# Mixtures of AR(1) Components: Sieve–Whittle Estimation, Support Localization, and a Closed–Form EM Weight Update

Zhaoyang Xu

October 28, 2025

**Abstract**

We model persistence by representing the spectral density of a stationary series as a nonnegative mixture of AR(1) kernels plus an optional white atom. We study a sieve–Whittle estimator that places probability mass on a grid of poles $\rho \in (0, 1)$ and develop a Gaussian EM routine with a closed–form M–step for the mixture weights under a unit–variance constraint. Our theory delivers: (i) a *uniform* law of large numbers for the Whittle criterion over the AR(1)+white sieve—both on a fixed boundary and on a *shrinking* boundary that allows $\rho \uparrow 1$ at rate $\delta_T \downarrow 0$—implying consistency of spectra and weak convergence of mixing measures; (ii) *support localization* under separated true poles, so simple barycentric averages of nearby grid points recover pole locations and weights; and (iii) an explicit one–dimensional dual update for the EM M–step that guarantees existence, uniqueness, and monotone ascent of the observed likelihood. Simulations verify the uniform LLN, localization, and EM monotonicity. In an application to quarterly U.S. unemployment changes, the mixture improves out-of-sample MSE and log score relative to a BIC-selected AR(2). The approach is positivity-preserving, interpretable, and computationally robust.

## 1 Introduction

Macroeconomic and financial series often combine heterogeneous persistence: slow-moving cycle components, medium-run dynamics, and high-frequency noise. We model this structure by writing the observed process as a sum of latent AR(1) states and, optionally, white noise. The second-order implications are transparent: the autocovariance is a finite sum of exponentials and the spectral density is a nonnegative mixture of AR(1) (Poisson) kernels plus a white atom. This positivity-preserving representation is interpretable, adapts naturally to near-unit-root behavior, and yields simple likelihoods in both the frequency and state-space domains.

**This paper.** We develop estimation, identification, and inference for AR(1)+white mixtures via a sieve–Whittle approach and a complementary Gaussian EM routine. Our goal is a method that (i) remains well-behaved near $\rho = 1$, (ii) produces components with a clear persistence interpretation, and (iii) is easy to compute at scale.

**Contributions.**

1. **Uniform Whittle LLN on fixed and shrinking boundaries.** We approximate the spectrum by a convex combination of AR(1) atoms on a grid and index the empirical spectral

process by $\{1/f : f \in \mathcal{F}_{\hat{K}}\}$. We prove a *uniform* law of large numbers for the Whittle criterion over the sieve not only under a fixed boundary $\rho \le 1-\delta_\star$ but also under a *shrinking boundary* $\rho \le 1-\delta_T \downarrow 1$, provided $\delta_T$ declines slowly enough. The discrete Fourier version is uniformly equivalent to the population criterion under the periodogram scaling we use. These results imply consistency of the spectral fit and weak convergence of the estimated mixing measures.

2. **Support localization and barycentric pole estimates.** When true poles are separated, sieve mass concentrates in small neighborhoods of each pole. A simple *barycentric* average of grid points within the neighborhood consistently recovers both pole locations and weights, furnishing a direct sieve→parametric mapping.

3. **Closed-form EM weight update with a sum constraint.** For the Gaussian state-space likelihood with fixed poles, the EM M–step for the weights (and, when present, the observation-noise variance) under the unit-variance constraint reduces to a single dual variable. Each coordinate admits an explicit positive branch and a global bracket, yielding a unique solution and *monotone ascent* of the observed likelihood.

4. **Evidence.** Simulations confirm the uniform LLN and localization and show monotone EM paths. On quarterly U.S. unemployment changes, the mixture improves out-of-sample MSE and log score relative to a BIC-selected AR(2), indicating practical gains from the positivity-preserving decomposition.

**Identification and interpretation.** The model is identified from second-order structure. The covariance generating function is rational with simple poles at $1/\rho_k$, so the pole locations and residues (hence weights) are unique; with additive white noise, lags $h \ge 1$ identify the mixture and lag 0 recovers the noise variance. Under Gaussianity, the full law is identified. The nonnegativity of weights yields a meaningful variance decomposition across persistence sources and avoids sign indeterminacy common in factor-type models.

**Related work.** Our analysis builds on uniform limit theory for the Whittle likelihood via the empirical spectral process (e.g. Dahlhaus, 1988; Brillinger, 2001; Dahlhaus, 2009) and on classical identification for sums of exponentials (Carathéodory–Fejér/Prony–Pisarenko) in Toeplitz/Vandermonde decompositions. Relative to semiparametric or Bayesian spectral mixtures (e.g., spline or Dirichlet-process priors), our sieve is explicitly *positivity-preserving*, accommodates a *shrinking-boundary* analysis toward the unit root, and provides an elementary *barycentric* localization result. On the likelihood side, we give an *explicit* EM M–step under a unit-variance constraint with a global bracket, which to our knowledge has not been recorded for this model family.

**Organization.** Section 2 formalizes the model and identification. Section 3 presents the sieve–Whittle estimator, the uniform LLN on fixed and shrinking boundaries, and the discrete approximation. Section 4 proves support localization and barycentric pole estimation. Section 5 derives the closed-form EM update and shows monotonicity. Section 6 reports simulations and the unemployment application.

## 2 Model and identification

We consider $K$ latent AR(1) states,

$$X_{k,t+1} = \rho_k X_{k,t} + \varepsilon_{k,t}, \qquad 0 < \rho_1 < \cdots < \rho_K < 1, \tag{1}$$

with innovations mutually independent across $k$ and $t$, $\mathbb{E}\varepsilon_{k,t} = 0$, and independent of the initial states. We impose the standardization

$$\mathrm{Var}(\varepsilon_{k,t}) = 1 - \rho_k^2, \tag{2}$$

so each $X_{k,\cdot}$ is strictly stationary with $\mathrm{Var}(X_{k,t}) = 1$. Let $w_k := a_k^2 > 0$ and define scaled states

$$\alpha_{k,t} := a_k X_{k,t}\,, \qquad \alpha_{k,t+1} = \rho_k \alpha_{k,t} + \eta_{k,t}\,, \quad \eta_{k,t} := a_k \varepsilon_{k,t} \sim \mathcal{N}\big(0, (1-\rho_k^2) w_k\big).$$

We analyze two measurement settings:

**(A) Noise–free measurement:**

$$y_t = \sum_{k=1}^{K} \alpha_{k,t}. \tag{3}$$

**(B) Measurement with white noise:**

$$y_t = \sum_{k=1}^{K} \alpha_{k,t} + u_t, \qquad u_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \tag{4}$$

independent of the states. In (A), $\mathrm{Var}(y_t) = \sum_k w_k$. In (B), $\mathrm{Var}(y_t) = \sum_k w_k + \sigma^2$.

Throughout we also use the *unit–variance* normalization, obtained by rescaling $y$ so that $\mathrm{Var}(y_t) = 1$: in (A), $\sum_k w_k = 1$; in (B), $\sum_k w_k + \sigma^2 = 1$.

Write the autocovariance function as $\gamma(h) = \mathrm{Cov}(y_t, y_{t+h})$. Under (3),

$$\gamma(h) = \sum_{k=1}^{K} w_k \rho_k^{|h|}, \qquad h \in \mathbb{Z}. \tag{5}$$

Under (4), white noise affects only lag 0:

$$\gamma(h) = \begin{cases} \sum_{k=1}^{K} w_k \rho_k^h, & h \geq 1, \\ \sum_{k=1}^{K} w_k + \sigma^2, & h = 0. \end{cases} \tag{6}$$

**Theorem 2.1** (Identification without measurement noise)**.** *Under* (1)–(3)*, the* second–order structure *of* $\{y_t\}$*, equivalently the autocovariances* $\{\gamma(h)\}_{h\geq 0}$*, identifies* $(K, \{\rho_k\}, \{w_k\})$*; the signs of* $a_k$ *are not identified. If in addition* $\{y_t\}$ *is Gaussian, then the full law is identified.*

**Theorem 2.2** (Identification with additive white noise)**.** *Under* (1) *and* (4)*, the* second–order structure *of* $\{y_t\}$ *identifies* $(K, \{\rho_k\}, \{w_k\}, \sigma^2)$ *up to signs of* $a_k$*. Equivalently,* $\{\gamma(h)\}_{h\geq 1}$ *identifies* $(K, \{\rho_k\}, \{w_k\})$*, and then* $\sigma^2 = \gamma(0) - \sum_k w_k$*. If in addition* $\{y_t\}$ *is Gaussian, then the full law is identified.*

**Remark 2.3** (Identification is classical)**.** *Theorems 2.1–2.2 restate a classical fact: the covariance generating function* $G(z) = \sum_{h\geq 0} \gamma(h) z^h$ *of a finite AR(1) mixture is a rational function whose poles and residues determine* $(\rho_k, w_k)$*. This is consistent with Toeplitz/Vandermonde decompositions of positive semidefinite Toeplitz matrices; see, e.g., monographs on spectral estimation and sums of exponentials.*

## 3 Spectral representation and sieve–Whittle estimation

For a (mean–zero) Gaussian stationary process with spectral density $f$, the Whittle criterion is

$$\mathcal{L}_T(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\log f(\omega) + \frac{I_T(\omega)}{f(\omega)}\right\} d\omega,$$

where $I_T$ is the periodogram. Under (1) and either measurement setting, the spectral density has the *mixture* form

$$f(\omega) = \frac{1}{2\pi}\sum_{j=0}^{K}\theta_j\,\phi_j(\omega), \qquad \phi_0(\omega)\equiv 1, \quad \phi_j(\omega) := \frac{1-\rho_j^2}{1+\rho_j^2-2\rho_j\cos\omega}\ (j\geq 1), \tag{7}$$

with $\theta_j \geq 0$ and $\sum_j \theta_j = \mathrm{Var}(y_t)$. Note that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_j(\omega)\,d\omega = 1 \qquad \text{for every } j\geq 0,$$

so under the *unit–variance* normalization $\mathrm{Var}(y_t) = 1$ we have $\sum_j \theta_j = 1$ and $f$ is a convex combination of the atoms $\phi_j/(2\pi)$.

