Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend

Katsumi Shimotsu
Queen’s University

Department of Economics
Queen’s University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

3-2006
Exact Local Whittle Estimation of Fractional Integration with Unknown Mean and Time Trend

Katsumi Shimotsu*

Department of Economics
Queen’s University
Kingston, Ontario K7L 3N6, Canada

March 30, 2006

Abstract

Recently, Shimotsu and Phillips (2005) developed a new semiparametric estimator, the exact local Whittle (ELW) estimator, of the memory parameter \( d \) in fractionally integrated processes. The ELW estimator has been shown to be consistent and have the same \( N(0, \frac{1}{4}) \) limit distribution for all values of \( d \) if the optimization covers an interval of width less than \( \frac{9}{2} \) and the mean of the process is known. With the intent to provide an efficient semiparametric estimator suitable for economic data, we extend the ELW estimator so that it accommodates an unknown mean and a polynomial time trend. We show that the resulting feasible ELW estimator is consistent and has a \( N(0, \frac{1}{4}) \) limit distribution for \( d \in (-\frac{1}{2}, 2) \) (or \( d \in (-\frac{1}{2}, \frac{7}{4}) \) when the data has a polynomial trend) except for a few negligible intervals. We also develop a two-step feasible ELW estimator that avoids the exclusion of these intervals. A simulation study shows that the feasible ELW estimator inherits the desirable properties of the ELW estimator.

JEL Classification: C22

*E-mail: shimotsu@qed.econ.queensu.ca. The author thanks the Editor, Associate Editor, and referees for very helpful comments. The author also thanks Peter C. B. Phillips and Morten Ø. Nielsen for helpful comments and the Cowles Foundation for hospitality during his stay from January 2002 to August 2003. This research was supported by ESRC under Grant R000223629.
1 Introduction

Fractionally integrated (I(d)) processes have attracted growing attention among empirical researchers in economics and finance. In part this is because I(d) processes provide an extension to the classical dichotomy of I(0) and I(1) time series and equip us with more general alternatives for modelling long-range dependence. Empirical research continues to find evidence that I(d) processes can provide a suitable description of certain long range characteristics of economic and financial data (for a survey, see Henry and Zaffaroni 2002). Because of their flexibility in modeling temporal dependence, I(d) processes can also help to reconcile implications from economic models with observed data. Indeed, their use have provided solutions for many empirical “puzzles” in economics and finance, e.g., consumption (Diebold and Rudebusch 1991, Haubrich 1993), term structure (Backus and Zin 1993), international finance (Maynard and Phillips 2001), and economic growth (Michelacci and Zaffaroni 2000).

The memory parameter, d, plays a central role in the definition of fractional integration and is often the focus of empirical interest. Semiparametric estimation of d is appealing in empirical work because it is agnostic about the short-run dynamics of the process and hence is robust to its misspecification. Two common statistical procedures in this class are log periodogram regression and local Whittle estimation (Robinson 1995a, 1995b). Although these estimators are consistent for $d \in (\frac{1}{2}, 1]$ and asymptotically normally distributed for $d \in (\frac{1}{2}, \frac{3}{4})$, they are also known to exhibit nonstandard behavior when $d > \frac{3}{4}$. For instance, they have a nonnormal limit distribution for $d \in [\frac{3}{4}, 1]$, and they converge to unity in probability and are inconsistent for $d > 1$ (Kim and Phillips 1999, Phillips 1999b, Phillips and Shimotsu 2005). To avoid inconsistency and an unreliable basis for inference when $d$ may be larger than $\frac{3}{4}$, a simple and commonly used procedure is to estimate $d$ by taking first differences of the data, estimating $d - 1$, and adding one to the estimate $\hat{d} - 1$. However, if the data is trend stationary, i.e., I(d) with $d \in [0, \frac{1}{2})$ around a linear time trend, taking a first difference of a time series reduces it to $I(d)$ with $d \in [-1, -\frac{1}{2})$. In this case, the local Whittle estimator converges either to the true parameter value or to 0 depending on the number of frequencies used in estimation (Shimotsu and Phillips, 2006).

Data tapering has been suggested (Velasco, 1999, Hurvich and Chen, 2000) as a solution to extend the range of consistent estimation of $d$. Tapered estimators are invariant to a linear (and possibly higher order) time trend and asymptotically normal for $d \in (-\frac{1}{2}, \frac{3}{2})$ (and for larger values of $d$ if higher-order tapers are used), but they have a larger variance (1.5 times or more) than the untapered estimator. As a result, there is currently no general purpose efficient estimation procedure when the value of $d$ may take on values in the nonstationary zone beyond $\frac{3}{4}$.

Many economists and econometricians took part in the debate on whether economic time series are trend stationary or difference stationary. This debate remains inconclusive partly because of the low power and discontinuity in the data-generating model of the unit root tests. In the context of I(d) processes, these questions are translated into whether $d \geq \frac{1}{2}$ or $d < \frac{1}{2}$, because I(d) processes become nonstationary
when \( d \geq \frac{1}{2} \). Gil-Alaña and Robinson (1997) applied Robinson’s (1994) LM test to macroeconomic data to test the null hypothesis that \( d = d_0 \) for various values of \( d_0 \), including \( d = \frac{1}{2} \), and found that the results depend on how the short-run dynamics of the data is specified. Therefore, it is of great interest to investigate this issue using the semiparametric approach which is agnostic about short-run dynamics. However, neither using the raw data, differenced data, or combining the two can answer whether \( d \geq \frac{1}{2} \), because these procedures must assume either \( d < \frac{3}{4} \) or \( d > \frac{1}{2} \) prior to estimation.

Recently Shimotsu and Phillips (2005) developed a new semiparametric estimator, the exact local Whittle (ELW) estimator, which seems to offer a good general purpose estimation procedure for the memory parameter that applies throughout the stationary and nonstationary regions of \( d \). The ELW estimator is consistent and has the same \( N(0, \frac{1}{4}) \) limit distribution for all values of \( d \) if the optimization covers an interval of width less than \( \frac{9}{2} \) and the mean (initial value) of the process is known. As such, it provides a basis for constructing valid asymptotic confidence intervals for \( d \) that are valid regardless of the true value of the memory parameter.

Economic time series are often modeled with an unknown mean and a polynomial time trend. First, we examine the effect of an unknown mean (initial value) on ELW estimation. It is shown that (i) if an unknown mean is replaced by the sample average, then the ELW estimator is consistent for \( d \in (-\frac{1}{2}, 1) \) and asymptotically normal for \( d \in (-\frac{1}{2}, \frac{3}{4}) \), but simulations suggest that the estimator is inconsistent for \( d > 1 \), and (ii) if an unknown mean is replaced by the first observation, then the ELW estimator is consistent for \( d > 0 \) and asymptotically normal for \( d \in (0, 2) \), but the consistency and asymptotic normality for \( d \in (0, \frac{1}{2}) \) requires a strong assumption on the number of periodogram ordinates used in estimation, and simulations suggest the estimator is inconsistent for \( d \leq 0 \). An unknown mean needs to be estimated carefully in the ELW estimation.

In view of the above undesirable effect of unknown mean on the ELW estimation, we extend the ELW estimator so that it accommodates an unknown mean and a polynomial time trend. One approach, which we call feasible ELW estimation, appears promising. It combines two estimators of the unknown mean of the process, the sample average and the first observation, depending on the value of \( d \). The presence of a linear and/or quadratic time trend is dealt with by prior detrending of the data. The feasible ELW estimator is shown to be consistent for \( d > -\frac{1}{2} \) and have the same \( N(0, \frac{1}{4}) \) limit distribution for \( d \in (-\frac{1}{2}, 2) \) (\( d \in (-\frac{1}{2}, \frac{7}{2}) \) when the data are detrended) excluding arbitrary small intervals around 0 and 1. We also show that the two-step estimator, which is based on the objective function of the feasible ELW estimator and uses a tapered estimator in the first stage, does not require the exclusion of these intervals and has the same limit distribution as the feasible ELW estimator. The finite sample performance of the feasible ELW estimator inherits the desirable property of the ELW estimator, apart from a small increase in bias and variance when the data are detrended.

The remainder of the paper is organized as follows. Section 2 briefly reviews ELW estimation. In Section 3, two estimators for the unknown mean are compared, and
the asymptotic properties of the feasible ELW estimator are demonstrated. Two-step estimation is discussed in Section 4. Section 5 reports some simulation results and gives an empirical application using the extended Nelson-Plosser data. Section 6 concludes the paper. Proofs and some technical results are collected in Appendices A and B.

2 A model of fractional integration and ELW estimation

First we briefly review the exact local Whittle (ELW) estimation developed by Shimotsu and Phillips (2005) as it serves as the basis for the following analysis. Consider the fractionally integrated process

\[ \Delta^d X_t = (1 - L)^d X_t = u_t \{ t \geq 1 \}, \quad t = 0, \pm 1, \ldots \]  

where \( \Delta \{ \cdot \} \) denotes the indicator function. \( u_t \) is assumed to be stationary with zero mean and spectral density \( f_u(\lambda) \) satisfying \( f_u(\lambda) \sim G \) for \( \lambda \sim 0 \). Inverting and expanding the binomial in (1) gives a representation of \( X_t \) in terms of \( u_1, \ldots, u_n \), which is valid for all values of \( d \):

\[ X_t = \Delta^{-d} u_t \{ t \geq 1 \} = (1 - L)^{-d} u_t \{ t \geq 1 \} = \sum_{k=0}^{t-1} \frac{(d)_k}{k!} u_{t-k}, \quad t = 0, \pm 1, \ldots \]

where \((d)_k = \Gamma(d + k)/\Gamma(d) \) and \( \Gamma(\cdot) \) is the gamma function.

Define the discrete Fourier transform (dft) and the periodogram of a time series \( a_t \) evaluated at the fundamental frequencies as

\[ w_a(\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^{n} a_t e^{it\lambda_j}, \quad \lambda_j = \frac{2\pi j}{n}, \quad j = 1, \ldots, n, \]

\[ I_a(\lambda_j) = |w_a(\lambda_j)|^2. \]

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

\[ Q_m(G, d) = \frac{1}{m} \sum_{j=1}^{m} \left[ \log \left( G\lambda_j^{-2d} \right) + \frac{1}{G} I_{\Delta^d x}(\lambda_j) \right]. \]

Concentrating \( Q_m(G, d) \) with respect to \( G \), Shimotsu and Phillips (2005) define the ELW estimator as

\[ \tilde{d} = \arg \min_{d \in [\Delta_1, \Delta_2]} R(d), \]

where \( \Delta_1 \) and \( \Delta_2 \) are the lower and upper bounds of the admissible values of \( d \) and

\[ R(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d x}(\lambda_j). \]

In what follows, we distinguish the true value of \( d \) and \( G \) by \( d_0 \) and \( G_0 \). The ELW estimator has been shown to be consistent and asymptotically normally distributed for any \( d_0 \in (\Delta_1, \Delta_2) \) if \( \Delta_2 - \Delta_1 \leq \frac{9}{2} \) and under fairly mild assumptions on \( m \) and the stationary component \( u_t \):
Assumption 1 \( f_u(\lambda) \sim G_0 \in (0, \infty) \) as \( \lambda \to 0 + \).

Assumption 2 In a neighborhood \((0, \delta)\) of the origin, \( f_u(\lambda) \) is differentiable and \( \frac{d}{d\lambda} \log f_u(\lambda) = O(\lambda^{-1}) \) as \( \lambda \to 0 + \).

Assumption 3 \( u_t = C(L) \varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \) with \( \sum_{j=0}^{\infty} c_j^2 < \infty \), where \( E(\varepsilon_t|F_{t-1}) = 0 \), \( E(\varepsilon_t^2|F_{t-1}) = 1 \) a.s., \( t = 0, \pm 1, \ldots \), in which \( F_t \) is the \( \sigma \)-field generated by \( \varepsilon_s, s \leq t \), and there exists a random variable \( \varepsilon \) such that \( E\varepsilon^2 < \infty \) and for all \( \eta > 0 \) and some \( K > 0 \), \( \Pr(|\varepsilon| > \eta) \leq K \Pr(|\varepsilon| > \eta) \).

Assumption 4 \( m^{-1} + m(\log m)^{1/2} n^{-1} + m^{-\gamma} \log n \to 0 \) for any \( \gamma > 0 \).

Assumption 5 \( \Delta_2 - \Delta_1 \leq \frac{9}{2} \).


Lemma 1 (Shimotsu and Phillips 2005, Theorem 2.1). Suppose \( X_t \) is generated by (1) with \( d_0 \in [\Delta_1, \Delta_2] \) and Assumptions 1-5 hold. Then \( \tilde{d} \to_d d_0 \) as \( n \to \infty \).

Assumption 1' Assumption 1 holds and also for some \( \beta \in (0, 2] \), \( f_u(\lambda) = G_0(1 + O(\lambda^\beta)) \), as \( \lambda \to 0 + \).

 Assumption 2' In a neighborhood \((0, \delta)\) of the origin, \( C(e^{i\lambda}) \) is differentiable and \( \frac{d}{d\lambda} C(e^{i\lambda}) = O(\lambda^{-1}) \) as \( \lambda \to 0 + \).

Assumption 3' Assumption 3 holds and also \( E(\varepsilon_t^3|F_{t-1}) = \mu_3, E(\varepsilon_t^4|F_{t-1}) = \mu_4 \), a.s., \( t = 0, \pm 1, \ldots \), for finite constants \( \mu_3 \) and \( \mu_4 \).

Assumption 4' As \( n \to \infty \), \( m^{-1} + m^{1+2\beta}(\log m)^2 n^{-2\beta} + m^{-\gamma} \log n \to 0 \) for any \( \gamma > 0 \).

Assumption 5' Assumption 5 holds.

Lemma 2 (Shimotsu and Phillips 2005, Theorem 2.2). Suppose \( X_t \) is generated by (1) with \( d_0 \in (\Delta_1, \Delta_2) \) and Assumptions 1'-5' hold. Then \( m^{1/2}(\tilde{d} - d_0) \to_d N(0, \frac{1}{4}) \) as \( n \to \infty \).

3 ELW estimation with unknown mean

The asymptotic properties of the ELW estimator in Section 2 are derived under the assumption that \( X_t \) is generated by (1). However, when a researcher models an economic time series, typically its mean/initial condition is assumed to be unknown and it is often accompanied by a linear time trend. In this section, we analyze the
effect of an unknown mean/initial condition on the ELW estimation and extend the ELW estimator to accommodate it.

3.1 Two choices of \( \hat{\mu} \): \( \overline{X} \) and \( X_1 \)

We consider estimating \( \hat{\mu} = \mathbb{E}[X_t] \) when the data \( X_t \) are generated by

\[
X_t = \mu_0 + X^0_t, \quad X^0_t = (1 - L)^{-d_0} u_t 1_{\{t \geq 1\}},
\]

where \( \mu_0 \) is a non-random unknown finite number. Because \( EU_t = 0 \), the initial condition \( \mu_0 \) is also the mean of the process \( X_t \) in the sense \( EX_t = \mu_0 \). Consider estimating \( \mu_0 \) by \( \hat{\mu} \). One candidate for \( \hat{\mu} \) is the sample average \( \overline{X} = n^{-1} \sum_{t=1}^{n} X_t \).

For \( d_0 > -\frac{1}{2} \), the error in estimating \( \mu_0 \) by \( \overline{X} \) is

\[
\hat{\mu} - \mu_0 = \overline{X} - \mu_0 = n^{-1} (1 - L)^{-d_0} u_n 1_{\{t \geq 1\}} = O_p(n^{-1/2}).
\]

Because the magnitude of the error increases as \( d_0 \) increases, the sample average is not a good estimate of \( \mu_0 \) for large \( d_0 \).

Note that, when \( d_0 \geq \frac{1}{2} \), the variance of \( X^0_t \) tends to infinity as \( t \to \infty \), and the magnitude of \( X^0_t \) dominates that of \( \mu_0 \). Consequently, if \( d_0 \geq \frac{1}{2} \), the signal on the value of \( d \) from \( X^0_t \) dominates the noise from \( \mu_0 \), and one can estimate \( d \) consistently from \( X_1 \) without correcting for \( \mu_0 \). In other words, there is no need to estimate \( \mu_0 \). In a finite sample, however, it would be sensible to reduce the adverse effect of large \( \mu_0 \) (10,000, say) by using the first observation \( X_1 \) as a proxy of \( \mu_0 \). This leads to

\[
\hat{\mu} = X_1, \quad \text{whose error in estimating } \mu_0 \text{ is}^3
\]

\[
\hat{\mu} - \mu_0 = X_1 - \mu_0 = (1 - L)^{-d_0} u_1 1_{\{t \geq 1\}} = u_1 = O_p(1).
\]

Therefore, \( X_1 \) serves as another estimator of \( \mu_0 \) for large \( d_0 \) and complements \( \overline{X} \).

We state the results formally. Estimate \( \mu_0 \) by \( \hat{\mu} \), and define the resulting estimator as

\[
\hat{d} = \arg \min_{d \in \Theta} R^0(d),
\]

where \( \Theta \) is the space of the admissible values of \( d \) and

\[
R^0(d) = \log \hat{G}(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j, \quad \hat{G}(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^d(x-\hat{\mu})}(\lambda_j),
\]

and \( I_{\Delta^d(x-\hat{\mu})}(\lambda_j) \) is the periodogram of \( X_t - \hat{\mu} \). Define \( v_t = 1_{\{t \geq 1\}} \), then

\[
w_{\Delta^d(x-\hat{\mu})}(\lambda_j) = w_{\Delta^d(x-\mu_0)}(\lambda_j) + (\mu_0 - \hat{\mu})w_{\Delta^d v}(\lambda_j).
\]

The asymptotics of the estimator depend on the relative magnitude of \( w_{\Delta^d(x-\mu_0)}(\lambda_j) \) and \( (\mu_0 - \hat{\mu})w_{\Delta^d v}(\lambda_j) \). The ELW estimator with \( \hat{\mu} = \overline{X} \) is consistent for \( d_0 \in (-\frac{1}{2}, 1) \) and asymptotically normally distributed for \( d_0 \in (-\frac{1}{2}, \frac{3}{4}) \), while the ELW estimator with \( \hat{\mu} = X_1 \) is consistent and asymptotically normally distributed for \( d_0 > 0 \). The following theorems establish these results.

\footnote{Using an arbitrary random variable \( Z \) as \( \hat{\mu} \) results in the same order of the error.}
Assumption 6a  \( \Theta = [\Delta_1, \Delta_2] \) with \(-\frac{1}{2} < \Delta_1 < \Delta_2 < 1\).

**Theorem 1a.** Suppose \(X_t\) is generated by (5) with \(d_0 \in [\Delta_1, \Delta_2]\), Assumptions 1-5 and 6a hold, and \(\hat{\mu} = \bar{X}\). Then \(\hat{d} \to_p d_0\) as \(n \to \infty\).

**Theorem 1b.** Suppose \(X_t\) is generated by (5) with \(d_0 \in (\Delta_1, \Delta_2)\) and \(\Delta_2 \leq \frac{3}{4}\), Assumptions 1’-5’ and 6a hold, and \(\mu = X_{\infty}\). Then \(m^{1/2}(\hat{d} - d_0) \to_d N(0, \frac{1}{4})\) as \(n \to \infty\).

Assumption 6b  \( \Theta = [\Delta_1, \Delta_2] \) with \(0 < \Delta_1 < \Delta_2 < \infty\).

