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

JASON R. BLEVINS

The Ohio State University

September 23, 2025

Abstract. Fractionally integrated time series exhibiting long memory are commonly in economics, finance, and related fields. Semiparametric methods for estimating the memory parameter d have proven to be effective and robust, but practitioners face difficulties arising from the availability multiple estimators with different valid parameter ranges and the choice of bandwidth parameter m. This paper provides a comprehensive evaluation of local Whittle methods from Robinson's (1995) foundational estimator through the exact local Whittle approaches of Shimotsu and Phillips (2005) and Shimotsu (2010), where theoretical advances have expanded the feasible range of memory parameters and improved efficiency. Using a new implementation in Python, PyELW, we replicate key empirical and Monte Carlo results from the literature, providing external validation for both the original findings and the software implementation. We extend these empirical applications to demonstrate how method choice can affect substantive conclusions about persistence. Based on comprehensive simulation comparisons and empirical evidence, we provide practical guidance for applied researchers on how and when to use each method.

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

JEL Classification: C22, C63, C87.

1. Introduction

Estimation of the memory parameter d of fractionally integrated time series is a fundamental problem in econometrics, with applications 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 persistence between the extremes of stationary and unit root behavior, a literature on semiparametric estimation of the memory parameter based on frequency domain methods has emerged.

Early work by Geweke and Porter-Hudak (1983) and Robinson (1995b) was based on log-periodogram regressions. The local Whittle (LW) approach developed by Künsch (1987) and Robinson (1995a) achieved the semiparametric efficiency bound for stationary models with $d \in (-\frac{1}{2}, \frac{1}{2})$ by instead maximizing a frequency-domain quasi-likelihood function based on the Whittle likelihood.

To extend the valid range of parameter values to include nonstationary models, Velasco (1999), Hurvich and Chen (2000), and Phillips and Shimotsu (2004) developed LW modifications based on tapering. Shimotsu and Phillips (2005) later introduced the exact local Whittle (ELW) estimator, extending the method to accommodate the entire parameter space, including both stationary and nonstationary models, by using exact fractional differencing, provided that the optimization interval has width less than $\frac{9}{2}$. The two-step ELW procedure of Shimotsu (2010) (2ELW) generalized the approach to handle processes with unknown mean and time trend, while Shimotsu (2012) extended the framework to fractionally cointegrated systems.

Recent advances have further expanded the scope of local Whittle methods. Arteche (2020) extended ELW estimation to multiple poles to jointly estimate memory parameters at standard, seasonal, and other cyclical frequencies. Extensions to multivariate systems through fractional cointegration (Shimotsu, 2012), spatial-temporal data (Chen and Wang, 2017), functional time series (Li, Robinson, and Shang, 2021), and high-dimensional settings (Baek, Düker, and Pipiras, 2023) have broadened applicability, while computational improvements including debiased likelihood methods (Sykulski, Olhede, Guillaumin, Lilly, and Early, 2019) improve finite-sample performance.

The rich theoretical development of local Whittle methods has yielded several estimators, each with its own advantages, but this creates challenges for applied researchers in understanding the trade-offs, in choosing among them, and implementing them properly. Existing software implementations often cover only subsets of the available methods or lack validation against original results. This paper addresses these gaps through a comprehensive comparison of local Whittle methods that serves three primary objectives. First, we review the theoretical development of local Whittle methods, providing a survey of major results. Second, we systematically replicate the key empirical and Monte Carlo results from several foundational papers in the literature using PyELW, a new cross-platform library for LW, ELW, and 2ELW estimation in Python (Blevins, 2025). This provides external validation for both the original results and the Python implementation. Third, we conduct a comprehensive series of Monte Carlo simulations and empirical applications with non-ideal data—including processes with short-run dynamics, unknown means, and time trends—to compare methods and provide guidance for applied researchers on their application.

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 main results from the literature for each method. Section 3 presents a new, cross-method Monte Carlo comparison on a variety of simulated processes with controlled contaminations. Section 4 presents an extended comparative empirical analysis with several macroeconomic and financial time series. Section 5 provides practical advice for practitioners on selecting and applying local Whittle methods to real data. Section 6 concludes. All datasets and replication code is available at `https://github.com/jrblevin/lws`.

2. Semiparametric Estimation of Fractional Integration

This section surveys five major local Whittle estimators for fractional integration, examining their theoretical foundations, computational properties, and empirical performance. The progression of methods reflects a series of improvements, each overcoming limitations of previous estimators: Robinson's (1995a) original LW estimator handled only stationary processes, leading Velasco (1999) and Hurvich and Chen (2000) to develop tapered variants that extended the parameter range at the cost of efficiency. Shimotsu and Phillips (2005) restored efficiency through exact fractional differencing, while Shimotsu (2010) added robustness to unknown means and trends. For each method, we first introduce the estimator and its key theoretical results and then present focused Monte Carlo replications that verify the theoretical properties, reproduce original published results from the literature, and validate the new Python implementation. These simulations will set the stage for our subsequent comparative analysis in Section 3.

Consider the fractionally integrated process X_t defined as

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

where L is the lag operator, $d_0 \in \mathbb{R}$ denotes the memory parameter, $1\{\cdot\}$ denotes the indicator function imposing the initial condition $X_t = 0$ for $t \leq 0$, and u_t is a stationary, mean zero process with spectral density $f_u(\lambda)$. The memory parameter d_0 determines the fundamental properties of X_t: standard $I(0)$ behavior when $d_0 = 0$, stationary long memory for $d_0 \in (0, 0.5)$, the stationarity boundary at $d_0 = 0.5$, nonstationary but mean-reverting dynamics for $d_0 \in (0.5, 1)$, and the unit root case when $d_0 = 1$. Negative values $d_0 < 0$ generate antipersistent processes, while and $d_0 > 1$ yields explosive nonstationarity.

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

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

where $a \sim b$ means that $a/b \to 1$ and $G = f_u(0) > 0$.[1] This asymptotic 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 where (2) holds most accurately, avoiding contamination from short-run dynamics at higher frequencies. The discrete Fourier transform of a time series a_t at frequencies $\lambda_j = 2\pi j/n$ is

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

with periodogram $I_a(\lambda_j) = |w_a(\lambda_j)|^2$. The semiparametric approach substitutes the power-law approximation $f_X(\lambda_j) \approx G\lambda_j^{-2d}$ for the unknown spectral density in a localized version of the Whittle criterion. Treating G as a nuisance parameter and maximizing over G for each fixed d yields $\hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m} \lambda_j^{2d} I_X(\lambda_j)$. Substituting this concentrated estimator back gives the 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. The objective function $R(d)$ is convex in d (Baum, Hurn, and Lindsay, 2020, Appendix A), ensuring that any local minimum is a global minimum and making optimization reliable. This allows the use of efficient optimization algorithms such as golden section search to reliably locate the global minimum. Robinson (1995a) established regularity conditions under which for $d_0 \in (-0.5, 0.5)$, the local Whittle estimator is consistent, converges at the optimal $\sqrt{m}$ rate, and achieves the semiparametric efficiency bound:

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

Furthermore, the periodogram I_X in the objective function is independent of d and can be pre-computed, meaning that the objective function requires only $O(m)$ operations per evaluation.

The bandwidth parameter m controls the bias-variance trade-off: larger m reduces

[1] For nonstationary processes with $|d_0| \geq 0.5$, $f_X(\lambda)$ represents a pseudo-spectral density since the classical spectral density is undefined when the process has infinite variance.

variance but increases bias as higher frequencies contaminated by short-run dynamics enter the estimation. The result above requires $m \to \infty$ and $m/n \to 0$ as $n \to \infty$ to balance these concerns. Henry and Robinson (1996) show that the theoretical optimal rate is $m \propto n^{4/5}$ for ARFIMA processes, which minimizes the mean squared error of the estimator. However, this optimal rate can be sensitive to model misspecification and short-run dynamics, often leading practitioners to use more conservative choices $m = \lfloor n^{\alpha} \rfloor$ with $\alpha \in (0.6, 0.8)$. See Henry (2001), Baillie, Kapetanios, and Papailias (2014), and particularly the discussion in Arteche and Orbe (2017) for further developments on optimal bandwidth selection.

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

	SP (2005)			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$(0, d, 0)$, $m = n^{0.65} = 56$ frequencies, 10,000 replications.

In Table 1 we present results replicating the right panel of Table 1 from Shimotsu and Phillips (2005), illustrating the finite-sample performance of the local Whittle estimator. 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 across a range of d values. The number of frequencies used was $m = n^{0.65} = 56$. The "SP (2005)" columns report the original results while the "Replication" columns contain our results obtained with PyELW. Our results achieve near-exact correspondence with the original results across all specifications.