**Remark 3.1** (Discrete Whittle scaling and periodogram normalization)**.** *In computations we evaluate the discrete Whittle criterion at Fourier frequencies* $\omega_j = 2\pi j/T$ *using the classical periodogram*

$$I_T(\omega_j) \;=\; \frac{1}{2\pi T}\left|\sum_{t=1}^{T} y_t\, e^{-i\omega_j t}\right|^2.$$

*With this convention the discrete objective is*

$$\widehat{\mathcal{L}}_T(f) = \frac{1}{M}\sum_{j=1}^{M}\left\{\log s(\omega_j) + \frac{2\pi\, I_T(\omega_j)}{s(\omega_j)}\right\}, \qquad s(\omega) = 2\pi f(\omega),$$

*i.e. the integrand* $\log f + I_T/f$ *becomes* $\log s + (2\pi)I_T/s$ *in terms of* $s$*. Omitting the factor* $2\pi$ *in finite samples can erroneously favor near–unit–root atoms; the scaling above is consistent with* (7) *and the standard periodogram normalization.*

Fix a deterministic grid $\mathcal{G}_{\hat{K}} = \{\hat\rho_1,\dots,\hat\rho_{\hat K}\}\subset(0,1)$ and include the white–noise atom $\phi_0\equiv 1$. We work either on a fixed boundary $\hat\rho_j \leq 1-\delta_\star$ or a *shrinking boundary* $\hat\rho_j \leq 1-\delta_T$ with $\delta_T \downarrow 0$. Define the sieve of unit–variance spectra

$$\mathcal{F}_{\hat K} := \left\{ f_{\hat p}(\omega) = \frac{1}{2\pi}\sum_{j=0}^{\hat K}\hat p_j\,\phi_j(\omega)\ :\ \hat p_j \geq 0,\ \sum_{j=0}^{\hat K}\hat p_j = 1\right\}, \tag{8}$$

with $\phi_0 \equiv 1$ and $\phi_j(\omega) = \phi(\hat\rho_j,\omega)$ for $j\geq 1$.

Define the population functional

$$J(f) := \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\log f(\omega) + \frac{f^*(\omega)}{f(\omega)}\right\} d\omega, \tag{9}$$

where $f^*$ is the true spectral density.

The kernels satisfy, for $\rho\in[0,1)$ and $\omega\in[-\pi,\pi]$,

$$\min_\omega \phi(\rho,\omega) = \frac{1-\rho}{1+\rho}, \qquad \max_\omega \phi(\rho,\omega) = \frac{1+\rho}{1-\rho}. \tag{10}$$

Let $\Delta_{\hat K} := \max_j(\hat\rho_{j+1}-\hat\rho_j)$ denote the mesh size of $\mathcal{G}_{\hat K}$.

### 3.1 A uniform Whittle LLN via the empirical spectral process

The naive bound using $\int |I_T - 2\pi f^*|$ fails for uniformity (the raw periodogram is classically inconsistent in $L^p$ without smoothing; see, e.g., Brillinger, 2001, §5.3). The correct approach indexes the empirical spectral process by

$$\mathcal{G} := \left\{ g_f(\cdot) := \frac{1}{f(\cdot)} \; : \; f \in \mathcal{F}_{\hat{K}} \right\}$$

and invokes a Glivenko–Cantelli property for bounded–variation test functions.

**Theorem 3.2** (Empirical spectral process: uniform LLN on bounded–variation classes)**.** *Let $\{y_t\}$ be (mean–zero) Gaussian stationary with spectral density $f^*$ and periodogram $I_T$. For a class $\mathcal{G}$ of real functions on $[-\pi, \pi]$ assume*

$$M := \sup_{g \in \mathcal{G}} \|g\|_\infty < \infty, \qquad V := \sup_{g \in \mathcal{G}} \mathrm{Var}_{\mathrm{Tot}}(g) < \infty.$$

*Then*

$$\sup_{g \in \mathcal{G}} \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\omega) \big( I_T(\omega) - 2\pi f^*(\omega) \big) \, d\omega \right| \; = \; O_{\mathbb{P}}\!\left( \|f^*\|_\infty \, (M + V) \sqrt{\frac{\log T}{T}} \right). \tag{11}$$

**Lemma 3.3** (Uniform Whittle LLN on the sieve)**.** *Under Theorem 3.2 and on the fixed–boundary sieve,*

$$\sup_{f \in \mathcal{F}_{\hat{K}}} \big| \mathcal{L}_T(f) - J(f) \big| \;\xrightarrow{\mathbb{P}}\; 0.$$

*On the shrinking–boundary sieve, the same conclusion holds provided $\delta_T^{-5} \sqrt{\log T / T} \to 0$.*

**Lemma 3.4** (Uniform LLN for the discrete Whittle criterion)**.** *Let $\omega_j = 2\pi j/T$ and $M = \lfloor (T-1)/2 \rfloor$. Define*

$$\widehat{\mathcal{L}}_T(f) = \frac{1}{M} \sum_{j=1}^{M} \left\{ \log(2\pi f(\omega_j)) + \frac{2\pi I_T(\omega_j)}{2\pi f(\omega_j)} \right\}.$$

*On the sieve $\mathcal{F}_{\hat{K}}$ of (8), under a fixed boundary $\hat{\rho}_j \le 1 - \delta_\star$ or a shrinking one $\hat{\rho}_j \le 1 - \delta_T \downarrow 1$ with $\delta_T^{-5} \sqrt{\log T / T} \to 0$,*

$$\sup_{f \in \mathcal{F}_{\hat{K}}} \big| \widehat{\mathcal{L}}_T(f) - J(f) \big| = O_{\mathbb{P}}\big( \delta_T^{-5} \sqrt{\tfrac{\log T}{T}} \big). \tag{12}$$

### 3.2 Consistency on a fixed or shrinking boundary

**Assumption 3.5** (True support covered by the sieve)**.** *The true process is generated by (1) and either (3) or (4), Gaussian, with unit variance; its spectral density $f^*$ admits the mixture representation (7) with support $\{\rho_k^*\} \subset (0, 1 - \delta_\star]$ (for some fixed $\delta_\star \in (0,1)$). The sieves use grids $\mathcal{G}_{\hat{K}} \subset (0, 1 - \delta_\star]$ with mesh $\Delta_{\hat{K}} \to 0$.*

Let $\hat{f}_T \in \mathcal{F}_{\hat{K}}$ minimize $\mathcal{L}_T(f)$ and denote the corresponding mixing distribution on $\{0, \hat{\rho}_1, \dots, \hat{\rho}_{\hat{K}}\}$ by $\hat{F}_T = \sum_{j=0}^{\hat{K}} \hat{p}_{j,T} \delta_{\hat{\rho}_j}$ (with $\hat{\rho}_0 := 0$ for the white–noise atom).

**Theorem 3.6** (Consistency on a fixed boundary)**.** *Under Assumption 3.5, $\hat{f}_T \xrightarrow{\mathbb{P}} f^*$ in the sup–norm. Moreover, every subsequence of $\hat{F}_T$ has a further subsequence converging weakly to the true mixing distribution $F^*$; hence $\hat{F}_T \Rightarrow F^*$ in probability.*

**Assumption 3.7** (Shrinking coverage and rates)**.** *There exists $\delta_T \downarrow 0$ such that: (i) for all large $T$, the true poles satisfy $\max_k \rho_k^* \le 1 - \delta_T$; (ii) the sieves use grids $\mathcal{G}_{\hat{K}} \subset (0, 1 - \delta_T]$ with mesh $\Delta_{\hat{K}} \to 0$; (iii) the empirical spectral process envelope (Theorem 3.2) applied to $\mathcal{G} = \{1/f : f \in \mathcal{F}_{\hat{K}}\}$ vanishes at the rate*

$$\delta_T^{-5}\sqrt{\log T/T} \ \to \ 0;$$

*(iv) the grid resolves the shrinking boundary: $\Delta_{\hat{K}}\,\delta_T^{-2} \to 0$; (v) (for the discrete Whittle equivalence in Lemma 3.4) $T\,\delta_T \to \infty$.*

**Theorem 3.8** (Consistency with a shrinking boundary)**.** *Under Assumption 3.7, any Whittle minimizer $\hat{f}_T \in \mathcal{F}_{\hat{K}}$ satisfies $\|\hat{f}_T - f^*\|_\infty \xrightarrow{\mathbb{P}} 0$; the associated mixing distributions satisfy $\hat{F}_T \Rightarrow F^*$ in probability.*

**Corollary 3.9** (Consistency of the discrete Whittle estimator)**.** *Let $\hat{f}_T^{\text{disc}} \in \mathcal{F}_{\hat{K}}$ minimize $\widehat{\mathcal{L}}_T(f)$. Then under Assumption 3.5 or 3.7,*

$$\|\hat{f}_T^{\text{disc}} - f^*\|_\infty \xrightarrow{\mathbb{P}} 0, \qquad \hat{F}_T^{\text{disc}} \Rightarrow F^*.$$

### 3.3 Asymptotic normality on a fixed grid (finite dimension)

For completeness we record the finite–dimensional limit theory when the grid (including the white–noise atom) is fixed and $f^* \in \mathcal{F}_{\hat{K}}$ with all mixture weights strictly interior.