**Theorem 2a.** Suppose \(X_t\) is generated by (5) with \(d_0 \in [\Delta_1, \Delta_2]\), Assumptions 1-5 and 6b hold, \(n^{1-2d_0}m^{-1+\eta} \log m \to 0\) for some \(\eta > 0\), and \(\hat{\mu} = X_1\). Then \(\hat{d} \to_p d_0\) as \(n \to \infty\).

**Theorem 2b.** Suppose \(X_t\) is generated by (5) with \(d_0 \in (\Delta_1, \min\{\Delta_2, 2\})\), Assumptions 1’-5’ and 6b hold, \(n^{1-2d_0}m^{-1/2} \log n \to 0\), and \(\hat{\mu} = X_1\). Then \(m^{1/2}(\hat{d} - d_0) \to_d N(0, \frac{1}{4})\) as \(n \to \infty\).

**Remark 1.** We assume \(\Delta_1 > -\frac{1}{2}\), because the order of \(\bar{X} - \mu_0\) is not given by (6) (indeed, it becomes \(O_p(n^{-1} \log n)\)) if \(d_0 \leq -\frac{1}{2}\). For practical applications this assumption is innocuous because the ELW estimation does not require prior differencing of the data and the cases with \(d_0 < 0\) do not occur in practice. Theorems 1a-2b hold even if \(\mu_0\) is assumed to be an \(O_p(1)\) random variable.

**Remark 2.** The additional assumptions on \(m\) in Theorems 2a and 2b are automatically satisfied when \(d_0 \geq \frac{1}{2}\). When \(d_0 \in (0, \frac{1}{2})\), these conditions require \(m\) to grow fast, and they become stronger for smaller \(d_0\). This phenomenon occurs because, when \(d_0 \in [0, \frac{1}{2})\), both \(X_t\) and \(\hat{\mu} - \mu_0\) are \(O_p(1)\), but the leakage from the dft of \(\hat{\mu} - \mu_0\) has a nonnegligible effect to the behavior of the periodogram ordinates for extremely small \(\lambda_j\)‘s. Trimming the first \(\ell = 5m\) periodogram ordinates for arbitrary small \(\delta > 0\) will relax the condition to \(n^{1-2d_0}m^{2d_0-2} \to 0\) for consistency.

Shimotsu and Phillips (2006) report a similar phenomenon with untapered local Whittle estimation; when the local Whittle estimator is applied to an \(I(d_0)\) process with \(d_0 \in [-\frac{1}{2}, 0)\), the consistency of the estimator requires \(m\) to grow fast. They report that Monte Carlo simulation bias can be as large as 0.25 when \(d_0 = -1\), \(n = 200\), and \(m = 10\). The magnitude of the bias of the ELW estimator for \(d_0 = 0\) in Table 1 is smaller, but the bias does manifest itself in some cases; for example, the bias is 0.148 when \(d_0 = 0, n = 4096\), and \(m = 30\).

Intriguingly, when \(d_0 \in [\frac{3}{4}, 1)\), \(\hat{d}\) with \(\hat{\mu} = \bar{X}\) is still consistent, although \(\bar{X}\) is not a consistent estimate of \(\mu_0\). Table 1 shows the finite sample performance of the above two estimators. We generate the data according to (5) with \(u_t \sim iidN(0, 1)\) and \(\mu_0 = 0\). \(\Delta_1\) and \(\Delta_2\) are set to -1 and 3. Sample size and \(m\) are chosen to be \(n = 256\) and \(m = n^{-0.65} = 36\), and 10,000 replications are used. The ELW estimator with \(\hat{\mu} = \bar{X}\) becomes negatively biased for large \(d_0\), whereas the estimator with \(\hat{\mu} = X_1\) appears to be inconsistent when \(d_0\) is negative. Consequently, the ELW estimator can become inconsistent if the error in estimating \(X_0\) is not controlled properly.
3.2 Feasible ELW estimation

The above results indicate that

1. $\bar{X}$ is an acceptable estimator of $\mu_0$ for small $d_0$;
2. $X_1$ is an acceptable estimator of $\mu_0$ for large $d_0$;
3. for $d_0 \in [\frac{1}{2}, \frac{3}{4}]$, both $\bar{X}$ and $X_1$ are acceptable estimators of $\mu_0$.

Therefore, one promising approach for estimating $d$ consistently for a wide range of $d$ is to estimate $\mu_0$ with a certain combination of $\bar{X}$ and $X_1$. We propose to estimate $\mu_0$ by the following function:

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

where $w(d)$ is a twice continuously differentiable weight function such that $w(d) = 1$ for $d \leq \frac{1}{2}$ and $w(d) = 0$ for $d \geq \frac{3}{4}$. With this estimate of $\mu_0$, we define the feasible ELW (FELW) estimator as

$$\hat{d}_F = \arg\min_{d \in \Theta} R_F(d),$$

where $\Theta$ is the space of the admissible values of $d$ and

$$R_F(d) = \log \hat{G}_F(d) - 2d \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j,$$

$$\hat{G}_F(d) = \frac{1}{m} \sum_{j=1}^{m} I_{\Delta^c(x - \bar{\mu}(d))}(\lambda_j),$$

The FELW estimator is consistent for $d_0 > -\frac{1}{2}$, although we need to exclude a small interval around 0 and 1.

**Assumption 6c** For arbitrary small $\nu > 0$, $\Theta = [\Delta_1, \Delta_2] \setminus ((-\nu, \nu) \cup (1 - \nu, 1 + \nu))$ with $-\frac{1}{2} < \Delta_1 < \Delta_2 < 2$.

**Theorem 3a.** Suppose $X_i$ is generated by (5) with $d_0 \in \Theta$, Assumptions 1-5 and 6c hold. Then $\hat{d}_F \to_p d_0$ as $n \to \infty$.

The exclusion of $(-\nu, \nu) \cup (1 - \nu, 1 + \nu)$ is necessary because of the difficulty in proving the global consistency of the estimator. The consistency is proven by showing $R_F(d) - R_F(d_0)$ is uniformly bounded away from 0 when $d \neq d_0$. When $d$ is close to $d_0$, $R_F(d) - R_F(d_0)$ converges to a non-random function whose minimum is achieved at $d_0$. When $d$ is not close to $d_0$, in particular when $|d - d_0| \geq \frac{1}{2}$, $R_F(d) - R_F(d_0)$ does not converge to a non-random function, and we need an alternate way to bound it away from zero.\(^2\) One of the necessary steps in proving the lower bound is to show, for some $\zeta > 0$,

$$m^{-1} \sum_{j=|\kappa|}^{m} |A_j - B_j|^2 \geq \zeta \left\{ m^{-1} \sum_{j=|\kappa|}^{m} (|A_j|^2 + |B_j|^2) \right\},$$

\(^2\)In the proof of Theorem 3a, we use the fact $d \notin (-\nu, \nu) \cup (1 - \nu, 1 + \nu)$ in showing the necessary results for $\Theta^*_1$, even though $|\theta|$ may be smaller than $\frac{1}{2}$ in $\Theta^*_1$. We can prove the necessary results for $\Theta^*_2$ without using $d \notin (-\nu, \nu) \cup (1 - \nu, 1 + \nu)$, although the derivation is more tedious.
where \( \kappa \) is a fixed number between 0 and 1, \( A_j \) is a function of \( \Delta^d X^0_t \), and \( B_j \) is a function of \( w_{\Delta^d_\ell}(\lambda_j) \). Their explicit formula is given by (38). Note that (10) does not hold if \( A_j = B_j \neq 0 \). For (10) to hold, \( A_j \) and/or \( B_j \) must vary sufficiently as \( j \) changes, so that \( A_j - B_j \) is bounded away from 0 for sufficiently many \( j \)'s. When \( d \) is close to 0, the two leading terms of \( w_{\Delta^d_\ell}(\lambda_j) \), \( (1 - e^{i\lambda_j})^d \) and \( -n^{-d}/\Gamma(1 - d) \) (see Lemma B.2 (a)), are both close to 1, which makes it very hard to establish that \( w_{\Delta^d_\ell}(\lambda_j) \) has sufficient variation. A similar difficulty arises when \( d \) is close to 1.

Shimotsu and Phillips (2005) also needed to use a non-standard approach to show \( \inf_d R(d) - R(d_0) > 0 \) for \( |d - d_0| \geq \frac{1}{2} \) but were able to show it for \( \frac{1}{2} \leq |d - d_0| \leq \frac{9}{2} \). In a way, the presence of \( w_{\Delta^d_\ell}(\lambda_j) \) aggravates the difficulty in showing the global consistency in Shimotsu and Phillips (2005).

The following theorem establishes the asymptotic normality of the feasible ELW estimator. Because \( d_0 < 2 \) for most, if not all, economic data, the FELW estimator is consistent and asymptotically normally distributed for any value of \( d_0 \) encountered in practice.

**Theorem 3b.** Suppose \( X_t \) is generated by (5) with \( d_0 \in \text{Int}(\Theta) \), Assumptions 1′-5′ and 6c hold. Then \( n^{1/2}(\tilde{d}F - d_0) \rightarrow_d N(0, \frac{1}{4}) \) as \( n \rightarrow \infty \).

### 3.3 ELW estimation with unknown mean and time trend

In this subsection, we extend the FELW estimation to cases where the data have a polynomial time trend as well as an unknown mean:

\[
X_t = \mu_0 + \beta_{10} t + \beta_{20} t^2 + \cdots + \beta_{k0} t^k + X^0_t; \quad X^0_t = (1 - L)^{-d_0} u_t 1 \{t \geq 1\}. \tag{11}
\]

We propose to estimate \( d \) by regressing \( X_t \) on \( (1, t, \cdots, t^k) \) and applying the FELW estimation to the residuals \( \tilde{X}_t \). As shown in the proof, the residuals can be expressed as

\[
\tilde{X}_t = X^0_t + \Xi_{0n}(d_0) + \Xi_{1n}(d_0) t + \cdots + \Xi_{kn}(d_0) t^k,
\]

where \( \Xi_{kn}(d_0) \) are random variables. When we apply the feasible ELW estimator to the residuals, the estimate of \( \mu_0 \) takes the form

\[
\varphi(d) = w(d) \overline{\Xi} + (1 - w(d)) \tilde{X}_t.
\]

The following theorem establishes the asymptotics. Now the asymptotic normality requires \( d_0 \) to be smaller than \( \frac{7}{4} \), because the initial condition of \( \tilde{X}_t, \Xi_{0n}(d_0), \) is a random variable whose order of magnitude depends on \( d_0 \). \( \Xi_{1n}(d_0) t, \cdots, \Xi_{kn}(d_0) t^k \) have the same order of magnitude as \( \Xi_{0n}(d_0) \).

**Theorem 4.** Suppose \( X_t \) is generated by (11) and \( \tilde{X}_t - \varphi(d) \) is used in place of \( X_t - \tilde{\mu}(d) \) in defining \( R_F(d) \) in (9). Then, (a) If Assumptions 1-5 and 6c hold and \( d_0 \in \Theta \), then \( \tilde{d}F \to_d d_0 \) as \( n \to \infty \). (b) If Assumptions 1′-5′ and 6c hold, \( d_0 \in \text{Int}(\Theta) \), and \( d_0 < \frac{7}{4} \), then \( n^{1/2}(\tilde{d}F - d_0) \rightarrow_d N(0, \frac{1}{4}) \) as \( n \to \infty \).
4 Two-step estimation

4.1 Two-step feasible ELW estimator

The feasible ELW estimator excludes holes around 0 and 1 from the domain of optimization. Although one can make these holes as small as one desires, they may cause trouble for inference in some cases. In this section, we apply two-step estimation to address this problem. Two-step estimation enables us to circumvent the difficulties in proving the global consistency of the estimator discussed in Section 3.2.3

Two-step estimation has a long history, dating back to the work by Fisher (1925). It has been analyzed by many authors, including LeCam (1956), Pfanzagl (1974), Janssen et al. (1985), and Robinson (1988). In the context of long-memory processes, Lobato (1999) and Lobato and Velasco (2000) use the two-step estimation method to simplify inference and avoid the problems associated with proving the consistency of the considered estimators for certain values of $d$.

Two-step estimation requires a $\sqrt{m}$-consistent first step estimator. We propose to use the tapered local Whittle estimators of Velasco (1999) and Hurvich and Chen (2000) as an initial estimator. The asymptotic theory of these estimators are derived under “Type I” long-range dependent processes that are defined as an infinite order moving average of short-memory innovations for $d \in [-\frac{1}{2}, \frac{1}{2}]$ and as its partial sums for larger values of $d$. We need to extend their theory to the case where $X_t$ is generated by (11) (“Type II” processes) using the results in Robinson (2005). This result may be of interest itself, since the asymptotic properties of these estimators have not been studied under Type II processes in the literature. Phillips and Shimotsu (2004) and Shimotsu and Phillips (2006) analyze the untapered local Whittle estimator under Type II processes.

First we discuss the taper used by Velasco (1999). Let $h_t$ denote a taper of order $p$ generated by Kolmogorov’s proposal. Then $h_t$ satisfies the regularity conditions in Velasco (1999) and Robinson (2005), and the tapered estimator is invariant to a polynomial time trend of order $p - 1$. We do not discuss the properties of the tapers in details here; see Velasco (1999) and Robinson (2005) for further discussion. Define the tapered dft and periodogram as

$$w_T^{xp} (\lambda_j) = (2\pi n)^{-1/2} \sum_{t=1}^n h_t X_t e^{it\lambda_j}, \quad I_T^{xp} (\lambda_j) = |w_T^{xp} (\lambda_j)|^2.$$

As in Velasco (1999, page 99), define the tapered local Whittle estimator as

$$\hat{d}_p = \arg \min_{d \in [\Delta_1, \Delta_2]} R_p(d),$$

where $-\frac{1}{2} < \Delta_1 < \Delta_2 < \infty$, $R_p(d) = \log \hat{G}_p(d) - 2dpm^{-1} \sum_{j=p,2p,\ldots,m} \log \lambda_j$, and $\hat{G}_p(d) = pm^{-1} \sum_{j=p,2p,\ldots,m} \lambda_j^{2d} I_T^{xp} (\lambda_j)$.3

3The idea of two-step estimation was originally suggested by an anonymous reader of Shimotsu and Phillips (2005), albeit in a different context.
The tapered estimator by Hurvich and Chen (2000) takes the difference of the data and applies a complex-valued taper \( h_t^{HC} = 0.5[1 - \exp(2\pi i(t - 1/2)/n)] \) to \( \Delta X_t \). This taper reduces periodogram bias induced by possible overdifferencing of \( \Delta X_t \) found by Hurvich and Ray (1995). This tapered periodogram, \( I_{\Delta x}^{HC}(\lambda_j) = |(2\pi n)^{-1/2} \sum_{t=1}^n h_t^{HC} \Delta X_t|^2 \), may be viewed as an estimator of the spectral density of \( \Delta X_t \) at the frequency \( \lambda_{j+1/2} \). The objective function is defined in terms of \( \Delta X_t \), and the estimator is defined as

\[
\hat{d}_{HC} = \arg \min_{d \in [\Delta_1', \Delta_2']} R_{HC}(d),
\]

where \(-\frac{3}{2} < \Delta_1' < \Delta_2' < \frac{1}{2}\), \( R_{HC}(d) = \log \tilde{G}_{HC}(d) - 2(d - 1)m^{-1} \sum_{j=1}^m \log \lambda_{(j+1/2)} \), and \( \tilde{G}_{HC}(d) = m^{-1} \sum_{j=1}^m \lambda_j^{2(d-1)} I_{\Delta x}^{HC}(\lambda_j) \). Hurvich and Chen (2000) propose to use the powers of \( h_t^{HC} \) as a taper with the higher-order differences of \( X_t \) to allow for larger values of \( d \), albeit at the cost of inflation in variance. To save space, we restrict the range of \( d \) to be \((-\frac{1}{2}, \frac{3}{2})\) and allow only a linear trend with this estimator.

Following Robinson (2005), we need to impose an additional assumption on \( f_u(\lambda) \).

**Assumption 7'** \( f_u(\lambda) \) is bounded for \( \lambda \in [0, \pi] \).

Not allowing \( f_u(\lambda) \) to have poles outside the origin certainly restricts the class of the spectral density. However, it imposes no additional restrictions with respect to the smoothness of \( f_u(\lambda) \) beyond Assumptions 1'-5'. The following propositions establish the limiting distribution of the tapered estimators. \( \widehat{d}_2 \) and \( \hat{d}_{HC} \) are asymptotically normally distributed for \( d_0 \in (-\frac{1}{2}, \frac{3}{2}) \), and \( \widehat{d}_3 \) allows \( d_0 \in (-\frac{1}{2}, \frac{3}{2}) \).

**Proposition 1.** Suppose \( X_t \) is generated by (11) with \( d_0 \in (\Delta_1, \Delta_2) \) and \( \beta_{\mu_0} = \cdots = \beta_{\mu_0} = 0 \). Suppose \( p \geq \max\{[\Delta_2 + \frac{1}{2}] + 1, 2\} \) and Assumption 1'-5' and 7' hold. Then \( m^{1/2}(\hat{d}_p - d_0) \to_d N(0, \Phi / 4) \) as \( n \to \infty \), where \( \Phi \) is defined in equation (10) in Velasco (1999, p. 101).

**Proposition 2.** Suppose \( X_t \) is generated by (11) with \( d_0 \in (\Delta_1', \Delta_2') \) and \( \beta_{\mu_0} = \cdots = \beta_{\mu_0} = 0 \). Suppose Assumption 1'-5' and 7' hold and \( f_u(\lambda) = G_0 + E_{\beta} \lambda^\beta + o(\lambda^\beta) \) with \( \beta \in (1, 2] \) and \( E_{\beta} < \infty \). Then \( m^{1/2}(\hat{d}_{HC} - d_0) \to_d N(0, (1.5)/4) \) as \( n \to \infty \).

\( \Phi \) takes the value of 1.0500 and 1.00354 when \( p = 2 \) and 3, respectively. Thus, \( \hat{d}_{HC} \) has a smaller limiting variance than \( \hat{d}_2 \). In Proposition 2, additional assumptions on \( f_u(\lambda) \) are necessary in order to satisfy Assumption A1 in Hurvich and Chen (2000). Similar assumptions were also imposed in Velasco (1999), but later Lobato and Velasco (2000, p. 415) show that they are unnecessary for the tapers considered in Velasco (1999).

With the \( \sqrt{m} \)-consistency of the tapered estimators in hand, we are now ready to derive the limiting distribution of the two-step estimator. We focus on \( \hat{d}_3 \) as the first stage estimator, because of its weaker assumption on \( f_u(\lambda) \) and the possibility of \( d_0 \geq \frac{3}{2} \). Define the two-step FELW estimator, \( \hat{d}_{F2} \), as

\[
\hat{d}_{F2} = \hat{d}_3 - R_F(\hat{d}_3)'/R_F(\hat{d}_3)''.
\]
where \( R_F(d) \) is the objective function of the FELW estimator defined in (9). Iterating the above procedure and updating the estimator by
\[
\hat{d}^{(2)}_{F_2} = \hat{d}_{F_2} - R_F(\hat{d}_{F_2})',
\]
and similarly for \( \hat{d}^{(3)}_{F_2} \) does not change the asymptotic distribution of the estimator, but we find that iterating procedure can substantially improve its finite sample properties.