The results correspond closely with 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 $1/(2\sqrt{m}) = 0.067$, and low mean squared errors. The performance is particularly impressive for $d \in [-0.3, 0.3]$. However, the estimator's limitations become apparent for $d = -1.3$ and $d = 1.3$, where the theoretical validity breaks down and the bias begins to dominate the mean squared error.

2.2. Extending the Range via Tapering

Although the LW estimator is efficient for $d_0 \in (-0.5, 0.5)$, this range excludes important processes beyond the nonstationary boundary, preventing analysis of macroeconomic and financial time series with unit root or near-unit root behavior. Velasco (1999) demonstrated that tapering—multiplying the series by weights that decay smoothly to zero at the sample boundaries—can extend consistency to $d_0 \in (-0.5, 1)$ and asymptotic normality to $d_0 < 0.75$.

Formally, tapering involves scaling the original series X_t by a weight sequence h_t that is symmetric around $\lfloor n/2 \rfloor$, satisfies $\max h_t = 1$, and decays smoothly to zero at the sample boundaries. The tapered periodogram $I_h(\lambda_j) = |w_h(\lambda_j)|^2$, where

$$w_h(\lambda_j) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} h_t X_t e^{-i\lambda_j t}$$

replaces the standard periodogram $I_X(\lambda_j)$ in the objective function. For Velasco tapers, the periodogram uses frequencies $\lambda_j = 2\pi j/n$ with $j = p, 2p, 3p, \ldots, m$, where p is the taper order, effectively subsampling every p-th frequency. As with the standard LW estimator, evaluating these tapered objective functions remains computationally inexpensive and the functions are amenable to simple scalar minimization algorithms.

Using tapers of higher orders p extends the feasible range of d_0 values and provides robustness to trends of order $p-1$. The triangular Bartlett window taper ($p = 2$) is valid for $d_0 \in (-0.5, 1.5)$ and robust to linear trends, while the cosine bell and Zhurbenko-Kolmogorov tapers ($p = 3$) are valid for $d_0 \in (-0.5, 2.5)$ and robust to linear and quadratic time trends. However, this robustness comes at the cost of increased variance, which is at least 2.1 times the efficient LW variance. In general, the variance of the limiting distribution is $p\Phi_p/4$ 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 efficient LW variance while for $p = 3$ with $\Phi_3 = 1.00354$ the variance is inflated by a factor of 3.01.

Table 2 presents our replication of the right panel of Table 2 from Shimotsu and Phillips (2005), demonstrating the finite-sample performance of the Velasco tapered local Whittle estimator using the Bartlett taper. As before, our Python implementation yields results are very close to the original results across all specifications. Within the range $d \in [-1.7, 1.7]$, the estimator shows good performance with low bias and standard deviations of around 0.12, close to the theoretical asymptotic approximation of $\sqrt{2.1/(4m)} = 0.0968$ in this case. The variance inflation factor of 2.1 is associated with $p = 2$ for the Bartlett taper. At more extreme values of d, outside the method's validity range, we see large biases and mean squared errors.

Since PyELW implements all three tapers discussed in Velasco (1999), in Table 3 we

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

	SP (2005)			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$(0, d, 0)$, $m = n^{0.65} = 56$ 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	2.7450	0.1254	0.1644	0.0427	0.2207	0.1796	0.0810
−2.3	0.2173	0.1734	0.0773	0.0698	0.1644	0.0319	0.0757	0.1648	0.0329
−1.7	0.0279	0.1211	0.0154	0.0515	0.1643	0.0296	0.0542	0.1649	0.0301
−1.3	0.0090	0.1203	0.0146	0.0419	0.1642	0.0287	0.0431	0.1648	0.0290
−0.7	−0.0048	0.1200	0.0144	0.0315	0.1640	0.0279	0.0313	0.1645	0.0280
−0.3	−0.0097	0.1201	0.0145	0.0274	0.1636	0.0275	0.0268	0.1640	0.0276
0.0	−0.0111	0.1201	0.0145	0.0258	0.1632	0.0273	0.0255	0.1635	0.0274
0.3	−0.0103	0.1201	0.0145	0.0257	0.1629	0.0272	0.0259	0.1631	0.0273
0.7	−0.0051	0.1203	0.0145	0.0281	0.1625	0.0272	0.0296	0.1626	0.0273
1.3	0.0170	0.1222	0.0152	0.0410	0.1628	0.0282	0.0432	0.1622	0.0282
1.7	0.0485	0.1258	0.0182	0.0722	0.1654	0.0326	0.0597	0.1620	0.0298
2.3	−0.1774	0.1410	0.0514	0.2316	0.2010	0.0940	0.1020	0.1636	0.0372
3.5	−1.4529	0.1378	2.1298	−0.3994	0.1041	0.1704	−0.4160	0.1835	0.2067

Notes: PyELW results for Velasco tapered LW estimators with different tapers, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = n^{0.65} = 56$ frequencies, 10,000 replications. Bartlett taper ($p = 2$), Cosine bell taper ($p = 3$), Kolmogorov taper ($p = 3$).

compare their finite-sample performance across the full range of d values. These results clearly demonstrate the fundamental trade-off between bias reduction and variance inflation inherent in higher-order tapering. Within the range $d \in [-0.7, 0.7]$, all three tapers perform reasonably well, but the Bartlett taper ($p = 2$) achieves the lowest variance with standard deviations around 0.12, close to the theoretical value of $\sqrt{2.1/(4m)} = 0.097$.

The higher-order tapers demonstrate their advantages in the nonstationary range. For $d \in [1.3, 2.3]$, the cosine bell and Kolmogorov tapers ($p = 3$) show substantially lower bias than the Bartlett taper, though at the cost of increased variance with standard deviations around 0.16, consistent with the theoretical value of $\sqrt{3.01/(4m)} = 0.116$. At extreme values like $d = -3.5$, the third-order tapers maintain reasonable performance while the second-order Bartlett taper exhibits large bias. However, all three tapers eventually break down outside their theoretical validity ranges, as evidenced by the large biases at $d = 3.5$.

2.3. Efficiency Gains via Complex-Valued Tapering

Hurvich and Chen (2000) developed an alternative tapered local Whittle estimator that achieved better efficiency than the Velasco approach while extending the valid range to $d_0 \in (-0.5, 1.5)$. The Hurvich-Chen (HC) method applies a complex-valued taper to first-differenced data $\Delta X_t = X_t - X_{t-1}$, then estimates the memory parameter of the differenced series and adds back one degree of integration. The method uses a complex exponential taper $h_t = \exp(i\pi t/n)$ applied to the differenced data, and frequencies are shifted to $\lambda_{\tilde{j}} = 2\pi(j + 0.5)/n$ to avoid the singularity at zero. Their 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_{HC} has the same form as $R(d)$ but uses the tapered periodogram of ΔX_t.

The HC estimator achieves asymptotic variance of only 1.5/4 compared to 2.1/4 for second-order Velasco tapers, representing a substantial efficiency gain. This approach achieves robustness to linear trends by first differencing the data before applying the taper. The complex taper then mitigates bias from potential overdifferencing while maintaining good finite-sample performance across the extended parameter range.

In Table 4 we replicate the HC complex tapered LW estimator results from Shimotsu and Phillips (2005), demonstrating very close agreement with the original results and confirming the HC estimator's strong finite-sample performance across the extended parameter range $d \in (-3.5, 3.5)$. Bias remains minimal for $d \in (-0.7, 0.7)$, with standard deviations closely matching the theoretical predictions. The complex taper successfully mitigates overdifferencing bias while maintaining efficiency gains relative to alternative tapered approaches. We defer a full cross-method comparison until Section 3.

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

	SP (2005)			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: HC tapered LW estimator, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = n^{0.65} = 56$ frequencies, 10,000 replications.

2.4. Exact Local Whittle Estimation

Shimotsu and Phillips (2005) fundamentally reshaped the local Whittle approach by working directly with fractionally differenced data rather than relying on approximations that fail outside of $d \in (-0.5, 0.5)$. The exact local Whittle (ELW) estimator computes the periodogram of $\Delta^d X_t = (1-L)^d X_t$ for each candidate value of d, eliminating the approximation error inherent in the original method.

The fractional differencing operator, defined through the binomial expansion

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

with coefficients computed recursively via

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

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

The ELW objective function has the same functional form as that of $R(d)$ in (3) for LW, but uses the periodogram of the fractionally differenced series:

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

where $I_{\Delta^d X}(\lambda_j)$ denotes the periodogram of $\Delta^d X_t$. The ELW objective function is no longer convex due to the complex dependence on d through the fractional differencing operation, so more care must be taken during optimization to avoid local minima. The advantage of the ELW estimator is that we can consistently estimate d_0 whether it lies in the stationary region, the nonstationary region, or exactly at the boundary $d_0 = 0.5$, provided that the optimization interval is smaller than 9/2 in width and contains the true value. Furthermore, the ELW estimator retains the efficient limiting distribution of the LW estimator despite this expanded range:

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

The computational cost of ELW is higher than for the LW estimator, which only required one precomputed periodogram calculation and no fractional differencing operations. For the ELW estimator, for each candidate value of d during optimization we must recompute the fractional difference $\Delta^d X_t$ and the periodogram $I_{\Delta^d X}$. While computing $\Delta^d X_t$ naïvely requires $O(n^2)$ operations, the PyELW implementation employs the fast fractional differencing algorithm of Jensen and Nielsen (2014) which reduces this to $O(n \log n)$ using the fast Fourier transform.

