CHAPTER 3

Low-Frequency Econometrics

Ulrich K. Müller and Mark W. Watson

Many questions in economics involve long-run or "trend" variation and covariation in time series. Yet, time series of typical lengths contain only limited information about this long-run variation. This paper suggests that long-run sample information can be isolated using a small number of low-frequency trigonometric weighted averages, which in turn can be used to conduct inference about long-run variability and covariability. Because the low-frequency weighted averages have large sample normal distributions, large sample valid inference can often be conducted using familiar small sample normal inference procedures. Moreover, the general approach is applicable for a wide range of persistent stochastic processes that go beyond the familiar $I(0)$ and $I(1)$ models.

1 INTRODUCTION

This paper discusses inference about trends in economic time series. By "trend" we mean the low-frequency variability evident in a time series after forming moving averages such as low-pass (cf. Baxter and King, 1999) or Hodrick and Prescott (1997) filters. To measure this low-frequency variability we rely on projections of the series onto a small number of trigonometric functions (e.g., discrete Fourier, sine, or cosine transforms). The fact that a small number of projection coefficients capture low-frequency variability reflects the scarcity of low-frequency information in the data, leading to what is effectively a "small-sample" econometric problem. As we show, it is still relatively straightforward to conduct statistical inference using the small sample of low-frequency data summaries. Moreover, these low-frequency methods are appropriate for both weakly and highly persistent processes. Before getting into the details, it is useful to fix ideas by looking at some data.

We thank Frank Schorfheide for a thoughtful discussion at the 2015 World Congress in Montréal. Support was provided by the National Science Foundation through grant SES-1226464. Replication files are available at www.princeton.edu/~mwatson.

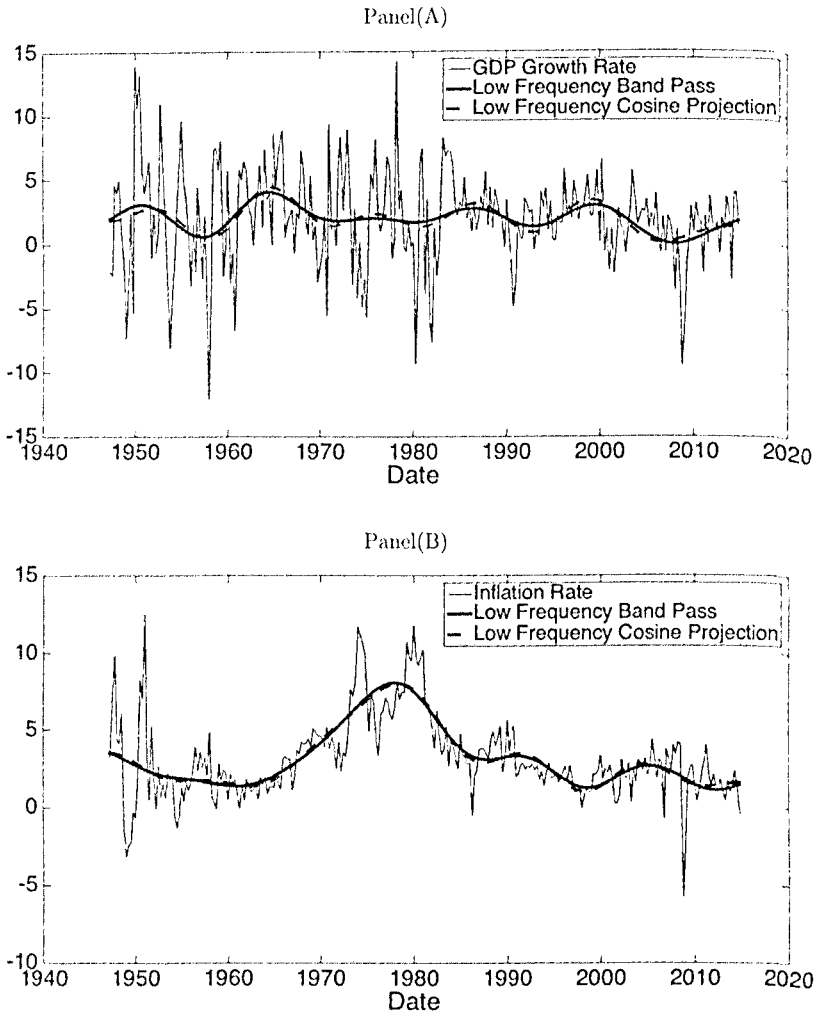

Figure 1 Trends in US Time Series

Figure 1 plots the value of per-capita GDP growth rates (panel A) and price inflation (panel B) for the United States using quarterly data from 1947 through 2014, and where both are expressed in percentage points at an annual rate.[1] The plots show the raw series and two "trends." The first trend was constructed using a band-pass moving average filter designed to pass cyclical components

[1] Three data series are used in this paper. All are quarterly series from 1947:Q1 to 2014:Q4 for the United States. Real per capita gross domestic product is available from FRED as series A939RX0Q048SBEA. Inflation is measured using the price deflator for personal consumption expenditures (FRED series DPCERD3Q086SBEA). Total factor productivity is from Fernald (2014) and is available at www.frbsf.org/economic-research/economists/john-fernald/.

with periods longer than $T/6 \approx 11$ years, and the second is the full-sample projection of the series onto a constant and twelve cosine functions with periods $2T/j$ for $j = 1, \ldots, 12$, also designed to capture variability for periods longer than 11 years.[2] A glance at the figure shows that the band-pass and projection trends essentially coincide both for GDP, for which there is only moderate trend variability, and inflation, for which there is substantial trend variability. This paper focuses on trends computed by projection methods because, as we will see, they give rise to simple methods for statistical inference.

Figure 1 raises several empirical questions. For example, the trend in GDP growth has been low for the last decade. Does this portend low average growth over the next decade? Or, is there mean reversion in the trend so that the next decade is more likely to exhibit faster than average growth? And, what *is* the value of the population mean growth rate of GDP? The plot for inflation shows large persistent variation in its trends over the past sixty years. Does this suggest that the inflation process is $I(1)$? Or is this behavior consistent with an $I(0)$ process? Alternatively, what about something "between" the $I(0)$ and $I(1)$ processes? More generally, the average value of inflation varies considerably over non-overlapping 10- or 25-year time periods; average GDP growth rates show less, but still substantial variability. How much variability should we expect to see in future 10- and 25-year samples? Section 3 takes up these questions, along with several other questions that can be answered using univariate time series methods.

Figure 2 plots both total factor productivity (TFP) and GDP growth rates. The trend components of the series move together, suggesting (for good reason) that long-run variation in GDP and TFP growth rates are closely related. But, exactly how closely are they related? And, by how much is the trend growth rate in GDP predicted to increase if the trend growth rate of TFP increases by, say, 10 basis points? Section 4 takes up questions like these that involve multiple (here two) time series.

The paper begins, in Section 2, with notation and (finite and large-sample) properties of discrete "cosine transforms," the trigonometric projections used in our analysis.[3] Sections 3 and 4 show how these cosine transforms can be used to answer trend-related inference questions. "Low-frequency" variability conjures up spectral analysis, and Section 5 uses frequency-domain concepts to explain several facets of the analysis. To keep the focus on key ideas and concepts, the analysis in the body of the paper uses relatively simple stochastic processes, and Section 6 briefly discusses some extensions and concludes.