**Theorem 5.** (a) Suppose \( X_t \) is generated by (5) with \( d_0 \in (\Delta_1, \Delta_2) \) and \( \Delta_2 < 2 \), and Assumptions 1'-5' and 7 hold. Then \( m^{1/2}(\hat{d}_{F_2} - d_0) \rightarrow_d N(0, \frac{1}{4}) \) as \( n \rightarrow \infty \). (b) Suppose \( X_t \) is generated by (11) with \( d_0 \in (\Delta_1, \Delta_2) \) and \( \Delta_2 \leq \frac{7}{4} \), Assumptions 1'-5' and 7 hold, and \( \hat{X}_t - \varphi(d) \) is used in place of \( X_t - \hat{\mu}(d) \) in defining \( R_F(d) \) in (8). Then \( m^{1/2}(\hat{d}_{F_2} - d_0) \rightarrow_d N(0, \frac{1}{4}) \) as \( n \rightarrow \infty \).

**Proof** From the standard proof of the two-step estimator, the stated result follows if (i) \( R_F(\hat{d})'' \rightarrow_p 4 \) for all \( \hat{d} \) such that \( |\hat{d} - d_0| \leq |\hat{d}_3 - d_0| \) and (ii) \( m^{1/2}R_F(d_0)' \rightarrow_d N(0,4) \). For (i), define \( M = \{ d : |d - d_0| \leq (\log n)^{-1} \} \), then we have \( \Pr(\hat{d}_3 \notin M) \rightarrow 0 \) from Proposition 1 and \( \sup_{d \in M} |R_F(d)'' - 4| \rightarrow_p 0 \) from the proof of Theorem 3a, giving (i). (ii) is shown in the proof of Theorem 3b. Note that Lemma B.3 is used only in showing the global consistency of \( \hat{d} \) in Theorem 3a, and the proof of Theorem 3b does not use Lemma B.3. Therefore, we do not need the assumption \((-\nu, \nu, 1 - \nu, 1 + \nu) \notin \Theta \).

Theorem 5 holds if the Hessian is replaced with 4, because \( R_F(\hat{d}_3)'' \rightarrow_p 4 \). In the simulations reported below, we replaced \( R_F(\hat{d}_3)'' \) with \( \max\{R_F(\hat{d}_3)'' ,2\} \) and found that it improves the finite sample performance of the estimator. The lower bound on \( R_F(\hat{d}_3)'' \) prevents the occurrence of extraordinary large values of \( \hat{d}_{F_2} \).

### 4.2 Feasible ELW estimation under Type I processes

In this subsection, we discuss the effect of the specification of \( I(d) \) processes on the asymptotics of the FELW estimators. Suppose \( Y_t \) is generated by a Type I \( I(d_0) \) process plus an initial condition:

\[
Y_t = Y_t^0 + \mu_0, \quad Y_t^0 = (1 - L)^{-s}U_t^{(s)}1\{t \geq 1\}, \quad U_t^{(s)} = (1 - L)^{-d_0+s}u_t,
\]

where \( u_t \) satisfies Assumptions 1'-3', \( d_0 > -\frac{1}{2} \), and \( s = [d_0 + \frac{1}{2}] \).

Consider the case where \( \mu_0 = 0 \) first. We conjecture that the 2-step FELW estimator has the same asymptotic properties under Type I processes, albeit a rigorous proof is beyond the scope of this paper. First, it is known that Type I and Type II processes with \( |d| < \frac{1}{2} \) are asymptotically equivalent (Marinucci and Robinson, 1999) and that the effect of their difference in their initialization becomes negligible as \( t \rightarrow \infty \). Second, the untapered LW estimator has \( N(0, \frac{1}{4}) \) asymptotic distribution both under Type I (Robinson, 1995b) and Type II (Shimotsu and Phillips, 2006) processes. Therefore, we conjecture that the asymptotic equivalence between these processes will also apply to the asymptotic distribution of the semiparametric estimators.

Note that the FELW estimator uses the periodograms of the \( d \)th difference of the data with truncation at \( t = 0 \). In the following, we show the \( d \)th difference of Type
I and Type II \( I(d_0) \) processes truncated at \( t = 0 \) are asymptotically equivalent for \( d \in [d_0 - \varepsilon, d_0 + \varepsilon] \) and small \( \varepsilon > 0 \). It suffices to consider this range of \( d \) because we use a two-step method. For illustration, focus on the case when \( d_0 \in (-\frac{1}{2}, \frac{1}{2}) \) and \( Y_t = (1 - L)^{-d_0} u_t \). Taking the \( d \)th difference of \( Y_t \) with truncation gives

\[
(1 - L)^d Y_t \mathbf{1}\{t \geq 1\} = \sum_{k=0}^{t-1} \frac{(-d)_k}{k!} Y_{t-k} = \sum_{k=0}^{\infty} \frac{(-d)_k}{k!} Y_{t-k} - \sum_{k=t}^{\infty} \frac{(-d)_k}{k!} Y_{t-k}.
\]

The first term on the right is \( (1 - L)^d (1 - L)^{-d_0} u_t = (1 - L)^{-d_0} u_t \), which is a Type I \( I(d_0 - d) \) process. This is asymptotically equivalent to a Type II \( I(d_0 - d) \) process, which is the \( d \)th difference of a Type II \( I(d_0) \) process truncated at \( t = 0 \) by definition. For the second term, let \( \gamma_k \) denote the \( k \)th autocovariance of \( Y_t \) and assume it satisfies \( \gamma_k = O(k^{2d_0-1}) \) for \( d_0 \neq 0 \) and \( \sum_{k=1}^{\infty} |\gamma_k| < \infty \) for \( d_0 = 0 \). Then, a tedious but routine calculation gives

\[
E \left[ \sum_{k=t}^{\infty} \frac{(-d)_k}{k!} Y_{t-k} \right]^2 = O \left( \sum_{k=t}^{\infty} k^{d-1} \sum_{l=t}^{\infty} l^{d-1} |\gamma_{l-k}| \right) = O \left( t^{-2d-1} + t^{2d_0-2d-1} \right).
\]

Since \( d_0 \in (-\frac{1}{2}, \frac{1}{2}) \) and \( |d - d_0| \leq \varepsilon \), this is \( o(1) \) as \( t \to \infty \), and the asymptotic equivalence of the two \( d \)th differenced processes follows.

When \( \mu_0 \neq 0 \), the FELW estimation estimates \( \mu_0 \) by a linear combination of the sample average and the first observation. Using Type I specification does not affect the asymptotic behavior of the FELW estimator, because Type I and Type II processes have the same stochastic order, and the basic intuition used in Type II specification carries through. Specifically, if we estimate \( \mu_0 \) by the sample average of \( Y_t \), then the estimation error is

\[
n^{-1} \sum_{t=1}^{n} Y_t - \mu_0 = n^{-1} \sum_{t=1}^{n} Y_t^0 = n^{-1} (1 - L)^{-s-1} U_t^{(s)} \mathbf{1}\{t \geq 1\}.
\]

The right hand side is \( n^{-1} \) times a Type I \( I(d_0 + 1) \) process and is \( O_p(n^{-d_0-1/2}) \) under weak regularity conditions; see Marinucci and Robinson (1999) and the references therein. Therefore, the order of the error is the same as in (6). If we estimate \( \mu_0 \) by \( Y_1 \), then \( Y_1 - \mu_0 = Y_1^0 = U_1^{(s)} = O_p(1) \), and the order of the error is the same as in (7). Similarly, the effect of detrending polynomial trends is the same, because the partial sums of Type I and Type II processes have the same stochastic order.

5 Simulations and an empirical application

This section reports some simulation results. \( X_t \) is generated by (5) with \( \mu_0 = 0 \). \( \Delta_1 \) and \( \Delta_2 \) are set to \(-1\) and \(3\). The form of the weight function \( w(d) \) for \( d \in [\frac{1}{2}, \frac{3}{2}] \) is chosen to be \((1/2)(1 + \cos(4\pi d))\). We use 10,000 replications. In two-step estimation, analytic derivatives are used to compute \( R_F(d)' \) and \( R_F(d)'' \). The terms involving \( \partial \tilde{\mu}(d)/\partial d \) are omitted from the derivatives, because they are negligible in the limit. The procedure (12) is iterated (with updating) 10 times.
We compare the ELW estimator and varieties of FELW estimators with another state-of-art semiparametric estimator. Among the existing tapered local Whittle estimators, the version by Hurvich and Chen (2000) discussed in the previous section has the smallest limiting variance, \(1.5/(4m)\) for \(d \in (-\frac{1}{2}, \frac{3}{2})\). We focus on \(d \in (-\frac{1}{2}, \frac{3}{2})\) in most of our simulations, because this is the range of \(d\) that is relevant for many economic applications.

Table 2 compares \(\hat{d}_{F_2}\) with the tapered estimator for value of \(d \in (-\frac{1}{2}, \frac{3}{2})\) and with varying short-run dynamics of \(u_t\). The sample size and \(m\) are chosen to be \(n = 512\) and \(m = n^{0.65} = 57\), and \(u_t\) is modeled as an AR(1) with the parameter \(\rho\). This table corresponds to Table 1 of Hurvich and Chen (2000). The bias of the two estimators is very similar and not affected by the changes in \(d\) for a given value of \(\rho\). For a given value of \(d\), the bias of both estimators increases as \(\rho\) increases. The variance of \(\hat{d}_{F_2}\) is smaller than that of the tapered estimator for any parameter combination, corroborating the theoretical result.

Tables 3 and 4 compare the ELW estimator, two-step FELW estimator with and without linear detrending, and the tapered estimator.\(^4\) The estimation of the mean has little negative effect on the bias and standard deviation of the ELW estimator. Also, the MSE of the ELW estimator and \(\hat{d}_{F_2}\) are virtually the same for \(n = 512\). If the data are detrended prior to estimation, \(\hat{d}_{F_2}\) suffers from a mild increase in standard deviation and a small negative bias for \(d = 0.0 \sim 0.8\). Overall, the finite sample performance of both the feasible ELW estimator and feasible ELW estimator with detrending is very close to that of the ELW estimator except for a few cases. On the other hand, the tapered estimator has substantially larger standard deviations and MSE compared with the ELW estimator for all values of \(d\). In sum, the simulation evidence shows that the feasible ELW estimator’s performance is comparable to the ELW estimator’s.

Table 5 shows the performance of the ELW estimator and the two-step FELW estimator under Type I processes with \(n = 128\) and 512 to examine the conjecture in Section 4.2. When \(n = 128\), the variance of both estimators appears to be slightly larger than their variance under Type II processes reported in Table 3. The results with \(n = 512\) are very similar to the corresponding ones in Tables 4.

As an empirical illustration, the feasible ELW estimator, \(\hat{d}_{F_2}\), with detrending was applied to the historical economic times series considered in Nelson and Plosser (1982) and extended by Schotman and van Dijk (1991). For comparison, we also estimate \(d\) by first taking the difference of the data, estimating \(d - 1\) by the local Whittle estimator, and adding unity to the estimate \(\hat{d} - 1\). This procedure is invariant to the linear trend. For the feasible ELW estimates, 95% asymptotic confidence intervals are constructed by adding and subtracting \(1.96 \times 1/\sqrt{4m}\) to the estimates. Table 6 shows the results based on \(m = n^{0.7}\). The feasible ELW estimate and the local Whittle estimate from the differenced data are fairly close to each other. For real measures such as real GNP, real per capita GNP, and employment, the estimates are close to 1. For price variables such as the GNP deflator, CPI, and nominal wage, the

\(^4\)For the tapered estimator, the results for \(d = 1.6\) is only for reference, because the tapered estimator with taper of order 1 is asymptotically normal only for \(d < \frac{2}{3}\).
estimates are substantially larger than 1. This confirms previous empirical results (Hassler and Wolters, 1995) that inflations are $I(d)$ with $d \in (0, 1)$. Interestingly, the null of trend stationarity $H_0 : d = 0$ is accepted in none of the series. Crato and Rothman (1994) obtained a similar result using the ARFIMA model, therefore it appears that the case for trend stationarity is weaker than has been suggested from the KPSS test by Kwiatkowski et al. (1992).

6 Conclusion

By tailoring the ELW estimator developed by Shimotsu and Phillips (2005) to accommodate an unknown mean/initial condition and a polynomial time trend, this paper develops a general purpose tool for estimation and inference of the memory parameter of typical economic time series. The new estimator, the feasible ELW estimator, covers a range of values of $d$ that are commonly encountered with economic data and makes it possible to construct valid confidence intervals in a standard and simple way. Both in asymptotics and in small samples, the feasible ELW estimator inherits the desirable properties of the ELW estimator.

The restrictions on $d$ ($d < \frac{7}{4}$ for asymptotic normality and small intervals around 0 and 1) are somewhat bothersome. However, other semiparametric estimators are also liable to restrictions, and this estimator covers a wider range of $d$ with the smallest variance for the same $m$. Two-step estimation removes the exclusion of these intervals, at the cost of a stronger assumption on $f_u(\lambda)$, in particular, the global boundedness.
Appendix A: proofs

In this and the following section, \(x^*\) denotes the complex conjugate of \(x\). \(C\) and \(\varepsilon\) denote generic constants such that \(C \in (1, \infty)\) and \(\varepsilon \in (0, 1)\) unless specified otherwise, and they may take different values in different places. Henceforth, let \(I_{\Delta xj}\) denote \(I_{\Delta x}(\lambda_j)\), \(w_{uj}\) denote \(w_u(\lambda_j)\), and similarly for other dft’s and periodograms.

A.1 Proof of Theorem 1a

Assume \(\mu_0 = 0\) without loss of generality. We follow the approach developed by Shimotsu and Phillips (2005), hereafter simply SP. Define \(S^\circ(d) = R^\circ(d) - R^\circ(d_0)\). For arbitrary small \(0 < \Delta < \frac{1}{8}\), define \(\Theta_1 = \{d_0 - \frac{1}{2} + \Delta \leq d \leq d_0 + \frac{1}{2}\}\) and \(\Theta_2 = \{d \in [\Delta_1, d_0 - \frac{1}{2} + \Delta] \cup [d_0 + \frac{1}{2}, \Delta_2]\}\), \(\Theta_2\) being possibly empty. For \(\frac{1}{2} > \rho > 0\), define \(N_\rho = \{d : |d - d_0| < \rho\}\). From SP pages 1900-01, we have

\[
\Pr\left(\left|d - d_0\right| \geq \rho\right) \leq \Pr\left(\inf_{d \in \Theta_1 \setminus N_\rho} S^\circ(d) \leq 0\right) + \Pr\left(\inf_{\Theta_2} S^\circ(d) \leq 0\right). \tag{14}
\]

As in SP (page 1902, between equations (13) and (14)), define \(\theta = d - d_0\) and

\[Y_t(\theta) = (1 - L)^d X_t = (1 - L)^{d-d_0} (1 - L)^{d_0} X_t = (1 - L)^\theta u_t 1\{t \geq 1\}.
\]

Note that \(R^\circ(d)\) is constructed by replacing \(I_{\Delta xj}\) in the objective function of SP, \(R(d)\), with \(I_{\Delta x(x-\bar{\mu})}\). Since \(w_{\Delta x(x-\bar{\mu})} = w_{\Delta xj} - \bar{\mu} w_{\Delta vj} = w_{yj} - \bar{\mu} w_{\Delta vj}\), the theorem is proven by replacing \(w_{yj}\) in SP with \(w_{yj} - \bar{\mu} w_{\Delta vj}\) and showing the results in SP carry through. We only state the main steps and refer the readers to SP for further details.

As in equation (15) of SP, define \(A(d) = (2(d-d_0)+1)m^{-1} \sum_{j=1}^{m} (j/m)^{2\theta} |\lambda_j|^{-2\theta} I_{yj} - G_0\). In order to show the first probability on the right of (14) tends to 0, we need to replace \(I_{yj}\) in \(A(d)\) with \(|w_{yj} - \tilde{\mu} w_{\Delta vj}|^2\) and show \(\sup_{\Theta_1} |A(d)| \to 0\) still holds. Because

\[
I_{yj} - |w_{yj} - \tilde{\mu} w_{\Delta vj}|^2 = 2\bar{\mu}\Re\left[ w_{yj} w_{\Delta vj}^* \right] - \tilde{\mu}^2 I_{\Delta vj}, \tag{15}
\]

it suffices to show

\[
\sup_{\theta \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^{m} \left( \frac{j}{m} \right)^{2\theta} \lambda_j^{-2\theta} w_{yj} w_{\Delta vj}^* \right| = O_p\left( m^{(d_0-1)/2} \log m + m^{-\Delta} \log m \right), \tag{16}
\]

\[
\sup_{\theta \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^{m} \left( \frac{j}{m} \right)^{2\theta} \lambda_j^{-2\theta} I_{\Delta vj} \right| = O_p\left( m^{d_0-1} \log m + m^{-2\Delta} \log m \right). \tag{17}
\]

We proceed to derive the order of \(\tilde{\mu} w_{\Delta vj}\) and show (16) and (17). Since \(w_{\Delta vj} = O(j^{d-1} n^{1/2-d})\) for \(d \geq 0\) and \(O(j^{1/2} n^{1/2-d})\) for \(d \leq 0\) from Lemma B.2, we obtain

\[
\lambda_j^{-\theta} w_{\Delta vj} = \begin{cases} O\left( n^{1/2-d_0 j^{-d_0-1}} \right), & d \in [0, \Delta_2], \\
O\left( n^{1/2-d_0 j^{-d_0-1}} \right), & d \in [-1 + \varepsilon, 0], 
\end{cases} \tag{18}
\]
uniformly in $d$ and $j = 1, \ldots, m$. Observe that
\[
\hat{\mu} = n^{-1} \sum_{t=1}^{n} X_t = n^{-1} (1 - L)^{-d_0-1} u_t \mathbf{1}\{t \geq 1\}
\]
with $d_0 > \frac{1}{2}$. We can show $E[(1 - L)^{-d_0-1} u_t \mathbf{1}\{t \geq 1\}]^2 = O(n^{2d_0+1})$ easily from Lemma A.5 (a2) of Phillips and Shimotsu (2004), and it follows that $E\hat{\mu}^2 = O(n^{2d_0-1})$ and
\[
\hat{\mu} \cdot \lambda_j^\theta w_{\Delta^4 v j} = \begin{cases}
\xi_n \cdot O\left(\frac{J^{d_0-1}}{2^n}\right), & d \geq 0, \\
\xi_n \cdot O\left(\frac{J^{-\theta-1}}{2^n}\right), & d \leq 0,
\end{cases}
E|\xi_n| < \infty,
\]
where $O(\cdot)$ terms are uniform in $d$ and in $j = 1, \ldots, m$. We also have, uniformly in $\alpha \in [-C, C]$, (note that $\Theta_1 = \{-\frac{1}{2} + \Delta \leq \theta \leq \frac{1}{2}\}$)
\[
\sup_{\theta \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^{m} \left( \frac{j}{m} \right)^{2\theta} j^{\alpha} \right| \leq m^{\alpha} \sum_{j=1}^{m} \left( \frac{j}{m} \right)^{2\Delta - 1 + \alpha} = O(m^\alpha \log m + m^{-2\Delta} \log m),
\]
where the order of magnitude follows from considering the cases where $2\Delta - 1 + \alpha \geq -1$ and $2\Delta - 1 + \alpha \leq -1$ separately. Therefore, (17) follows from (19), (20), and the fact that $d_0 < 1$ and $|\theta| \leq \frac{1}{2}$ in $\Theta_1$. For (16), its left hand side is bounded by
\[
\sup_{\theta \in \Theta_1} \left| \frac{1}{m} \sum_{j=1}^{m} \left( \frac{j}{m} \right)^{2\theta} \left| w_{yjj} - \hat{\mu} w_{\Delta^4 v j} \right| \right| \leq \frac{1}{m} \sum_{j=1}^{m} \left( \frac{j}{m} \right)^{2\Delta - 1 + \alpha} = O(m^\alpha \log m + m^{-2\Delta} \log m),
\]
where the order of magnitude follows from considering the cases where $2\Delta - 1 + \alpha \geq -1$ and $2\Delta - 1 + \alpha \leq -1$ separately. Moreover, (20) follows from (17).