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 across 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 $d \in [-3.5, 3.5]$. As with other estimators, the results of our Python implementation are strikingly similar to the original results across all d values.

The results demonstrate the key advantage of the ELW estimator over the standard LW estimator: excellent performance across the entire range of d values tested. Throughout $d \in [-3.5, 3.5]$, the estimator exhibits minimal bias, standard deviations near the theoretical asymptotic value of $1/\sqrt{4m} = 0.067$, and uniformly low mean squared errors. This stability contrasts with the LW estimator's breakdown outside the stationary region $|d| \geq 0.5$.

2.5. Robustness to Unknown Mean and Trend

The previous estimators assumed that the stochastic process X_t has zero mean and does not contain a trend. However, real economic time series typically have unknown means

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

	SP (2005)			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: ELW estimator, $n = 500$ observations from ARFIMA$(0, d, 0)$, $m = n^{0.65} = 56$ frequencies, 10,000 replications.

or trends that can bias fractional integration estimates. Shimotsu (2010) developed the two-step exact local Whittle (2ELW) estimator to address these issues while preserving the optimal $\mathrm{N}(0, 1/4)$ limiting distribution.

Consider the model with 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, because the best way to handle the unknown mean depends on the unknown value of d_0 itself. For stationary processes ($d_0 < 0.5$), the sample average $\bar{X}$ provides the best mean estimate, but for highly persistent processes ($d_0 > 0.75$) the first observation X_1 is better, as the sample mean becomes biased. To address this, Shimotsu (2010) proposed an adaptive mean estimation step inside the ELW objective 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:

$$w(d) = \begin{cases} 1 & \text{if } d \leq 0.5, \\ \frac{1}{2}[1 + \cos(4\pi d)] & \text{if } 0.5 < d < 0.75, \\ 0 & \text{if } d \geq 0.75. \end{cases} \tag{6}$$

The intuition behind this weight function is as follows: for stationary processes ($d \leq 0.5$), the sample mean $\bar{X}$ is consistent and efficient for estimating μ_0, so $w(d) = 1$ places full weight on $\bar{X}$. For highly persistent processes ($d \geq 0.75$), the sample mean exhibits bias due to the strong dependence structure, making the initial value X_1 a more reliable estimator, so $w(d) = 0$. The smooth cosine transition for $0.5 < d < 0.75$ gradually shifts weight from the sample mean to the initial observation as persistence increases.

Estimation of the mean in this way renders direct asymptotic theory difficult, so Shimotsu (2010) instead proposed a two-step approach based on a $\sqrt{m}$-consistent first-step estimator—one of the tapered local Whittle estimators discussed above—to obtain an initial estimate $\hat{d}_T$. This estimate need not be efficient, only consistent. 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 $R''_F(\hat{d}_T)$ takes very small values, resulting in extremely large updates. To address this instability, he recommends using $\max\{R''_F(\hat{d}_T), 2\}$ instead of the raw Hessian in the Newton-Raphson step, with the lower bound preventing the occurrence of extremely large values of $\hat{d}_{2\text{ELW}}$. Robust implementations may alternatively employ quasi-Newton methods such as BFGS or bounded optimization approaches to ensure numerical stability.

For data thought to contain a polynomial trend of order k, the procedure first removes the trend via OLS regression on $(1, t, t^2, \ldots, t^k)$, then applies the two-step estimator to the residuals. The estimator is shown to be consistent following this detrending, but the valid parameter range is restricted to $d_0 \in (-0.5, 1.75)$, compared to $d_0 \in (-0.5, 2)$ for the unknown mean case. The procedure achieves the optimal $\mathrm{N}(0, 1/4)$ limiting distribution across these expanded ranges while maintaining robustness to both unknown means and polynomial trends.

Table 6 presents our replication of the 2ELW estimator results from Shimotsu (2010), demonstrating close agreement with the original simulation across the parameter range. The 2ELW estimator maintains excellent finite-sample properties with minimal bias and variances near the asymptotic value of $1/(4m) = 0.0044$ for $m = 57$. Performance remains stable across stationary ($d = 0.0, 0.4$) and nonstationary ($d = 0.8, 1.2$) regions. The two-step procedure successfully combines consistency of the first-stage tapered estimator with efficiency of the second-stage exact approach.

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

	Shimotsu (2010)		Replication	
d	Bias	Var	Bias	Var
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: 2ELW estimator, $n = 512$ observations from ARFIMA$(0, d, 0)$, $m = n^{0.65} = 57$ frequencies, 10,000 replications.

In this experiment we simply verify the basic theoretical properties, only considering specifications from Table 2 of Shimotsu (2010) with $\rho = 0$. In Section 3, we will compare all of the estimators discussed across a more thorough set of benchmarks including data with AR(1) short-term dynamics ($\rho > 0$), unknown means, and time trends. In these simulations, we will further investigate 2ELW's robustness to such non-ideal data.

2.6. Extensions and Related Methods

While this paper focuses on univariate local Whittle methods for single-frequency estimation, several important extensions should be mentioned. For multivariate time series, Nielsen (2007) developed local Whittle estimation for fractionally cointegrated systems under the assumption of stationarity and long-run exogeneity conditions. Shimotsu (2012) later extended exact local Whittle estimation to fractionally cointegrated systems, enabling joint modeling of long-memory processes with common stochastic trends. For seasonal and cyclical long memory, Arteche (2020) generalized the ELW framework to handle multiple spectral poles. Other recent advances have improved finite-sample performance through bias reduction (Sykulski et al., 2019) and extended local Whittle methods to high-dimensional settings (Baek et al., 2023). Practitioners requiring these specialized extensions will also benefit from the comparative analysis of univariate methods that follows, since these methods provide a foundation for the aforementioned extensions.

While we focus on standard asymptotic approximations in this paper, bootstrap methods by Arteche and Orbe (2016) provide improved inference when asymptotic approximations may be unreliable. Hou and Perron (2014) modified estimators to handle low-frequency contaminations including random level shifts, deterministic level shifts, and deterministic trends, while Wingert, Leschinski, and Sibbertsen (2020) handle seasonal contamination by omitting affected frequencies.

3. Monte Carlo Comparisons Across Methods

This section presents the results of a new and extensive Monte Carlo study comparing the finite-sample performance of multiple local Whittle estimators under various data-generating processes and model mis-specifications. We examine the impact of short-run dynamics, unknown means, time trends, and varying degrees of fractional integration to provide practical guidance on the relative merits and limitations of each method. These simulations, along with our extended empirical results in Section 4, serve to inform the practical guidance for researchers in Section 5.

3.1. Comprehensive Estimator Comparison with Short-Run Dynamics

To serve as a baseline, comprehensive, cross-estimator comparison, we revisit the Monte Carlo design from Table I of Hurvich and Chen (2000). The experiment uses 500 replications of ARFIMA$(1, d, 0)$ processes with AR(1) short-run dynamics with autocorrelation parameter ρ, sample size $n = 500$, and bandwidth $m = 56$ (using the standard $m = \lfloor n^{0.65} \rfloor$ rather than Hurvich-Chen's choice of $m = 36 = \lfloor 0.25 n^{0.8} \rfloor$). We extend beyond the original study by comparing all five estimators we have considered: LW, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW, and 2ELW. Furthermore, we evaluate these estimators over a wider range of 33 parameter combinations $(d, \rho) \in \{-2.2, -1.8, -1.2, -0.6, -0.3, 0.0, 0.3, 0.6, 1.2, 1.8, 2.2\} \times \{0.0, 0.5, 0.8\}$, providing a more comprehensive evaluation across the full range of fractional integration from deep antipersistence to strong persistence. The results, reported in Table 7, reveal several key patterns regarding the relative performance of estimators under short-run AR(1) dynamics.

In the absence of short-run dynamics ($\rho = 0$), the ELW estimator demonstrates superior performance across all values of d, exhibiting minimal bias and consistently achieving the lowest MSE values. The ELW estimator maintains bias within ± 0.003 and MSE below 0.007 across the entire range of $d \in [-2.2, 2.2]$. The 2ELW estimator has similar behavior, but fails with large bias for antipersistent processes with $d < -0.6$. In contrast, the standard LW estimator exhibits more severe deterioration at the boundaries, with bias reaching -0.738 and MSE of 0.560 at $d = 1.8$, and further degradation to bias of -1.160 and MSE of 1.358 at the extreme value $d = 2.2$. The Velasco (V) estimator consistently demonstrates positive bias an order of magnitude higher than other estimators in the valid range, while the Hurvich-Chen (HC) estimator shows intermediate performance with generally smaller bias than V but larger MSE than the exact methods.

