233)
Panel B: Bootstrap-MSE-optimal bandwidth m^*								
Pre-oil shocks	194	91	0.356	−0.013	0.205	0.542	0.415	0.356
			(0.059)	(0.091)	(0.082)	(0.076)	(0.052)	(0.059)
Oil shocks	139	32	0.453	0.112	0.433	0.620	0.491	0.453
			(0.114)	(0.153)	(0.144)	(0.121)	(0.088)	(0.114)
Disinflation	158	58	0.110	−0.347	−0.083	0.175	0.180	0.069
			(0.074)	(0.114)	(0.103)	(0.127)	(0.066)	(0.091)

Notes: Estimates of d with standard errors in parentheses. French inflation is monthly diff-log CPI from January 1957 to December 1997. Structural breaks detected following Bai and Perron (2003): 1973-03, 1984-10. Mean inflation: 5.0%/yr (pre-1973), 10.3%/yr (1973–1984), 2.5%/yr (post-1984). Panel A uses the same power rule as the full sample. Panel B uses bootstrap MSE-optimal bandwidth selection (Arteche and Orbe, 2017).

(given its robustness to an unknown mean), the three regimes yield estimates $\hat{d}$ in the ranges 0.275–0.415, 0.453–0.491, and 0.180–0.251, respectively, indicating genuine weak-to-moderate long memory within regimes.

The results suggest there is more ambiguity about memory during the disinflation period. Here, estimates collapse further and in some cases become negative. In Panel A, four of the six 95% confidence intervals (LW, V, HC, LWLFC) contain zero, consistent with $I(0)$ dynamics. There is a large discrepancy between ELW and 2ELW (−0.205 vs. 0.251). ELW's 95% confidence interval lies entirely below zero while 2ELW's lies entirely above zero. Panel B, which uses bootstrap-MSE-optimal bandwidths ($m = 58$), reveals similar disagreement. The 95% confidence intervals for LW, HC, ELW, and LWLFC contain zero, while V ($\hat{d} = -0.347$) is significantly negative and 2ELW ($\hat{d} = 0.180$) remains significantly positive.

We draw two main conclusions from the subsample analysis. First, there is evidence of genuine long memory within the earlier regimes. The pre-oil-shocks and oil-shocks periods both produce statistically significant estimates of persistence ($\hat{d} \approx 0.275$–0.491 using 2ELW). However, the disinflation period shows weaker and more ambiguous memory estimates that

vary across estimators. Second, regime shifts further inflate the full-sample estimates, with the range $\hat{d} \in [0.450, 0.819]$ for the non-break-robust estimators being due to a combination of the genuine within-regime persistence and spurious memory from structural breaks.

We conclude our discussion by applying the Qu (2011) test to assess whether the observed long memory is genuine or spurious. The null hypothesis is true fractional integration and the alternative is that the process is affected by structural breaks or a smoothly varying trend. For the full sample, the test statistic is $W = 0.659$, well below the 10% critical value of 1.022, so we fail to reject the null hypothesis of true long memory. We also apply the test to each subsample, producing $W = 0.751$ (pre-oil-shocks), $W = 0.656$ (oil shocks), and $W = 0.434$ (disinflation). In each of the three cases, we fail to reject the null at conventional levels.

These results can be reconciled as follows. The Qu test fails to reject true long memory for both the full sample and each subsample, suggesting that the observed persistence is not entirely spurious. Several estimators, including LWLFC, indicate positive memory within the pre-oil-shocks and oil-shocks regimes. At the same time, the full-sample LWLFC estimate ($\hat{d} = 0.263$) and the subsample estimates are consistently lower than full-sample estimates from non-robust estimators. Taken together, these findings suggest genuine but weak long memory combined with structural breaks that inflate non-robust full-sample estimates. For time series containing structural breaks, LWLFC combined with subsample analysis provides a useful diagnostic.

5. Practical Guidance for Practitioners

This section distills the preceding theoretical, Monte Carlo, and empirical sections to provide practical recommendations for selecting and implementing local Whittle estimators in applied work. The Monte Carlo evidence confirms that all estimators behave according to their expected theoretical properties under ideal conditions. However, no single estimator dominates in all scenarios we considered, and so practitioners should take into account the characteristics of their data when choosing among the methods.