We now show that the second probability on the right of (14) tends to 0. As in SP, let $\kappa \in (0, 1)$ and let $\sum_{j=1}^{m} j$ denote the sum over $j = [\kappa m], \ldots, m$. From the argument on pages 1904–05 of SP that leads to their equation (23), the second probability on the right of (14) tends to zero if there exists $\delta > 0$ such that
\[
\Pr \left( \inf_{\Theta_2} \left( m^{-1} \sum_{j=1}^{m} \left( \frac{j}{p} \right)^{2\theta} \left| w_{yjj} - \hat{\mu} w_{\Delta^4 v j} \right|^2 - G_0 \right) \leq -3\delta G_0 \right) \to 0,
\]
where $p = \exp(m^{-1} \sum_{j=1}^{m} \log j) \sim m/e$ as $m \to \infty$.

We show (21) for subsets of $\Theta_2$. Define $\eta = 1 - d_0 > 0$ and split $\Theta_2$ into two, $\Theta_2^a = \{\theta \geq -1 + \eta/2\} \cap \Theta_2$ and $\Theta_2^b = \{\theta \leq -1 + \eta/2\} \cap \Theta_2$. First, SP show (equation (23) on page 1905) that (21) holds if $|w_{yjj} - \hat{\mu} w_{\Delta^4 v j}|^2$ is replaced by $I_y(\lambda_j)$. Second, if $\theta \in \Theta_2^a$ or $d > 0$, we have
\[
\sup_{\Theta_2} \left| \hat{\mu} m^{-1} \sum_{j=1}^{m} \left( \frac{j}{p} \right)^{2\theta} \lambda_j^{-2\theta} w_{yjj} w_{\Delta^4^4 v j} \right| \leq \sup_{\Theta_2} \left| \hat{\mu} m^{-1} \sum_{j=1}^{m} \left( \frac{j}{p} \right)^{2\theta} \lambda_j^{-2\theta} I_{\Delta^4 v j} \right| = o_p(1),
\]
from using the bound in (19) with $d_0 \leq 1 - \eta$ and $-\theta - 1 \leq -\eta/2$ and proceeding as in the proof of (16) and (17) with Lemma 5.4 of SP. Thus, (21) holds for $\theta \in \Theta_2^a$ or $d > 0$.

If $\theta \in \Theta_2^b$ and $d < 0$, we cannot use (19) because $-\theta - 1$ may take a positive value and its left hand side is not $o_p(1)$. Note that $|\theta| = |d - d_0| \leq \frac{1}{2}$ because $d, d_0 \in (\frac{1}{2}, 1)$. Define $\Theta_2^b = \{-\frac{3}{2} \leq \theta \leq -\frac{1}{2}\}$ as in SP page 1909, then $\Theta_2^b$ is a subset of $\Theta_2^b$. We show the required result by replacing $\lambda_j^{-\theta} w_{yjj}$ in the corresponding proof for $\Theta_2^b$ in SP (equation (45) on page 1909) with $\lambda_j^{-\theta} w_{yjj} - \hat{\mu} w_{\Delta^4 v j}$ and showing their argument
carries through. Replacing $\lambda_j^\theta (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta)$ on the right of (45) of SP with

$$\lambda_j^\theta (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta) - \lambda_j^\theta \hat{\mu} w_{\Delta v_j},$$

we find that (47) in SP needs to be replaced with

$$m^{-1} \sum' (j/p)^{2\theta} \lambda_j^{-2\theta} (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta) - \hat{\mu} w_{\Delta v_j}|^2,$$

and their equations (49) and (50) have additional terms

$$+2Re[m^{-1} \sum' (j/p)^{2\theta} U_{nj}(\theta) \lambda_j^{-\theta} \hat{\mu} w_{\Delta v_j}],$$

$$-2Re[m^{-1} \sum' (j/p)^{2\theta} D_{nj}(\theta)^* w_{uj} \lambda_j^{-\theta} \hat{\mu} w_{\Delta v_j}],$$

where $D_{nj}(\theta)$ and $U_{nj}(\theta)$ are defined on page 1909 of SP. Then, in view of the bounds of (48)-(50) in SP provided in page 1910 of SP, (21) holds by Lemma B.1 if

$$(23) \geq \zeta m^{-2\theta-2} n^{2\theta+1} Y_n(\theta)^2 + \zeta n^{1-2d+2\theta} m^{-2-2\theta} \hat{\mu}^2,$$

for some $\zeta > 0$ and

$$(24) + (25) = n^{1/2-d+\theta} m^{-1-\theta} \hat{\mu} \cdot O_p(m^{-\eta/2} \log n + mn^{-1}).$$

Loosely speaking, if (26) and (27) hold, then (24) and (25) are dominated by (23). It remains to show (26) and (27). Note that $d \leq -\nu$. For (26), applying Lemma B.3 (b) with $Q_2 = Y_n(\theta)$, $Q_1 = 0$, and $Q_0 = -\hat{\mu}$ gives

$$(23) = p^{-2\theta} (2\pi)^{-2\theta} n^{2\theta} m^{-1} \sum' |(2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} Y_n(\theta) - \hat{\mu} w_{\Delta v_j}|^2
\geq \zeta m^{-2\theta} n^{2\theta} [nm^{-2} Y_n(\theta)^2 + n^{-1/2} m^{-2} \hat{\mu}^2],$$

giving (26). (27) follows from applying Lemma B.4 with $\alpha = d$ to (24) and (25), because $U_{nj}(\theta)$ and $D_{nj}(\theta)$ satisfy the assumptions of Lemma B.4 from equation (39) and (31) of SP, respectively. Thus (21) holds, and we complete the proof. □

**A.2 Proof of Theorem 1b**

Assume $\mu_0 = 0$ without loss of generality. Theorem 1a holds under the current conditions and implies that with probability approaching 1, as $n \rightarrow \infty$, $\tilde{d}$ satisfies

$$0 = R^\circ (\tilde{d})' = R^\circ (d_0)' + R^\circ (\tilde{d})''(\tilde{d} - d_0),$$

where $|\tilde{d} - d_0| \leq |\tilde{d} - d_0|$. Again the theorem is proven by replacing $w_{uj}$ in SP with $w_{uj} - \hat{\mu} w_{\Delta v_j}$ and showing the results in SP carry through. Fix $\rho > 0$ and let $M = \{d : (\log n)^4|d - d_0| < \rho\}$. Note that $\sup_{\theta_1} |A(d)| = o_p((\log n)^{-10})$ still holds even if we replace $I_j(\lambda_j)$ in $A(d)$ with $|w_{uj} - \hat{\mu} w_{\Delta v_j}|^2$, because the order of the additional terms shown in (16) and (17) are smaller than $(\log n)^{-10}$. Therefore, $\Pr(\tilde{d} \notin M)$ tends to zero in view of equation (55) of SP and the argument surrounding it. Thus we assume $d \in M$ in the following.
First we show $R^\varnothing(d)'\to p 4$. Define $\hat{G}(d) = m^{-1} \sum_{j=1}^m I_{\Delta_{d}v_j} = m^{-1} \sum_{j=1}^m I_{y_j}$ as in SP, and define

$$a_n(d) = \frac{1}{m} \sum_{j=1}^m \left\{ -2\tilde{\mu} \text{Re} \left[ w_{uj} w_{\Delta_{d}v_j}^* \right] + \tilde{\mu}^2 I_{\Delta_{d}v_j} \right\},$$

(29)

so that $G^\varnothing(d) = \hat{G}(d) + a_n(d)$. Then $\hat{G}_0(d), \hat{G}_1(d)$ and $\hat{G}_2(d)$ defined in page 1913 of SP have additional terms $(2\pi/n)^{-2\theta} a_n(d), (2\pi/n)^{-2\theta} \partial a_n(d)/\partial d$, and $(2\pi/n)^{-2\theta} \partial^2 a_n(d)/\partial d^2$, respectively. In view of the results in SP pages 1915–16 leading to their equation (60), $R^\varnothing(d)'\to p 4$ holds if we show these three terms are all $o_p((\log n)^{-2})$ uniformly in $d \in M$. First, $\sup_M |(2\pi/n)^{-2\theta} a_n(d)| = o_p((\log n)^{-10})$ follows from (16), (17), and $\sup_{\theta \in M} m^{2|\theta|} < \infty$. For $(2\pi/n)^{-2\theta} \partial a_n(d)/\partial d$, note that

$$\frac{\partial}{\partial d} a_n(d) = \frac{1}{m} \sum_{j=1}^m \left\{ -2\tilde{\mu} \frac{\partial}{\partial d} \text{Re} \left[ w_{uj} w_{\Delta_{d}v_j}^* \right] + \tilde{\mu}^2 \frac{\partial}{\partial d} I_{\Delta_{d}v_j} \right\}.$$

From Lemma B.2 (a) and (b), the order of $\partial w_{\Delta_{d}v_j}/\partial d = -w_{\log((1-L)\Delta_{d}v_j)}$ is no larger than $\log n$ times the order of $w_{\Delta_{d}v_j}$. Furthermore, from Lemma 5.9 (a) of SP, the order of $\partial w_{y_j}/\partial d$ is no larger than $(\log n)^2$ times the order of $w_{y_j}$. Therefore, the order of $(2\pi/n)^{-2\theta} \partial a_n(d)/\partial d$ is no larger than $(\log n)^2$ times that of $(2\pi/n)^{-2\theta} a_n(d)$. Similarly, the order of $(2\pi/n)^{-2\theta} \partial^2 a_n(d)/\partial d^2$ is no larger than $(\log n)^4$ times that of $(2\pi/n)^{-2\theta} a_n(d)$ in view of Lemma 5.9 (c) of SP and Lemma B.2 (c). Therefore, the three additional terms are all $o_p((\log n)^{-2})$ uniformly in $d \in M$, and we establish $R^\varnothing(d)'\to p 4$.

The proof is completed by showing $m^{1/2} R^\varnothing(d_0)' \to_d N(0, 4)$. Since $G^\varnothing(d) = \hat{G}(d) + a_n(d)$, we have

$$m^{1/2} R^\varnothing(d_0)' = m^{1/2} \left[ \frac{\partial G^\varnothing(d)/\partial d|_{d_0}}{G^\varnothing(d_0)} - 2 \frac{1}{m} \sum_{j=1}^m \log \lambda_j \right],$$

$$= m^{1/2} \left[ \frac{\partial G(d)/\partial d|_{d_0} + \partial a_n(d)/\partial d|_{d_0}}{G(d_0) + a_n(d_0)} - 2 \frac{1}{m} \sum_{j=1}^m \log \lambda_j \right].$$

Because SP shows $\hat{G}(d_0) \to_p G_0, m^{1/2} \{\partial G(d)/\partial d|_{d_0}/\hat{G}(d_0) - 2m^{-1} \sum_{j=1}^m \log \lambda_j \} \to_d N(0, 4)$, and $m^{-1} \sum_{j=1}^m \log \lambda_j = O(\log n)$, the required result follows if

$$a_n(d_0) = o_p(m^{-1/2}(\log n)^{-1}), \quad \partial a_n(d)/\partial d|_{d_0} = o_p(m^{-1/2}).$$

Note that $a_n(d_0) = m^{-1} \sum_{j=1}^m \{ -2\tilde{\mu} \text{Re}[w_{uj} w_{\Delta_{d_0}v_j}^*] + \tilde{\mu}^2 I_{\Delta_{d_0}v_j} \}$. Using $w_{uj} = C(e^{i\lambda_j}) w_{\xi_j} + r_{nj}$ with $E|r_{nj}|^2 = O(j^{-1} \log n)$ uniformly in $j = 1, \ldots, m$ (Robinson, 1995b), and
the order of $\hat{\mu}w_{\Delta i}w_{\Delta j}$ given in (19) with $\theta = 0$, we have

$$\left| \frac{a_n(d_0)}{2} \right| \leq \left| \frac{1}{m} \sum_{j=1}^{m} C(e^{i\lambda_j}) w_{\Delta i}w_{\Delta j}^* \right| + \left| \frac{1}{m} \sum_{j=1}^{m} r_{nj}w_{\Delta i}w_{\Delta j}^* \right| + \hat{\mu}_n \left| \frac{1}{m} \sum_{j=1}^{m} I_{\Delta i}w_{\Delta j} \right|$$

$$= O_p \left( \left( \frac{1}{m^2} \sum_{j=1}^{m} (j^{2d_0-2} + j^{-2}) \right)^{1/2} \right) + O_p \left( \frac{1}{m} \sum_{j=1}^{m} (j^{d_0-3/2} + j^{-3/2}) \right) + O_p \left( \frac{1}{m} \sum_{j=1}^{m} (j^{2d_0-2} + j^{-2}) \right) = O_p((m^{d_0-3/2} + m^{-1} + m^{2d_0-2}) \log m).$$

This is $o_p(m^{-1/2}(\log n)^{-1})$ because $d_0 < \frac{3}{4}$. For $\partial a_n(d)/\partial d|_{d_0}$, we have

$$\frac{\partial}{\partial d} a_n(d) \big|_{d_0} = 2\hat{\mu}_n \left[ \sum_{j=1}^{m} \text{Re} \left[ w_{\log(1-L)uj}w_{\Delta i}w_{\Delta j}^* \right] \right]$$

$$+ 2\hat{\mu}_n \left[ \sum_{j=1}^{m} \text{Re} \left[ w_{uj}w_{\log(1-L)\Delta i}w_{\Delta j}^* \right] \right] + \hat{\mu}_n^2 \left[ \sum_{j=1}^{m} \partial I_{\Delta i}w_{\Delta j} \right] \big|_{d_0}. $$

It follows easily from Lemma B.2 that the order of $w_{\log(1-L)\Delta i}w_{\Delta j}^*$ and $\partial I_{\Delta i}w_{\Delta j}/\partial d|_{d_0}$ are log $n$ times the order of $w_{\Delta i}w_{\Delta j}^*$ and $I_{\Delta i}w_{\Delta j}$, respectively. Therefore, the second and third terms on the right are $o_p(m^{-1/2})$ in view of the order of $a_n(d_0)$. For the first term on the right, SP Lemma 5.9 (a) shows that $w_{\log(1-L)uj} = -J(e^{i\lambda_j})w_{uj} + R_{uj}$ with $J(e^{i\lambda_j}) = O(\log n)$ and $E|R_{uj}|^2 = O(j^{-1}(\log n)^4)$ uniformly in $j = 1, \ldots, m$. Therefore, it follows from a similar argument as above that the first term on the right is $o_p(m^{-1/2})$, thus $\partial a_n(d)/\partial d|_{d_0} = o_p(m^{-1/2})$ and we complete the proof. □

### A.3 Proof of Theorems 2a and 2b

From (18) and the fact that $d \geq 0$, we have $\lambda_j^-\theta w_{\Delta i}w_{\Delta j} = O(n^{1/2-d_0}\cdot j^{-d_0-1})$. Combining it with $E\hat{\mu}^2 = E|u_j|^2 < \infty$, we have, in place of (19),

$$\hat{\mu} \cdot \lambda_j^-\theta w_{\Delta i}w_{\Delta j} = \xi_n \cdot O(n^{1/2-d_0}\cdot j^{-d_0-1}), \quad E|\xi_n| < \infty, \quad (30)$$

uniformly in $d$. If $d_0 \geq \frac{1}{2}$, then $\hat{\mu}\lambda_j^-\theta w_{\Delta i}w_{\Delta j} = \xi_n \cdot O((j/n)^{d_0-1/2}j^{-d_0-1/2}) = \xi_n \cdot O(j^{-1/2})$, whose order is no larger than that of $\lambda_j^-\theta \cdot (2\pi n)^{-1/2} \tilde{U}_{\theta} \cdot \hat{\lambda}_n(\theta)$ in equation between (20) and (21) of SP and that of $\tilde{U}_{\theta} \cdot \hat{\lambda}_n(\theta)$ in (30) and (39) of SP. Therefore, if we replace $w_{uj}$ in SP with $w_{yj} - \hat{\mu}\lambda w_{\Delta i}w_{\Delta j}$, the proof of the consistency of SP carries through. For the asymptotic normality for $d_0 \geq \frac{1}{2}$, we can use the proof of Theorem 2b without changes, because $O(j^{-1/2})$ is no larger than the maximum of the right hand side of (19).

To show the consistency for $d_0 \in (0, \frac{1}{2})$, we need to modify the proof of Theorem 1a. Split $\Theta_1$ into two, $\Theta_1^\eta = \Theta_1 \cap \{ \theta : |\theta| \leq \eta \}$ and $\Theta_1^\eta = \Theta_1 \setminus \Theta_1^\eta$, where $\eta$ is the
constant specified in the statement of the theorem. Then the consistency of $\hat{d}$ follows if we show
\[
Pr \left( \inf_{\Theta_1} S^c(d) \leq 0 \right) + Pr \left( \inf_{\Theta_1 \cup \Theta_2} S^c(d) \leq 0 \right) \to 0, \quad as \ n \to \infty. \tag{31}
\]
For the set $\Theta_1^0$, we can strengthen the bound in (20) to
\[
\sup_{\theta \in \Theta_1^0} |m^{-1} \sum_{j=1}^{m} (j/m)^{2\theta} | = O(m^{\alpha} \log m + m^{-1+2\alpha} \log m), \tag{32}
\]
uniformly in $\alpha \in [-C,C]$. Then, it follows from (30), (32) and $d_0 < \frac{1}{2}$ that
\[
\sup_{\theta \in \Theta_1^0} \left| \mu^2 m^{-1} \sum_{j=1}^{m} \frac{j}{m} \lambda_j^{-2\theta} I_{\Delta^2_{yj}} \right| = O_p(n^{1-2d_0}m^{-1+2\eta} \log m). \tag{33}
\]
Therefore, the first probability in (31) tends to zero by applying the argument of the proof of Theorem 1a for $\Theta_1$.

The second probability of (31) tends to zero if there exists $\delta > 0$ such that
\[
Pr \left( \inf_{\Theta_1 \cup \Theta_2} \left( m^{-1} \sum' \frac{j}{p} \left( \lambda_j^{-2\theta} w_{yj} - \hat{\mu} w_{\Delta^2_{yj}} \right)^2 - G_0 \right) \right) \to -3\delta G_0 \to 0. \tag{34}
\]
This is because the algebra on pages 1904–05 of SP leading to (22) remains unchanged even if $\Theta_2$ is replaced with $\Theta_1^0 \cup \Theta_2$ and we can replace the equation between (22) and (23) in SP with $\inf_{\Theta_1^0 \cup \Theta_2} G_0(m^{-1} \sum' \frac{j}{p} \lambda^{-2\theta} - 1) \geq 4\delta G_0$ using Lemma B.5.

We proceed to show (34) for subsets of $\Theta_1^0 \cup \Theta_2$. First, note that, it follows from (30) and Lemma 5.4 of SP that
\[
\sup_{\Theta_1^0 \cup \Theta_2} \left| \mu^2 m^{-1} \sum' \frac{j}{p} \lambda_j^{-2\theta} I_{\Delta^2_{yj}} \right| = O_p(n^{1-2d_0}m^{2d_0-2}) = O_p(m^{-2\eta}). \tag{35}
\]
Consequently, we can show (34) holds for $\Theta_2$ by applying the proof of Theorem 1a for $\Theta_2$.