The introduction of moderate AR(1) persistence ($\rho = 0.5$) alters the relative performance landscape. All estimators experience upward bias, with the ELW and 2ELW methods maintaining their advantage but now exhibiting bias around 0.100. The standard LW

Table 7. Comprehensive Estimator Comparison

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

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

estimator shows improved performance relative to the no-autocorrelation case for stationary processes ($d < 0.5$) but continues to deteriorate for nonstationary processes, particularly at $d = 1.8$ where bias remains at -0.735, and at the extreme value $d = 2.2$ where bias reaches -1.160. The Velasco estimator exhibits bias of approximately 0.150 across d values, maintaining its stability but at an elevated level. The HC estimator again shows intermediate performance with bias generally increasing with $|d|$ as before, but with larger bias for large negative d values.

Under strong AR(1) persistence ($\rho = 0.8$), all estimators suffer performance degradation, with bias for central $|d|$ values exceeding 0.40 for most methods and MSE values increasing by an order of magnitude. The ELW estimator maintains the smallest bias uniformly at approximately 0.416, while the Velasco estimator exhibits bias around 0.500 for most values. The LW estimator demonstrates erratic behavior, with bias ranging from -1.158 at $d = 2.2$ to 1.250 at $d = -2.2$, exhibiting instability at the boundaries. As before, the 2ELW estimator performs similarly to ELW across central values of d, but fails for large negative d.

Regarding robustness to AR(1) dynamics, the exact methods (ELW and 2ELW) demonstrate stability, degrading gracefully as autocorrelation increases. The Velasco estimator exhibits consistent but systematically biased performance, while the HC estimator shows intermediate robustness. The standard LW estimator proves most vulnerable to both high values of d and autocorrelation, suggesting limited applicability when short-run dynamics are present. At extreme parameter values ($d \in [-2.2, 2.2]$), performance differences become magnified: the exact methods maintain robustness, while LW becomes unreliable and even the Velasco and HC estimators show degradation. These findings indicate that practitioners should favor exact methods when AR(1) dynamics may be present, unless strong anti-persistence is expected, in which case the 2ELW estimator should be avoided.

3.2. Robustness to Unknown Mean

Next we examine estimator sensitivity to unknown population means, comparing performance across mean correction strategies. Table 8 presents results for ARFIMA$(0, d, 0)$ processes across $d \in [-2.2, 2.2]$ with zero mean ($\mu = 0$) or nonzero mean ($\mu = 5$), evaluating three mean correction approaches: no correction ($\hat{\mu} = 0$), sample mean ($\hat{\mu} = \bar{X}$), and first-observation ($\hat{\mu} = X_1$). Note that 2ELW is applied in its standard form to the mean-corrected data, which means that it still applies its own adaptive mean correction in all cases (including the case of no mean correction).

In the top panel, without mean correction, the nonzero mean reveals differences in robustness. The Velasco estimator demonstrates remarkable stability, with MSE values virtually identical across all d values regardless of μ, suggesting inherent robustness to level shifts. Similarly, the LW, HC, and 2ELW estimators are largely unaffected by the presence

Table 8. Robustness to Unknown Mean

	$\mu = 0$					$\mu = 5$				
d	LW	V	HC	ELW	2ELW	LW	V	HC	ELW	2ELW
Mean correction: $\hat{\mu} = 0$ (None)										
-2.2	2.359	0.032	1.064	0.006	1.505	2.363	0.032	1.056	8.525	1.499
-1.8	1.074	0.030	0.351	0.006	0.615	1.074	0.029	0.352	6.308	0.615
-1.2	0.139	0.028	0.024	0.006	0.091	0.136	0.028	0.024	3.287	0.090
-0.6	0.007	0.027	0.010	0.006	0.006	0.007	0.027	0.010	0.835	0.006
-0.3	0.006	0.028	0.010	0.006	0.006	0.006	0.027	0.010	0.206	0.006
0.0	0.006	0.026	0.010	0.006	0.006	0.006	0.027	0.010	0.035	0.006
0.3	0.006	0.027	0.010	0.006	0.006	0.006	0.027	0.010	0.085	0.006
0.6	0.007	0.027	0.010	0.006	0.006	0.006	0.027	0.010	0.015	0.006
1.2	0.022	0.028	0.010	0.006	0.006	0.022	0.028	0.010	0.006	0.006
1.8	0.557	0.031	0.009	0.006	0.006	0.556	0.031	0.009	0.006	0.006
2.2	1.357	0.035	0.011	0.006	0.006	1.355	0.035	0.011	0.006	0.006
Mean correction: $\hat{\mu} = \bar{X}$										
-2.2	2.343	0.032	1.063	2.089	1.497	2.354	0.031	1.054	2.096	1.492
-1.8	1.075	0.029	0.353	0.979	0.618	1.065	0.029	0.354	0.973	0.618
-1.2	0.136	0.027	0.024	0.120	0.090	0.138	0.028	0.024	0.121	0.091
-0.6	0.007	0.027	0.010	0.006	0.006	0.007	0.027	0.010	0.006	0.006
-0.3	0.006	0.028	0.010	0.006	0.006	0.006	0.027	0.010	0.006	0.006
0.0	0.006	0.027	0.010	0.006	0.006	0.006	0.027	0.010	0.006	0.006
0.3	0.006	0.027	0.010	0.006	0.006	0.006	0.027	0.010	0.006	0.006
0.6	0.006	0.028	0.010	0.006	0.006	0.006	0.027	0.010	0.006	0.005
1.2	0.022	0.028	0.010	0.012	0.006	0.022	0.028	0.009	0.012	0.006
1.8	0.559	0.030	0.009	0.427	0.006	0.558	0.031	0.009	0.425	0.006
2.2	1.356	0.035	0.011	1.126	0.006	1.356	0.035	0.011	1.122	0.006
Mean correction: $\hat{\mu} = X_1$										
-2.2	2.364	0.031	1.071	4.845	1.510	2.346	0.032	1.053	4.826	1.495
-1.8	1.062	0.030	0.348	3.130	0.609	1.068	0.030	0.348	3.122	0.611
-1.2	0.135	0.029	0.024	1.297	0.090	0.137	0.028	0.024	1.296	0.090
-0.6	0.007	0.028	0.010	0.279	0.006	0.007	0.028	0.010	0.279	0.006
-0.3	0.006	0.027	0.010	0.058	0.006	0.006	0.027	0.010	0.057	0.006
0.0	0.006	0.027	0.010	0.003	0.006	0.006	0.026	0.009	0.003	0.006
0.3	0.006	0.028	0.010	0.010	0.006	0.006	0.027	0.010	0.009	0.006
0.6	0.006	0.026	0.010	0.006	0.006	0.006	0.027	0.010	0.006	0.006
1.2	0.022	0.028	0.010	0.006	0.006	0.022	0.027	0.010	0.006	0.006
1.8	0.556	0.031	0.009	0.006	0.006	0.559	0.030	0.009	0.006	0.006
2.2	1.353	0.034	0.010	0.006	0.006	1.358	0.036	0.011	0.006	0.006

Notes: MSE results for ARFIMA$(0, d, 0)$ with $n = 500$, $m = 56$, $10,000$ replications. LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW.

of the nonzero mean, but HC has large bias for antipersistent processes and LW is biased for large $|d|$ in both cases. Recall that 2ELW includes its own mean correction, but as before it is biased for antipersistent processes (for zero mean and positive mean). In contrast, ELW exhibits severe asymmetric sensitivity: severe degradation for antipersistent processes (MSE increasing from 0.006 to 8.525 at $d = -2.2$) but full robustness for $d \geq 0.6$.

In the middle panel, we consider sample mean correction ($\hat{\mu} = \bar{X}$). We can see that applying this correction when unnecessary ($\mu = 0$) can be harmful. For ELW, it damages performance at both extremes: MSE increases above 2.089 at $d = -2.2$ and 1.126 at $d = 2.2$, while central values remain robust. Even when the true mean is nonzero ($\mu = 5$), demeaning is detrimental to ELW performance when $d \geq 1.2$. The other estimators appear to be robust to this correction.

In the bottom panel, we see the effects of first-observation correction ($\hat{\mu} = X_1$). Again, all estimators except ELW are largely unaffected. For ELW, the negative effects in this case are asymmetric: when the correction is unnecessary ($\mu = 0$), subtracting the first observation is harmful only for $d < 0$, with MSE reaching 4.845 at $d = -2.2$, but performance is unaffected for $d \geq 0$. This correction strategy avoids the problems with $\bar{X}$ for large positive d.

Overall, these results demonstrate that the Velasco tapered LW estimator provides remarkable robustness to unknown means—corrected or not—across 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.