Let $\boldsymbol{p} = (\hat{p}_0, \ldots, \hat{p}_{\hat{K}})'$ denote the weights with $\sum_j p_j = 1$. Parameterize by $\tilde{\boldsymbol{p}} = (p_0, \ldots, p_{\hat{K}-1})'$ and $p_{\hat{K}} = 1 - \mathbf{1}'\tilde{\boldsymbol{p}}$; define basis functions $b_j(\omega) := \phi_j(\omega) - \phi_{\hat{K}}(\omega)$ for $j = 0, \ldots, \hat{K} - 1$. Then

$$f_{\tilde{\boldsymbol{p}}}(\omega) = \frac{1}{2\pi}\left[\phi_{\hat{K}}(\omega) + \sum_{j=0}^{\hat{K}-1} \tilde{p}_j\, b_j(\omega)\right], \qquad \frac{\partial \log f_{\tilde{\boldsymbol{p}}}(\omega)}{\partial \tilde{p}_j} = \frac{b_j(\omega)}{f_{\tilde{\boldsymbol{p}}}(\omega)}.$$

**Theorem 3.10** (Asymptotic normality on a fixed grid)**.** *Assume $f^* = f_{\tilde{\boldsymbol{p}}^*} \in \mathcal{F}_{\hat{K}}$ with $\tilde{\boldsymbol{p}}^*$ strictly interior. Then any Whittle minimizer $\hat{\tilde{\boldsymbol{p}}}_T$ satisfies*

$$\sqrt{T}\,(\hat{\tilde{\boldsymbol{p}}}_T - \tilde{\boldsymbol{p}}^*) \ \xrightarrow{\ \mathcal{D}\ } \ \mathcal{N}\big(0,\ I^{-1}(\tilde{\boldsymbol{p}}^*)\big),$$

*with Fisher information*

$$I(\tilde{\boldsymbol{p}}^*) = \frac{1}{4\pi}\int_{-\pi}^{\pi} \boldsymbol{s}(\omega)\,\boldsymbol{s}(\omega)'\,d\omega, \qquad \boldsymbol{s}(\omega) := \left(\frac{b_j(\omega)}{f_{\tilde{\boldsymbol{p}}^*}(\omega)}\right)_{j=0}^{\hat{K}-1}.$$

## 4 Support localization and barycentric pole estimation

Assume

$$f^*(\omega) = \frac{1}{2\pi}\sum_{k=0}^{K} \theta_k^*\,\phi(\rho_k^*, \omega), \qquad 0 < \rho_1^* < \cdots < \rho_K^* < 1.$$

Let $\tau = \min_{k \ne \ell} |\rho_k^* - \rho_\ell^*| > 0$.

**Theorem 4.1** (Support localization)**.** *Assume*

$$f^*(\omega) = \frac{1}{2\pi}\sum_{k=0}^{K} \theta_k^*\,\phi(\rho_k^*, \omega), \qquad 0 = \rho_0^* < \rho_1^* < \cdots < \rho_K^* < 1,$$

*with separation* $\tau := \min_{k\neq\ell} |\rho_k^* - \rho_\ell^*| > 0$*. Suppose the sieve/grid and boundary conditions of Theorem 3.6 or Theorem 3.8 hold so that* $\hat{F}_T \Rightarrow F^*$ *in probability. Fix* $\varepsilon \in (0, \tau/2)$*. Then, for* $k = 0, 1, \ldots, K$*,*

$$\hat{F}_T\big(B_\varepsilon(\rho_k^*)\big) \xrightarrow{\mathbb{P}} \theta_k^*.$$

*Equivalently,* $\hat{F}_T(C) \to F^*(C)$ *in probability for every continuity set* $C$ *of* $F^*$*.*

**Corollary 4.2.** *Under the assumptions of Theorem 4.1, for every* $\delta \in (0, \rho_1^*/2)$*,*

$$\hat{F}_T\big([0, \delta]\big) \xrightarrow{\mathbb{P}} \theta_0^*.$$

**Corollary 4.3** (Barycentric poles)**.** *Fix* $\varepsilon \in (0, \tau/2)$ *and suppose the sieve grid satisfies* $\Delta_{\hat{K}} \to 0$*. For* $k = 1, \ldots, K$ *define*

$$\hat{\rho}_{k,T} = \frac{\sum_{j:\,|\hat{\rho}_j - \rho_k^*| \le \varepsilon} \hat{p}_{j,T}\, \hat{\rho}_j}{\sum_{j:\,|\hat{\rho}_j - \rho_k^*| \le \varepsilon} \hat{p}_{j,T}}, \quad \hat{\theta}_{k,T} = \sum_{j:\,|\hat{\rho}_j - \rho_k^*| \le \varepsilon} \hat{p}_{j,T}.$$

*Then* $\hat{\rho}_{k,T} \to \rho_k^*$ *and* $\hat{\theta}_{k,T} \to \theta_k^*$ *in probability.*

# 5 Gaussian pseudo–likelihood and EM for the weights

We work with the Gaussian state–space likelihood for the model (1) and either (3) or (4). Let $T = \mathrm{diag}(\rho_1, \ldots, \rho_K)$, $Z = (1, \ldots, 1)$, $Q = \mathrm{diag}((1-\rho_k^2)w_k)$, and in the noisy case let $H = \sigma^2$. The state vector is $\alpha_t = (\alpha_{1,t}, \ldots, \alpha_{K,t})'$.

At iteration $m$ of EM we keep $(K, \boldsymbol{\rho})$ fixed and update $W = (w_1, \ldots, w_K)'$ and, in the noisy setting, $\sigma^2$.

## 5.1 Complete–data log–likelihood and the EM $Q$–function

Let $Y_T = (y_1, \ldots, y_T)'$ and denote the initial state by $\zeta = (\zeta_1, \ldots, \zeta_K)'$ with $\zeta_k := \alpha_{k,1}$. The complete–data log–likelihood (up to constants) decomposes into independent scalar contributions in $k$, and (in the noisy case) an independent contribution from $u_t$:

$$\begin{aligned}\log p_\theta(\alpha_{1:T}, Y_T) = &-\frac{1}{2}\sum_{k=1}^{K}\left\{\log w_k + \frac{\zeta_k^2}{w_k} + \sum_{t=2}^{T}\Big(\log\big((1-\rho_k^2)w_k\big) + \frac{\eta_{t-1,k}^2}{(1-\rho_k^2)w_k}\Big)\right\}\\ &-\frac{1}{2}\sum_{t=1}^{T}\left(\log\sigma^2 + \frac{u_t^2}{\sigma^2}\right),\end{aligned}$$

where $\eta_{t-1,k}$ are the state disturbances and $u_t := y_t - Z\alpha_t$ is the observation disturbance (present only under (4)). Dropping constants in $(W, \sigma^2)$ and taking conditional expectations given $Y_T$ under the current parameter (superscript $(m)$) yields

$$Q(W, \sigma^2 \mid \theta^{(m)}) = -\frac{1}{2}\sum_{k=1}^{K}\Big(T\log w_k + \frac{A_k^{(m)}}{w_k}\Big) \;-\; \frac{1}{2}\Big(T\log\sigma^2 + \frac{B^{(m)}}{\sigma^2}\Big) + C, \tag{13}$$

with

$$A_k^{(m)} := \mathbb{E}_{\theta^{(m)}}\left[\zeta_k^2 \,\middle|\, Y_T\right] + \frac{1}{1-\rho_k^2}\sum_{t=2}^{T}\mathbb{E}_{\theta^{(m)}}\left[\eta_{t-1,k}^2 \,\middle|\, Y_T\right], \tag{14}$$

$$B^{(m)} := \sum_{t=1}^{T}\mathbb{E}_{\theta^{(m)}}\left[u_t^2 \,\middle|\, Y_T\right]. \tag{15}$$

The conditional expectations are computed by standard disturbance smoothing for both state and observation disturbances; formulas are recalled in Appendix B. In the noise–free case, $B^{(m)} \equiv 0$ and the last term in (13) is absent.