It remains to show (34) for $\Theta_1^0$. Write
\[
m^{-1} \sum' \left( \frac{j}{p} \lambda_j^{-2\theta} w_{yj} - \hat{\mu} w_{\Delta^2_{yj}} \right)^2 - G_0 = L_{1n}(d) + L_{2n}(d) + L_{3n}(d), \tag{36}
\]
where $L_{1n}(d) = m^{-1} \sum' \left( \frac{j}{p} \lambda_j^{-2\theta} w_{yj} - G_0 \right)$, $L_{2n}(d) = -2\hat{\mu} m^{-1} \sum' \left( \frac{j}{p} \lambda_j^{-2\theta} w_{yj} \right)$, and $L_{3n}(d) = \hat{\mu} m^{-1} \sum' \left( \frac{j}{p} \lambda_j^{-2\theta} \right)$. For $L_{1n}(d)$, we can apply the argument from line 7, page 1905 of SP without change to conclude $\sup_{\Theta_1^0} |L_{1n}(d)| = o_p(1)$. We have $\sup_{\Theta_1^0} |L_{3n}(d)| = o_p(1)$ from (35), and the bound of $L_{2n}(d)$ follows from the bound of $L_{1n}(d)$, $L_{3n}(d)$ and Cauchy-Schwartz inequality. This completes the proof of consistency for $d_0 \in (0, \frac{1}{2})$.

Proof of the asymptotic normality for $d_0 \in (0, \frac{1}{2})$ follows the proof of Theorem 1b. We use the bound (33) in place of (17) to show the $R^c(d)^{\prime} \to_d N(0,4)$. To show $m^{1/2} R^c(d_0)^{\prime} \to_d N(0,4)$, we simply repeat the proof of Theorem 1b with replacing (19) with (30), then the stated result follows by $n^{1-2d_0}m^{-1} = o(m^{-1/2})(\log n)^{-1}$. \(\square\)
A.4 Proof of Theorem 3a

Take \( \nu \) to be smaller than \( 2 - \Delta_2 > 0 \) without the loss of generality. We need to treat the cases for different values of \( d_0 \) and \( d \) separately. When \( d_0 \in \left[ \frac{1}{2}, 1 \right) \), the required result follows from the proof of Theorems 1a and 2a, because \( \tilde{d} \) is consistent both under \( \hat{\mu} = X_1 \) and \( \hat{\mu} = X \). When \( d_0 < \frac{1}{2} \) and \( d \in [\Delta_1, \frac{1}{2}] \), the proof of Theorem 1a applies because \( \hat{\mu}(d) = X \). When \( d_0 \geq 1 \) and \( d \in [\frac{3}{4}, \Delta_2] \), the proof of Theorem 2a applies because \( \hat{\mu}(d) = X_1 \). It leaves us with the consideration of the two cases:

(i) \( d_0 < \frac{1}{2} \) and \( d \in \left[ \frac{1}{2}, \Delta_2 \right] \), (ii) \( d_0 \geq 1 \) and \( d \in [\Delta_1, \frac{3}{4}] \). \hspace{1cm} (37)

Note that (i) implies \( \theta = d - d_0 \geq \frac{1}{2} - d_0 > 0 \) and (ii) implies \( \theta \leq \frac{3}{4} - 1 \leq -\frac{1}{4} \). With a slight abuse of notation, define \( \eta = \min\{\frac{1}{2} - d_0, \frac{1}{4}\} > 0 \) and define \( \Theta_1^\delta = \Theta_1 \cap \{\theta > \eta\} \) as in the proof of Theorem 2a for \( d_0 < \frac{1}{2} \). Because \( \theta \in \Theta_1^\delta \cup \Theta_2 \) if (i) or (ii) is true, \( \Pr(\inf_{\Theta_1^\delta \cup \Theta_2} S_F(d) \leq 0) \rightarrow 0 \) suffices for the consistency of \( \hat{d}_F \).

Consider \( \Theta_1^\delta \) first. We show \( \Pr(\inf_{\Theta_1^\delta} S_F(d) \leq 0) \rightarrow 0 \) by using the proof of Theorem 2a for \( \Theta_1^\delta \) and showing that (34) holds for \( \Theta_1^\delta \) if \( |w_{ij} - \tilde{\mu}w_{\Delta_i^d v_j}|^2 \) in (34) is replaced by \( |w_{ij} - \hat{\mu}(d)w_{\Delta_i^d v_j}|^2 \). We obtain a decomposition similar to (36) with \( \hat{\mu}(d) \) replacing \( \tilde{\mu} \) and \( \sup_{\Theta_1^\delta} |L_{3n}(d)| = o_p(1) \). For \( L_{2n}(d) \), it follows from equation (14) of SP, Lemma 5.2 (b) of SP, and the equation between (20) and (21) of SP that \( \lambda_j^\theta w_{ij} = D_{nj}(\theta)w_{ij} + U_{nj}(\theta) \), where \( D_{nj}(\theta) \) and \( U_{nj}(\theta) \) satisfy the assumptions of Lemma B.4. Therefore, applying Lemma B.4 with \( \alpha = d \) gives

\[
L_{2n}(d) = \hat{\mu}(d) \cdot \left\{ \begin{array}{ll}
 n^{1/2-d_0}m^{d_0-1} \cdot O_p(n^{-1}m^{1-\nu} \log n), & d \geq \nu, \\
 n^{1/2-d_0}m^{-\theta-1} \cdot O_p(n^{-1}m^{\log \log n}), & d \leq \nu.
\end{array} \right.
\]

Define \( D^- = [\Delta_1, -\nu] \) and \( D^+ = [\nu, 1 - \nu] \cup [1 + \nu, \Delta_2] \), so that \( \Theta \subset D^- \cup D^+ \). Applying Lemma B.3 (a) to \( L_{3n}(d) \) with \( Q_2 = Q_1 = 0 \) and \( Q_0 = \hat{\mu}(d) \), we find \( L_{3n}(d) \) is bounded from below by, for some \( \eta > 0 \),

\[
\eta \hat{\mu}(d)^2 n^{1-2d_0}m^{2d_0-2} \text{ for } d \in D^+ , \quad \eta \hat{\mu}(d)^2 n^{1-2d_0}m^{-\theta-2} \text{ for } d \in D^- .
\]

Hence, \( \Pr(\inf_{\Theta_1^\delta} [L_{2n}(d) + L_{3n}(d)] \leq \xi) \rightarrow 0 \) for any \( \xi > 0 \) from Lemma B.1, and (34) follows.

For \( \Theta_2^\delta = \left\{ -\frac{3}{2} \leq \theta \leq -\frac{1}{2} \right\} \), we show \( \Pr(\inf_{\Theta_2^\delta} S_F(d) \leq 0) \rightarrow 0 \) by using the argument of the proof of Theorem 1a in pages 17–18 and showing that (21) holds for \( \Theta_2^\delta \) if \( |w_{ij} - \tilde{\mu}w_{\Delta_i^d v_j}|^2 \) in (21) is replaced by \( |w_{ij} - \hat{\mu}(d)w_{\Delta_i^d v_j}|^2 \). The algebra leading to (23)–(25) still holds with \( \hat{\mu}(d) \) in place of \( \tilde{\mu} \). Thus (21) holds for \( \Theta_2^\delta \) if we can replace (26) with

\[
 m^{-1} \sum_j (j/p)^{2\theta} \lambda_j^{-2\theta} |2\pi n|^{-1/2} e^{i\lambda_j (1 - e^{i\lambda_j})^{-1}Y_n(\theta) - \hat{\mu}(d)w_{\Delta_i^d v_j}|^2 \\
 \geq \left\{ \begin{array}{ll}
 \zeta m^{-2d+2d}n^{2+1}Y_n(\theta)^2 + \zeta n^{-1-2d+2\theta}m^{2d-2-2\theta} \hat{\mu}(d)^2, & d \in D^+, \\
 \zeta m^{-2d+2d}n^{2+1}Y_n(\theta)^2 + \zeta n^{-1-2d+2\theta}m^{-2-2\theta} \hat{\mu}(d)^2, & d \in D^-.
\end{array} \right.
\]

and replace (27) with

\[
 m^{-1} \sum_j (j/p)^{2\theta} \lambda_j^{-2\theta} \hat{\mu}(d)w_{\Delta_i^d v_j} \] 
\] 
\[+ m^{-1} \sum_j (j/p)^{2\theta} D_{nj}^\ast w_{nj}^\ast \lambda_j^{-\theta} \hat{\mu}(d)w_{\Delta_i^d v_j} | m^{-1} \sum_j (j/p)^{2\theta} D_{nj}(\theta)^* w_{nj}^\ast \lambda_j^{-\theta} \hat{\mu}(d)w_{\Delta_i^d v_j} | 
\]
\[
= \left\{ \begin{array}{ll}
 n^{1/2-d+1-\theta}m^{d-1} \hat{\mu}(d) \cdot O_p(m^{-\nu} \log n + mn^{-1}), & d \in D^+, \\
 n^{1/2-d+1-\theta}m^{-1} \hat{\mu}(d) \cdot O_p(m^{-\nu} \log n + mn^{-1}), & d \in D^-.
\end{array} \right.
\]

(39)
Since \( d \) is bounded away from 0, 1, and 2 by \( \nu > 0 \), applying Lemma B.3 (b) with
\( Q_2 = Y_n(\theta), Q_1 = 0, \) and \( Q_0 = -\tilde{\mu}(d) \) gives (38). (39) follows from Lemma B.4.

For the other subsets of \( \Theta_2, \) \( \Pr(\inf_\theta S_F(d) \leq 0) \to 0 \) is shown by showing that
(23) in the consistency proof of SP holds for those subsets if \( I_{yj} \) in (23) is replaced
with \( |w_{yj} - \tilde{\mu}(d)w_{\Delta^4 dj}|^2 \). For example, for \( \Theta_2^0 = \{ \frac{3}{2} \leq \theta \leq \frac{5}{2} \} \), the proof in SP begins
from page 1910. If we replace \( \lambda_j^{-\theta} w_{yj} \) in line 9, page 1910 of SP with \( \lambda_j^{-\theta} w_{yj} -
\lambda_j^{-\theta} \tilde{\mu}(d)w_{\Delta^4 dj} \), then, in place of (52) of SP, we have

\[
m^{-1} \sum (j/p)^{2\theta} \lambda_j^{-\theta} (2\pi n)^{-1/2} e^{i\lambda_j (1-e^{i\lambda_j})} \sum Z_t(\theta) + (2\pi n)^{-1/2} e^{i\lambda_j Z_n(\theta)} + \tilde{\mu}(d)w_{\Delta^4 dj}.
\]  

Applying Lemma B.3 (b) with \( Q_2 = \sum Z_t(\theta), Q_1 = Z_n(\theta), \) and \( Q_0 = -\tilde{\mu}(d) \) gives
the lower bound of (40). The terms involving the cross products of \( w_{yj} \), \( U_{nj}(\theta) \)
and \( \tilde{\mu}(d)w_{\Delta^4 dj} \) are dominated by (40) from Lemma B.4. For the other terms in
\( m^{-1} \sum (j/p)^{2\theta} \lambda_j^{-\theta} I_{yj} \), the result in pages 1910–11 of SP holds without change, and
(23) of SP holds. \( \square \)

A.5 Proof of Theorem 3b

From Theorem 3a, \( \Pr(|\hat{d}_F - d_0| > \varepsilon) \to 0 \) for any \( \varepsilon > 0 \). Therefore, the cases (i) or
(ii) in (37) occur with probability approaching zero, and we can apply the proof of
Theorem 1a,1b, 2a, and 2b hereafter.

As in the proof of Theorem 1b, define \( M = \{ d : (\log n)^4|d - d_0| < \rho \} \) for a fixed
\( \rho > 0 \). Then \( \Pr(d \notin M) \to 0 \) from the proof of Theorems 1b and 2b. For the limit of
\( R(d_0)' \) and \( R(\hat{d})'' \), observe that \( w_{\Delta^4(x-\tilde{\mu}(d))} = w_{\Delta^4(x-\mu_0)} + (\mu_0 - \tilde{\mu}(d))w_{\Delta^4 dj}, \) hence

\[
\frac{\partial}{\partial d} w_{\Delta^4(x-\tilde{\mu}(d))} = \frac{\partial}{\partial d} w_{\Delta^4(x-\mu_0)} + (\mu_0 - \tilde{\mu}(d))\frac{\partial}{\partial d} w_{\Delta^4 dj} + [(\partial/\partial d)\tilde{\mu}(d)]w_{\Delta^4 dj}.
\]

Note that \( \tilde{\mu}(d) \) is a weighted average of \( X_1 \) and \( \bar{X} \). The second term on the
right of (41) does not affect the limiting distribution of \( \hat{d}_F \), because we simply need
to replace \( \tilde{\mu}(d) \) with \( \bar{X} \) or \( X_1 \) or their linear combination and apply the proof of
Theorems 1b and 2b.

For the third term on the right of (41), observe that

\[
\frac{\partial}{\partial d} \tilde{\mu}(d) = (\hat{\partial}/\partial d)w(d)X - (\hat{\partial}/\partial d)w(d)X_1 = (\hat{\partial}/\partial d)w(d)\bar{X} - (\hat{\partial}/\partial d)w(d)(X_1 - \mu_0).
\]

For \( d \notin (\frac{1}{2}, \frac{3}{2}) \), we have \( (\hat{\partial}/\partial d)w(d) = 0 \) and hence \( (\hat{\partial}/\partial d)\tilde{\mu}(d) = 0 \). For \( d \in (\frac{1}{2}, \frac{3}{2}) \), first, \( (\hat{\partial}/\partial d)\tilde{\mu}(d) \) is bounded by \( C|X - \mu_0| + C|X_1 - \mu_0| \) because \( (\hat{\partial}/\partial d)w(d) \) is
uniformly bounded. Second, the order of \( w_{\Delta^4 dj} \) is bounded by that of \( (\hat{\partial}/\partial d)w_{\Delta^4 dj} = w_{\log(1-L)}\Delta^4 dj \) from Lemma B.2 (a) and (b). Therefore, the order of the third term
on the right of (41) is bounded by the order of the second term on the right of (41),
and it does not affect the limit of \( R(d_0)' \) and \( R(\hat{d})'' \). A similar argument applies to
the second derivatives of \( \tilde{\mu}(d) \), and the required follows from the proof of Theorems
1b and 2b. \( \square \)
A.6 Proof of Theorem 4

To simplify the notation, let $\Xi$ denote $\Xi(d_0)$, suppressing their dependence on $d_0$. We give the proof only for $k = 2$. The proof for larger $k$ follows the same argument, apart from more tedious algebra. A routine calculation gives

$$\hat{X}_t = X_t^0 - T_{kn} M_{kn}^{-1} X_{kn},$$

(42)

where $T_{kn} = (1, (t/n), (t/n)^2)$ and

$$M_{kn} = \begin{pmatrix}
1 & n^{-2} \sum_1^n t & n^{-3} \sum_1^n t^2 \\
-1 & n^{-3} \sum_1^n t^2 & n^{-4} \sum_1^n t^3 \\
0 & n^{-3} \sum_1^n t^3 & n^{-5} \sum_1^n t^4
\end{pmatrix}, \quad X_{kn} = \begin{pmatrix}
n^{-1} \sum_1^n X_t^0 \\
-2 \sum_1^n t X_t^0 \\
3 \sum_1^n t^2 X_t^0
\end{pmatrix}.
$$

First we derive the order of $X_{kn}$. Recall $\sum_{t=1}^n X_t^0 = (1-L)^{-1} X_n^0 = (1-L)^{-d_0-1} u_n \{ t \geq 1 \}$. Since $d_0 > -\frac{1}{2}$, clearly $E[\sum_{t=1}^n X_t^0] = O(n^{d_0-1})$. Summation by parts gives

$$\sum_{t=1}^n t X_t^0 = \sum_{t=1}^n (k-(k+1)) X_t^0 = -\sum_{k=1}^{n-1} \sum_{t=1}^n X_t^0 +\sum_{k=1}^{n-1} X_{k+1}^0, \quad \text{and it follows that } E[I_{t=1}^n t X_t^0] = O(n^{d_0+3}).$$

Similarly, we can derive $E[I_{t=1}^n t^\alpha X_t^0] = O(n^{d_0+3+\alpha})$ for any positive integer $\alpha$, and $E[||X_{kn}||^2] = O(n^{2d_0-4})$ follows. Since $M_n$ converges to a finite and invertible matrix, we can express $\hat{X}$ as

$$\hat{X}_t = X_t^0 + \Xi_0 n + \Xi_1 n t + \Xi_2 n t^2,$$

where $E[\Xi_0 n] = O(n^2)$, $E[\Xi_1 n] = O(n^{d_0-3})$, and $E[\Xi_2 n] = O(n^{d_0-5})$. Taking the dft of $\Delta^d(\hat{X}_t - \varphi(d))$ gives

$$w_{\Delta^d(\hat{X}_t - \varphi(d))|j} = w_{\Delta^d v_0} + [\Xi_0 n - \varphi(d)] w_{\Delta^d v_0} + \Xi_1 n w_{\Delta^d v_0} + \Xi_2 n w_{\Delta^d v_0},$$

(43)

The proof proceeds by (i) replacing $\pm d$ in the proof of Theorems 1a and 3b with $\Xi_0 n - \varphi(d)$ and checking the proof carries through, and (ii) checking the order of $\Xi_1 n w_{\Delta^d v_0}$ and $\Xi_2 n w_{\Delta^d v_0}$. First we derive the order of $\Xi_0 n - \varphi(d)$. Note that, since $\Xi = 0$,

$$\Xi_0 n - \varphi(d) = \begin{cases}
\Xi_0 n - \Xi_0 = \Xi_0, & \text{for } \varphi(d) = \bar{\Xi}_0 \\
\Xi_0 - \bar{\Xi}_1 = -\Xi_1 - \Xi_1 - \Xi_2, & \text{for } \varphi(d) = \bar{\Xi}_1.
\end{cases}$$

The order of $\Xi_0 n w_{\Delta^d v_0}$ is given by (19), because the order of $\Xi_0 n$ is the same as that of $\Xi_0$. When $\varphi(d) = \bar{\Xi}_1$ and $d_0 \geq \frac{1}{2}$, it follows from the order of $\Xi_n$ and $\lambda_j^\theta w_{\Delta^d v_0}$ provided in (18) that

$$[\Xi_0 n - \varphi(d)] \lambda_j^\theta w_{\Delta^d v_0} = \xi_n \cdot O((1+n^{d_0-3/2})n^{1/2} - d_0 j^{d_0-1}) = \xi_n \cdot O(j^{-1/2} + j^{d_0-2}),$$

(44)

with $E|\xi_n|^2 < \infty$. Therefore, $[\Xi_0 n - \varphi(d)] \lambda_j^\theta w_{\Delta^d v_0}$ can be handled in the same manner as $-\mu(d) \lambda_j^\theta w_{\Delta^d v_0}$ in the proof of Theorems 1a, 2a and 3a.