5.1. Expected Parameter Range

The first consideration for choosing an estimator should be the expected parameter range. If the researcher expects that the process is stationary ($|d| < 0.5$), the standard LW and ELW estimators achieve the lowest variance. However, LW exhibits severe bias for $|d| > 1$, with estimates clustering near the unit root boundary when true persistence is stronger. For applications where low-frequency contamination from level shifts is suspected, the LWLFC estimator provides robustness without loss of efficiency, but it is restricted to stationary

processes with $d \in [0, \frac{1}{2})$.

For possibly nonstationary series with $d > 1$, practitioners should consider using the ELW and 2ELW estimators. ELW is valid for all $d \in (-\infty, \infty)$, as long as the optimization interval has width less than $\frac{9}{2}$. The downside is that ELW is quite sensitive to the specification of the unknown mean and trend. 2ELW includes adaptive mean estimation and optionally robustness to deterministic trends, though with a more restricted range of valid parameter values: $d \in (-\frac{1}{2}, 2)$. Tapered LW estimators offer an intermediate solution. They extend the range of valid parameter values compared to LW, but at the cost of higher variance.

The choice is more difficult for antipersistent processes, with $d < 0$. ELW, and to a large extent Velasco-tapered LW, maintain consistency even for strongly antipersistent processes in ideal settings. However, the antipersistent region is precisely where ELW is extremely sensitive to unknown deterministic components, and the 2ELW estimator fails entirely for $d < -0.6$ in our simulations. Among tapered methods, HC has large MSE and positive bias for $d < -1.2$. Velasco maintains stable MSE values across memory parameters, although with positive bias.

When the researcher suspects a process is antipersistent, we suggest comparing the ELW and Velasco estimates as a diagnostic. If they agree, it suggests genuine antipersistence. However, if ELW produces substantially higher estimates than Velasco, this may indicate contamination from an unknown deterministic components. This would be most informative for strongly antipersistent processes ($d < -1$), where ELW's sensitivity to deterministic components is greatest. For researchers who are uncertain about the nature of deterministic components, the Velasco estimator serves as a good baseline. Its performance is stable across antipersistent memory parameter values, offering robustness to misspecification in exchange for reduced efficiency.

5.2. Robustness Considerations

The presence of AR(1) short-run dynamics affects all local Whittle estimators, inducing a positive bias that increases with the degree of autocorrelation ρ. Under moderate autocorrelation ($\rho = 0.5$), ELW remains the best estimator. 2ELW performs similarly for $d \geq -0.6$ but fails for more strongly antipersistent processes. With even stronger persistence ($\rho = 0.8$), ELW maintains the smallest uniform bias. The Velasco tapered LW estimator has reasonably uniform behavior across d values, but is more vulnerable to AR(1) contamination relative to the exact methods. Practitioners who suspect non-trivial short-run dynamics should expect to see upward bias in all estimates and prefer ELW, 2ELW, or LWLFC for relative robustness (though 2ELW and LWLFC should be avoided when strong antipersistence is suspected). More conservative bandwidth choices (i.e., smaller values of m) can help mitigate contamination from short-run dynamics affecting higher frequencies.

We saw that the ELW estimator is highly sensitive to nonzero means and linear trends, particularly for antipersistent processes ($d < 0$). On the other hand, the Velasco and HC tapered estimators demonstrate robustness to both unknown means and trends without any preprocessing, though HC's baseline performance is poor for $d < -1.2$. When analyzing data with potential nonzero means or trends, practitioners should consider using the 2ELW estimator, noting that its range is restricted to $d \in (-0.5, 1.75)$ when detrending. For stationary processes with $d \in [0, \frac{1}{2})$, LWLFC is robust to level shifts. Finally, we reiterate that demeaning or detrending when it is not necessary can also be detrimental, especially for ELW when $|d|$ is large.

5.3. Bandwidth Selection

The bandwidth parameter m is an important tuning parameter that must be selected by the researcher. Applied work typically adopts power rules $m = \lfloor n^\alpha \rfloor$ with $\alpha \in \{0.5, 0.65, 0.7, 0.8\}$, with $\alpha = 0.65$ emerging as a practical standard (Shimotsu and Phillips, 2005; Shimotsu, 2010; Baum et al., 2020). Data-driven methods such as the bootstrap MSE minimization procedure of Arteche and Orbe (2017) can be used to estimate the optimal bandwidth for LW, though this procedure often suggests much larger bandwidths than power rules. We recommend carrying out a sensitivity analysis at multiple bandwidths, particularly when qualitative conclusions (e.g., $d < 1$ vs. $d \geq 1$) change with reasonable bandwidth choices.