[2] The low-pass values were computed using an ideal low-pass filter truncated after $T/2$ terms applied to the padded series using forecasts/backcasts constructed from an $AR(4)$ model. The cosine projections are the fitted values from the regression of the series onto a constant term and $\sqrt{2}\cos\left(j\left(t-\frac{1}{2}\right)/T\right)$ for $j = 1, \ldots, 12$.

[3] As discussed below, analogous properties hold for other transforms such as discrete Fourier transforms.

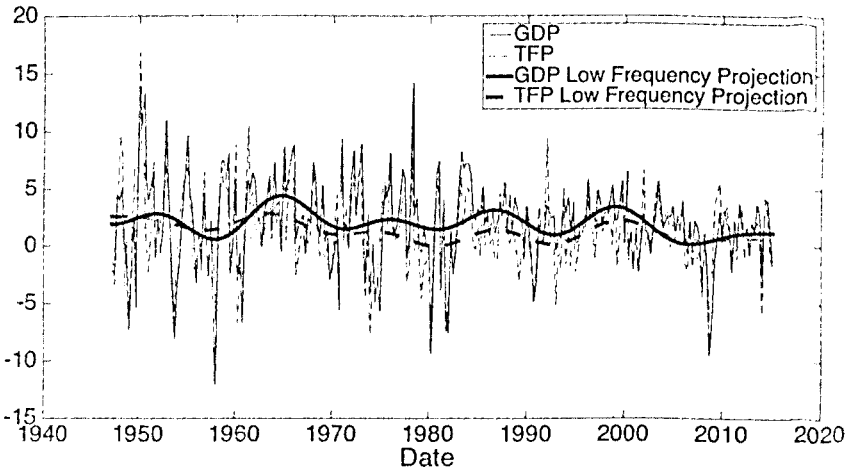

Figure 2 Trends in GDP and TFP Growth Rates

2 SOME PROPERTIES OF LOW-FREQUENCY WEIGHTED AVERAGES

Let x_t denote a scalar time series that is observed for $t = 1, \ldots, T$, and let $\Psi_j(s) = \sqrt{2}\cos(js\pi)$, so that $\Psi_j(t/T)$ has period $2T/j$. Let $\Psi(s) = [\Psi_1(s), \Psi_2(s), \ldots, \Psi_q(s)]'$, a $\mathbb{R}^q$ valued function, and let $\Psi_T = [\Psi((1 - 1/2)/T), \Psi((2-1/2)/T), \ldots, \Psi((T-1/2)/T)]'$ be the $T \times q$ matrix obtained by evaluating $\Psi(\cdot)$ at $s = (t - \frac{1}{2})/T, t = 1, \ldots, T$. The low-frequency projections shown in Figures 1 and 2 are the fitted values from the OLS regression of $[x_1, x_2, \ldots, x_T]$ onto a constant and Ψ_T; that is, $\widehat{x}_t = \bar{x} + \Psi((t - 1/2)/T)'X_T$, where $\bar{x} = T^{-1}\sum_{t=1}^{T} x_t$ and X_T are the OLS regression coefficients. Because the columns of Ψ_T are orthogonal, sum to zero, and have length T (that is, $T^{-1}\Psi_T'\Psi_T = I_q$ and $\Psi_T'\ell_T = 0$, where ℓ_T denotes a $T \times 1$ vector of 1s), the OLS regression coefficients have a simple form

$$X_T = T^{-1}\sum_{t=1}^{T} \Psi((t - 1/2)/T)x_t.$$

The jth regression coefficient, X_{jT}, is called the jth cosine transform of $[x_1, x_2, \ldots, x_T]'$.

2.1 Large-Sample Properties of X_T

Suppose that x_t can be represented as $x_t = \mu + u_t$, where μ is the mean of x and u_t is a zero mean stochastic process. Because the cosine weights sum to zero, the value of μ has no effect on X_T, so cosine weighted averages of x_t and of u_t coincide ($T^{-1}\sum_{t=1}^{T} \Psi((t - 1/2)/T)x_t = T^{-1}\sum_{t=1}^{T} \Psi((t - 1/2)/T)u_t$). Scaled versions of these weighted averages are normally distributed in large

samples if u_t satisfies certain moment and persistence properties. In Section 5.3 we present a central limit theorem that relies on assumptions about the spectral density of u_t. Here we use a simpler argument from Müller and Watson (2008) that relies on the assumption that suitably scaled partial sums of u_t behave like a Gaussian process in large samples.

Specifically, suppose that for some number κ, the linearly interpolated partial sum process of u_t scaled by $T^{-\kappa}$, $G_T(s) = T^{-\kappa}\sum_{t=1}^{\lfloor sT\rfloor} u_t + T^{-\kappa}(sT - \lfloor sT\rfloor)u_{\lfloor sT\rfloor+1}$, converges to a Gaussian process, $G_T(\cdot) \Rightarrow G(\cdot)$. Using an identity for the cosine weights ($\int_{(t-1)/T}^{t/T}\Psi_j(s)ds = l_{jT}^{-1}T^{-1}\Psi_j((t-1/2)/T)$ with $l_{jT} = \sin(j\pi/(2T))/(j\pi/(2T))$, the following representation for the jth cosine transform follows directly

$$
\begin{aligned}
T^{1-\kappa}X_{jT} &= T^{-\kappa}\sum_{t=1}^{T}\Psi_j((t-1/2)/T)u_t \\
&= l_{jT}T^{1-\kappa}\sum_{t=1}^{T}\int_{(t-1)/T}^{t/T}\Psi_j(s)u_{\lfloor sT\rfloor+1}ds \qquad (1)\\
&= l_{jT}\sum_{t=1}^{T}\left[G_T\left(\tfrac{t}{T}\right)\Psi_j\left(\tfrac{t}{T}\right) - G_T\left(\tfrac{t-1}{T}\right)\Psi_j\left(\tfrac{t-1}{T}\right)\right. \\
&\quad \left. - \int_{(t-1)/T}^{1/T}\psi_j(s)G_T(s)ds\right] \\
&= l_{jT}[\Psi_j(1)G_T(1) - \int_0^1 \psi_j(s)G_T(s)ds] \\
&\Rightarrow \Psi_j(1)G(1) - \int_0^1 \psi_j(s)G(s)ds \\
&= \int_0^1 \Psi_j(s)dG(s)
\end{aligned}
$$

where the first two lines use the properties of Ψ_j, the third uses $G_T(s) = T^{1-\kappa}\int_0^s u_{\lfloor rT\rfloor+1}dr$ and integration by parts with $\psi_j(s) = d\Psi_j(s)/ds$, the fifth uses the continuous mapping theorem and $l_{jT} \to 1$, and the final line uses the stochastic calculus version of integration by parts. This representation holds jointly for the elements of X_T, so

$$
T^{1-\kappa}X_T \Rightarrow X = \int_0^1 \Psi(s)dG(s) \sim \mathcal{N}(0, \Sigma). \qquad (2)
$$

The elements of the covariance matrix Σ follow directly from the covariance kernel of G and the cosine weights (see Müller and Watson, 2008). The key idea of our approach to low-frequency econometrics is to conduct inference based on the large sample approximation (2), $T^{1-\kappa}X_T \overset{a}{\sim} \mathcal{N}(0, \Sigma)$.