### 5.2 Closed–form M–steps and their properties

**(a) Unconstrained updates.** Maximizing (13) over $w_k > 0$ and (if present) $\sigma^2 > 0$ gives:

**Proposition 5.1** (Unconstrained EM M–steps)**.** *For fixed $(A_k^{(m)})_{k\le K}$ and $B^{(m)} \ge 0$,*

$$w_k^{(m+1)} = \frac{A_k^{(m)}}{T}, \qquad \sigma^{2(m+1)} = \begin{cases} \frac{B^{(m)}}{T}, & \textit{under } (4), \\ \textit{(absent)}, & \textit{under } (3). \end{cases}$$

*Each update is unique.*

**(b) Unit–variance (sum) constraint.** Under the unit–variance normalization, in the noise–free case we impose $\sum_k w_k = S$ (typically $S = 1$); with measurement noise we impose

$$\sum_{k=1}^{K} w_k + \sigma^2 = S \quad \text{(typically } S = 1\text{)}.$$

Both cases admit the same one–dimensional dual solution:

**Theorem 5.2** (Closed–form EM M–step under a sum constraint)**.** *Fix $S > 0$ and consider the EM Q–function in (13) with fixed $(A_k^{(m)})_{k\le K}$ and $B^{(m)} \ge 0$. Maximize $Q$ over $w_k > 0$ and, if present, $\sigma^2 > 0$, subject to the unit–variance constraint*

$$\sum_{k=1}^{K} w_k = S \quad \textit{(noise–free)} \qquad \textit{or} \qquad \sum_{k=1}^{K} w_k + \sigma^2 = S \quad \textit{(with noise)}.$$

*Let $\lambda$ be the multiplier for the equality constraint. The KKT conditions reduce for each scalar parameter $x \in \{w_k\}$ (and $x = \sigma^2$ if present) to the quadratic*

$$2\lambda x^2 + Tx - C = 0, \qquad C \in \{A_k^{(m)}\} \textit{ or } C = B^{(m)}. \tag{16}$$

*For any*

$$\lambda \;\ge\; \lambda_{\min} := -\frac{T^2}{8\max\{A_{\max}^{(m)},\, B^{(m)}\}}, \qquad A_{\max}^{(m)} := \max_{1\le k\le K} A_k^{(m)},$$

*the unique positive solution of (16) is*

$$x(\lambda) = \frac{2C}{T + \sqrt{T^2 + 8\lambda C}} \quad \textit{(with the convention } x(0) = C/T\textit{)}. \tag{17}$$

*Define $F(\lambda) := \sum_{k=1}^{K} w_k(\lambda)$ in the noise–free case and $F(\lambda) := \sum_{k=1}^{K} w_k(\lambda) + \sigma^2(\lambda)$ when noise is present. Then:*

1. *$x(\lambda)$ is continuous and strictly decreasing on $[\lambda_{\min}, \infty)$; hence $F(\lambda)$ is continuous and strictly decreasing with*

$$\lim_{\lambda\uparrow\infty} F(\lambda) = 0, \qquad F(0) = \frac{\sum_k A_k^{(m)} + B^{(m)}}{T},$$

*and*

$$F(\lambda_{\min}) = \sum_j \frac{2C_j}{T + T\sqrt{1 - C_j/C_{\max}}} \in \Big(F(0),\ 2\textstyle\sum_j \frac{C_j}{T}\Big],$$

*where $C_{\max} = \max_j C_j$ and $C_j$ ranges over $\{A_k^{(m)}\}$ (and $B^{(m)}$ if present).*

2. *For any $S \in \big(0,\ F(\lambda_{\min})\big]$ there exists a unique $\lambda^* \in [\lambda_{\min}, \infty)$ such that $F(\lambda^*) = S$. In particular, if $S \le F(0)$ then $\lambda^* \ge 0$, while if $S > F(0)$ then $\lambda^* \in (\lambda_{\min}, 0)$.*

3. *The associated $(w^{(m+1)}, \sigma^{2(m+1)})$ given by (17) at $\lambda^*$ is the unique global maximizer of $Q$ under the constraint.*

Existence/uniqueness note. *Since $F(0) = \sum_j \frac{C_j}{T}$ and the E-step is computed under parameters satisfying $\sum_k w_k + \sigma^2 = S$, we typically have $F(0) \approx S$. As $F(\lambda_{\min}) > F(0)$, it follows that $S \in (0,\ F(\lambda_{\min})]$, so an interior KKT solution exists and is unique.*

**Proposition 5.3** (EM monotonicity)**.** *Each EM iteration using the M–steps in Propositions 5.1 and Theorem 5.2 increases the observed Gaussian log–likelihood; equality holds if and only if $(W^{(m+1)}, \sigma^{2(m+1)}) = (W^{(m)}, \sigma^{2(m)})$.*

# 6 Simulations and an Empirical Application

This section documents (i) a simulation study that mirrors our theory and (ii) an empirical illustration on quarterly U.S. unemployment (changes). The experiments use the positivity–preserving AR(1) + white dictionary and the Whittle criterion; the periodogram is computed after demeaning and omits the DC and (if $T$ even) Nyquist ordinates.[1] All experiments use the *unit–variance* normalization; mixture weights and $\sigma^2$ therefore sum to 1.

## 6.1 Simulation design and verification of the uniform Whittle LLN

We consider two boundary regimes for the sieve $\mathcal{F}_{\hat{K}}$:

- **Fixed boundary:** grid over $\rho \in [0, 1 - \delta_\star]$ with $\delta_\star = 0.02$ (mesh $\approx 0.005$).
- **Shrinking boundary:** $\rho \le 1 - \delta_T$ with $\delta_T = 1/(1 + \log T)$ and a mesh satisfying $\Delta_{\hat{K}}\, \delta_T^{-2} \to 0$.

For each $T$ we simulate a three–component mixture with white noise (unit variance) and approximate $\sup_{f\in\mathcal{F}_{\hat{K}}} \big|\, \widehat{\mathcal{L}}_T(f) - J_d(f) \,\big|$ by drawing random simplex points (and adding the sieve minimizer). Table 1 shows a sharp decrease on the fixed boundary. On the shrinking boundary (Table 2) the deviations remain very small (below $1.5 \times 10^{-2}$ at $T = 2048$ and below $3 \times 10^{-3}$ by $T = 32768$), broadly consistent with the $O_{\mathbb{P}}\big(\delta_T^{-5}\sqrt{\log T/T}\big)$ envelope from Lemma C.2 and Lemma 3.3.[2]

[1] Implementation follows the scripts archived with the paper: the `periodogram` de–means $y_t$ and evaluates $I_T(\omega) = (2\pi T)^{-1}\big|\sum_t y_t e^{-i\omega t}\big|^2$ at Fourier frequencies, dropping DC and Nyquist; the discrete Whittle objective uses $s(\omega) = 2\pi f(\omega)$ so $\log s + (2\pi) I_T/s$ is minimized. See §3 and the remark therein for the scaling.

[2] The envelope is deliberately conservative: with $\delta_T = 1/(1 + \log T)$ it scales as $(1 + \log T)^5\sqrt{\log T/T}$; constants are not optimized.

Table 1: Uniform Whittle LLN on the fixed-boundary sieve: absolute deviations between the discrete criterion $\widehat{\mathcal{L}}_T$ and its discrete "population" counterpart $J_d$ evaluated at the same Fourier grid and the sieve minimizer.

| $T$ | $\sup_{p\in\mathcal{F}_{\hat{K}}} \lvert\widehat{\mathcal{L}}_T - J_d\rvert$ | mean | median | 90% | 99% |
|---|---|---|---|---|---|
| 2048 | 0.0458 | 0.0379 | 0.0378 | 0.0413 | 0.0444 |
| 8192 | 0.0054 | 0.0046 | 0.0046 | 0.0048 | 0.0051 |
| 32768 | 0.0018 | 0.0016 | 0.0015 | 0.0017 | 0.0018 |

Here $J_d(f) := \frac{1}{M}\sum_{j=1}^{M}\{\log s(\omega_j) + s^*(\omega_j)/s(\omega_j)\}$ with $s = 2\pi f$; see Lemma 3.4.

Table 2: Uniform Whittle LLN with a shrinking boundary $\rho \le 1 - \delta_T$ ($\delta_T = 1/(1 + \log T)$).

| $T$ | $\sup_{p\in\mathcal{F}_{\hat{K}}} \lvert\widehat{\mathcal{L}}_T - J_d\rvert$ | mean | median | 90% | 99% |
|---|---|---|---|---|---|
| 2048 | 0.0149 | 0.0040 | 0.0038 | 0.0056 | 0.0070 |
| 8192 | 0.0055 | 0.0051 | 0.0051 | 0.0052 | 0.0052 |
| 32768 | 0.0028 | 0.0025 | 0.0025 | 0.0026 | 0.0027 |

### 6.2 Support localization and barycentric pole estimates

On a fixed boundary with well–separated true poles, sieve mass concentrates in small neighborhoods (Theorem 4.1). Table 3 reports the *barycentric* pole estimates $\hat{\rho}_k$ for a representative run ($T =$ 16,000) with window $\varepsilon = 0.05$; the estimated poles are accurate to $\approx 10^{-3}$–$10^{-2}$.[3]

Table 3: Support localization on a fixed boundary (simulation, $T =$ 16,000). For each true pole $\rho_k^*$ we report the barycentric estimate $\hat{\rho}_k$ computed from sieve mass within a $\pm\varepsilon$ window around $\rho_k^*$, and the local mass (which need not equal the true weight).

| $k$ | $\rho_k^*$ | $\hat{\rho}_k$ | Mass in $\pm\varepsilon$ | $\varepsilon$ |
|---|---|---|---|---|
| 1 | 0.200 | 0.201 | 0.148 | 0.05 |
| 2 | 0.600 | 0.596 | 0.072 | 0.05 |
| 3 | 0.950 | 0.945 | 0.067 | 0.05 |

Figure 1 overlays the true spectrum and the sieve–Whittle fit; Figure 2 shows the estimated mixing measure on $\{\text{white}\} \cup \mathcal{G}_{\hat{K}}$. Both figures are generated by the simulation script shipped with the paper.

[3]The column "Mass in $\pm\varepsilon$" reports the sieve mass captured inside each window; it need not equal the true weight at finite $T$.