3.3. Robustness to Linear Time Trends

Finally, the Monte Carlo results in Table 9 examine the sensitivity of estimators to the presence of linear time trends, comparing performance with and without trend contamination across the extended parameter range $d \in [-2.2, 2.2]$. The experiment considers ARFIMA$(1, d, 0)$ processes with no AR dynamics ($\rho = 0$) under two scenarios: no trend ($\beta = 0.0$) and linear trend ($\beta = 0.05$), evaluating performance both without trend correction and with linear OLS detrending.

With no trend correction (top panel), the presence of a linear trend ($\beta = 0.05$) affects the LW estimator most severely, followed by ELW. For antipersistent processes, both estimators show catastrophic performance degradation, with MSE at $d = -2.2$ increasing from 2.359 to 10.168 for LW and from 0.006 to 8.977 for ELW. The sensitivity of both estimators to the trend diminishes as d increases. This asymmetric pattern mirrors the unknown mean results, confirming that ELW's vulnerability to these specification errors is primarily for antipersistent processes. In contrast, the Velasco and Hurvich-Chen estimators maintain

Table 9. Robustness to Time Trend

	$\beta = 0.0$					$\beta = 0.05$				
d	LW	V	HC	ELW	2ELW	LW	V	HC	ELW	2ELW
Trend correction: None										
-2.2	2.359	0.032	1.064	0.006	1.505	10.168	0.032	1.071	8.977	1.641
-1.8	1.072	0.030	0.351	0.006	0.613	7.765	0.030	0.351	6.720	0.691
-1.2	0.136	0.028	0.024	0.006	0.089	4.757	0.028	0.024	3.891	0.145
-0.6	0.007	0.027	0.010	0.006	0.006	2.464	0.028	0.010	1.792	0.086
-0.3	0.006	0.027	0.010	0.006	0.006	1.592	0.027	0.010	1.038	0.080
0.0	0.006	0.028	0.010	0.006	0.006	0.905	0.027	0.010	0.489	0.075
0.3	0.006	0.028	0.010	0.006	0.006	0.409	0.027	0.010	0.149	0.072
0.6	0.006	0.027	0.010	0.006	0.006	0.110	0.026	0.010	0.014	0.019
1.2	0.022	0.028	0.010	0.006	0.006	0.023	0.028	0.010	0.006	0.006
1.8	0.556	0.031	0.009	0.006	0.006	0.558	0.031	0.009	0.006	0.006
2.2	1.357	0.035	0.011	0.006	0.006	1.363	0.035	0.011	0.006	0.006
Trend correction: linear OLS detrending										
-2.2	2.535	0.032	1.063	2.427	1.523	2.556	0.032	1.056	2.448	1.524
-1.8	1.198	0.030	0.349	1.163	0.619	1.209	0.029	0.353	1.173	0.625
-1.2	0.180	0.028	0.024	0.177	0.102	0.176	0.028	0.023	0.173	0.100
-0.6	0.008	0.027	0.010	0.006	0.006	0.008	0.027	0.010	0.007	0.007
-0.3	0.007	0.027	0.010	0.007	0.007	0.006	0.027	0.010	0.006	0.006
0.0	0.007	0.027	0.010	0.007	0.007	0.007	0.027	0.010	0.007	0.007
0.3	0.008	0.027	0.010	0.007	0.007	0.007	0.027	0.010	0.007	0.007
0.6	0.008	0.027	0.010	0.007	0.007	0.008	0.027	0.010	0.007	0.007
1.2	0.015	0.028	0.010	0.009	0.006	0.014	0.027	0.009	0.009	0.006
1.8	0.324	0.030	0.009	0.322	0.006	0.319	0.031	0.009	0.321	0.006
2.2	0.878	0.035	0.011	1.024	0.011	0.880	0.036	0.011	1.027	0.011

Notes: MSE results for ARFIMA$(1, d, 0)$ with $\rho = 0.0$, $n = 500$, $m = 56$, $10,000$ replications.
LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW.

remarkable stability, with MSE values remaining essentially unchanged in the presence of the trend.

Detrending by OLS (bottom panel) effectively neutralizes the impact of the linear trend for the standard LW estimator. Velasco and HC were already largely unaffected by it, and their performance is not hurt by detrending. However, the ELW estimator continues to exhibit sensitivity at the extreme boundaries. In the presence of a trend ($\beta = 0.05$), detrending improves ELW performance for antipersistent processes ($d < 0$), but significant bias remains. However, detrending hurts ELW performance for highly persistent processes ($d \geq 1.8$), even when no trend is present ($\beta = 0$). As with the unknown mean, 2ELW performs well for most values of d but struggles with antipersistence.

These results highlight a crucial practical finding: the Velasco and Hurvich-Chen tapered estimators provide exceptional robustness to linear deterministic trends without requiring any preprocessing, while the standard and exact local Whittle methods require careful trend correction strategies that depend on the unknown memory parameter.

3.4. Sampling Distribution Comparison

To conclude our examination of the finite sample behavior of local Whittle estimators across a wide range of memory parameters, we revisit the extensive baseline Monte Carlo analysis from Section 3.1. Figure 1 presents kernel density estimates of the sampling distributions for five estimators—LW, Velasco, 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 reveal systematic patterns in estimator performance across different persistence regimes, verifying the theoretical results discussed in Section 2. For stationary values ($|d| < 0.5$), all estimators exhibit approximately symmetric distributions centered near the true parameter, with the tapered estimators (Velasco and HC) showing increased dispersion relative to the efficient LW, ELW, and 2ELW estimators. The Velasco tapered estimator exhibits the highest variance and also shows some positive bias. Even for unit root processes ($d = 1.0$), the standard LW estimator remains well-behaved. For the nonstationary, explosive process ($d = 1.6$), the LW estimator is inconsistent and its distribution remains centered around $d = 1.0$, while the exact methods (ELW and 2ELW) maintain their concentration around the true values. In this simulation, with zero mean and no time trend, the ELW estimator exhibits the most robust performance across all regimes, including the strongly nonstationary case with $d = 1.6$ and the antipersistent case with $d = -1.2$. For the $d = -1.2$ case specifically, we see a deterioration of performance due to a breakdown of theoretical validity of each of the other estimators. LW has a large positive bias and high dispersion. 2ELW is also biased upward, with a mean around -0.9. The Velasco and HC tapered LW estimators perform better, but also exhibit some bias. Only the ELW estimator remains largely unbiased and tightly concentrated around the true value of $d = -1.2$.

While these Monte Carlo experiments provide controlled evidence about the performance of these estimators under known data-generating processes, we next evaluate the methods with real economic data, where the memory parameter is unknown and the data may violate the ideal assumptions. The following section turns to empirical applications, replicating key results from the literature to demonstrate how each method behaves with real-world data and to validate our Python implementation.

4. Extended Empirical Applications

This section demonstrates the practical aspects of local Whittle method selection, and validates the PyELW implementation, by replicating key empirical analyses from the literature. We revisit results from Hurvich and Chen (2000), Shimotsu (2010), and Baum et al. (2020), examining how different estimators yield varying conclusions about persistence in economic, financial, and environmental time series.

4.1. Replication of Hurvich and Chen (2000) Table III

As our first empirical comparison, we revisit the analysis in Table III of Hurvich and Chen (2000). The authors apply their complex tapered estimator to seven real-world datasets including both economic and environmental time series as shown in Table 10. The global temperature data consist of seasonally adjusted monthly temperatures for the northern hemisphere from 1854–1989, representing deviations from monthly averages over 1950–1979 (Beran, 1994). The S&P 500 series contains natural logarithms of daily composite stock index levels from July 1962 to December 1995. The remaining series are monthly economic indicators from January 1957 to December 1997 sourced from the International Monetary Fund's International Financial Statistics, including CPI-based inflation rates for the U.S., U.K., and France (computed as differences of the logarithms of CPI), the logarithm of U.S. real manufacturing wages, and the logarithm of U.S. industrial production. All economic series are seasonally adjusted.