Figure 3 plots the first 12 cosine transforms for GDP growth rates and inflation. Roughly speaking, the cosine transforms for GDP growth rates look like

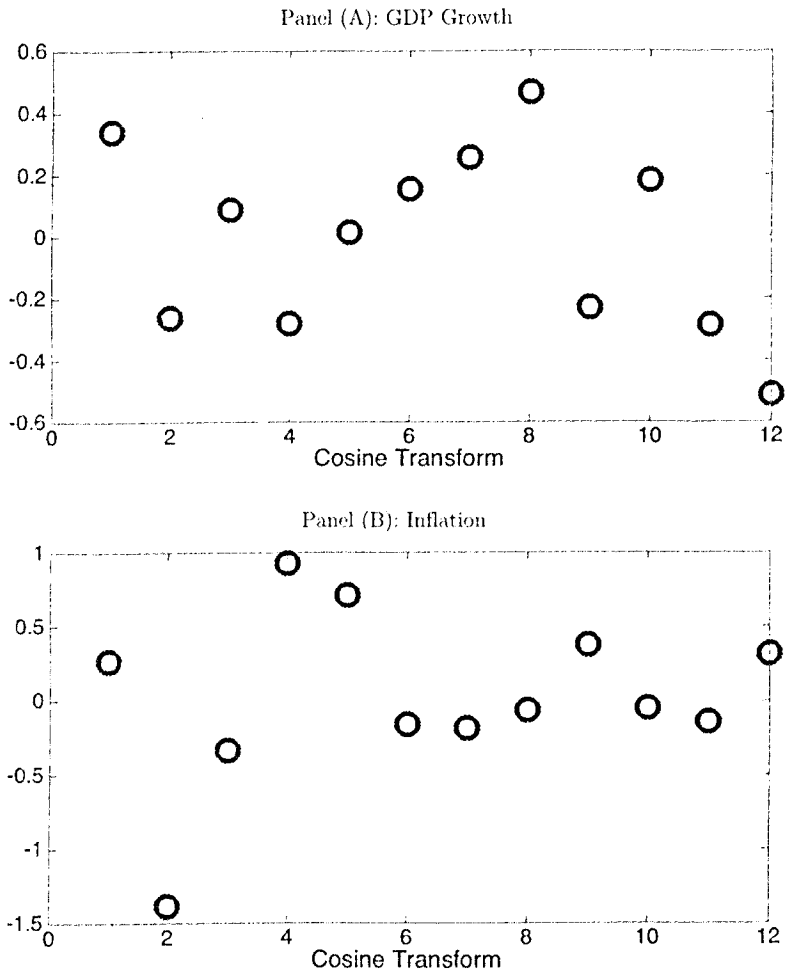

Figure 3 Cosine Transforms of GDP Growth and Inflation

a sample of *i.i.d.* random variables, while the cosine transforms for inflation appear to be heteroskedastic, with larger variance for the first few cosine transforms. The value of the covariance matrix Σ in the $I(0)$ and $I(1)$ models helps explain these differences. In the $I(0)$ model $T^{-1/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow \omega W(\cdot)$ with W a standard Wiener process and $\omega^2 > 0$ the long-run variance, $\Sigma = \omega^2 I_q$, so X_j is $i.i.d.\mathcal{N}(0, \omega^2)$. In the $I(1)$ model $T^{-3/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow \omega \int_0^{\cdot} W(s)ds$, $\Sigma = \omega^2 D$, where D is a diagonal matrix with elements $D_{jj} = 1/(j\pi)^2$, so the X_js are independent but heteroskedastic.[4] In the following section we

[4] The asymptotic covariance matrix Σ is diagonal in both the $I(0)$ and $I(1)$ models because the cosine functions are the eigenfunctions of the covariance kernel of a demeaned Wiener process; cf. Phillips (1998).

present tests for both the $I(0)$ and $I(1)$ null hypotheses. Given these scatter plots, it is not too surprising that the $I(0)$ null is not rejected for GDP growth, but the $I(1)$ null is rejected, and just the opposite result obtains for inflation.

While Σ has a simple form for the $I(0)$ and $I(1)$ models, it can be more complicated for other stochastic processes, so it is useful to have a simple numerical method for obtaining accurate approximations for Σ. As representation (2) suggests, Σ can be approximated by the large-T finite sample covariance matrix $\mathrm{Var}[T^{1-\kappa}X_T] = T^{-2\kappa}\Psi_T'\Xi_T\Psi_T$ where Ξ_T is the $T \times T$ covariance matrix of $[u_1, u_2, \ldots, u_T]'$ under any (not too heavy-tailed) process for u_t that induces the same Gaussian process limit of its partial sums. For example, if u_t is $I(0)$, then Σ can be approximated using the large-T finite-sample covariance matrix of the cosine transform of an *i.i.d.* process with variance ω^2, $\Sigma \approx \omega^2 T^{-1}\Psi_T'\Psi_T = \omega^2 I_q$. Indeed, because $\Sigma = \omega^2 I_q$ for the $I(0)$ model, this approximation is exact. When u_t is $I(1)$ Σ can be approximated using the large-T finite sample covariance matrix of the cosine transforms of a random walk process, $\Sigma \approx T^{-3}\Psi_T'\Xi_T\Psi_T$, where Ξ_T has (i, j)th element $\omega^2 \min(i, j)$. In Section 3, we consider the local-to-unity AR model and fractional $I(d)$ model, and the covariance matrices for these processes can be approximated in analogous fashion.

While this discussion has focused on discrete cosine transforms, the representation in (1) is seen to hold for *any* set of smooth set of weights, $\Psi(s)$. Because our interest is focused on trends, it is important that the weighted averages extract low-frequency variability from the data, and Figure 1 shows that the cosine transforms do just that (also see Section 5.1 below). But, low-frequency Fourier or sine transforms are equally well suited and can be used instead of the cosine transforms after appropriate modification of the covariance matrix Σ.[5]

And finally, it is worth stressing that in the asymptotic analysis presented in this section the dimension of X_T, which we have denoted by q, is held fixed as T grows large. Such asymptotics reflect the notion that there is only a relatively modest amount of information about low-frequency phenomena in, say, 65-year realizations of macro time series. In practice, this translates into using a relatively small number of low-frequency averages from a long time series. Our empirical analysis uses $q = 12$ from $T = 272$ quarterly observations because, from Figures 1 and 2, this produces a sensible notion of the "trend" in inflation and growth rates in GDP and TFP, corresponding to the periodicities longer than 11 years. The econometric challenge then is to draw valid conclusions from these q observations. Further discussion of the choice of q may be found in Section 5.

[5] For example, while the cosine transforms of an $I(1)$ process have the diagonal covariance matrix discussed in the last paragraph, the covariance matrix of Fourier transforms is somewhat more complicated (cf. Akdi and Dickey, 1998).

3 INFERENCE EXAMPLES: UNIVARIATE TIME SERIES

This section takes up five examples of statistical inference involving the low-frequency properties of a univariate stochastic process. In each case, the normal limit for X_T discussed in the last section yields a simple, and standard, small-sample inference problem involving a normal random vector.