Now we derive the order of $w_{\Delta^d v_0}$ and $w_{\Delta^d v_0}$. Observe that $t^{\alpha} = (1-L)^{-\alpha} v_t$ for any positive integer $\alpha$. Hence $w_{\Delta^d v_0} = w_{\Delta^{d-1} v_0}$ and $w_{\Delta^d v_0} = w_{\Delta^{d-2} v_0}$, and applying
Lemma B.2 (a) gives (recall $d \leq 2 - \nu$)

$$w_{\Delta t_{ij}} = \begin{cases} \Theta(n^{\frac{3}{2} - d}2^{d-2}), & d \geq 1 + \nu, \\ -e^{1/2 - 1}(2\pi n)^{-1/2}\Gamma(2 - d)^{-1}n^{1-d}[\Theta(1 + O(j^{-\nu}))], & d \leq 1 - \nu, \end{cases}$$

Therefore, in view of the order of $\Xi_{1n}$ and $\Xi_{2n}$, we obtain

$$\Xi_{1n}\lambda_j^{-\theta}w_{\Delta t_{ij}} = \xi_{1n} \cdot O(j^{d-2}) \text{ if } d \geq 1 + \nu, \quad \xi_{1n} \cdot O(j^{-\theta-1}) \text{ if } d \leq 1 - \nu,$$

$$\Xi_{2n}\lambda_j^{-\theta}w_{\Delta t_{ij}} = \xi_{2n} \cdot O(j^{d-1}), \quad E|\xi_{1n}|^2, E|\xi_{2n}|^2 < \infty.$$

We consider three cases separately, (i) $\theta \geq \frac{1}{2}$, (ii) $\theta \leq \frac{1}{2}$ and $d \geq 1 + \nu$, and (iii) $\theta \leq \frac{1}{2}$ and $d \leq 1 - \nu$. Under (i), both $\Xi_{1n}\lambda_j^{-\theta}w_{\Delta t_{ij}}$ and $\Xi_{2n}\lambda_j^{-\theta}w_{\Delta t_{ij}}$ are $o_p(1)$, and the proof of Theorem 3a applies without a change. Under (ii), $\Xi_{1n}\lambda_j^{-\theta}w_{\Delta t_{ij}}$ is still $o_p(1)$ but $\Xi_{2n}\lambda_j^{-\theta}w_{\Delta t_{ij}}$ may not be $o_p(1)$, and we need to treat it separately.

Recall the $O(j^{-\nu})$ term above is $O(m^{-\nu})$ uniformly in $j \geq [\kappa n]$. Therefore, for $-\frac{3}{2} \leq \theta \leq \frac{1}{2}$, we replace $\lambda_j^{-\theta}(1 - e^{1/2 - 1})(2\pi n)^{-1/2}e^{1/2}Y_n(\theta) - \lambda_j^{-\theta-1}w_{\Delta t_{ij}}$ in (22) in the proof of Theorem 1a with

$$\lambda_j^{-\theta}(1 - e^{1/2 - 1})(2\pi n)^{-1/2}e^{1/2}Y_n(\theta) - \Xi_{2n}\Gamma(3 - d)^{-1}n^{2-d}[1 + O(m^{-\nu})],$$

and obtain a lower bound similar to (38), which dominates the cross-products. The proof for $\theta \leq -\frac{3}{2}$ follows a similar argument. Under (iii), we only need to replace $Y_n(\theta) - \Xi_{2n}\Gamma(3 - d)^{-1}n^{2-d}[1 + O(m^{-\nu})]$ in the above with

$$Y_n(\theta) - [\Xi_{1n}\Gamma(2 - d)^{-1}n^{1-d} + \Xi_{2n}\Gamma(3 - d)^{-1}n^{2-d}][1 + O(m^{-\nu})].$$

When $k \geq 2$, the same argument gives the required result because $w_{\Delta t_{ij}} = -e^{1/2 - 1}(2\pi n)^{-1/2}\Gamma(k + 1 - d)^{-1}n^{1-d-k}[1 + O(j^{-\nu})].$

For part (b), first define $\Psi_n(d) = [\Xi_{0n} - \varphi(d)]w_{\Delta t_{ij}} + [\Xi_{1n}w_{\Delta t_{ij}} + \Xi_{2n}w_{\Delta t_{ij}}]$ so that $\lambda_j^{-\theta}w_{\Delta t_{ij}} = \lambda_j^{-\theta}w_{\Delta t_{ij}} + \lambda_j^{-\theta}(1 - e^{1/2 - 1})(2\pi n)^{-1/2}e^{1/2}Y_n(\theta)$ (see (43)). We replace $\mu_2^{-\theta}w_{\Delta t_{ij}}$ in the proof of Theorem 1b and 2b with $\lambda_j^{-\theta}\Psi_n(d)$ and confirm the result still holds. We can obtain a similar approximation for $w_{(log(1-L))\Delta t_{ij}}$ and $w_{(log(1-L))2\Delta t_{ij}}$, and the asymptotic normality follows if $\lambda_j^{-\theta}\Psi_n(d), \lambda_j^{-\theta}\partial\Psi_n(d)/\partial d$, and $\lambda_j^{-\theta}\partial^2\Psi_n(d)/\partial d^2$ can be written as $\xi_n \cdot O(1)$, where $E|\xi_n|^2 < \infty, \alpha < -\frac{1}{4}$, and $O(1)$ is uniformly in $d \in M$. The above conditions are satisfied if $d_0 < \frac{1}{4}$, and the required result follows.

\[A.7\] Proof of Proposition 1

Let $s$ be the integer part of $d_0 + \frac{1}{2}$ and $Y_t$ be a Type I $I(d_0)$ process with the sth order polynomial time trend:

$$Y_t = (1 - L)^sU_t \{t \geq 1\} + \mu_{t}^{(0)} + \mu_{t}^{(1)}t + \cdots + \mu_{t}^{(s)}t^s, \quad U_t^{(s)} = (1 - L)^{s-d_0}u_t,$$
where \( u_t \) satisfies Assumptions 1′–3′. The proof consists of two steps. First, we show \( \hat{d}_p \) is consistent and has the stated limiting distribution if the objective function is constructed using \( Y_t \). Second, we show that replacing \( Y_t \) with \( X_t \) in the objective function does not change the limiting behavior of \( \hat{d}_p \).

The first part is proven by checking \( Y_t \) satisfies the assumptions of Theorems 5 and 6 of Velasco (1999) (hereafter Vel for short). As discussed in Lobato and Velasco (2000, page 414), the asymptotic normality of the tapered estimator still holds even if Assumption 8 of Vel is weakened to

\[
f_{U(s)}(\lambda) - G\lambda^{-2(d-s)} = O(\lambda^{-2(d-s)+\beta}) \text{ for } \beta \in (0,2].
\]  

(45)

\( U_t^{(s)} \) satisfies (45) by Assumption 1′ and \(|1-e^{i\lambda}|^{2s-2d_0} = (4\sin^2(\lambda/2))^{s-d_0} = \lambda^{-2(d_0-s)} + O(\lambda^{-2(d_0-s)+2}) \). Therefore, it suffices to check \( Y_t \) and \( U_t^{(s)} \) satisfy Assumptions 5, 7, 9, and 10 of Vel.

For Assumption 5 of Vel, define \( d(\lambda) = \sum_{k=0}^{\infty} d_k e^{ik\lambda} = (1-e^{i\lambda})^{-d_0} \) and \( \alpha(\lambda) = \sum_{k=0}^{\infty} \alpha_k e^{ik\lambda} \), then we have \( \alpha(\lambda) = \alpha(\lambda)d(\lambda). \) Now \( \partial \alpha(\lambda) / \partial \lambda = O(|\alpha(\lambda)|/|\lambda|) \) follows from Assumption 2′, and Assumption 5 is satisfied. Assumption 7 of Vel follows from (45). Assumption 9 is satisfied because \( \partial |1-e^{i\lambda}|^{2s-2d_0} / \partial \lambda = O(\lambda^{-1-2(d_0-s)}) \) and \( \partial f_\alpha(\lambda) / \partial \lambda = O(1) \) from Assumption 2′ and \( f_\alpha(\lambda) > 0 \) for \( \lambda \) sufficiently small.

For Assumption 10 of Vel, since \( U_t^{(s)} = \sum_{k=0}^{\infty} d_k v_{t-k} \) with \( d_k = (d_0-s)_k/k! \), we can rewrite \( U_t^{(s)} \) as a linear process as \( U_t^{(s)} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} d_k c_{l-k} e^{i\lambda \xi_{l-k}} = \sum_{l=0}^{\infty} \alpha_l e^{i\lambda \xi_{l-k}} \) with \( \alpha_l = \sum_{j=0}^{l} d_j c_{l-j} \). Assumption 10 is satisfied because \( \sum_{k=0}^{\infty} \alpha_k^2 = E(U_t^{(s)})^2 < \infty \), and we complete the first part of the proof.

Second, we use the results on the difference between \( w_{xp}(\lambda_j) \) and \( w_{yp}(\lambda_j) \) by Robinson (2005) to show that \( \hat{d}_p \) has the stated limit distribution when the objective function is constructed with \( X_t \). Note that \( (d+q,d) \) in Robinson corresponds to our \( (d_0,d_0-s) \) and the statement of Theorem in Robinson has a typo: \( d \in (-\frac{1}{2},\frac{1}{2}] \) should be replaced with \( d \in [-\frac{1}{2},\frac{1}{2}] \). From Theorem of Robinson (2005), we have

\[
E \left\{ \lambda_j^{2d_0} |w_{xp}(\lambda_j) - w_{yp}(\lambda_j)|^2 \right\} \leq C_j^{-2\eta - 1} \log n,
\]

where \( 2\eta = 2(d_0-s)-1 < 0 \). It follows from the triangle inequality, Cauchy-Schwartz inequality, and \( E\lambda_j^{2d_0} I_{yp}(\lambda_j) = O(1) \) for \( j = p,2p,\ldots,m \) (Vel, Theorem 4) that

\[
E\lambda_j^{2d_0} \left| I_{xp}(\lambda_j) - I_{yp}(\lambda_j) \right| \leq E\lambda_j^{2d_0} \left| w_{xp}(\lambda_j) - w_{yp}(\lambda_j) \right|^2 + 2E\lambda_j^{2d_0} \left| w_{xp}(\lambda_j) - w_{yp}(\lambda_j) \right| \left| w_{yp}(\lambda_j) \right| = O(j^{-\eta - 1/2} \log n), \quad \text{for } j = p,2p,\ldots,m.
\]

Note that the periodogram \( I_j \) in Vel is equal to \( I_{yp}(\lambda_j) \) in our notation. Therefore, if we replace \( I_j \) in \( A_p(d) \) in line 3, page 112 of Vel with \( I_{xp}(\lambda_j) \), then the right hand side of (A14) in Vel has an additional term whose order is \( O_p(m^{-\xi} \log m \log n) \) for some \( \xi > 0 \), and the proof of consistency is not affected.

For the asymptotic normality (Theorem 6 in Vel), if we replace \( I_j \) in (A23) on page 116 of Vel with \( I_{xp}(\lambda_j) \), then the right hand side of (A23) has an additional term.
$O_p(r^{-\eta+1/2} \log n)$. Consequently, the left hand side of the equation in line 14, page 117 of Vel has an additional term $O_p(m^{-\eta+1/2} \log n)$. Since this is $o_p(1)$, the right hand side of that equation remains unchanged and their argument carries through. Finally, the equation in line 18, page 117 of Vel has an additional term $O_p(m^{-\eta} \log n)$, which is $o_p(1)$, and the asymptotic normality follows. □

A.8 Proof of Proposition 2

Applying the first part of the proof of Proposition 1, we can easily show that our Assumptions 1-5 and 1’-5’ imply that $m$ and $\Delta Y_t = (1 - L)^{-d_0 + 1}u_t$ satisfy Assumptions A1’-A4’ of Hurvich and Chen (2000). Therefore, both the consistency and asymptotic normality of $\hat{d}_{HC}$ follow if we show

$$E\{\lambda_j^{2(d_0-1)}|w_{\Delta y}^{HC}(\lambda_j) - w_{\Delta x}^{HC}(\lambda_j)|^2\} \leq C j^{-2\eta-1} \log n, \; j = 1, \ldots, m$$

(46)

for some $\eta > 0$, because then the proof of Theorems 1 and 2 of Hurvich and Chen (2000) (hereafter HC) carries through if we replace their $I_T^T$ (that corresponds to our $I_{\Delta y}^{HC}(\lambda_j)$) with $I_{\Delta x}^{HC}(\lambda_j)$. Specifically, Lemma 1 and equation (8) of HC still holds, and Lemma 6 of HC has an additional $O_p(r^{-\eta+1/2})$ term that does not affect the validity of their Theorem 2.

We proceed to show (46). First, observe that Hurvich-Chen taper satisfies the bounds (2.1)-(2.3) in page 286 of Robinson (2005) with $p = 1$. The other two conditions on $h(t)$ in page 286 do not matter for Theorem of Robinson (2005) to hold. Therefore, for $d_0 \in \left[\frac{1}{2}, \frac{3}{2}\right]$, applying (2.6) in Theorem of Robinson (2005) to $(\Delta X_t, \Delta Y_t)$ gives

$$E\{\lambda_j^{2(d_0-1)}|w_{\Delta y}^{HC}(\lambda) - w_{\Delta x}^{HC}(\lambda)|^2\} \leq C|\log \lambda|^{1\{d_0=1/2\}}(n\lambda)^{2(d_0-1)-2}, \; 0 < \lambda \leq \pi.$$  

Hence, (46) holds with $\eta = \frac{3}{2} - d_0 > 0$.

For $d_0 \in \left[-\frac{1}{2}, \frac{1}{2}\right]$, summation by parts gives

$$\sum_{t=1}^{n} h_t^{HC}(\Delta Y_t - \Delta X_t)e^{it\lambda} = \sum_{t=1}^{n-1} (h_t^{HC}e^{it\lambda} - h_{t+1}^{HC}e^{i(t+1)\lambda})(Y_t - X_t) + h_n^{HC}e^{in\lambda}(Y_n - X_n).$$

(47)

Because $h_{t+1}^{HC} = h_t^{HC}e^{2\pi i/n} + 0.5(1 - e^{2\pi i/n})$, routine algebra gives $h_t^{HC}e^{it\lambda} - h_{t+1}^{HC}e^{i(t+1)\lambda} = h_t^{HC}e^{it\lambda}(1 - e^{i(\lambda+2\pi/n)}) + 0.5(e^{2\pi i/n} - 1)e^{i(t+1)\lambda}$. It follows that $w_{\Delta y}^{HC}(\lambda) - w_{\Delta x}^{HC}(\lambda) = A_\lambda + B_\lambda + R_\lambda$, where

$$A_\lambda = \left(1 - e^{i(\lambda+2\pi/n)}\right)|w_y^{HC}(\lambda) - w_x^{HC}(\lambda)|, \quad B_\lambda = 0.5(e^{2\pi i/n} - 1)e^{i\lambda}w_y(\lambda) - w_x(\lambda),$$

and $R_\lambda = (2\pi n)^{-1/2}h_\lambda^{HC}e^{in\lambda}(Y_n - X_n)e^{i(\lambda+2\pi/n)} - (2\pi n)^{-1/2}(e^{2\pi i/n} - 1)e^{i\lambda}e^{in\lambda}(Y_n - X_n)$. For $A_\lambda$, it follows from (2.7) in Theorem of Robinson (2005) and $\lambda_j^{-2}|1 - e^{i(\lambda_j+2\pi/n)}|^2 < C$ that, for $1/n \leq \lambda_j \leq \pi$,

$$E\{\lambda_j^{2(d_0-1)}|A_j|^{2}\} \leq C(n\lambda_j)^{2d_0-2} \log n \leq Cj^{-2(1/2-d_0)-1} \log n,$$  

27
with $\frac{1}{2} - d_0 > 0$. For $B_{\lambda_j}$, using (2.6) of Theorem of Robinson (2005) and $e^{2\pi i/n} - 1 = O(n^{-1})$, we have $E\{\lambda^2(d_0 - 1)B_{\lambda_j}^2\} \leq C(n\lambda_j)^{-3} \log n = C\lambda_j^{-3} \log n$ for $1/n \leq \lambda_j \leq \pi$. Finally, for $R_{\lambda_j}$, it follows from $|h_n^{HC}| \leq Cn^{-1}$ and $E(Y_n - X_n)^2 = O(n^{2d_0-1})$ (Marinucci and Robinson, 1999, page 119) that $E\{\lambda^2(d_0 - 1)|R_{\lambda_j}|^2\} \leq C(n\lambda_j)^{2d_0 - 2} n^{-2} \leq C\lambda_j^{2(1/2 - d_0)} n^{-2}$ for $1/n \leq \lambda_j \leq \pi$. Therefore, (46) holds with $\frac{1}{2} - d_0 > 0$ and the proof is completed. □

Appendix B: technical lemmas

Lemma B.1 is a generalized restatement of equation (40) in SP and stated as a lemma because it is repeatedly used in the proofs. Lemma B.2 gives the approximation formula for the dft of the deterministic process $v_t = 1 \{t \geq 1\}$. Lemmas B.3 and B.5 extend Lemmas 5.10 and 5.5 of SP, respectively, and are used in the proof of Theorem 3a. Lemma B.3 is used in the proof of consistency for establishing the lower bound of the objective function when $|d - d_0| \geq \frac{1}{2}$.

**Lemma B.1.** Let $\theta \in \Theta$ be a parameter and assume $m \rightarrow \infty$ as $n \rightarrow \infty$. Suppose two random variables $A_n(\theta)$ and $B_n(\theta)$ satisfy (i) $A_n(\theta) \geq \eta X_n(\theta)^2$ uniformly in $\theta \in \Theta$ for some $\eta > 0$, and (ii) $B_n(\theta) = X_n(\theta) R_n(\theta)$, where $\sup_\theta |R_n(\theta)| = O_p(k_n)$ with $k_n \log m \rightarrow 0$. Then, for any $\zeta > 0$,

$$\Pr \left( \inf_{\theta \in \Theta} [A_n(\theta) + B_n(\theta)] \leq -\zeta \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

**Proof.** This result is a generalized restatement of equation (40) in SP. $A_n(\theta)$ and $X_n(\theta)$ correspond to (35) and $m^{-\theta} n^{\theta - 1/2} Z_n(\theta)$ in SP, respectively. $B_n(\theta)$ corresponds to (37)+(38) in SP. The proof follows from repeating the argument in pages 1908-1909 of SP. □

**Lemma B.2.** Let $v_t = 1 \{t \geq 1\}$. Then the following holds uniformly in $j = 1, \ldots, m$ with $m = o(n)$ and in $d$:

(a) $w_{\Delta v_j}^d$

$$= \begin{cases} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} (2\pi n)^{-1/2} (1 - 1 - (1 - e^{i\lambda_j})^d - n^{-d} / \Gamma(1 - d) + O(n^{-d} j^{-1})), & d \in \left[ -\frac{1}{2}, C \right], \\ -e^{i\lambda_j} (1 - e^{i\lambda_j})^{-1} (2\pi n)^{-1/2} \Gamma(1 - d)^{-1} n^{-d} [1 + O(j^{-1/2})], & d \in \left[ -C, -\frac{1}{2} \right], \end{cases}$$

(b) $w_{\log(1-L)\Delta v_j}^d$

$$= \begin{cases} J_n(e^{i\lambda_j}) w_{\Delta v_j}^d + O\left( j^{-1} n^{1/2 - d} \log n \right), & d \in [-C, 1], \\ J_n(e^{i\lambda_j}) w_{\Delta v_j}^d + O\left( n^{1/2 - d} \log n \right), & d \in [1, 2], \end{cases}$$

(c) $w_{(\log(1-L))^2\Delta v_j}^d$

$$= \begin{cases} J_n(e^{i\lambda_j})^2 w_{\Delta v_j}^d + O\left( j^{-1} n^{1/2 - d} (\log n)^2 \right), & d \in [-C, 1], \\ J_n(e^{i\lambda_j})^2 w_{\Delta v_j}^d + O\left( n^{1/2 - d} (\log n)^2 \right), & d \in [1, 2], \end{cases}$$

where $J_n(e^{i\lambda_j}) = \sum_{k=1}^n k^{-1} e^{i\lambda_j} = O(\log n)$. 28
Proof For part (a), first, from Lemma 5.1 (b) of SP, we have
\[ w_{\Delta^j} = (1 - e^{i\lambda_j})^{-1} \left[ w_{\Delta^{j+1}} - e^{i\lambda_j} (2\pi n)^{-1/2} \Delta^d v_n \right]. \] (48)
From the proof of Lemma A.7 of Phillips and Shimotsu (2004, page 676, line 10), we have
\[ \Delta^{d+1} v_t = (1 - L)^{d+1} v_t = (-d)_{t-1} / (t - 1)! . \] (49)
Therefore, the first term in the bracket in (48) can be expressed as
\[ w_{\Delta^{d+1}} v_j = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} \frac{(-d)_{t-1}}{(t - 1)!} e^{it\lambda_j} = \frac{e^{i\lambda_j}}{\sqrt{2\pi n}} \sum_{k=0}^{n-1} \frac{(-d)_k}{k!} e^{ik\lambda_j}. \]
Define \( D_n(e^{i\lambda}; d) = \sum_{k=0}^{n} (-d)_k/k! e^{ik\lambda} \) as defined in Lemma 5.1 of SP, then it follows from (48) that
\[ w_{\Delta^d} v_j = \frac{e^{i\lambda_j}}{1 - e^{i\lambda_j}} (2\pi n)^{-1/2} \left[ D_{n-1} \left( e^{i\lambda}; d \right) - \Delta^d v_n \right]. \]
The stated result for \( d \in [-\frac{1}{2}, C] \) follows from the approximation of \( D_n(e^{i\lambda}; d) \) shown by Lemma A.2 of Phillips and Shimotsu (2004) and the fact that (see (49))
\[ \Delta^d v_n = \frac{(1 - d)_{n-1}}{(n - 1)!} = \frac{\Gamma(n - d)}{\Gamma(n) \Gamma(1 - d)} = \frac{1}{\Gamma(1 - d)} n^{-d} \left( 1 + O \left( n^{-1} \right) \right). \] (50)
For \( d \in \left[ -\frac{3}{2}, -\frac{1}{2} \right] \), it follows from (48), the result for \( d \in [-\frac{1}{2}, C] \) and (50) that
\[ w_{\Delta^d} v_j = \frac{1}{1 - e^{i\lambda_j}} \left[ O \left( j^d n^{-d-1/2} \right) + O \left( j^{-1} n^{-d-1/2} \right) - \frac{e^{i\lambda_j}}{\sqrt{2\pi n}} \Delta^d v_n \right] \]
\[ = - \frac{1}{1 - e^{i\lambda_j}} \frac{n^{-d}}{\sqrt{2\pi n} \Gamma(1 - d)} \left[ 1 + O(j^{-1/2}) \right]. \]
The results for smaller \( d \) follow from (48) and induction.