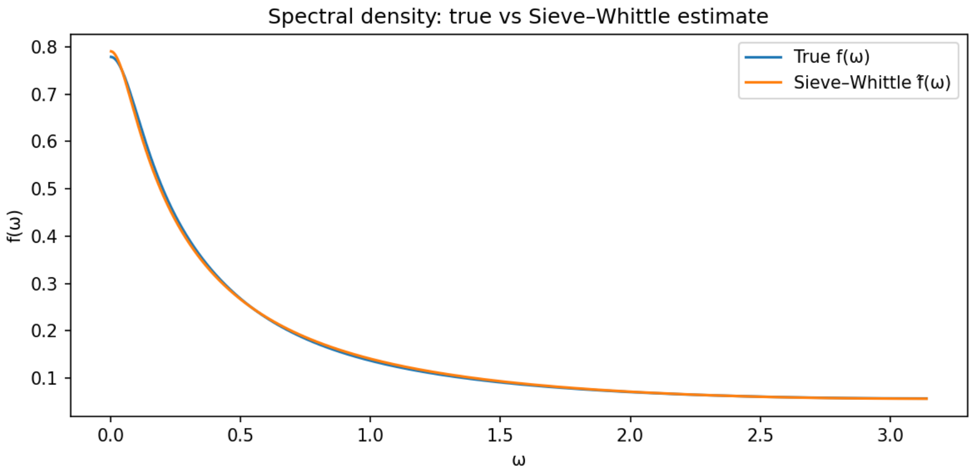


Figure 1: Simulation: spectral density, true vs. sieve–Whittle estimate ($s = 2\pi f$ converted back to $f$).

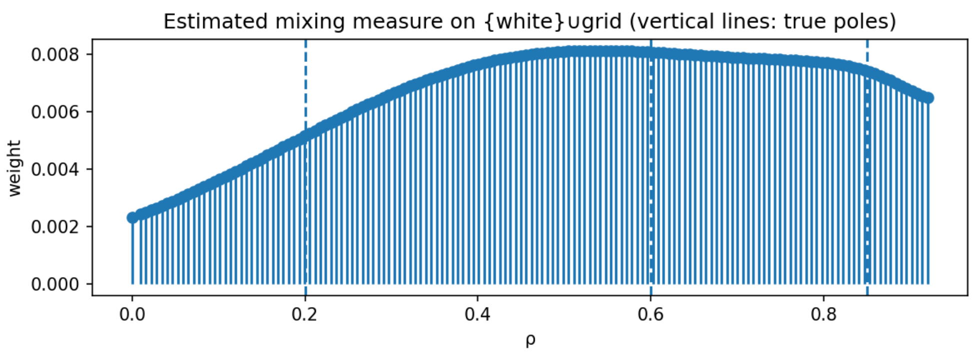


Figure 2: Simulation: estimated mixing measure on $\{\text{white}\} \cup \mathcal{G}_{\hat{K}}$ (dashed vertical lines: true poles).

### 6.3 Closed–form EM M–step: monotonicity in practice

With $(K, \boldsymbol{\rho})$ fixed, the EM update for $(W, \sigma^2)$ under $\sum_k w_k + \sigma^2 = 1$ has the closed form in Theorem 5.2. Figure 3 plots the observed Gaussian log–likelihood across EM iterations for a typical run; the path is *monotone increasing*, in line with Proposition 5.3.

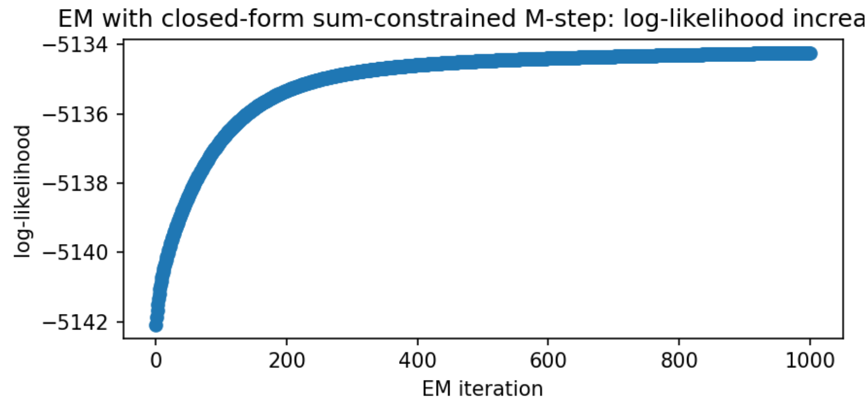


Figure 3: Simulation: EM with closed–form sum–constrained M–step—observed log–likelihood increases monotonically.

### 6.4 Empirical application: U.S. unemployment (quarterly change)

We fit the sieve–Whittle mixture on the standardized training portion (80%) of the U.S. macro data bundle in `statsmodels`; we then initialize EM by peak–picking and barycentric mass (Sec. 4), and run the *closed–form* EM M–step on the training set. Out–of–sample (20%) we compare against a Yule–Walker AR($p$) selected by BIC.

Table 4 reports one–step predictive scores on the test set (higher log score and lower MSE are better). The mixture improves test MSE by about 4.2% and the test log score by 0.584 points relative to AR(2); the Diebold–Mariano statistic on squared errors is negative ($-1.694$; $p\!=\!0.090$), favoring the mixture in this small sample.

Table 4: U.S. unemployment change (`statsmodels macrodata`, standardized; 80% train / 20% test). Higher test log score and lower MSE are better. DM statistic is for squared–error loss, negative values favor the mixture.

| Model | $K$ or $p$ | Test log score | Test MSE | DM stat | $p$–value | $\bar{d}$ |
|---|---|---|---|---|---|---|
| Mixture (AR(1) sum) | 2 | $-43.635$ | 0.510 | $-1.694$ | 0.090 | $-0.022$ |
| AR(2) YW-BIC | 2 | $-44.219$ | 0.532 | | | |

**Takeaways.** (i) The sieve–Whittle *uniform* LLN is visible in finite samples (Tables 1–2); (ii) support is well localized and barycentric pole estimates are accurate (Table 3); (iii) the *closed–form* EM update delivers a monotone likelihood path (Fig. 3); and (iv) on a canonical macro time series ($\Delta$UNEMP) the mixture achieves better out–of–sample accuracy than a BIC–selected AR($p$).

## Conclusions

The AR(1)+white sieve provides a simple positivity–preserving dictionary on which a uniform Whittle LLN (including shrinking boundaries) can be verified by explicit envelope and variation bounds. This supports consistent nonparametric recovery of the spectrum; for discrete mixtures, sieve mass concentrates near the true poles and a *barycentric* rule yields consistent pole estimates. The *closed–form* EM M–step under the unit–variance sum constraint gives a robust weight/noise update. While parametric Whittle asymptotics are classical, the sieve→localization→parametric pipeline with the EM weight update is a practical recipe validated by our simulations.

# A Proofs for main results

*Proof of Theorem 2.1.* Consider the power series (covariance generating function) $G(z) = \sum_{h=0}^{\infty} \gamma(h) z^h$ for $|z| < 1$. Then

$$G(z) = \sum_{k=1}^{K} \frac{w_k}{1 - \rho_k z}.$$

Thus $G$ is a rational function analytic on $|z| < 1$ with simple poles at $z = 1/\rho_k > 1$ and residues $-w_k/\rho_k$. If two parameter triples produce the same $\gamma(h)$, then the corresponding rational functions $G$ agree on the open unit disc (since their power series coincide there). Two rational functions that agree on a nonempty open set are identical on $\mathbb{C}$, hence their partial fraction decompositions have the same simple poles $\{1/\rho_k\}$ and the same residues $-w_k/\rho_k$. This uniquely identifies $K$, the ordered $\{\rho_k\}$, and the $\{w_k\}$; signs of $a_k$ are not identified. Gaussianity then identifies the full law. □

*Proof of Theorem 2.2.* From (6), the sequence $\{\gamma(h)\}_{h \geq 1}$ equals $\sum_k w_k \rho_k^h$ and is independent of $\sigma^2$. By Theorem 2.1, $(K, \{\rho_k\}, \{w_k\})$ are identified from $\{\gamma(h)\}_{h \geq 1}$. Then $\sigma^2 = \gamma(0) - \sum_k w_k$ is identified from the lag–0 variance identity. □

**Lemma A.1** (Uniqueness of the population minimizer)**.** *Let $f^*$ be the true spectral density. For any candidate $f > 0$ a.e.,*

$$J(f) - J(f^*) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left\{ \log \frac{f(\omega)}{f^*(\omega)} + \frac{f^*(\omega)}{f(\omega)} - 1 \right\} d\omega \;\geq\; 0,$$

*with equality if and only if $f = f^*$ a.e. In particular, $J$ is uniquely minimized at $f^*$.*

*Proof of Lemma A.1.* Write $t(\omega) = f(\omega)/f^*(\omega) > 0$ and $g(t) = \log t + t^{-1} - 1$. Then $g'(t) = (t-1)/t^2$, so $t = 1$ is the unique stationary point and global minimum with $g(1) = 0$. Hence $g(t) \geq 0$ for all $t > 0$, with equality only at $t = 1$. Integrating $g(t(\omega))$ over $\omega$ yields the claim. □