3.1 I(0) inference

The first two examples assume the $I(0)$ model and concern inference about the long-run variance (ω^2) and mean (μ) of x. These problems are related as both involve ω^2, in the first instance because it is the parameter of interest, and in the second because it characterizes the variability of the sample mean. There is a large literature on consistent estimation of ω^2, including the important contributions of Newey and West (1987) and others, and earlier work on spectral estimation (see Priestley, 1981 for a classic textbook treatment). As is now widely understood, consistent long-run variance estimators can perform poorly in finite samples even for only moderately persistent series. Motivated by these poor finite-sample properties, a burgeoning literature studies inconsistent long-run variance estimators (see Kiefer, Vogelsang, and Bunzel, 2000 and Kiefer and Vogelsang, 2002, 2005; Jansson, 2004; Phillips, 2005b; Sun, Phillips, and Jin, 2008; Phillips, Sun and Jin, 2006, 2007; and Gonçalves and Vogelsang, 2011; and Müller, 2014 for a recent survey), and estimators constructed from the cosine transforms as suggested by Müller (2004, 2007) provide a leading example.

Thus, consider using X_T to learn about the value of ω^2. Because $\sqrt{T}X_T \Rightarrow X \sim \mathcal{N}(0, \omega^2 I_q)$, asymptotically justified inference proceeds as in the familiar small sample *i.i.d.* normal model. Letting $\mathrm{SSX}_T = TX_T'X_T \Rightarrow \mathrm{SSX} = X'X$ denote the (scaled) sum of squares, the large-sample log-likelihood is proportional to $-\frac{1}{2}[q \ln(\omega^2) + \mathrm{SSX}/\omega^2]$ and the corresponding maximum likelihood estimator (MLE) for ω^2 is $\hat{\omega}_T^2 = \mathrm{SSX}_T/q$. Because $\hat{\omega}_T^2$ is computed using only q observations (and in the asymptotics q is held fixed as $T \to \infty$), $\hat{\omega}_T^2$ is not consistent. Indeed, using $q = 12$, X_T contains limited information about ω^2, so there is considerable uncertainty about its value. Figure 4 shows the log-likelihood for ω constructed using X_T for real GDP growth rates. The MLE is $\hat{\omega}_T = 4.8$, but the plot shows that values of ω as small as 3.5 and as large as 6.5 have likelihood values within one log-point of the maximum. Because $X'X/\omega^2 \sim \chi_q^2$ confidence intervals for ω are readily constructed ($P(\mathrm{SSX}_T/\chi^2_{q,1-\alpha/2} \le \omega^2 \le \mathrm{SSX}_T/\chi^2_{q,\alpha/2}) \to 1-\alpha$, with $\chi^2_{q,p}$ the pth quantile of the Chi-squared distribution with q degrees of freedom), and the resulting confidence intervals for the long-run standard deviation of GDP growth rates and inflation are shown in Table 1.

Next, consider inference about μ, the mean of x_t. The same arguments used in Section 2 show that $\sqrt{T}[(\bar{x}-\mu), X_T']' \Rightarrow (X_0, X')' \sim \mathcal{N}(0, \omega^2 I_{q+1})$, so inference about μ can be obtained as in the standard small-sample normal

Table 1 *Descriptive statistics for GDP growth rates and Inflation: 1947:Q2–2014:Q4*

	GDP	Inflation
Sample mean $\bar{x}$	1.94	3.12
Sample long-run standard deviation $\hat{\omega}_T$	4.76	9.25
90% Confidence interval for ω in $I(0)$ model	[3.60;7.21]	[6.99;14.02]
90% Confidence interval for μ in $I(0)$ model	[1.43;2.45]	[2.12;4.12]

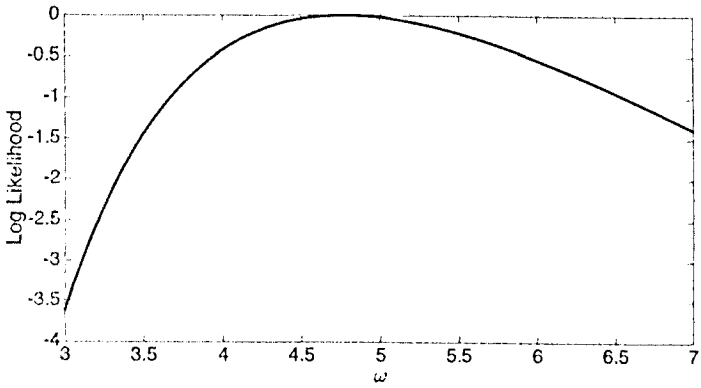

Figure 4 Low-Frequency Log-Likelihood for the Long-Run Standard Deviation of GDP Growth

problem. In particular, $\sqrt{T}(\bar{x} - \mu)/\hat{\omega}_T \Rightarrow t_q$ (the Student-t distribution with q degrees of freedom), so inference about μ follows directly. Table 1 shows 90 percent confidence intervals for the mean GDP growth rate and the mean inflation rate constructed as $\bar{x} \pm 1.78\hat{\omega}_T/\sqrt{T}$, where 1.78 is the 95th percentile of the t_{12} distribution.

These confidence intervals for ω^2 and μ are predicated on the assumption that $x_t \sim I(0)$ and, particularly for inflation, this assumption is suspect. The next inference examples concern the persistence of the x_t process.

3.2 Inference About Persistence Parameters

As discussed in Section 2, the normal limit for X_T holds under conditions more general than the $I(0)$ model, and the parameters characterizing the persistence of x_t determine the covariance matrix Σ. Thus, inference about these parameters can be conducted using methods for inference about the covariance matrix of a multivariate normal random variable. This section discusses three classes of examples. First a few preliminaries.

Some Parametric Models of Persistence. Section 2 discussed the covariance matrix Σ for $I(0)$ and $I(1)$ processes. Here we describe three other widely used parametric models for persistence in economic time series. The first is a sum of $I(0)$ and $I(1)$ processes: with appropriate scaling for the

components this is a version of Harvey's (1989) "local-level-model" (LLM) $x_t = v_{1t} + (bT)^{-1}\sum_{s=1}^{t} v_{2s}$ where (v_{1t}, v_{2t}) is bivariate $I(0)$ with long-run covariance matrix $\omega^2 I_2$. Another is the "local-to-unity AR" model (LTUM) in which $u_t = \rho_T u_{t-1} + v_t$, with AR coefficient $\rho_T = 1 - c/T$ and v_t an $I(0)$ process, as introduced by Cavanagh (1985), Chan and Wei (1987) and Phillips (1987). A third is the "fractional" model (FRM) $(1-L)^d x_t = v_t$ where v_t is $I(0)$ with $-\frac{1}{2} < d < \frac{3}{2}$ (cf. Baillie, 1996 and Robinson, 2003 for surveys). Each of these processes exhibits different degrees of persistence that depend on the value of the model's parameter value. For example, $I(0)$ behavior follows when b is very large in LLM, $d = 0$ in FRM, and c very large in LTUM. The processes exhibit $I(1)$ persistence when $b = 0$ (LLM), $c = 0$ (LTUM) and $d = 1$ (FRM). Alternative parameter values yield persistence between $I(0)$ and $I(1)$, and the FRM with $d < 0$ or $d > 1$ allows for persistence beyond these extremes. Excluding their $I(0)$ and $I(1)$ parameterizations, the models yield subtly different low-frequency behavior as evidenced by their (pseudo-) spectra discussed in Section 5.2. For now we simply note that this behavior results in different values of the covariance matrix Σ.[6]

Scale Invariance/Equivariance. In many inference problems, interest is focused on the persistence of the process (say b, c, or d) and not on the scale of the process (ω). This motivates basing inference on statistics that are invariant to scale transformations, which in our setting corresponds to basing inference on $X_T^s = X_T/\sqrt{X_T' X_T}$.[7] The continuous mapping theorem implies that $X_T^s \Rightarrow X^s = X/\sqrt{X'X}$, and the density $f_{X^s}(x^s|\Sigma)$ of X^s is explicitly computed in King (1980) and proportional to $f_{X^s}(x^s|\Sigma) \propto |\Sigma|^{-1/2}(x^{s\prime}\Sigma^{-1}x^s)^{-q/2}$.