For part (b), first we find a uniform bound for \( d \in [-C, 1] \). Define \( J_n(L) = \sum_{k=1}^{n} \frac{1}{k} L^k \). Lemma 5.7 (a) of SP gives
\[ -\log(1 - L) \Delta^d v_t = J_n(L) \Delta^d v_t = J_n(e^{i\lambda_j}) \Delta^d v_t + \tilde{J}_{n\lambda_j}(e^{-i\lambda_j} L)(e^{-i\lambda_j} L - 1) \Delta^d v_t, \]
where \( \tilde{J}_{n\lambda_j}(e^{-i\lambda_j} L) = \sum_{p=0}^{n-1} \frac{1}{k} e^{-ip\lambda_j} L^p \) and \( \tilde{J}_{\lambda_p} = \sum_{p=1}^{n} \frac{1}{k} e^{ik\lambda_j} \). Taking its dft leaves us with
\[ -w_{\log(1-L) \Delta^d} v_j = J_n(e^{i\lambda_j}) w_{\Delta^d} v_j - (2\pi n)^{-1/2} \tilde{J}_{n\lambda_j}(e^{-i\lambda_j} L) \Delta^d v_n. \] (51)
Define \( |x|_+ = \max\{x, 1\} \). Since \( \Delta^d v_n \) is \( O((n-p)^{-d}) \) from (50) and \( \tilde{J}_{\lambda_p} = O(|p|_+ n^{-1}) \) from Lemma 5.8 (b) of SP, the second term on the right of (51) is
\[ - (2\pi n)^{-1/2} \sum_{p=0}^{n-1} \frac{1}{k} e^{-ip\lambda_j} \Delta^d v_{n-p} \]
\[ = O \left( n^{-1/2} \sum_{p=0}^{n-1} |p|_+ n^{-1} (n-p)^{-d} \right) = O \left( j^{-1} n^{1/2} \sum_{p=0}^{n-1} |p|_+ (n-p)^{-d} \right). \]
Uniformly in \( d \in [-C, 1] \), we have
\[
\sum_{p=0}^{n-1} \frac{1}{p+1} (n-p)^{-d} \leq \sum_{p=0}^{n/2} \frac{1}{p+1} (n-p)^{-d} + \sum_{p=n/2}^{n-1} \frac{1}{p+1} (n-p)^{-d} \\
\leq C n^{-d} \sum_{p=0}^{n/2} \frac{1}{p+1} + (n/2)^{-1} \sum_{p=n/2}^{n-1} (n-p)^{-d} \\
= O(n^{-d} \log n).
\] (52)

Therefore, the second term on the right of (51) is \( O(j^{-1} n^{1/2-d} \log n) \), and the stated result follows. The order of \( J_n(e^{i \lambda_j}) \) is shown in Lemma 5.8 (a) of SP.

For \( d \in [1, 2] \), Lemma 5.1 (b) of SP gives
\[
-w_{\Delta_d v_j} = -(1 - e^{i \lambda_j}) w_{\Delta_{d-1} v_j} - e^{i \lambda_j} (2\pi n)^{-1/2} \Delta^{-1} v_n.
\] (53)

Differentiating it with respect to \( d \), we find
\[
-w_{\log (1-L) \Delta_d v_j} = -(1 - e^{i \lambda_j}) w_{\log (1-L) \Delta_{d-1} v_j} - e^{i \lambda_j} (2\pi n)^{-1/2} \log (1-L) \Delta^{-1} v_n.
\] (54)

From (50), (52) and the fact that \( d - 1 \leq 1 \), the second term on the right of (54) is bounded by
\[
n^{-1/2} \sum_{p=1}^{n-1} p^{-1} \Delta^{d-1} v_{n-p} = O \left( n^{-1/2} \sum_{p=1}^{n-1} \frac{1}{p} (n-p)^{1-d} \right) = O \left( n^{1/2-d} \log n \right).
\]

Substituting the result for \( d \in [-C, 1] \) to \( w_{\log (1-L) \Delta_{d-1} v_j} \) on the right of (54) and then applying (53) gives the stated result.

For part (c), first, for \( d \in [-C, 1] \), we have from Lemma 5.7 (a) of SP
\[
J_n(L)^2 = \left[ J_n(e^{i \lambda}) + \tilde{J}_{n\lambda}(e^{-i \lambda} L)(e^{-i \lambda} L - 1) \right]^2 \\
= J_n(e^{i \lambda})^2 + J_n(e^{i \lambda}) \tilde{J}_{n\lambda}(e^{-i \lambda} L)(e^{-i \lambda} L - 1) + J_n(L) \tilde{J}_{n\lambda}(e^{-i \lambda} L)(e^{-i \lambda} L - 1).
\]

It follows that
\[
w_{\log (1-L)^2 \Delta_d v_j} = w_{\log (1-L) \Delta_d v_j} \\
= J_n(e^{i \lambda})^2 w_{\Delta_d v_j} - J_n(e^{i \lambda})(2\pi n)^{-1/2} \tilde{J}_{n\lambda}(e^{-i \lambda} L) \Delta_d v_n \\
- J_n(L)(2\pi n)^{-1/2} \tilde{J}_{n\lambda}(e^{-i \lambda} L) \Delta_d v_n.
\]

The second term is \( J_n(e^{i \lambda}) \) times the second term on the right of (51), hence it is \( O(j^{-1} n^{1/2-d} (\log n)^2) \). The third term is
\[
- (2\pi n)^{-1/2} \sum_{q=1}^{n} \frac{1}{q} \sum_{p=0}^{n-q-1} \tilde{J}_{\lambda q} e^{-ip \lambda} \Delta_d v_{n-p-q} \\
\leq O \left( j^{-1} n^{1/2} \sum_{q=1}^{n} \frac{1}{q} \sum_{p=0}^{n-q-1} |p|^{-1} (n-q-p)^{-d} \right) \\
= O \left( j^{-1} n^{1/2} \sum_{q=1}^{n} \frac{1}{q} (n-q)^{-d} \log n \right) = O \left( j^{-1} n^{1/2-d} (\log n)^2 \right).
\] (55)
For $d \in [1, 2]$, taking the second derivative of $(-53)$ with respect to $d$ gives

$$w_{(\log(1-L))^2\Delta^d v_j} = (1 - e^{i\lambda_j})w_{(\log(1-L))^2\Delta^{d-1}v_j} - e^{i\lambda_j}(2\pi n)^{-1/2}(\log(1 - L))^2\Delta^{d-1}v_n.$$  

(56)

The second term on the right of (56) is

$$O \left( n^{-1/2} \sum_{p=1}^{n-1} p^{-1} \sum_{q=1}^{n-p-1} q^{-1} (n-p-q)^{-1-d} \right) = O \left( n^{1/2-d} (\log n)^2 \right),$$

and the first term on the right of (56) is, from the result for $d \in [-C, 1]$,

$$J_n(e^{i\lambda_j})^2(1 - e^{i\lambda_j})w_{\Delta^{d-1}v_j} + O(n^{1/2-d}(\log n)^2) = J_n(e^{i\lambda_j})^2w_{\Delta^dv_j} + O(n^{1/2-d}(\log n)^2),$$

giving the stated result. □

**Lemma B.3.** Let $Q_k, k = 0, 1, 2$, be any real numbers, $\nu_t = 1 \{ t \geq 1 \}$, $\kappa \in (0, 1/8)$, and $m = o(n)$. Then, there exists $\eta > 0$ not depending on $Q_k$ such that, uniformly in $d \in \{ [-\frac{1}{2}, -\epsilon] \cup \{ 1 + \epsilon, 2 - \epsilon \} \}$ and for sufficiently large $n$,

(a) $m^{-1} \sum_{j=\lceil km \rceil}^m |(2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})Q_2 + (2\pi n)^{-1/2} e^{i\lambda_j} Q_1 + w_{\Delta^dv_j}Q_0|^2 \geq \left\{ \begin{array}{ll}
\eta(n^{-3} m^2 Q_2^2 + n^{-1} Q_1^2 + n^{-2} m^2 Q_0^2), & d \in \{ 1 - \epsilon \cup [1 + \epsilon, 2 - \epsilon] \}, \\
\eta(n^{-3} m^2 Q_2^2 + n^{-1} Q_1^2 + n^{-2} m^2 Q_0^2), & d \in \{-\frac{1}{2}, -\epsilon\}.
\end{array} \right.$

(b) $m^{-1} \sum_{j=\lceil km \rceil}^m |(2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-2}Q_2 + (2\pi n)^{-1/2} e^{i\lambda_j} (1 - e^{i\lambda_j})^{-2}Q_1 + w_{\Delta^dv_j}Q_0|^2 \geq \left\{ \begin{array}{ll}
\eta(n m^{-4} Q_2^2 + n^{-1} Q_1^2 + n^{-2} m^2 Q_0^2), & d \in \{ 1 - \epsilon \cup [1 + \epsilon, 2 - \epsilon] \}, \\
\eta(n m^{-4} Q_2^2 + n^{-1} Q_1^2 + n^{-2} m^2 Q_0^2), & d \in \{-\frac{1}{2}, -\epsilon\}.
\end{array} \right.$

**Proof** The proof follows the approach of the proof of Lemma 5.10 of SP. For part (a), first define

$$A(\lambda_j) = (1 - e^{i\lambda_j})Q_2 + Q_1 + (2\pi n)^{1/2} e^{i\lambda_j} w_{\Delta^dv_j}Q_0,$$

so the right hand side of (a) is $m^{-1} \sum_{j=\lceil km \rceil}^m (2\pi n)^{-1/2} e^{i\lambda_j} A(\lambda_j)$. Then part (a) for $d \geq \epsilon$ follows if, for sufficiently large $n$,

$$m^{-1} \sum_{j=\lceil km \rceil}^m |A(\lambda_j)|^2 \geq \eta(n^{-2} m^2 Q_2^2 + Q_1^2 + n^{-2} m^2 Q_0^2).$$

(57)

We consider the case with $d \in [1 + \epsilon, 2 - \epsilon]$ in details. The other cases follow the same line of argument. Because $d \geq \epsilon$ implies that $j^d = o(1)$ as $m \rightarrow \infty$ uniformly in $j \geq \lceil km \rceil$, we can refine the approximation of $w_{\Delta^dv_j}$ in Lemma B.2 (a) as

$$w_{\Delta^dv_j} = e^{i\lambda_j}(2\pi n)^{-1/2}(1 - e^{i\lambda_j})^{-1} \left( 1 + o(1) \right) = e^{i\lambda_j}(2\pi n)^{-1/2} e^{-\frac{\pi}{(d-1)}(d-1)^{-1}}(1 + o(1)),$$

(58)

uniformly in $j = \lceil km \rceil, \ldots, m$, where the second equality follows from Lemma 5.2 of SP. Define $\bar{c}(d) = \cos(-\pi(d-1)/2)$ and $\bar{s}(d) = \sin(-\pi(d-1)/2)$, then it follows that

$$A(\lambda_j) = -i\lambda_j Q_2 + o(\lambda_j)Q_2 + Q_1 + \bar{c}(d)\lambda_j^{d-1}Q_0 + i\bar{s}(d)\lambda_j^{d-1}Q_0 + o(\lambda_j^{d-1})Q_0,$$

$$|A(\lambda_j)|^2 = \left[ Q_1 + \bar{c}(d)\lambda_j^{d-1}Q_0 \right]^2 + \left[ \lambda_j Q_2 - i\bar{s}(d)\lambda_j^{d-1}Q_0 \right]^2 + r_{nj},$$
where \( r_{nj} = o(\lambda_j^2 Q_2^2 + o(\lambda_j)Q_1 Q_2 + o(\lambda_j^d)Q_2Q_0 + o(\lambda_j^d - 1)Q_2Q_0 + o(\lambda_j^{2d-2})Q_0^2) \). Note that \( m^{-1}\sum_{j=km}^m r_{nj} = o(m^2 n^{-2}Q_2^2 + Q_1^2 + m^{2d-2}n^{-2}dQ_0^2) \). Therefore, (57) follows if we show that, either for \( j = [km], \ldots, [m/4] \) or \( j = [3m/4], \ldots, m \),

\[
\left[ Q_1 + c(d)\lambda_j^{d-1}Q_0 \right]^2 \geq \eta \left[ Q_1^2 + c(d)^2 \lambda_{m/2}^{2d-2}Q_0^2 \right],
\]

and either for \( j = [km], \ldots, [m/4] \) or \( j = [3m/4], \ldots, m \),

\[
\left[ \lambda_j Q_2 - \bar{s}(d)\lambda_j^{d-1}Q_0 \right]^2 \geq \eta \left[ \lambda_j^2 Q_2^2 + \bar{s}(d)^2 \lambda_{m/2}^{2d-2}Q_0^2 \right].
\]

We proceed to show (59). Assume \( \bar{c}(d) \geq 0 \) without the loss of generality. When \( \text{sgn}(Q_1) = \text{sgn}(Q_0) \), the result follows immediately, so assume \( Q_1 < 0 \) and \( Q_0 > 0 \) without the loss of generality. Now suppose \( Q_1 + \bar{c}(d)\lambda_m^{d-1}Q_0 \geq 0 \) and consider \( j \geq [3m/4] \). Since \( d - 1 \geq \varepsilon \), we have \( \lambda_{3m/4}^{d-1} = (3/2)^{d-1} \lambda_{m/2}^{d-1} \geq (1 + 2\xi)\lambda_{m/2}^{d-1} \) for some \( \xi > 0 \) uniformly in \( d \), thus \( Q_1 + \bar{c}(d)\lambda_{3m/4}^{d-1}Q_0 \geq 2\xi\bar{c}(d)\lambda_{m/2}^{d-1}Q_0 \geq -2\xi Q_1 > 0 \). Since \( \lambda_j^{d-1} \) is an increasing function of \( j \), we have, for \( j = [3m/4], \ldots, m \),

\[
Q_1 + \bar{c}(d)\lambda_j^{d-1}Q_0 \geq \xi \left( -Q_1 + \bar{c}(d)\lambda_m^{d-1}Q_0 \right),
\]

and (59) follows because both \( -Q_1 \) and \( \bar{c}(d)\lambda_m^{d-1}Q_0 \) are positive. Now suppose \( Q_1 + \bar{c}(d)\lambda_m^{d-1}Q_0 < 0 \) and consider \( j \leq [m/4] \). Then \( \lambda_{m/4}^{d-1} = (1/2)^{d-1} \lambda_{m/2}^{d-1} \leq (1 - 2\xi)\lambda_{m/2}^{d-1} \) for some \( \xi \in (0, 1/4) \) uniformly in \( d \), and it follows that \( Q_1 + \bar{c}(d)\lambda_m^{d-1}Q_0 \leq Q_1 + (1 - 2\xi)\bar{c}(d)\lambda_{m/2}^{d-1}Q_0 \leq \xi(Q_1 - \bar{c}(d)\lambda_m^{d-1}Q_0) < 0 \). Therefore, we have, for \( j = [km], \ldots, [m/4] \),

\[
Q_1 + \bar{c}(d)\lambda_j^{d-1}Q_0 \leq \xi \left( Q_1 - \bar{c}(d)\lambda_m^{d-1}Q_0 \right),
\]

and (59) follows. Since \( Q_1 + \bar{c}(d)\lambda_m^{d-1}Q_0 \) can be only \( \geq 0 \) or \( < 0 \), we established (59).

(60) is obtained by writing down \( [\lambda_j Q_2 - \bar{s}(d)\lambda_j^{d-1}Q_0]^2 = \lambda_j^2[Q_2 - \bar{s}(d)\lambda_j^{d-2}Q_0]^2 \) and proceeding in the same manner with \( d = 2 \leq \varepsilon \).

The other cases in part (a) follow the same argument. The essential element is that there is sufficient variation in \( Q_1 + \bar{c}(d)\lambda_j^{d-1}Q_0 \) and \( Q_2 - \bar{s}(d)\lambda_j^{d-2}Q_0 \) as \( j \) changes, which is guaranteed by bounding away \( |d - 1| \) and \( |d - 2| \) from 0. Part (b) follows from the same argument, because there is sufficient variation in \( \lambda_j^{-1} \), \( \lambda_j^{-2} \), and \( \lambda_j^{d-1} \) if \( |d - 2| \geq \varepsilon \), \( |d - 1| \geq \varepsilon \) and \( |d| \geq \varepsilon \).