**Lemma A.2** (Uniform lower and upper bounds)**.** *On the fixed–boundary sieve with $\hat\rho_j \leq 1 - \delta_\star$, there exist constants $0 < c_\star < C_\star < \infty$ (depending only on $\delta_\star$) such that $c_\star \leq f(\omega) \leq C_\star$ for all $f \in \mathcal{F}_{\hat K}$ and all $\omega$. One may take*

$$c_\star = \frac{1}{2\pi} \frac{\delta_\star}{2 - \delta_\star}, \qquad C_\star = \frac{1}{2\pi} \frac{2 - \delta_\star}{\delta_\star}.$$

*On the shrinking boundary sieve with $\hat\rho_j \leq 1 - \delta_T$, the same holds with $c_T := \frac{1}{2\pi} \frac{\delta_T}{2 - \delta_T}$ and $C_T := \frac{1}{2\pi} \frac{2 - \delta_T}{\delta_T}$.*

*Proof of Lemma A.2.* By (10), each $\phi(\hat\rho_j, \cdot)$ is bounded between $\frac{1 - \hat\rho_j}{1 + \hat\rho_j}$ and $\frac{1 + \hat\rho_j}{1 - \hat\rho_j}$, and the white–noise atom equals 1. Convex combinations preserve the bounds. The stated constants follow by taking $\hat\rho_j \leq 1 - \delta$. □

**Lemma A.3** (Denseness)**.** *Let $\Phi(\rho, \omega) := (1 - \rho^2)/(1 + \rho^2 - 2\rho \cos \omega)$.*

*(i) For any fixed $\delta \in (0,1)$, $(\rho, \omega) \mapsto \Phi(\rho, \omega)$ is uniformly continuous on $[0, 1 - \delta] \times [-\pi, \pi]$.*

*(ii) If $F_n \Rightarrow F$ weakly and* $\operatorname{supp}(F_n) \cup \operatorname{supp}(F) \subset [0, 1-\delta]$*, then*

$$\sup_{\omega \in [-\pi,\pi]} \left| \int \Phi(\rho,\omega)\, dF_n(\rho) - \int \Phi(\rho,\omega)\, dF(\rho) \right| \to 0.$$

*(iii) If the grid mesh $\Delta_{\hat{K}} \to 0$ and the sieve boundary covers the true support (i.e.,* $\max_k \rho_k^* \leq 1-\delta$ *with $\delta = \delta_\star$ or $\delta = \delta_T$), then $\bigcup_{\hat{K}} \mathcal{F}_{\hat{K}}$ is dense (in sup–norm) in the class of unit–variance spectra of the form (7) with discrete support.*

*Proof of Lemma A.3.* Part (i) follows from continuity on the compact rectangle $[0, 1-\delta] \times [-\pi, \pi]$. For (ii), the family $\{g_\omega(\cdot) := \Phi(\cdot, \omega) : \omega \in [-\pi, \pi]\} \subset C([0, 1-\delta])$ is uniformly bounded and equicontinuous by (i), hence totally bounded in $\|\cdot\|_\infty$. Approximate the class by a finite $\varepsilon$–net and use weak convergence plus the uniform envelope to conclude uniform convergence in $\omega$. For (iii), send each true mass to the nearest grid point; (ii) plus $\Delta_{\hat{K}} \to 0$ yields $\|f_{\hat{F}_{\hat{K}}} - f^*\|_\infty \to 0$. □

*Proof of Theorem 3.2.* This is a direct specialization of the maximal exponential inequality and Glivenko–Cantelli/CLT for the empirical spectral process indexed by bounded–variation classes; see Dahlhaus (2009, Thm. 2.6, Cor. 2.11). Apply those results to the stationary case with the indexing class $\mathcal{G}$; the stated rate follows from the entropy bound for BV classes (details as in Dahlhaus, 2009). □

*Proof of Lemma 3.3.* Write $\mathcal{L}_T(f) - J(f) = \frac{1}{2\pi} \int g_f(\omega) \big( I_T(\omega) - 2\pi f^*(\omega) \big)\, d\omega$ with $g_f = 1/f$. Apply (11) with $\mathcal{G} = \{g_f : f \in \mathcal{F}_{\hat{K}}\}$ and use Lemma C.2. □

*Proof of Lemma 3.4.* Write $\widehat{\mathcal{L}}_T(f) - J(f)$ as the sum of $\frac{1}{M} \sum_j \log(2\pi f(\omega_j)) - \frac{1}{2\pi} \int \log(2\pi f)$ and $\frac{1}{M} \sum_j \{2\pi I_T(\omega_j)\}/\{2\pi f(\omega_j)\} - \frac{1}{2\pi} \int \{2\pi I_T\}/\{2\pi f\}$. The first part is a deterministic Riemann error bounded by $\mathrm{Var}_{\mathrm{Tot}}(\log f)\, T^{-1} = O(\delta_T^{-3} T^{-1})$ uniformly on the sieve (Lemma C.2). For the second part, identify the discrete objective with a quadratic form in the Fourier ordinates and apply the maximal inequality in Theorem 3.2 to $\mathcal{G} = \{1/f : f \in \mathcal{F}_{\hat{K}}\}$, which yields a uniform bound of order $\|f^*\|_\infty (M+V) \sqrt{\log T/T} = O(\delta_T^{-5} \sqrt{\log T/T})$; this dominates the Riemann error. □

*Proof of Theorem 3.6.* By Lemma 3.3 (fixed boundary) we have a uniform LLN on $\mathcal{F}_{\hat{K}}$, and by Lemma A.1, $J$ is uniquely minimized at $f^*$. By Lemma A.3(iii), for each $T$ there exists $g_T \in \mathcal{F}_{\hat{K}}$ with $\|g_T - f^*\|_\infty \to 0$, which implies $J(g_T) \to J(f^*)$ by continuity of $J$ over the uniformly bounded-away-from-zero sieve. Let $\hat{f}_T$ be any Whittle minimizer over $\mathcal{F}_{\hat{K}}$. The uniform LLN and the previous display yield $J(\hat{f}_T) \to J(f^*)$ in probability. Tightness of the associated mixing measures on $[0,1]$ gives subsequences $\hat{F}_T \Rightarrow \tilde{F}$; by Lemma A.3(ii), $f_{\tilde{F}} = f^*$. Matching all moments $\int \rho^h\, d\tilde{F} = \int \rho^h\, dF^*$ then gives $\tilde{F} = F^*$ (Hausdorff moment problem), hence $\hat{F}_T \Rightarrow F^*$ and $\|\hat{f}_T - f^*\|_\infty \to 0$ in probability. □

*Proof of Theorem 3.8.* By Theorem 3.2 with $\mathcal{G} = \{1/f : f \in \mathcal{F}_{\hat{K}}\}$ and Lemma C.2,

$$\sup_{f \in \mathcal{F}_{\hat{K}}} \left| \mathcal{L}_T(f) - J(f) \right| = o_{\mathbb{P}}(1)$$

under $\delta_T^{-5} \sqrt{\log T/T} \to 0$. By Lemma A.5 and Assumption 3.7(iv), there exists $g_T \in \mathcal{F}_{\hat{K}}$ with $\|g_T - f^*\|_\infty \to 0$, hence $J(g_T) \to J(f^*)$. Let $\hat{f}_T$ be any Whittle minimizer on $\mathcal{F}_{\hat{K}}$. The uniform LLN and Lemma A.1 imply $J(\hat{f}_T) \to J(f^*)$ in probability. Tightness of the mixing measures and Lemma A.3(ii) yield $f_{\tilde{F}} = f^*$ for any weak subsequential limit $\tilde{F}$, hence $\tilde{F} = F^*$ by the Hausdorff moment problem. Therefore $\hat{F}_T \Rightarrow F^*$ and $\|\hat{f}_T - f^*\|_\infty \to 0$ in probability.

□

*Proof of Corollary 3.9.* Uniform equivalence (12) and Lemma 3.3 imply $\sup_f |\widehat{\mathcal{L}}_T(f) - J(f)| = o_{\mathbb{P}}(1)$. Uniqueness of the population minimizer (Lemma A.1) then yields the same consistency as Theorems 3.6–3.8. $\square$

*Proof of Theorem 3.10.* This is standard Whittle MLE theory in finite dimension for Gaussian time series; differentiability and identification are immediate from the mixture representation and interiority; a score CLT for the empirical spectral process yields the limiting normal law with covariance $I^{-1}$; see, e.g., Dahlhaus (1988); Brillinger (2001). $\square$

*Proof of Theorem 4.1.* By Theorems 3.6–3.8, $\hat{F}_T \Rightarrow F^*$ in probability. Since $F^*$ is discrete with separated atoms, each $B_\varepsilon(\rho_k^*)$ is a continuity set, hence by Portmanteau $\hat{F}_T(B_\varepsilon(\rho_k^*)) \xrightarrow{\mathbb{P}} \theta_k^*$ for $k = 0, 1, \dots, K$. Moreover,

$$\hat{F}_T\Big([0,1] \setminus \bigcup_{k=0}^{K} B_\varepsilon(\rho_k^*)\Big) \ \xrightarrow{\mathbb{P}}\ 0,$$

because the union includes $k = 0$. $\square$

*Proof of Corollary 4.2.* Fix $\varepsilon \in (\delta, \rho_1^*/2)$ and let $U := \bigcup_{k=1}^{K} B_\varepsilon(\rho_k^*)$ (balls around nonzero poles only). Then $U^c$ is a continuity set with $F^*(U^c) = \theta_0^*$, hence $\hat{F}_T(U^c) \to \theta_0^*$ in probability by Portmanteau. The set $U^c \setminus [0,\delta]$ is a finite union of closed intervals whose boundaries carry zero $F^*$-mass; thus $\hat{F}_T(U^c \setminus [0,\delta]) \to 0$ in probability. Therefore

$$\hat{F}_T([0,\delta]) = \hat{F}_T(U^c) - \hat{F}_T(U^c \setminus [0,\delta]) \ \xrightarrow{\mathbb{P}}\ \theta_0^*.$$