Armed with these preliminaries we are now ready to tackle inference about persistence parameters. We will construct hypothesis tests, and in some cases invert these tests to construct confidence sets. We focus on three classes of efficient tests: (i) point-optimal, (ii) weighted average power (WAP) optimal and (iii) approximately optimal tests constructed using numerical approximations to least favorable distributions. Each is discussed in the context of a specific testing problem.

[6] The scaling factor for X_T is $T^{1/2}$ for LLM, $T^{-1/2}$ for LTUM, and is $T^{1/2-d}$ for FRM. The covariance matrix Σ for each of these models is given in Müller and Watson (2008), and can be approximated using the appropriate choice of Ξ_T for the finite-sample methods in Section 2: the covariance matrix for the LTUM can be approximated using the autocovariances of a stationary AR(1) process with coefficient $(1 - c/T)$; and the covariance matrix for the LLM is the sum of the $I(0)$ covariance matrix and $(bT)^{-2}$ times the $I(1)$ covariance matrix. The covariance matrix for FRM with $-1/2 < d < 1/2$ can be approximated using the autocovariances for $(1-L)^d u_t = \epsilon_t$ where ϵ_t is white noise with variance ω^2 and the autocovariances for u_t are given, for example, in Baillie (1996); for $1/2 < d < 3/2$, Σ can be approximated using the autocovariances of $\sum_{s=0}^{t} u_s$, where $(1-L)^{d-1}u_t = \epsilon_t$. Müller and Watson (2008) show that the resulting limiting covariance matrix Σ is continuous at $d = 1/2$ in the FRM.

[7] Note that the $T^{1-\kappa}$ scale factor cancels in X_T^s, so the value of κ no longer matters for the analysis.

Point-Optimal Tests. Let θ denote the value of a persistence parameter (say b, c, or d in the models described above), so that $\Sigma = \Sigma(\theta)$, and the scale of Σ does not matter because of the form of the density for X^s. Several problems of interest follow the classic Neyman-Pearson setup with a simple null and simple alternative, $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_a$, so the likelihood ratio statistic yields the most powerful test. Given the form of the density f_{X^s}, this means rejecting the null for large values of the ratio of generalized sums of squares: $(X_T'\Sigma(\theta_0)^{-1}X_T)/(X_T'\Sigma(\theta_a)^{-1}X_T)$. Such tests are "point-optimal" in the sense of having best power against the $\theta = \theta_a$ alternative.[8] Two examples are low-frequency "unit-root" (LFUR) and "stationarity" (LFST) tests, which respectively test the $I(1)$ and $I(0)$ null hypotheses, as previously derived in Müller and Watson (2008). We discuss them in turn.

Dufour and King (1991), Elliott, Rothenberg, and Stock (1996) and Elliott (1999) derive efficient Neyman–Pearson tests in the AR(1) Gaussian model with AR coefficient $\rho_0 = 1$ and $\rho = \rho_a$ under the alternative, and the latter two references extend these tests to more general settings using asymptotic approximations for LTUM where the unit root null corresponds to $c_0 = 0$ and the alternative to $c = c_a$. While a uniformly most powerful test does not exist, Elliott, Rothenberg, and Stock (1996) show that the point optimal test associated with a particular value of c_a exhibits near optimality for a wide range of values of c under the alternative. The low-frequency version of the point-optimal unit root test rejects for large values of

$$\text{LFUR} = \frac{X_T'\Sigma(c_0)^{-1}X_T}{X_T'\Sigma(c_a)^{-1}X_T} \tag{3}$$

where $\Sigma(c_0)$ is the covariance matrix under the null (the $I(1)$ model with $c_0 = 0$) and $\Sigma(c_a)$ is the covariance matrix using c_a. Table 2 shows the p-value of this LFUR test for GDP growth rates and inflation using Elliott's (1999) choice of $c_a = 10$. The $I(1)$ model is not rejected for inflation (p-value = 0.18), but is rejected for GDP growth rates (p-value = 0.00).

Nyblom (1989) and Kwiatkowski, Phillips, Schmidt, and Shin (1992) develop tests for the null of an $I(0)$ model versus an alternative in which the process exhibits LLM-type nonstationary behavior.[9] As in the unit root problem, a uniformly most powerful test does not exist, but a Neyman–Pearson point-optimal test against the alternative with $b = b_a$ follows directly. Given the special structure of the $I(0)$ and $I(1)$ Σ matrices, the point-optimal test has a particularly simple form:

$$\text{LFST} = \frac{\sum_{i=1}^{q} X_{iT}^2}{\sum_{i=1}^{q} X_{iT}^2(1 + 1/(b_a i\pi)^2)^{-1}}. \tag{4}$$

[8] Müller (2011) shows that these point optimal tests, viewed as a function of X_T, are asymptotically point optimal in the class of all scale invariant tests that control size for all processes satisfying (2).

[9] See Stock (1994) for a survey and additional references.

Table 2 *Persistence tests for GDP growth rates and inflation*

	GDP	Inflation
	p-values	
LFST	0.53	0.03
LFUR	0.00	0.18
Confidence intervals for d in $I(d)$ model		
67% level	[−0.24;0.27]	[0.48;1.01]
90% level	[−0.49;0.44]	[0.32;1.21]

Notes: LFST and LFUR are point-optimal tests for the I(0) and I(1) models, respectively. Confidence intervals for d in the I(d) model were constructed by inverting WAP-optimal tests for d. See the text for details.

Müller and Watson (2013) tabulate critical value for this test using $b_a = 1/10$, and show that the test has power close to optimal power for a wide range of values of b. Table 2 shows the p-values of this test for GDP growth rates (p-value = 0.53) and for inflation (p-value = 0.03).

Weighted Average Power (WAP) Tests. LFST and LFUR test the $I(d)$ model for $d = 0$ and $d = 1$, but what about other values of d? And, how can the tests be inverted to form a confidence set for the true value of d? One approach is to use point-optimal tests for the null and alternative $H_0 : d = d_0$ versus $H_a : d = d_a$ for various values of d_0; the collection of values of d_0 that are not rejected form a confidence set. But an immediate problem arises: what value of d_a should be used? A desirable test should have good power for a wide range of values of d_a, both larger and smaller than d_0, so that point-optimal tests (which specify a single value of d_a) are not appealing. A useful criterion in these situations is to consider the "weighted average power" (WAP) of tests, where the weights put mass on different values of d_a. As explained in Andrews and Ploberger (1994), for example, optimal WAP tests are easy to construct. The logic is as follows: consider an alternative in which d is a random variable with distribution function F_a. This is a simple alternative H_{a,F_a}: X^s has mixture density $\int f_{X^s}(x^s, \Sigma(d))dF_a(d)$, and the optimal test of H_0 versus H_{a,F_a} is the Neyman–Pearson likelihood ratio test. Rearranging the power expression for this test shows that it has greatest WAP for $H_a : d = d_a$ using F_a as the weight function.[10]

Figure 5 shows the large-sample log likelihood for d using X_T^s for real GDP growth rates and inflation. The MLE of d is near zero ($\hat{d} = -0.06$) for GDP, but much larger ($\hat{d} = 0.80$) for inflation. But the plots indicate substantial uncertainty about the value of d. Using a weighting function F_a that is uniform on $-0.5 < d < 1.5$, we construct WAP tests for a fine grid of values of

[10] As pointed out by Pratt (1961), the resulting confidence set $\mathrm{CS}(X) \subset \mathbb{R}$ for d has minimal expected weighted average length $E[\int \mathbf{1}[d \in \mathrm{CS}(X)]dF_a(d)]$ among all confidence sets of the prespecified level.