Table 10 compares the original Hurvich and Chen (2000) estimates (for HC only) with our results comparing all estimators produced using our PyELW-based implementation. For all estimators, we use the same number of frequencies m as in the original study. For the HC tapered LW estimator, the agreement between our results and the original estimates is excellent across all datasets, with differences typically at the third decimal place or

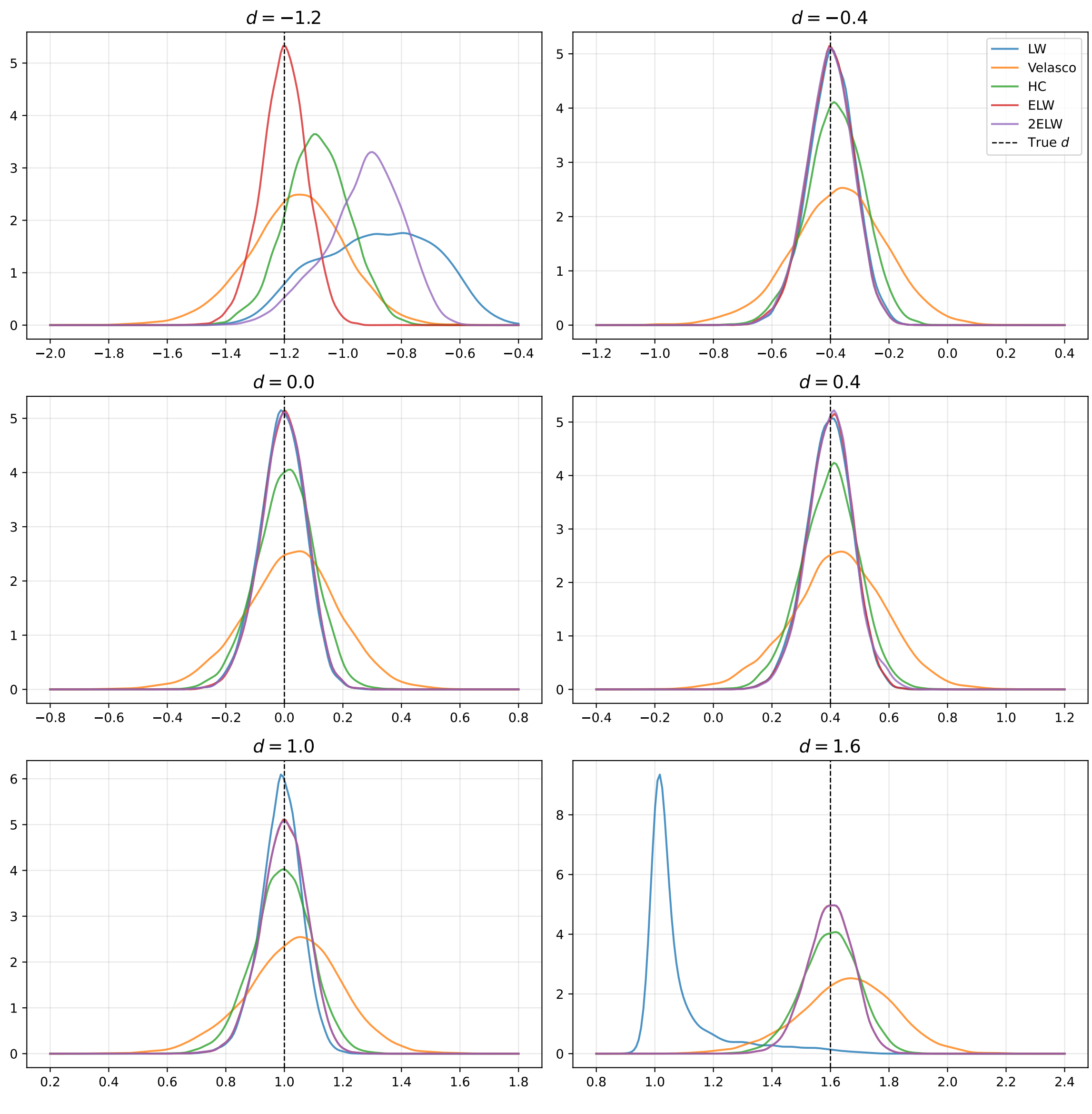

FIGURE 1. Sampling distributions of local Whittle estimators with $n = 500$ over $10,000$ replications

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

Series	n	m	HC (2000) HC	Replication LW	V	HC	ELW	2ELW
Global temp.	1632	130	0.450	0.496	0.442	0.451	0.495	0.471
			(0.060)	(0.046)	(0.076)	(0.060)	(0.045)	(0.044)
S&P 500	8432	1383	0.990	0.980	0.978	0.972	0.986	0.983
			(0.018)	(0.014)	(0.023)	(0.017)	(0.013)	(0.013)
Inflation, US	491	40	0.570	0.626	0.662	0.563	0.643	0.635
			(0.123)	(0.080)	(0.137)	(0.121)	(0.082)	(0.079)
Inflation, UK	491	40	0.330	0.462	0.494	0.339	0.467	0.468
			(0.123)	(0.072)	(0.137)	(0.121)	(0.074)	(0.079)
Inflation, FR	491	40	0.670	0.462	0.819	0.682	0.499	0.450
			(0.123)	(0.083)	(0.137)	(0.121)	(0.087)	(0.079)
Real wages, US	492	35	1.430	1.065	1.560	1.426	1.001	1.287
			(0.121)	(0.096)	(0.147)	(0.132)	(0.102)	(0.085)
Ind. prod., US	492	100	1.340	0.999	1.311	1.363	1.004	1.312
			(0.075)	(0.054)	(0.087)	(0.071)	(0.058)	(0.050)

HC (2000) column reports original results from Table III of Hurvich and Chen (2000).
Number of frequencies m match values used by Hurvich and Chen (2000).
Standard errors in parentheses based on Fisher information.
LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW.

smaller. For U.S. real wages, the small differences likely reflect our use of FRED data for average hourly earnings of production and nonsupervisory employees in manufacturing (CES3000000008) divided by CPI, as the original IMF real wages series is no longer available.

In terms of the economic implications of these estimates, there is broad agreement across estimators. Global temperatures exhibit moderate long memory, with $\hat{d} \in [0.442, 0.496]$ indicating persistent but stationary dynamics. The S&P 500 index displays near-unit root behavior, where $\hat{d} \in [0.972, 0.990]$ aligns with the efficient market hypothesis. The inflation series reveal heterogeneous persistence across countries: U.S. inflation has moderate persistence, with $\hat{d} \in [0.563, 0.662]$, while U.K. inflation shows somewhat weaker dependence, with $\hat{d} \in [0.330, 0.494]$. The five estimators give a mixed picture of inflation persistence in France, with a wide range of estimates $\hat{d} \in [0.450, 0.819]$.

U.S. real wages display strongly nonstationary behavior, with $\hat{d} \in [1.001, 1.560]$ suggesting that wage shocks have permanent effects. The lower estimates from LW (1.065) are consistent with the negative bias exhibited by LW when $d > 1$ as shown in our Monte Carlo results, while ELW's low estimate (1.001) may reflect convergence to a local minimum in its non-convex objective function rather than the global optimum found by the tapered methods and 2ELW. Similarly, industrial production shows evidence of nonstationarity, with $\hat{d} \in [0.999 - 1.363]$ indicating that productivity shocks generate persistent growth effects (again with LW and ELW as outliers). These results add nuance to the traditional classifications of economic series as either $I(0)$ or $I(1)$, underscoring the empirical relevance of fractional integration models.

4.2. Replication of Empirical Results of Shimotsu (2010)

Our second empirical comparison replicates results from Shimotsu (2010), who applied the 2ELW estimator to the classic Nelson and Plosser (1982) dataset extended through 1988 by Schotman and van Dijk (1991). This dataset contains 14 key U.S. macroeconomic series spanning 1860–1988, with sample sizes varying from 80 to 129 observations depending on data availability for each series.

As before, we extend the application by comparing all five estimators considered. To match the original analysis, we modify the LW estimation by first differencing the data, estimating $d - 1$ using the standard LW estimator, then adding unity to the resulting estimate. Additionally, for the 2ELW estimator, we follow Shimotsu (2010) and first use OLS to remove any nonzero mean and linear time trend and we use bandwidth $m = n^{0.7}$ as in the original paper.[2]

Table 11 presents the original estimates from Shimotsu (2010) along with the replication from our Python implementation. The replication accuracy is very high: we achieve exact matches (to three decimal places) for the LW estimates on 12 of 14 series and for the 2ELW estimator on 10 of 14 series. The most notable discrepancy occurs for the unemployment rate series. We observe $n = 99$ for unemployment while Shimotsu (2010) reports $n = 129$, but we were not able to determine the reason for this difference.

Our replication confirms the key empirical findings of Shimotsu (2010). The estimates for real GNP, real per capita GNP, employment, real wage, velocity of money, and stock prices are near $d = 1$, consistent with unit root behavior. For the GNP deflator, CPI, and nominal wage, estimates substantially exceed unity ($d > 1.3$), confirming that inflation series exhibit $I(d)$ behavior with $d \in (0, 1)$ when expressed in first differences. All series appear to exhibit fractional integration.

Although there is broad agreement across methods, it is apparent that ELW has some difficulty with certain series (e.g., Real GNP), often reporting spuriously low values. For these results, we did not apply any additional pre-processing. 2ELW is robust to unknown means and time trends, but ELW is not. To investigate this further, we plot the objective functions for each estimator in Figure 2. We can see that LW objective functions (LW, V, HC) are convex with a single global minimum. These differ in the location of the minimum—recall that LW is only valid for stationary models—but have the same overall shape. The ELW and 2ELW objective functions are fundamentally different. They are non-convex and have multiple local minima. Up to $d = 0.5$ they coincide exactly. At $d = 0.5$, the piecewise adaptive weight function defined in (6) begins to apply positive weight to $\hat{\mu} = X_1$. This

[2] We discovered that using `round`($n^{0.7}$) rather than `floor`($n^{0.7}$) was required to match the results.