$\square$

*Proof of Corollary 4.3.* Let $B := B_\varepsilon(\rho_k^*)$. By Theorem 4.1, $\hat{F}_T(B) \to \theta_k^* > 0$ in probability, and $F^*(\partial B) = 0$. Consider the restricted measures $\hat{F}_T^{(B)}(\cdot) := \hat{F}_T(\cdot \cap B)$ and $F^{*,(B)} := \theta_k^* \delta_{\rho_k^*}$. Since restriction to $B$ is continuous at $F^*$, we have $\hat{F}_T^{(B)} \Rightarrow F^{*,(B)}$ in probability. Because $\rho$ is bounded and continuous,

$$\int \rho \, d\hat{F}_T^{(B)} \ \xrightarrow{\mathbb{P}}\ \rho_k^* \theta_k^*.$$

Thus

$$\hat{\rho}_{k,T} = \frac{\int \rho \, d\hat{F}_T^{(B)}}{\int d\hat{F}_T^{(B)}} \ \xrightarrow{\mathbb{P}}\ \frac{\rho_k^* \theta_k^*}{\theta_k^*} = \rho_k^*, \qquad \hat{\theta}_{k,T} = \hat{F}_T(B) \ \xrightarrow{\mathbb{P}}\ \theta_k^*.$$

$\square$

*Proof of Proposition 5.1.* Fix a scalar parameter $x \in \{w_k, \sigma^2\}$ and write $g(x) := -\frac{1}{2}(T \log x + C/x)$ on $(0,\infty)$. Then $g'(x) = -\frac{T}{2x} + \frac{C}{2x^2} = 0 \Leftrightarrow x^\star = C/T$. Moreover $g''(x^\star) = -T^3/(2C^2) < 0$, and $\lim_{x\downarrow 0} g(x) = \lim_{x\uparrow\infty} g(x) = -\infty$. Hence $x^\star = C/T$ is the unique global maximizer. $\square$

*Proof of Theorem 5.2.* With Lagrangian $\mathcal{L} = Q + \lambda\{S - \sum x\}$, the scalar FOC is $-\frac{T}{2x} + \frac{C}{2x^2} - \lambda = 0$, which is (16). The discriminant condition yields $\lambda \geq -T^2/(8C)$; taking $C = \max\{A_{\max}^{(m)}, B^{(m)}\}$ gives the common lower bound $\lambda_{\min}$. The positive branch (17) follows by solving (16).

Differentiating (17) gives, for each $C > 0$,

$$x'(\lambda) \ = \ -\frac{4C^2}{\big(T + \sqrt{T^2 + 8\lambda C}\big)^2 \sqrt{T^2 + 8\lambda C}} \ < \ 0,$$

so $x(\lambda)$ and hence $F(\lambda)$ are strictly decreasing and continuous on $[\lambda_{\min}, \infty)$. The limits $\lim_{\lambda\uparrow\infty} F(\lambda) = 0$ and $F(0) = (\sum C_j)/T$ are immediate. At $\lambda_{\min} = -T^2/(8C_{\max})$ one has

$$x_j(\lambda_{\min}) = \frac{2C_j}{T + T\sqrt{1 - C_j/C_{\max}}} \in \left(\tfrac{C_j}{T},\, \tfrac{2C_j}{T}\right],$$

whence $F(\lambda_{\min}) = \sum_j x_j(\lambda_{\min}) \in \big(F(0),\, 2\sum_j C_j/T\big]$.

By strict monotonicity and the intermediate value theorem there is a unique $\lambda^* \in [\lambda_{\min}, \infty)$ such that $F(\lambda^*) = S$ for every $S \in (0, F(\lambda_{\min})]$; moreover $\lambda^* \geq 0$ iff $S \leq F(0)$ and $\lambda^* \in (\lambda_{\min}, 0)$ iff $S > F(0)$.

Finally, the Hessian of $Q$ in the free variables is diagonal with entries $g''(x) = \frac{T}{2x^2} - \frac{C}{x^3} = \frac{Tx-2C}{2x^3}$. At any feasible solution one has $x(\lambda^*) < 2C/T$ (since $x(\lambda) \uparrow 2C/T$ only as $\lambda \downarrow \lambda_{\min}$), whence $g''(x(\lambda^*)) < 0$. Therefore the bordered Hessian is negative definite on the tangent space $\{\delta x : \sum \delta x = 0\}$ and the KKT point is the unique global maximizer under the linear constraint. □

*Proof of Proposition 5.3.* Standard EM arguments: $Q(\theta^{(m+1)} \mid \theta^{(m)}) \geq Q(\theta^{(m)} \mid \theta^{(m)})$ by construction of the M–step; Jensen's inequality gives $\log p(Y_T \mid \theta^{(m+1)}) - \log p(Y_T \mid \theta^{(m)}) \geq Q(\theta^{(m+1)} \mid \theta^{(m)}) - Q(\theta^{(m)} \mid \theta^{(m)}) \geq 0$. Uniqueness of the constrained maximizer implies equality only at fixed points. □

**Lemma A.4** (Lipschitz in the pole)**.** *For any $\rho_1, \rho_2 \in [0, 1)$ and all $\omega$,*

$$\big|\phi(\rho_1, \omega) - \phi(\rho_2, \omega)\big| \;\leq\; \frac{2}{\left(1-\bar\rho\right)^2}\,|\rho_1 - \rho_2|, \qquad \bar\rho := \max\{\rho_1, \rho_2\}.$$

*Proof.* By the mean value theorem in $\rho$ and Lemma C.1,

$$|\phi(\rho_1, \omega) - \phi(\rho_2, \omega)| \leq \sup_{\tilde\rho\in[\rho_1,\rho_2]} |\partial_\rho \phi(\tilde\rho, \omega)| \cdot |\rho_1 - \rho_2| \leq \frac{2}{(1-\bar\rho)^2}\,|\rho_1 - \rho_2|.$$

□

**Lemma A.5** (Grid approximation under a shrinking boundary)**.** *There exists a universal constant $C > 0$ such that: if $F$ is a discrete mixing measure supported on $[0, 1-\delta_T]$, and $F_{\mathrm{grid}}$ is obtained by sending each atom of $F$ to its nearest gridpoint on a grid with mesh $\Delta_{\hat K}$, then*

$$\|f_F - f_{F_{\mathrm{grid}}}\|_\infty \;\leq\; C\,\frac{\Delta_{\hat K}}{\delta_T^2}.$$

*Proof of Lemma A.5.* Write $f_F(\omega) = \frac{1}{2\pi}\int \phi(\rho, \omega)\, dF(\rho)$. By Lemma A.4 and linearity in $F$,

$$\sup_\omega |f_F(\omega) - f_{F_{\mathrm{grid}}}(\omega)| \leq \frac{1}{2\pi}\int \sup_\omega |\phi(\rho, \omega) - \phi(\rho', \omega)|\, d\pi(\rho, \rho') \leq C\,\frac{\Delta_{\hat K}}{\delta_T^2},$$

for the coupling $\pi$ that pairs each true atom with its nearest gridpoint. The constant $C$ absorbs absolute multiplicative factors (including $1/(2\pi)$). □

**Lemma A.6** (Global curvature via a Bregman bound)**.** *For any $f, f^\dagger > 0$ a.e.,*

$$J(f) - J(f^\dagger) \;\geq\; \frac{1}{4\pi}\int_{-\pi}^{\pi} \frac{\big(f(\omega) - f^\dagger(\omega)\big)^2}{\max\{f(\omega), f^\dagger(\omega)\}^2}\, d\omega.$$

*In particular, writing $C_T := \sup_{f\in\mathcal{F}_{\hat K}} \|f\|_\infty$,*

$$\|f - f^\dagger\|_{L^2}^2 \;\leq\; 4\pi\, C_T^2\, \big\{J(f) - J(f^\dagger)\big\}.$$

*Proof.* For any $a, b > 0$, $\log(a/b) + b/a - 1 \geq \frac{1}{2}\frac{(a-b)^2}{\max\{a,b\}^2}$. Integrate the pointwise inequality with $a = f(\omega)$, $b = f^\dagger(\omega)$ over $\omega \in [-\pi, \pi]$ and multiply by $(2\pi)^{-1}$. □

**Proposition A.7** (Risk/approximation decomposition). *Let $f_T^\dagger \in \mathcal{F}_{\hat{K}}$ minimize $J$ over the sieve. Under Assumption 3.7,*

$$J(\hat{f}_T) - J(f_T^\dagger) \;=\; O_{\mathbb{P}}\left(\delta_T^{-5}\sqrt{\frac{\log T}{T}}\right),$$

*and*

$$\|f_T^\dagger - f^*\|_\infty \;\leq\; C\,\frac{\Delta_{\hat{K}}}{\delta_T^2} \quad \textit{(deterministic)}.$$

*Consequently,*

$$\|\hat{f}_T - f^*\|_\infty \;\leq\; C\,\frac{\Delta_{\hat{K}}}{\delta_T^2} \;+\; o_{\mathbb{P}}(1).$$