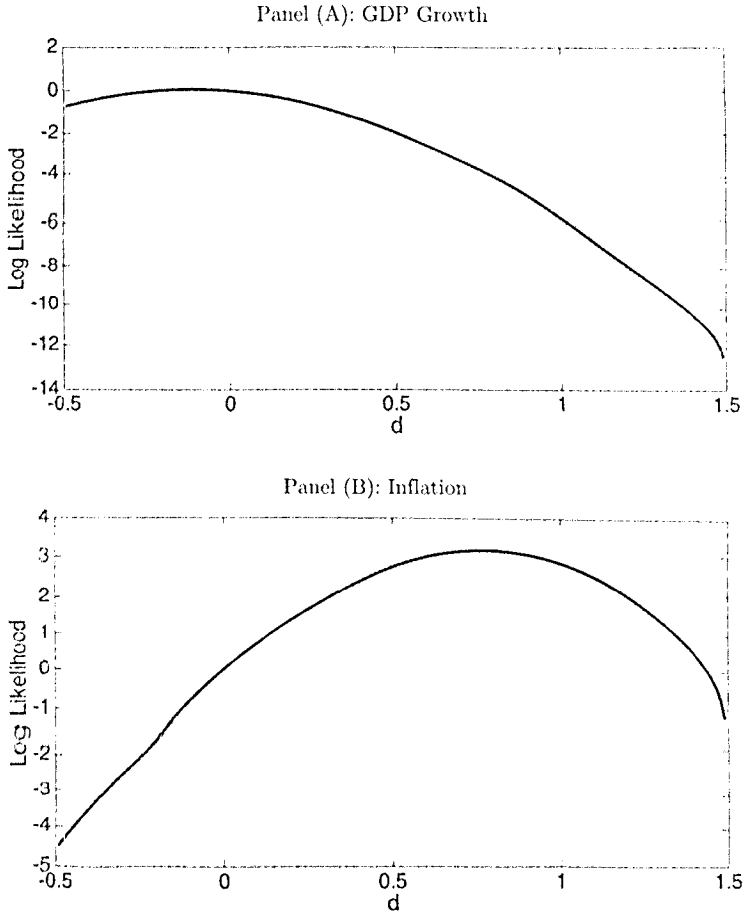

Figure 5 Low-Frequency Log-Likelihood in Fractional Model

d_0. Inverting these tests yields a confidence set for d, and these are shown in Table 2. For GDP, the 67 percent confidence sets for d ranges from -0.24 to 0.27, and for inflation it ranges from 0.48 to 1.01. Ninety percent confidence sets are wider.

Least Favorable Distributions. Our final example considers the following question: how much should we expect the sample average value of x_t (GDP growth rates or inflation) to vary over long periods of time, say a decade or a quarter of century? To be specific, let $\overline{x}_{1:h}$ denote the sample mean constructed using observations 1 through h and consider the variance $\sigma^2_{\Delta\overline{x}}(h) = \text{Var}[\overline{x}_{h+1:2h} - \overline{x}_{1:h}]$, for large values of h. The parameter $\sigma_{\Delta\overline{x}}(h)$ is the standard deviation of the change in the sample mean over adjacent non-overlapping sample periods of length h.[11] In the large sample framework, let

[11] Note that $\sigma_{\Delta\overline{x}}$ is well defined even for some infinite variance processes, such as the FRM with $1/2 < d < 3/2$.

$r = h/T$ and consider asymptotic approximations constructed with r held fixed as $T \to \infty$. To conduct inference about $\sigma^2_{\Delta\bar{x}}(rT)$ using X_T, we must determine how $\sigma^2_{\Delta\bar{x}}(rT)$ relates to the value of Σ, the limiting covariance matrix for X_T (appropriately scaled). This is straightforward: first an extension of the central limit result discussed in Section 2 yields

$$T^{1-\kappa}\begin{pmatrix} \bar{x}_{rT+1:2rT} - \bar{x}_{1:rT} \\ X_T \end{pmatrix} \Rightarrow \begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix}\right).$$

Second, because $\mathrm{Var}[(\bar{x}_{h+1:2h} - \bar{x}_{1:h})/\sigma_{\Delta\bar{x}}(rT)] = 1$, $X_T/\sigma_{\Delta\bar{x}}(rT) \Rightarrow \mathcal{N}(0, \Gamma)$ where $\Gamma = \Sigma_{XX}/\Sigma_{YY}$. A test of $H_0 : \sigma_{\Delta\bar{x}}(rT) = \sigma_{\Delta\bar{x},0}(rT)$ against $H_a : \sigma_{\Delta\bar{x}}(rT) \neq \sigma_{\Delta\bar{x},0}(rT)$ may thus be based on the statistic $X_\sigma = X_T/\sigma_{\Delta\bar{x},0}(rT)$. Under H_0, $X_\sigma \Rightarrow \mathcal{N}(0, \Gamma)$, and under H_1, $X_\sigma \Rightarrow \mathcal{N}(0, (\frac{\sigma_{\Delta\bar{x}}(rT)}{\sigma_{\Delta\bar{x},0}(rT)})^2\Gamma)$, so the problem again reduces to an inference problem about the covariance matrix of a multivariate normal.

The alternative can be reduced to a single alternative H_{a,F_a} by maximizing a weighted average power criterion, as above. The problem is still more challenging than the previous example, however, because the (large sample) *null* distribution of $X_\sigma \Rightarrow \mathcal{N}(0, \Gamma)$ depends on nuisance parameters that describe the persistence of the x_t process. In the FRM, for example, $\Gamma = \Gamma(d)$ with $d \in D$ for some range of values D. In the jargon of hypothesis testing, the null hypothesis is composite because it contains a set of probability distributions for X_σ indexed by the value of d. The challenge is to find a powerful test of H_0 that controls size under *all* values of $d \in D$.

One general solution to this type of problem uses the same simplification that was used to solve the weighted average power problem: introduce a weighting function, say $\Lambda(d)$ for the values of $d \in D$ and use the mixture density $H_{0,\Lambda} : X_\sigma \sim \int f_{X_\sigma}(x_\sigma, \Gamma(d))d\Lambda(d)$ as the density under the null. This yields a simple null hypothesis (that is, a single probability density for the data), so the best test is again given by the likelihood ratio. However, while the resulting test will control the probability of a false null rejection *on average* over the values of d with d drawn from Λ, it won't necessarily control the rejection probability for *all* values of d allowed under H_0; that is, the test may not have the correct size. But, because any test that controls the null rejection probability for *all* values of d automatically controls the rejection frequency *on average* over d, any test with size α under H_0 is also a feasible size α under under $H_{0,\Lambda}$, so it must have power less than or equal to the power-maximizing test for $H_{0,\Lambda}$. Thus, if a distribution Λ can be found that does control size for all d, the resulting likelihood ratio test for $H_{0,\Lambda}$ is guaranteed to be the optimal test for H_0. Such a Λ is called a "least favorable distribution" (LFD). In some problems, least favorable distributions can be found by clever reasoning (see Lehmann and Romano, 2005 for examples), but this is the exception rather than the rule. Numerical methods can alternatively be used to approximate the LFD (see Elliott, Müller, and Watson, 2015 for discussion and examples), and we will utilize that approach here.