Table 11. Extended Nelson and Plosser Data: LW Estimator Comparison

			Shimotsu (2010)		Replication				
Series	n	m	LW	2ELW	LW	V	HC	ELW	2ELW
Real GNP	80	21	1.077	1.126	1.077	1.424	1.268	0.135	1.126
			(0.109)	(0.109)	(0.109)	(0.189)	(0.134)	(0.109)	(0.109)
Nominal GNP	80	21	1.273	1.303	1.273	1.479	1.375	0.122	1.303
			(0.109)	(0.109)	(0.109)	(0.189)	(0.134)	(0.109)	(0.109)
Real per capita GNP	80	21	1.077	1.128	1.077	1.470	1.287	0.054	1.127
			(0.109)	(0.109)	(0.109)	(0.189)	(0.134)	(0.109)	(0.109)
Industrial production	129	30	0.821	0.850	0.821	0.975	0.946	0.919	0.850
			(0.091)	(0.091)	(0.091)	(0.158)	(0.112)	(0.091)	(0.091)
Employment	99	25	0.968	1.000	1.007	1.315	1.188	0.041	1.055
			(0.102)	(0.102)	(0.100)	(0.174)	(0.122)	(0.100)	(0.100)
Unemployment rate[a]	99	25	0.951	0.980	0.698	0.960	0.911	0.875	0.741
			(0.091)	(0.091)	(0.100)	(0.174)	(0.122)	(0.100)	(0.100)
GNP deflator	100	25	1.374	1.398	1.374	1.376	1.364	0.240	1.398
			(0.100)	(0.100)	(0.100)	(0.174)	(0.122)	(0.100)	(0.100)
CPI	129	30	1.273	1.287	1.273	1.375	1.308	1.112	1.287
			(0.091)	(0.091)	(0.091)	(0.158)	(0.112)	(0.091)	(0.091)
Nominal wage	89	23	1.300	1.351	1.300	1.409	1.435	0.165	1.351
			(0.104)	(0.104)	(0.104)	(0.181)	(0.128)	(0.104)	(0.104)
Real wage	89	23	1.047	1.089	1.047	1.298	1.282	0.138	1.089
			(0.104)	(0.104)	(0.104)	(0.181)	(0.128)	(0.104)	(0.104)
Money stock	100	25	1.460	1.501	1.460	1.528	1.578	0.760	1.501
			(0.100)	(0.100)	(0.100)	(0.174)	(0.122)	(0.100)	(0.100)
Velocity of money	120	29	0.953	0.993	0.969	0.753	0.722	1.004	1.001
			(0.095)	(0.095)	(0.093)	(0.161)	(0.114)	(0.093)	(0.093)
Bond yield	89	23	1.091	1.108	1.091	1.171	1.121	0.967	1.108
			(0.104)	(0.104)	(0.104)	(0.181)	(0.128)	(0.104)	(0.104)
Stock prices	118	28	0.900	0.958	0.900	0.808	0.922	0.930	0.958
			(0.095)	(0.095)	(0.094)	(0.164)	(0.116)	(0.094)	(0.094)

Shimotsu (2010) columns report original results from Table 8 of Shimotsu (2010).
Bandwidth $m = \lfloor n^{0.70} \rfloor$. Asymptotic standard errors in parentheses.
LW = Local Whittle, V = Velasco (Kolmogorov), HC = Hurvich-Chen, ELW = Exact Local Whittle, 2ELW = Two-step ELW.

[a] Shimotsu (2010) reports 129 observations for this series, while we observe 99.

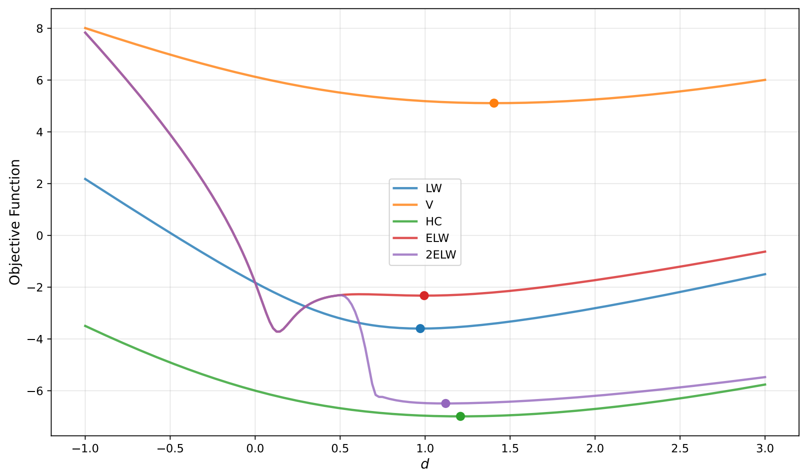

FIGURE 2. Local Whittle objective functions for U.S. Real GNP

leads the 2ELW objective function to diverge from the ELW objective function. The weight fully shifts away from $\hat{\mu} = \bar{X}$ to $\hat{\mu} = X_1$ at $d = 0.75$, precisely where the 2ELW objective function has another kink. The first-step estimation via HC provides a consistent estimator and a good starting value for optimization of the 2ELW objective function, which avoids the spurious ELW/2ELW local minimum at 0.135 and instead locates the 2ELW global minimum at 1.126. Despite the simple scalar nature of these objective functions, this highlights the difficulty of finding a guaranteed global minimum without an exhaustive search.

Overall, these findings support the empirical relevance of fractional integration models for macroeconomic time series and validate the practical utility of the local Whittle approach for applied research, while also highlighting the importance for accounting for unknown means and time trends in macroeconomic data.

4.3. Replication of Baum, Hurn, and Lindsay (2020)

As a final replication and validation of PyELW's results we reproduce the LW and ELW estimates reported in Baum et al. (2020), which introduced the `whittle` command for Stata. Their examples involve two classic long-memory datasets: annual Nile River minimum levels ($N = 663$, 622–1284) and monthly global sea levels ($N = 1,558$, 1880–2009). These datasets exhibit features of fractional integration, and they prove to be useful benchmarks.

Table 12 presents the original Stata results reported by Baum et al. (2020) along side our replicated results using PyELW. For each dataset and preprocessing[3] approach (demeaning only vs. demeaning and detrending), we compare both LW and ELW estimates using

[3]We note that the Stata whittle command always demeans the data (i.e., it is not optional) and will optionally also detrend the data when requested.

identical bandwidth selections. The replication achieves a high degree of accuracy across all specifications, with exact agreement for point estimates, standard errors, and asymptotic standard errors.

Table 12. Replication of empirical results from Baum, Hurn, and Lindsay (2020)

Specification			BHL (2020)			PyELW Results		
Transform	Estimator	m	$\hat{d}$	S.E.	Asy.S.E.	$\hat{d}$	S.E.	Asy.S.E.
Nile River Data ($N = 663$)								
Demeaned	LW	68	0.409	0.062	0.060	0.409	0.062	0.060
Demeaned	ELW	68	0.407	0.062	0.060	0.407	0.062	0.060
Detrended	LW	68	0.394	0.065	0.060	0.394	0.065	0.060
Detrended	ELW	68	0.397	0.066	0.060	0.397	0.066	0.060
Sea Level Data ($N = 1,558$)								
Demeaned	LW	118	0.859	0.044	0.046	0.859	0.044	0.046
Demeaned	ELW	118	0.802	0.035	0.046	0.802	0.035	0.046
Detrended	LW	39	0.551	0.081	0.080	0.551	0.081	0.080
Detrended	LW	118	0.454	0.042	0.046	0.454	0.042	0.046
Detrended	ELW	39	0.524	0.079	0.080	0.524	0.079	0.080
Detrended	ELW	118	0.486	0.044	0.046	0.486	0.044	0.046

The table compares LW and ELW estimates using identical datasets and specifications. Original results from Stata `whittle` command compared with PyELW results.

The sea level results raise important practical considerations for long-memory estimation. The demeaned specification yields high memory parameter estimates ($\hat{d} \approx 0.8$), reflecting the strong deterministic trend in the data. Detrending substantially reduces these estimates to more plausible values ($\hat{d} \approx 0.5$), consistent with stationary long-memory behavior. This demonstrates the importance of handling potentially nonzero means and time trends when analyzing real data, underscoring the problems leading Shimotsu (2010) to develop the 2ELW estimator. Finally, the exact agreement between the original Stata results and those of PyELW provide evidence of the accuracy of the implementations and underlying algorithms across both statistical computing platforms.