**Lemma B.4.** Suppose \( D_{nj}(\theta) \) and \( U_{nj}(\theta) \) satisfy

\[
\begin{align*}
D_{nj}(\theta) &= e^{-(\pi/2)\hat{\theta}_j} + O(\lambda_j) + O(j^{-1/2}), \quad \text{uniformly in } \theta \in [-\frac{1}{2}, \frac{1}{2}], \\
E\sup_{\theta \in [-\frac{1}{2}, \frac{1}{2}]} |U_{nj}(\theta)|^2 &= O(j^{-1}(\log n)^2).
\end{align*}
\]

Let \( \kappa \in (0, 1/8) \) and \( m = o(n) \). Then, uniformly in \( \theta \in [-\frac{1}{2}, \frac{1}{2}] \) and \( \alpha \in [-C, -\varepsilon] \cup \)
\[ [\varepsilon, 1 - \varepsilon] \cup [1 + \varepsilon, 2 - \varepsilon] \text{ with } \varepsilon < \frac{1}{2}, \]

\[
\frac{1}{m} \sum_{j=\lfloor km \rfloor}^{m} \left( \frac{j}{m} \right)^{2\theta} D_{nj}(\theta) w_{uj} \lambda_j^{-\theta} w_{\Delta vj} + \frac{1}{m} \sum_{j=\lfloor km \rfloor}^{m} \left( \frac{j}{m} \right)^{2\theta} U_{nj}(\theta) \lambda_j^{-\theta} w_{\Delta vj} \]

\[
= \begin{cases} 
    n^{1/2-\alpha+\theta} m^{\alpha-1-\theta} \cdot O_p(n^{-1}m + m^{-\varepsilon} \log n), & \alpha \geq \varepsilon, \\
    n^{1/2-\alpha+\theta} m^{\alpha-1-\theta} \cdot O_p(n^{-1}m + m^{-\varepsilon} \log n), & \alpha \leq -\varepsilon.
\end{cases}
\]

(62)

Prove Define \( A^+ = [\varepsilon, 1 - \varepsilon] \cup [1 + \varepsilon, 2 - \varepsilon] \) and \( A^- = [-C, -\varepsilon] \), so that \( A^+ \cup A^- \) covers the admissible value of \( \alpha \). For the first term on the left of (62), from (61) and Lemma B.2, we obtain

\[ D_{nj}(\theta) w_{\Delta vj} = \begin{cases} 
    C_n(\theta)n^{1/2-\alpha} j^{\alpha-1} [1 + O(\lambda_j) + O(j^{-\varepsilon})], & \alpha \in A^+, \\
    C_n(\theta)n^{1/2-\alpha} j^{\alpha-1} [1 + O(\lambda_j) + O(j^{-\varepsilon})], & \alpha \in A^-,
\end{cases} \]

(63)

where \( C_n(\theta) \) is a non-random function of \( \theta \) such that \( 0 < |C_n(\theta)| < \infty \) uniformly in \( \theta \). The required result follows from Lemmas 5.4 and 5.6 of SP, (63), and \( E|w_u(\lambda_j)|^2 < \infty \) (e.g., equation (19) in SP). For the second term on the left of (62), the stated bound follows straightforwardly from Lemma B.2, (61), and Lemma 5.4 of SP.

Lemma B.5. Let \( \eta > 0 \) be a fixed number and \( p \sim m/e \) as \( m \to \infty \). There exist \( \varepsilon \in (0, 0.1) \) and \( \bar{\kappa} \in (0, 1/4) \) such that, for all fixed \( \kappa \in (0, \bar{\kappa}] \) and sufficiently large \( m \),

\[
\inf_{\gamma \in [-C, -\eta] \cup [\eta, C]} \frac{1}{m} \sum_{j=\lfloor km \rfloor}^{m} \left( \frac{j}{p} \right)^{\gamma} \geq 1 + 2\varepsilon.
\]

Proof Lemma 5.5 of SP establishes the stated result for \( \gamma \in [-C, -1 + 2\Delta] \cup [1, C] \) with \( \Delta \in (0, 1/(2e)) \). Hence, we only need to show the stated result for \( \gamma \in [-1 + 2\Delta, -\eta] \cup [\eta, 1] \). From Lemma 5.4 of SP, we have, uniformly in \( \gamma \in [-1 + 2\Delta, 1] \),

\[
\frac{1}{m} \sum_{j=\lfloor km \rfloor}^{m} \left( \frac{j}{p} \right)^{\gamma} = \left( \frac{m}{p} \right)^{\gamma} \frac{1}{m} \sum_{j=\lfloor km \rfloor}^{m} \left( \frac{j}{m} \right)^{\gamma} = (e + o(1))^{\gamma} \left( \int_{\kappa}^{1} x^{\gamma} dx + o(1) \right) = (\gamma + 1)^{-1} e^{\gamma}(1 - \kappa^{\gamma}) + o(1), \quad \text{as } m \to \infty.
\]

(64)

Note that \( g(\gamma) = (\gamma + 1)^{-1} e^{\gamma} \) takes the value 1 when \( \gamma = 0 \), \( g'(\gamma) > 0 \) when \( \gamma \geq \eta \), and \( g'(\gamma) < 0 \) when \( \gamma \in [-1 + 2\Delta, -\eta] \). Therefore, choosing \( \kappa \) sufficiently small makes (64) larger than \( 1 + 2\varepsilon \) for \( \gamma \in [-1 + 2\Delta, -\eta] \cup [\eta, 1] \) and sufficiently large \( m \). \( \Box \)
References


Table 1. Monte Carlo simulation bias: \( n = 256, m = n^{0.65} = 36 \)

<table>
<thead>
<tr>
<th>( d_0 )</th>
<th>-0.4</th>
<th>0.0</th>
<th>0.4</th>
<th>0.8</th>
<th>1.2</th>
<th>1.6</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu} = X )</td>
<td>-0.0047</td>
<td>-0.0034</td>
<td>0.0000</td>
<td>0.0144</td>
<td>-0.0639</td>
<td>-0.3926</td>
<td>-0.8021</td>
</tr>
<tr>
<td>( \hat{\mu} = X_1 )</td>
<td>0.3066</td>
<td>0.0048</td>
<td>-0.0064</td>
<td>-0.0001</td>
<td>-0.0025</td>
<td>-0.0029</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 2. Simulation results: \( n = 512, m = n^{0.65} = 57 \)

<table>
<thead>
<tr>
<th>( d )</th>
<th>( \rho )</th>
<th>ELW-F2 bias</th>
<th>var</th>
<th>Tapered estimator bias</th>
<th>var</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
<td>-0.0022</td>
<td>0.0058</td>
<td>0.0020</td>
<td>0.0094</td>
</tr>
<tr>
<td>0.0</td>
<td>0.5</td>
<td>0.0994</td>
<td>0.0061</td>
<td>0.1130</td>
<td>0.0098</td>
</tr>
<tr>
<td>0.0</td>
<td>0.8</td>
<td>0.4133</td>
<td>0.0072</td>
<td>0.4404</td>
<td>0.0114</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0</td>
<td>0.0001</td>
<td>0.0058</td>
<td>-0.0030</td>
<td>0.0097</td>
</tr>
<tr>
<td>0.4</td>
<td>0.5</td>
<td>0.1003</td>
<td>0.0060</td>
<td>0.1055</td>
<td>0.0099</td>
</tr>
<tr>
<td>0.4</td>
<td>0.8</td>
<td>0.4160</td>
<td>0.0072</td>
<td>0.4381</td>
<td>0.0113</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0</td>
<td>-0.0003</td>
<td>0.0058</td>
<td>-0.0066</td>
<td>0.0095</td>
</tr>
<tr>
<td>0.8</td>
<td>0.5</td>
<td>0.0988</td>
<td>0.0060</td>
<td>0.1014</td>
<td>0.0098</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
<td>0.4125</td>
<td>0.0073</td>
<td>0.4325</td>
<td>0.0113</td>
</tr>
<tr>
<td>1.2</td>
<td>0.0</td>
<td>-0.0006</td>
<td>0.0057</td>
<td>-0.0057</td>
<td>0.0093</td>
</tr>
<tr>
<td>1.2</td>
<td>0.5</td>
<td>0.0990</td>
<td>0.0061</td>
<td>0.1022</td>
<td>0.0099</td>
</tr>
<tr>
<td>1.2</td>
<td>0.8</td>
<td>0.4117</td>
<td>0.0070</td>
<td>0.4302</td>
<td>0.0108</td>
</tr>
</tbody>
</table>

Table 3. Simulation results: \( n = 128, m = n^{0.65} = 23 \)

<table>
<thead>
<tr>
<th>( d )</th>
<th>( \text{bias} )</th>
<th>( \text{s.d.} )</th>
<th>( \text{MSE} )</th>
<th>( \text{bias} )</th>
<th>( \text{s.d.} )</th>
<th>( \text{MSE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>0.0043</td>
<td>0.1369</td>
<td>0.0188</td>
<td>-0.0004</td>
<td>0.1383</td>
<td>0.0191</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0007</td>
<td>0.1397</td>
<td>0.0195</td>
<td>0.0001</td>
<td>0.1385</td>
<td>0.0192</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0004</td>
<td>0.1404</td>
<td>0.0197</td>
<td>0.0052</td>
<td>0.1381</td>
<td>0.0191</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.0008</td>
<td>0.1395</td>
<td>0.0195</td>
<td>0.0031</td>
<td>0.1338</td>
<td>0.0179</td>
</tr>
<tr>
<td>1.0</td>
<td>0.0006</td>
<td>0.1405</td>
<td>0.0197</td>
<td>0.0015</td>
<td>0.1377</td>
<td>0.0190</td>
</tr>
<tr>
<td>1.2</td>
<td>-0.0004</td>
<td>0.1390</td>
<td>0.0193</td>
<td>-0.0003</td>
<td>0.1386</td>
<td>0.0192</td>
</tr>
<tr>
<td>1.6</td>
<td>0.0023</td>
<td>0.1381</td>
<td>0.0191</td>
<td>0.0031</td>
<td>0.1380</td>
<td>0.0191</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( d )</th>
<th>( \text{bias} )</th>
<th>( \text{s.d.} )</th>
<th>( \text{MSE} )</th>
<th>( \text{bias} )</th>
<th>( \text{s.d.} )</th>
<th>( \text{MSE} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>-0.0108</td>
<td>0.1340</td>
<td>0.0181</td>
<td>0.0434</td>
<td>0.1740</td>
<td>0.0322</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0444</td>
<td>0.1481</td>
<td>0.0239</td>
<td>0.0115</td>
<td>0.1757</td>
<td>0.0310</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.0426</td>
<td>0.1550</td>
<td>0.0258</td>
<td>-0.0042</td>
<td>0.1783</td>
<td>0.0318</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.0168</td>
<td>0.1536</td>
<td>0.0239</td>
<td>-0.0164</td>
<td>0.1787</td>
<td>0.0322</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.0034</td>
<td>0.1442</td>
<td>0.0208</td>
<td>-0.0163</td>
<td>0.1783</td>
<td>0.0321</td>
</tr>
<tr>
<td>1.2</td>
<td>-0.0002</td>
<td>0.1398</td>
<td>0.0195</td>
<td>-0.0193</td>
<td>0.1757</td>
<td>0.0312</td>
</tr>
<tr>
<td>1.6</td>
<td>0.0132</td>
<td>0.1342</td>
<td>0.0182</td>
<td>-0.0074</td>
<td>0.1732</td>
<td>0.0301</td>
</tr>
</tbody>
</table>
Table 4. Simulation results: \( n = 512, m = n^{0.65} = 57 \)

<table>
<thead>
<tr>
<th>(d)</th>
<th>ELW bias</th>
<th>s.d.</th>
<th>MSE</th>
<th>ELW-F2 bias</th>
<th>s.d.</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>-0.0023</td>
<td>0.0765</td>
<td>0.0059</td>
<td>-0.0039</td>
<td>0.0764</td>
<td>0.0059</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0021</td>
<td>0.0774</td>
<td>0.0060</td>
<td>-0.0020</td>
<td>0.0774</td>
<td>0.0060</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.0022</td>
<td>0.0772</td>
<td>0.0060</td>
<td>-0.0003</td>
<td>0.0765</td>
<td>0.0059</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.0016</td>
<td>0.0771</td>
<td>0.0059</td>
<td>-0.0008</td>
<td>0.0762</td>
<td>0.0058</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.0024</td>
<td>0.0768</td>
<td>0.0059</td>
<td>-0.0024</td>
<td>0.0767</td>
<td>0.0059</td>
</tr>
<tr>
<td>1.2</td>
<td>-0.0005</td>
<td>0.0768</td>
<td>0.0059</td>
<td>-0.0004</td>
<td>0.0769</td>
<td>0.0059</td>
</tr>
<tr>
<td>1.6</td>
<td>-0.0008</td>
<td>0.0772</td>
<td>0.0060</td>
<td>-0.0007</td>
<td>0.0772</td>
<td>0.0060</td>
</tr>
</tbody>
</table>

ELW-F2 with detrending Tapered estimator

<table>
<thead>
<tr>
<th>(d)</th>
<th>ELW bias</th>
<th>s.d.</th>
<th>MSE</th>
<th>ELW-F2 bias</th>
<th>s.d.</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>-0.0078</td>
<td>0.0759</td>
<td>0.0058</td>
<td>0.0131</td>
<td>0.0962</td>
<td>0.0094</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0214</td>
<td>0.0815</td>
<td>0.0071</td>
<td>0.0037</td>
<td>0.0977</td>
<td>0.0096</td>
</tr>
<tr>
<td>0.4</td>
<td>-0.0190</td>
<td>0.0818</td>
<td>0.0071</td>
<td>-0.0049</td>
<td>0.0984</td>
<td>0.0097</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.0059</td>
<td>0.0802</td>
<td>0.0065</td>
<td>-0.0069</td>
<td>0.0985</td>
<td>0.0097</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.0035</td>
<td>0.0774</td>
<td>0.0060</td>
<td>-0.0086</td>
<td>0.0973</td>
<td>0.0095</td>
</tr>
<tr>
<td>1.2</td>
<td>0.0001</td>
<td>0.0769</td>
<td>0.0059</td>
<td>-0.0058</td>
<td>0.0966</td>
<td>0.0094</td>
</tr>
<tr>
<td>1.6</td>
<td>0.0060</td>
<td>0.0770</td>
<td>0.0060</td>
<td>-0.0011</td>
<td>0.0957</td>
<td>0.0092</td>
</tr>
</tbody>
</table>

Table 5. Simulation results with Type I processes

\( n = 128, m = n^{0.65} = 23 \)

<table>
<thead>
<tr>
<th>(d)</th>
<th>ELW bias</th>
<th>s.d.</th>
<th>MSE</th>
<th>ELW-F2 bias</th>
<th>s.d.</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>0.0023</td>
<td>0.1473</td>
<td>0.0217</td>
<td>0.0129</td>
<td>0.1405</td>
<td>0.0199</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0004</td>
<td>0.1391</td>
<td>0.0193</td>
<td>0.0003</td>
<td>0.1386</td>
<td>0.0192</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0050</td>
<td>0.1559</td>
<td>0.0243</td>
<td>0.0110</td>
<td>0.1379</td>
<td>0.0191</td>
</tr>
<tr>
<td>0.8</td>
<td>0.0008</td>
<td>0.1418</td>
<td>0.0201</td>
<td>0.0048</td>
<td>0.1349</td>
<td>0.0182</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.0004</td>
<td>0.1393</td>
<td>0.0194</td>
<td>0.0003</td>
<td>0.1375</td>
<td>0.0189</td>
</tr>
<tr>
<td>1.2</td>
<td>0.0014</td>
<td>0.1401</td>
<td>0.0196</td>
<td>0.0016</td>
<td>0.1385</td>
<td>0.0192</td>
</tr>
<tr>
<td>1.6</td>
<td>-0.0013</td>
<td>0.1482</td>
<td>0.0220</td>
<td>-0.0010</td>
<td>0.1477</td>
<td>0.0218</td>
</tr>
</tbody>
</table>

\( n = 512, m = n^{0.65} = 57 \)

<table>
<thead>
<tr>
<th>(d)</th>
<th>ELW bias</th>
<th>s.d.</th>
<th>MSE</th>
<th>ELW-F2 bias</th>
<th>s.d.</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.4</td>
<td>0.0010</td>
<td>0.0794</td>
<td>0.0063</td>
<td>0.0057</td>
<td>0.0785</td>
<td>0.0062</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.0025</td>
<td>0.0781</td>
<td>0.0061</td>
<td>-0.0024</td>
<td>0.0781</td>
<td>0.0061</td>
</tr>
<tr>
<td>0.4</td>
<td>0.0022</td>
<td>0.0797</td>
<td>0.0064</td>
<td>0.0011</td>
<td>0.0756</td>
<td>0.0057</td>
</tr>
<tr>
<td>0.8</td>
<td>-0.0016</td>
<td>0.0775</td>
<td>0.0060</td>
<td>-0.0013</td>
<td>0.0766</td>
<td>0.0059</td>
</tr>
<tr>
<td>1.0</td>
<td>-0.0019</td>
<td>0.0765</td>
<td>0.0059</td>
<td>-0.0020</td>
<td>0.0766</td>
<td>0.0059</td>
</tr>
<tr>
<td>1.2</td>
<td>-0.0008</td>
<td>0.0774</td>
<td>0.0060</td>
<td>-0.0007</td>
<td>0.0774</td>
<td>0.0060</td>
</tr>
<tr>
<td>1.6</td>
<td>0.0001</td>
<td>0.0796</td>
<td>0.0063</td>
<td>0.0002</td>
<td>0.0795</td>
<td>0.0063</td>
</tr>
</tbody>
</table>
Table 6: Estimates of $d$ for US Economic Data: $m = n^{0.7}$

<table>
<thead>
<tr>
<th></th>
<th>$n$</th>
<th>LW</th>
<th>FEIW2</th>
<th>95% asy. CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real GNP</td>
<td>80</td>
<td>1.077</td>
<td>1.126</td>
<td>[0.912, 1.340]</td>
</tr>
<tr>
<td>Nominal GNP</td>
<td>80</td>
<td>1.273</td>
<td>1.303</td>
<td>[1.089, 1.517]</td>
</tr>
<tr>
<td>Real per capita GNP</td>
<td>80</td>
<td>1.077</td>
<td>1.128</td>
<td>[0.914, 1.342]</td>
</tr>
<tr>
<td>Industrial production</td>
<td>129</td>
<td>0.821</td>
<td>0.850</td>
<td>[0.671, 1.029]</td>
</tr>
<tr>
<td>Employment</td>
<td>99</td>
<td>0.968</td>
<td>1.000</td>
<td>[0.800, 1.200]</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>129</td>
<td>0.951</td>
<td>0.980</td>
<td>[0.801, 1.159]</td>
</tr>
<tr>
<td>GNP deflator</td>
<td>100</td>
<td>1.374</td>
<td>1.398</td>
<td>[1.202, 1.594]</td>
</tr>
<tr>
<td>CPI</td>
<td>129</td>
<td>1.273</td>
<td>1.287</td>
<td>[1.109, 1.466]</td>
</tr>
<tr>
<td>Nominal wage</td>
<td>89</td>
<td>1.300</td>
<td>1.351</td>
<td>[1.147, 1.555]</td>
</tr>
<tr>
<td>Real wage</td>
<td>89</td>
<td>1.047</td>
<td>1.089</td>
<td>[0.885, 1.293]</td>
</tr>
<tr>
<td>Money stock</td>
<td>100</td>
<td>1.460</td>
<td>1.501</td>
<td>[1.305, 1.697]</td>
</tr>
<tr>
<td>Velocity of money</td>
<td>120</td>
<td>0.953</td>
<td>0.993</td>
<td>[0.808, 1.179]</td>
</tr>
<tr>
<td>Bond yield</td>
<td>89</td>
<td>1.091</td>
<td>1.108</td>
<td>[0.903, 1.312]</td>
</tr>
<tr>
<td>Stock prices</td>
<td>118</td>
<td>0.900</td>
<td>0.958</td>
<td>[0.772, 1.143]</td>
</tr>
</tbody>
</table>