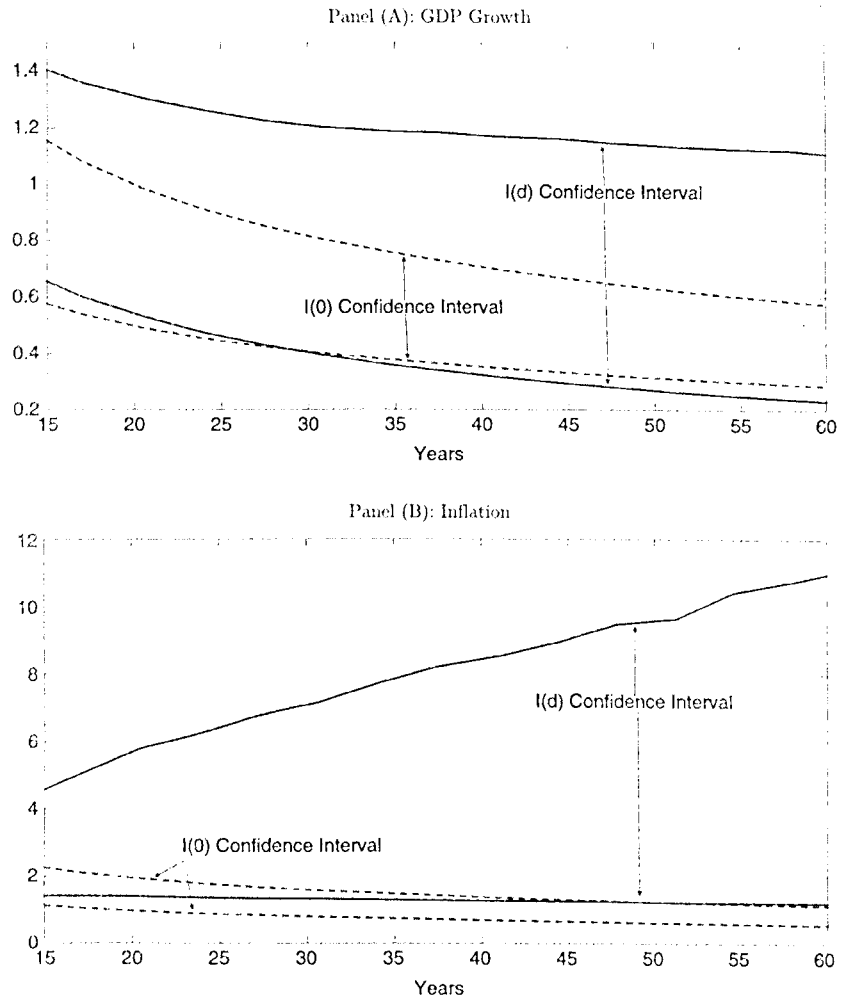

Figure 6 90% Confidence Sets for Standard Deviation of $\bar{x}_{h+1:2h} - \bar{x}_{1:h}$ as a function of h

With this background out of the way, we can now return to the problem of determining the value of $\sigma^2_{\Delta\bar{x}}(h)$. We do this in the $I(d)$ model for persistence, and allow d to take on any value between -0.4 to 1.4. The weighting in the WAP criterion is uniform on d, and, conditional on d, sets the alternative covariance matrix equal to $\Gamma_a = \Gamma(d)\, e^U$ where U is uniformly distributed on $(-5, 5)$.[12] Figure 6 summarizes the resulting (pointwise in h) 90% confidence sets for $\sigma_{\Delta\bar{x}}(h)$ for h ranging from 15 to 70 years. For averages of GDP growth rates computed over 20 years, the 90 percent confidence interval

[12] The numerical work uses fine discrete grids for the Γ_a and the approximate LFD.

for $\sigma_{\Delta\bar{x}}$ ranges from approximately 0.5 to 1.3 percentage points, and the range shifts down somewhat (to 0.25 to 1.2 percentage points) for averages computed over 60 years. The lower range of the confidence set essentially coincides with the corresponding $I(0)$ confidence set (which only reflects uncertainty about the value of the long-run variance of GDP growth), but the upper range is substantially larger than its $I(0)$ counterpart, reflecting that the data does not rule out some non-$I(0)$ persistence. The confidence intervals for inflation indicate both large and uncertain values for $\sigma_{\Delta\bar{x}}$, reflecting the clearly larger persistence in the series. The uncertainty about persistence in inflation leads to very wide confidence intervals for averages computed over long samples.

4 INFERENCE EXAMPLES: MULTIPLE TIME SERIES

In this section we use the same structure and tools discussed in the last section to study some inference problems involving multiple time series. Thus, now let x_t denote an $n \times 1$ vector of time series and X_{jT} denote the $n \times 1$ vector composed of the jth cosine transform for each of the n variables. Let X_T denote a $q \times n$ matrix with jth row given by X'_{jT}. In the multivariate model, the assumption of Section 2 about the large sample behavior of the partial sum process becomes $\Upsilon_T \sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow G(\cdot)$, where G is now an $n \times 1$ Gaussian process and Υ_T is an $n \times n$ scaling matrix.[13] Using the argument from Section 2, this implies $T\Upsilon_T X_T \Rightarrow X$ with $\text{vec}(X) \sim \mathcal{N}(0, \Sigma)$. In this section we discuss inference problems in the context of this limiting distribution. We first consider the multivariate $I(0)$ model and inference about its key parameters. We then relate these parameters to population properties of the trend projections, and return to the empirical question of the relationship between the trend in GDP and TFP growth rates discussed in the introduction. We end this section with a discussion of inference in cointegrated models; these are characterized by linear combinations of the data that are $I(0)$ and other linear combinations that are $I(1)$.

4.1 Inference in the $I(0)$ Model

In the multivariate $I(0)$ model $x_t = \mu + u_t$, we set $\Upsilon_T = T^{-1/2} I_q$ and assume $T^{-1/2} \sum_{t=1}^{\lfloor \cdot T \rfloor} u_t \Rightarrow \Omega^{1/2} W(\cdot)$, where Ω is the $n \times n$ long-run covariance matrix of u_t and W is a $n \times 1$ multivariate standard Wiener process. The covariance matrix of $\text{vec}(X)$ then becomes $\Sigma = \Omega \otimes I_q$, so that $X_j \sim i.i.d.\mathcal{N}(0, \Omega)$, where X'_j is the jth row of X.

Inference About Ω. With $\sqrt{T} X_{jT}$ approximately $i.i.d.\mathcal{N}(0, \Omega)$, standard methods for *i.i.d.* multivariate normal samples (see, for instance, Anderson,

[13] The scaling matrix Υ_T replaces the scale factor $T^{1-\kappa}$ of Section 2 to allow the partial sum of linear combination of u_t to converge at different rates, which occurs, for example, if the elements of u_t are cointegrated.