5. Practical Guidance for Practitioners

Building on the theoretical properties discussed in Section 2, the comprehensive Monte Carlo evidence presented in Section 3, and the extended empirical analysis of Section 4, this section distills these findings to provide practical recommendations for selecting and implementing local Whittle estimators in applied research. We emphasize that no single estimator dominates across 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 valid parameter range represents the primary consideration in method selection. For processes believed to be stationary ($d < 0.5$), the standard LW estimator provides optimal efficiency with minimal computational burden. However, our Monte Carlo evidence demonstrates that LW exhibits severe bias for $|d| > 1.0$. For potentially nonstationary series where $|d|$ may exceed 1.0, practitioners should employ either the ELW or 2ELW estimators. ELW maintains consistency across the unrestricted range $d \in (-\infty, \infty)$, provided that the optimization interval has width less than $9/2$, but is sensitive to the mean and trend specification. The 2ELW provides automatic mean and trend removal with better robustness to local minima, though with restricted range $d \in (-0.5, 2)$ and increased computational cost. The tapered estimators offer intermediate solutions: they extend the valid parameter range compared to LW, but at the cost of increased variance.

5.2. Robustness to Short-Run Dynamics

Our Monte Carlo evidence demonstrates that AR(1) dynamics affect all local Whittle estimators, inducing positive bias that increases with the autocorrelation parameter. For moderate autocorrelation ($\rho = 0.5$), the exact methods (ELW and 2ELW) maintain their relative advantage, while stronger dynamics ($\rho = 0.8$) push bias higher for all methods. The Velasco estimator exhibits predictable degradation, maintaining consistent bias levels across d values, while the standard LW estimator becomes particularly unreliable at the boundaries. When short-run dynamics are suspected, practitioners should: (1) expect all estimates to be biased upward; (2) favor ELW or 2ELW for their relative robustness, unless antipersistence is suspected (where 2ELW fails); and (3) use more conservative (smaller) bandwidth selections to reduce contamination from higher frequencies where short-run dynamics dominate.

5.3. Accounting for Unknown Means and Time Trends

Our Monte Carlo results reveal important differences in robustness to deterministic components across estimators. The ELW estimator exhibits high sensitivity to nonzero means and linear trends, particularly for antipersistent processes ($d < 0$), where MSE can increase by many orders of magnitude. This vulnerability appears to be asymmetric: catastrophic for $d < -0.6$ but negligible for $d > 0.0$. When analyzing data with potential nonzero means or trends, practitioners face several options: (1) use the 2ELW estimator with automatic mean and trend removal, with range restricted to $d \in (-0.5, 1.75)$ when detrending; (2) apply the Velasco or HC tapered estimators, which demonstrate exceptional robustness to both unknown means and trends without preprocessing, at the expense of efficiency; (3) for suspected trends, preprocess with OLS detrending before applying LW, which effectively restores LW's performance; or (4) preprocess before ELW, though this only partially mitigates sensitivity at extreme parameter values and can introduce problems when unnecessary. We reiterate that blindly demeaning or detrending when unnecessary can be detrimental, particularly for ELW with large values of $|d|$.

5.4. Diagnostic Strategy Using Multiple Estimators

Given the complementary strengths and weaknesses of different methods, we recommend estimating d using multiple approaches as a diagnostic strategy. Substantial disagreement between estimators often signals specific sample characteristics requiring attention. For instance, if LW yields $\hat{d} \approx 1.0$ while ELW produces $\hat{d} > 1.5$, this pattern suggests true nonstationarity beyond LW's valid range. Conversely, large negative ELW estimates (e.g., $\hat{d} < -1$) while other methods yield values closer to zero could indicate true strong antipersistence, as Figure 1 demonstrates that only ELW remains unbiased for extreme negative values like $d = -1.2$, where other estimators exhibit substantial upward bias. However, if such negative ELW estimates seem implausible given the economic context, they may instead signal contamination from nonzero means or trends, which our Monte Carlo results show can severely affect ELW for negative d.

5.5. Bandwidth Selection Considerations

The bandwidth parameter m determines the number of frequencies used in estimation and its choice affects the bias-variance trade-off in local Whittle estimation. While Henry and Robinson (1996) established the mean squared error-optimal rate of $m \propto n^{0.8}$, the constant of proportionality is not known. Furthermore, this assumes no contamination from short-run dynamics and may be too optimistic for real data. Monte Carlo and empirical work in the literature has converged on the more conservative choice $m = \lfloor n^{0.65} \rfloor$ as a practical standard. This rate, used by Shimotsu and Phillips (2005) and Shimotsu (2010) in their simulations and implemented as the default in both PyELW and Stata's `whittle` command (Baum et al., 2020), provides additional robustness to model misspecification. Economic time series typically exhibit short-run dynamics that contaminate higher frequencies in finite samples. In empirical examples, Shimotsu (2010) used the slightly less conservative $m = n^{0.7}$. We recommend that, at a minimum, researchers carry out a sensitivity analysis across $m = n^{\alpha}$ for $\alpha \in [0.6, 0.8]$. Beyond simple power rules, Baillie et al. (2014) proposed cross-validation approaches focused on forecasting performance while Arteche and Orbe (2017) developed a bandwidth selection strategy based on minimizing a bootstrap approximation of the mean square error.

5.6. Computational and Optimization Issues

The non-convex objective functions of ELW and 2ELW, as illustrated in Figure 2, present computational challenges not encountered with standard LW. These methods can exhibit multiple local minima and for 2ELW the adaptive weighting function introduces kinks. Large disagreement in estimates across methods, with spuriously low ELW estimates, could indicate convergence to a local optimum rather than mean or trend contamination which typically inflates ELW estimates. The 2ELW's preliminary tapered estimation provides a consistent starting value that helps avoid spurious local minima, though at the cost of additional computational time and complexity from the two-stage procedure.

5.7. Recommended Empirical Approach

Based on our analysis, we propose the following practical approach:

1. Apply multiple estimators to obtain initial estimates and identify potential issues through their disagreement patterns.

2. Interpret disagreements diagnostically: LW bounded near unity with higher ELW/2ELW suggests $d > 1$; large negative ELW with others near zero indicates either true antipersistence or contamination; low ELW relative to tapered methods indicates possible ELW local minima.

3. Select your primary estimator based on the diagnosed parameter range and data characteristics.

4. Conduct sensitivity analysis using different bandwidths ($m = n^{\alpha}$ for $\alpha \in [0.6, 0.8]$) and preprocessing choices.

5. Report multiple estimates when several methods are valid, acknowledging uncertainty in the memory parameter.

This approach leverages the strengths of the different estimators while providing diagnostic information about potential data issues, ultimately leading to more robust and reliable inference about long-memory parameters.

6. Conclusion

This paper has provided a comprehensive evaluation of local Whittle methods for estimating the memory parameter in fractionally integrated time series, addressing a gap in the applied literature where practitioners face multiple estimators without clear guidance on their relative merits.

Our extensive Monte Carlo experiments and empirical replications reveal fundamental trade-offs among methods. The standard LW estimator provides optimal efficiency for stationary processes but exhibits bias for $|d| > 1$. ELW achieves excellent performance across the entire parameter space under ideal conditions, but suffers extreme sensitivity to nonzero means and trends, particularly for antipersistent processes where MSE can increase by orders of magnitude. The 2ELW estimator mitigates this through adaptive mean correction but restricts the valid parameter range and also fails for strong antipersistence. The Velasco and Hurvich-Chen tapered estimators demonstrate remarkable robustness to both unknown means and deterministic trends without preprocessing, though with increased variance relative to the efficient methods.

Through systematic replication of key results from the literature, we validate the PyELW implementation while demonstrating how method selection can substantially affect economic conclusions. Our empirical applications show that the same series can appear stationary under one estimator yet strongly persistent under another, with disagreement patterns providing diagnostic information about data characteristics such as contamination or convergence to local minima.

In the future, a key challenge for theoretical research in this area is developing an estimator that combines ELW's efficiency and unrestricted parameter range with the robustness properties of tapered methods. If possible, such a method would eliminate the difficult trade-offs practitioners currently face.

References

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

Arteche, J. and J. Orbe (2016). A bootstrap approximation for the distribution of the local Whittle estimator. *Computational Statistics and Data Analysis 100*, 645–660.

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.

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.

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.

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 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.

Nelson, C. R. and C. I. Plosser (1982). Trends and random walks in macroeconomic time series: Some evidence and implications. *Journal of Monetary Economics 10*(2), 139–162.

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.

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

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.

Schotman, P. C. and H. K. van Dijk (1991). On Bayesian routes to unit roots. *Journal of Applied Econometrics 6*(4), 387–401.

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.

Sykulski, A. M., S. C. Olhede, A. P. Guillaumin, J. M. Lilly, and J. J. Early (2019). The debiased Whittle likelihood. *Biometrika 106*(2), 251–266.

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

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