5.4. Structural Breaks and Spurious Long Memory

When there are large discrepancies among estimators, researchers should consider whether structural breaks offer an explanation. Regime changes can produce spuriously high memory estimates (Diebold and Inoue, 2001; Granger and Hyung, 2004). Subsample analysis can help determine whether the persistence is due to genuine long memory within regimes or spurious inflation from structural change. First, apply formal break detection methods such as those of Bai and Perron (1998, 2003) to assess whether any structural breaks are present. Then, estimate d separately for each regime. If the full-sample estimates substantially exceed all subsample estimates, this may be due to spurious long memory due to level shifts. However, if subsample estimates remain positive and statistically significant, there may still be genuine persistence within regimes.

In such cases we recommend applying formal tests and break-robust estimators. Shimotsu (2006) exploits the property that genuine long memory produces stable d estimates across sample splits. Ohanissian, Russell, and Tsay (2008) tests whether d is invariant to temporal aggregation—a property satisfied by true long memory but violated by spurious memory from breaks. The test of Qu (2011) is based on whether estimates of d remain consistent

across different frequency bands; Sibbertsen, Leschinski, and Busch (2018) developed a multivariate version of the test. For robust estimation, researchers can turn to the trimmed log periodogram estimator of McCloskey and Perron (2013) or the LWLFC estimator of Hou and Perron (2014). Among local Whittle estimators, LWLFC explicitly models low-frequency contamination from level shifts, however, with a more restricted parameter range of $d \in [0, \frac{1}{2})$. For nonstationary series, or series with linear time trends, one of the tapered estimators or 2ELW with detrending should be considered.

5.5. Summary of Recommendations for Empirical Work

First, we recommend that applied researchers visually inspect their time series data for obvious trends, level shifts, or regime changes that may influence the choice of estimator. Second, in light of the strengths and weaknesses of various local Whittle estimators, we recommend applying multiple methods to estimate d, both as a robustness check and a diagnostic strategy. The specific disagreement patterns we highlighted can provide insights about characteristics of the data. However, such patterns do not provide clear deterministic rules, which would depend on unknown properties of the data. Third, we recommend conducting a bandwidth sensitivity analysis across $\alpha \in \{0.5, 0.6, 0.65, 0.7, 0.8\}$. Researchers should also consider using the bootstrap MSE-optimal selection procedure of Arteche and Orbe (2017). Fourth, when estimators disagree or visual inspection suggests regime changes, apply formal break detection methods and break-robust estimation of d as described above.

6. Conclusion

This paper carries out a systematic evaluation of six local Whittle methods for estimating the fractional integration parameter d. We make three contributions. First, we replicate Monte Carlo and empirical results from several seminal papers in the literature. We then extend these replications with new cross-method comparisons under short-run dynamics, unknown means, and deterministic time trends. Second, we show that cross-estimator disagreement can provide diagnostic information about contamination. Third, we illustrate these patterns through detailed empirical case studies.

As the preceding analysis shows, no single local Whittle estimator dominates in all settings. The ELW estimator attains the lowest variance under ideal conditions, but is quite sensitive to deterministic components. Tapered methods offer robustness to such components at the cost of increased variance and a more limited range of valid parameter values. The 2ELW estimator includes adaptive mean correction and detrending, also at the expense of the parameter range. Overall, the best choice depends on the researcher's expectations about both the range of likely d values and the characteristics of the data.

When estimators disagree, the pattern of disagreement itself can reveal contamination from deterministic components or structural breaks—features that would be missed by applying a single method in isolation.

References

Abadir, K. M., W. Distaso, and L. Giraitis (2007). Nonstationarity-extended local Whittle estimation. *Journal of Econometrics 141*, 1353–1384.

Andrews, D. W. K. and P. Guggenberger (2003). A bias-reduced log-periodogram regression estimator for the long-memory parameter. *Econometrica 71*, 675–712.

Andrews, D. W. K. and Y. Sun (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica 72*, 569–614.

Arteche, J. (2020). Exact local Whittle estimation in long memory time series with multiple poles. *Econometric Theory 36*, 1064–1098.

Arteche, J. and J. Orbe (2017). A strategy for optimal bandwidth selection in local Whittle estimation. *Econometrics and Statistics 4*, 3–17.

Baek, C., M.-C. Düker, and V. Pipiras (2023). Local Whittle estimation of high-dimensional long-run variance and precision matrices. *Annals of Statistics 51*(6), 2386–2414.