*Moreover, by the global Bregman curvature bound (Lemma A.6) and $\sup_{f\in\mathcal{F}_{\hat{K}}}\|f\|_\infty \asymp \delta_T^{-1}$,*

$$\|\hat{f}_T - f_T^\dagger\|_{L^2} \;=\; O_{\mathbb{P}}\left(\frac{1}{\delta_T^{3.5}}\left(\frac{\log T}{T}\right)^{1/4}\right).$$

*Proof.* Basic inequality: $\widehat{\mathcal{L}}_T(\hat{f}_T) \leq \widehat{\mathcal{L}}_T(f_T^\dagger)$ implies $J(\hat{f}_T) - J(f_T^\dagger) \leq 2\sup_{f\in\mathcal{F}_{\hat{K}}}|\widehat{\mathcal{L}}_T(f) - J(f)|$, and Theorem 3.2 with Lemma C.2 yields the displayed $O_{\mathbb{P}}$ bound. The sieve approximation bound is Lemma A.5. By Lemma A.6,

$$\|\hat{f}_T - f_T^\dagger\|_{L^2}^2 \;\lesssim\; C_T^2\,\{J(\hat{f}_T) - J(f_T^\dagger)\},$$

and the displayed $L^2$ rate follows from the risk bound above. □

# B Disturbance smoothing with observation noise

**Lemma B.1.** *A general linear Gaussian state space model can be represented in a matrix form as*

$$Y_T = Z\alpha + \epsilon, \qquad \epsilon \sim N(0, H), \tag{18}$$

$$\alpha = T(\alpha_1^* + R\eta), \qquad \eta \sim N(0, Q) \tag{19}$$

*where*

$$Y_T = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}, \qquad Z = \begin{pmatrix} Z_1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & Z_T & 0 \end{pmatrix}, \qquad \alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_T \\ \alpha_{T+1} \end{pmatrix},$$

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{pmatrix}, \qquad H = \begin{pmatrix} H_1 & & 0 \\ & \ddots & \\ 0 & & H_T \end{pmatrix},$$

$$T = \begin{pmatrix} I & 0 & 0 & 0 & \dots & 0 \\ T_1 & I & 0 & 0 & \dots & 0 \\ T_2T_1 & T_2 & I & 0 & \dots & 0 \\ & & & & \ddots & \vdots \\ T_T\dots T_1 & T_T\dots T_2 & T_T\dots T_3 & T_T\dots T_4 & \dots & I \end{pmatrix}, \qquad \alpha_1^* = \begin{pmatrix} \alpha_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$R = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ R_1 & 0 & & 0 \\ 0 & R_2 & & 0 \\ & & \ddots & \vdots \\ 0 & 0 & \cdots & R_T \end{pmatrix}, \qquad \eta = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_T \end{pmatrix}, \qquad Q = \begin{pmatrix} Q_1 & & \\ & \ddots & \\ & & Q_T \end{pmatrix}.$$

*Furthermore, we have*

$$Y_T \sim N(ZTa_1^*, ZTQ^*T'Z' + H),$$

*where*

$$a_1^* = \begin{pmatrix} a_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad P_1^* = \begin{pmatrix} P_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad Q^* = P_1^* + RQR',$$

*Proof.* See Koopman and Durbin (2012, section 4.13). □

**Lemma B.2.** *For a general linear Gaussian state space model, define*

$$\begin{aligned} a_t &= E(\alpha_t|Y_{t-1}), & P_t &= Var(\alpha_t|Y_{t-1}), \\ v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t' + H_t, \\ K_t &= T_t P_t Z_t' F_t^{-1}, \end{aligned}$$

*then the Kalman filter for the model takes the form*

$$a_{t+1} = T_t a_t + K_t v_t, \qquad P_{t+1} = T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t'.$$

*Define*

$$\begin{aligned} \hat{a}_t &= E(\alpha_t|Y_T), & V_t &= Var(\alpha_t|Y_T), \\ L_t &= T_t - K_t Z_t, \\ r_T &= 0, & N_T &= 0, \\ r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t, & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t' N_t L_t, \end{aligned}$$

*then the smoother takes the form*

$$\hat{\alpha}_t = a_t + P_t r_{t-1}, \qquad V_t = P_t - P_t N_{t-1} P_t.$$

*Define*

$$\hat{\epsilon}_t = E(\epsilon_t|Y_T), \qquad \hat{\eta}_t = E(\eta_t|Y_T),$$

*then the disturbance smoother takes the form*

$$\begin{aligned} \hat{\epsilon}_t &= H_t(F_t^{-1} v_t - K_t' r_t), & Var(\epsilon_t|Y_T) &= H_t - H_t(F_t^{-1} + K_t' N_t K_t) H_t, \\ \hat{\eta}_t &= Q_t R_t' r_t, & Var(\eta_t|Y_T) &= Q_t - Q_t R_t' N_t R_t Q_t. \end{aligned}$$

*Proof.* See, for example, Koopman and Durbin (2012, chapter 4). □

# C Derivative bound for the AR(1) kernel

**Lemma C.1** (Poisson kernel series bounds)**.** *For $\rho \in [0,1)$ and $\omega \in \mathbb{R}$,*

$$\phi(\rho,\omega) = \frac{1-\rho^2}{1-2\rho\cos\omega+\rho^2} \;=\; 1 + 2\sum_{h\geq 1}\rho^h\cos(h\omega).$$

*Consequently,*

$$\sup_{\omega}\bigl|\partial_\omega\phi(\rho,\omega)\bigr| = \sup_{\omega}\left|-2\sum_{h\geq 1} h\,\rho^h\sin(h\omega)\right| \;\leq\; 2\sum_{h\geq 1} h\rho^h \;=\; \frac{2\rho}{(1-\rho)^2} \;\leq\; \frac{2}{(1-\rho)^2}.$$

*Moreover,*

$$\sup_{\omega}\left|\frac{\partial}{\partial\rho}\phi(\rho,\omega)\right| = \sup_{\omega}\left|2\sum_{h\geq 1} h\,\rho^{h-1}\cos(h\omega)\right| \;\leq\; 2\sum_{h\geq 1} h\rho^{h-1} \;=\; \frac{2}{(1-\rho)^2}.$$

**Lemma C.2** (Uniform boundedness and variation of $1/f$ on the sieve)**.** *Let $\mathcal{F}_{\hat{K}}$ be the unit–variance sieve of (8) with grid poles $\hat{\rho}_j \leq 1-\delta$ for some $\delta \in (0,1)$ (fixed boundary) and include the white atom $\phi_0 \equiv 1$. Then, for every $f \in \mathcal{F}_{\hat{K}}$,*

$$\inf_{\omega} f(\omega) \;\geq\; \frac{1}{2\pi}\,\frac{\delta}{2-\delta}, \qquad \sup_{\omega} f(\omega) \;\leq\; \frac{1}{2\pi}\,\frac{2-\delta}{\delta}.$$

*Hence*

$$\left\|\frac{1}{f}\right\|_\infty \;\leq\; \frac{2\pi(2-\delta)}{\delta} \qquad \textit{and} \qquad \mathrm{Var}_{\mathrm{Tot}}\!\left(\frac{1}{f}\right) \;\leq\; C\,\delta^{-4},$$

*for a universal constant $C > 0$. The same bounds hold with $\delta$ replaced by $\delta_T$ on the shrinking boundary sieve.*

*Proof of Lemma C.2.* By Lemma A.2, $\inf_\omega f(\omega) \geq c$ with $c = c_\star$ (fixed boundary) or $c = c_T$ (shrinking), hence $\|1/f\|_\infty \leq 1/c = O(1/\delta)$. For the total variation, note that $f$ is continuously differentiable with

$$f'(\omega) = \frac{1}{2\pi}\sum_{j=1}^{\hat{K}}\hat{p}_j\,\partial_\omega\phi(\hat{\rho}_j,\omega), \qquad \partial_\omega\phi(\rho,\omega) = -(1-\rho^2)\frac{2\rho\sin\omega}{\bigl(1+\rho^2-2\rho\cos\omega\bigr)^2}.$$

A calculus bound (Appendix C) shows that

$$\sup_{\omega\in[-\pi,\pi]}\bigl|\partial_\omega\phi(\rho,\omega)\bigr| \;\leq\; \frac{C_0}{(1-\rho)^2}$$

for a universal constant $C_0 > 0$. Therefore $\|f'\|_\infty \leq \frac{1}{2\pi}\max_j C_0/(1-\hat{\rho}_j)^2 = O(1/\delta^2)$. Since $(1/f)' = -f'/f^2$, we have

$$\left\|\left(\frac{1}{f}\right)'\right\|_\infty \;\leq\; \frac{\|f'\|_\infty}{(\inf f)^2} \;=\; O\!\left(\frac{1}{\delta^2}\right)\cdot O\!\left(\frac{1}{\delta^2}\right) = O\!\left(\frac{1}{\delta^4}\right),$$

and the total variation over $[-\pi,\pi]$ is at most $(2\pi)\|(1/f)'\|_\infty$. $\square$

## References


D. R. Brillinger. *Time Series: Data Analysis and Theory.* SIAM, 2001.

R. Dahlhaus. Empirical spectral processes and their applications to time series analysis. *Stochastic Processes and their Applications*, 30(1):69–89, 1988.

Raimund Dahlhaus. Empirical spectral processes for locally stationary time series. *Stochastic Processes and their Applications*, 119(10):3454–3501, 2009.

S. Koopman and J. Durbin. *Time Series Analysis by State Space Methods: Second Edition.* Oxford University Press, 2012.