Bai, J. and P. Perron (1998). Estimating and testing linear models with multiple structural changes. *Econometrica 66*, 47–78.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1–22.

Baillie, R. T., G. Kapetanios, and F. Papailias (2014). Bandwidth selection by cross-validation for forecasting long memory financial time series. *Journal of Empirical Finance 29*, 129–143.

Baum, C. F., S. Hurn, and K. Lindsay (2020). Local Whittle estimation of the long-memory parameter. *Stata Journal 20*(3), 565–583.

Beran, J. (1994). *Statistics for Long-memory Processes.* New York: Chapman & Hall.

Blevins, J. R. (2025). PyELW: Exact local Whittle estimation for long memory time series in Python. Working paper, The Ohio State University.

Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory.* Holden-Day.

Caporale, G. M., L. A. Gil-Alana, and C. Poza (2022). Inflation in the G7 countries: persistence and structural breaks. *Journal of Economics and Finance 46*, 493–506.

Chen, K. and M. Wang (2017). Local Whittle likelihood estimators and tests for spatial lattice data. *Journal of Statistical Planning and Inference 191*, 25–42.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics 105*(1), 131–159.

Frederiksen, P., F. S. Nielsen, and M. O. Nielsen (2012). Local polynomial whittle estimation of perturbed fractional processes. *Journal of Econometrics 167*, 426–447.

García-Enríquez, J. and J. Hualde (2019). Local Whittle estimation of long memory: Standard versus bias-reducing techniques. *Econometrics and Statistics 12*, 66–77.

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

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

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

Henry, M. (2001). Robust automatic bandwidth for long memory. *Journal of Time Series Analysis 22*, 293–316.

Henry, M. and P. M. Robinson (1996). Bandwidth choice in Gaussian semiparametric estimation of long range dependence. In P. M. Robinson and M. Rosenblatt (Eds.), *Athens Conference on Applied Probability and Time Series Analysis, Volume II: Time Series Analysis In Memory of E.J. Hannan*, pp. 220–232. Springer.

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

Hou, J. and P. Perron (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics 182*(2), 309–328.

Hurvich, C. M. and W. W. Chen (2000). An efficient taper for potentially overdifferenced long-memory time series. *Journal of Time Series Analysis 21*(2), 155–180.

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

Künsch, H. R. (1987). Statistical aspects of self-similar processes. In Y. V. Prokhorov and V. V. Sazanov (Eds.), *Probability Theory and Applications: Proceedings of the 1st World Congress of the Bernoulli Society*, Volume 1, pp. 67–74. De Gruyter.

Li, D., P. M. Robinson, and H. L. Shang (2021). Local Whittle estimation of long-range dependence for functional time series. *Journal of Time Series Analysis 42*(5-6), 685–695.

McCloskey, A. and P. Perron (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory 29*, 1196–1237.

Nielsen, M. O. (2007). Local Whittle analysis of stationary fractional cointegration and the implied–realized volatility relation. *Journal of Business and Economic Statistics 25*(4), 427–446.

Nielsen, M. O. and P. H. Frederiksen (2005). Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration. *Econometric Reviews 24*, 405–443.

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

Phillips, P. C. B. and K. Shimotsu (2004). Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics 32*(2), 656–692.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics 29*(3), 423–438.

Robinson, P. M. (1995a). Gaussian semiparametric estimation of long range dependence. *Annals of Statistics 23*(5), 1630–1661.

Robinson, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *Annals of Statistics 23*(3), 1048–1072.

Robinson, P. M. and M. Henry (2003). Higher-order kernel semiparametric m-estimation of long memory. *Journal of Econometrics 114*, 1–27.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Queen's Economics Department Working Paper 1101, Queen's University.

Shimotsu, K. (2010). Exact local Whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory 26*(2), 501–540.

Shimotsu, K. (2012). Exact local Whittle estimation of fractionally cointegrated systems. *Journal of Econometrics 169*(2), 266–278.

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

Sibbertsen, P., C. Leschinski, and M. Busch (2018). A multivariate test against spurious long memory. *Journal of Econometrics 203*(1), 33–49.

Velasco, C. (1999). Gaussian semiparametric estimation for non-stationary time series. *Journal of Time Series Analysis 20*(1), 87–126.

Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist & Wiksells.

Wingert, S., C. Leschinski, and P. Sibbertsen (2020). Seasonality robust local Whittle estimation. *Applied Economics Letters 27*(18), 1489–1494.