1984) can be used to obtain inference. In particular, the limiting distribution of the multivariate sum of squares is Wishart: $\mathrm{SSX}_T = TX_T'X_T \Rightarrow \mathrm{SSX} = X'X \sim W_n(\Omega, q)$, and $\hat{\Omega} = \mathrm{SSX}_T/q$ is the (low-frequency) MLE of Ω. This can be used directly for inference about the $n \times n$ parameter Ω. Often it will be more interesting to conduct inference about the scalar correlation parameters $\rho_{ij} = \Omega_{ij}/\sqrt{\Omega_{ii}\Omega_{jj}}$. By the continuous mapping theorem, $\hat{\rho}_{ij} = \hat{\Omega}_{ij}/\sqrt{\hat{\Omega}_{ii}\hat{\Omega}_{jj}} \Rightarrow \mathrm{SSX}_{ij}/\sqrt{\mathrm{SSX}_{ii}\,\mathrm{SSX}_{jj}}$. This limiting distribution is known and is a function only of ρ_{ij} and q (see Anderson's 1984 Theorem 4.2.2), allowing the ready construction of confidence intervals for ρ_{ij} based on $\hat{\rho}_{ij}$. For example, for the GDP and TFP growth rates, we obtain an equal-tailed 90 percent confidence interval for the low-frequency correlation of [0.28; 0.86].

Now partition x_t into a scalar y_t and a $k \times 1$ vector z_t, $x_t = (y_t, z_t')'$ (so that $k = n - 1$), with corresponding cosine transforms $X_{jT} = (Y_{jT}, Z_{jT}')'$ and

$$\sqrt{T}\begin{pmatrix} Y_{jT} \\ Z_{jT} \end{pmatrix} \Rightarrow \begin{pmatrix} Y_j \\ Z_j \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \Omega_{yy} & \Omega_{yz} \\ \Omega_{zy} & \Omega_{zz} \end{pmatrix}\right).$$

A function of Ω of potential interest is the $k \times 1$ vector $\beta = \Omega_{zz}^{-1}\Omega_{zy}$. Since conditional on $Z = (Z_1, \ldots, Z_q)$, $\varepsilon_j = Y_j - Z_j'\beta$ is $i.i.d.\mathcal{N}(0, \sigma^2)$ with $\sigma^2 = \Omega_{yy} - \Omega_{yz}\Omega_{zz}^{-1}\Omega_{zy}$, β is the population regression coefficient in a regression of Y_j on Z_j, $j = 1, \ldots, q$. Substantively, $Z_j'\beta$ provides the best predictor of Y_j given Z, so it gauges how low-frequency variability in z_t predicts corresponding low-frequency variability of y_t. The population R^2 of this regression (that is, the square of the multiple correlation coefficient) is $\rho^2 = 1 - \sigma^2/\Omega_{yy}$ and it measures the fraction of the low-frequency variability in y_t that can be explained by the low-frequency variability in z_t.[14]

As long as $k < q$, inference about these regression parameters follow immediately from standard small sample results for a linear regression with *i.i.d.* normal errors:

$$\hat{\beta}_T = \left(\sum_{j=1}^{q} Z_{jT}Z_{jT}'\right)^{-1}\left(\sum_{j=1}^{q} Z_{jT}Y_{jT}\right) \Rightarrow \hat{\beta}$$

$$= \left(\sum_{j=1}^{q} Z_j Z_j'\right)^{-1}\left(\sum_{j=1}^{q} Z_j Y_j\right)$$

$$\mathrm{RSS}_T = T\sum_{j=1}^{q}(Y_{jT} - Z_{jT}'\hat{\beta}_T)^2 \Rightarrow \mathrm{RSS} = \sum_{j=1}^{q}(Y_j - Z_j'\hat{\beta})^2$$

[14] Regressions like these using discrete Fourier transforms are familiar in the literature on band-spectral regression (e.g., Engle, 1974). Much of the analysis presented here can be viewed as versions of these methods for a narrow frequency band in the $1/T$ neighborhood of zero.

with $\text{RSS}/\sigma^2 \sim \chi^2_{q-k}$, the standard error of the regression is $\hat{\sigma}_T^2 \approx \text{RSS}_T/(q-k) \Rightarrow \hat{\sigma}^2 = \text{RSS}/(q-k)$, the total sum of squares is $\text{TSS}_T = T\sum_{j=1}^q Y_{jT}^2 \Rightarrow \text{TSS} = \sum_{j=1}^q Y_j^2$ and the regression R^2 is $R_T^2 = 1 - \text{RSS}_T/\text{TSS}_T \Rightarrow R^2 = 1 - \text{RSS}/\text{TSS}$. By standard arguments (e.g., Anderson, 1984 section 4.4.3), the distribution of R^2 depends only on ρ^2 and q, so that a confidence set for ρ^2 is readily computed from R^2. Furthermore, with $\hat{S}_T = \hat{\sigma}_T^2(q^{-1}T\sum_{j=1}^q Z_{jT}Z_{jT}')^{-1} \Rightarrow \hat{S} \approx \hat{\sigma}^2(q^{-1}\sum_{j=1}^q Z_j Z_j')^{-1}$, inference about β can be performed by the usual F-statistic $F = q(\hat{\beta}_T - \beta)'\hat{S}_T^{-1}(\hat{\beta}_T - \beta) \Rightarrow F_{k,q-k}$ (with $F_{n,m}$ the central F-distribution with n and m degrees of freedom in the numerator and denominator, respectively). For scalar elements of β, β_i, usual t-statistic inference is applicable: $\sqrt{q}(\hat{\beta}_{Ti} - \beta)/\sqrt{\hat{S}_{Tii}} \Rightarrow \sqrt{q}(\hat{\beta}_i - \beta)/\sqrt{\hat{S}_{ii}} \sim$Student-$t_{q-k}$.

Inference About μ. Let $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)'$ be the $n \times 1$ vector of sample means. In exact analogy to the derivations of Section 3.1, we now have

$$\sqrt{T}\,\text{vec}\begin{pmatrix} (\bar{x}-\mu)' \\ X_T \end{pmatrix} \Rightarrow \mathcal{N}(0, \Omega \otimes I_{q+1})$$

so that in large samples, $\sqrt{T}((\bar{x}-\mu), X_{1T}, X_{2T}, \ldots, X_{qT})' \sim i.i.d.\mathcal{N}(0, \Omega)$. Thus, as previously suggested by Müller (2004), Hotelling's (1931) T^2 statistic $T(\bar{x}-\mu)'\hat{\Omega}_T^{-1}(\bar{x}-\mu) \Rightarrow \frac{nq}{q+1-n}F_{n,q+1-n}$ provides a basis for inference about μ. Figure 7 shows 67 percent and 90 percent confidence ellipses for the average growth rates of TFP and GDP based on the T^2 statistic.

Interpreting Elements of Ω *in Terms of Time-Domain Projections:* Figure 2 plotted low-frequency trends for GDP and TFP growth rates, and the two series appear to be highly correlated. Figure 8 plots the cosine transforms of the two series which are also highly correlated. These are closely related. Denoting the two series by y_t (GDP) and z_t (TFP), the projections are $\hat{y}_t = \bar{y} + Y_T'\Psi((t-1/2)/T)$ and $\hat{z}_t = \bar{z} + Z_T'\Psi((t-1/2)/T)$ so

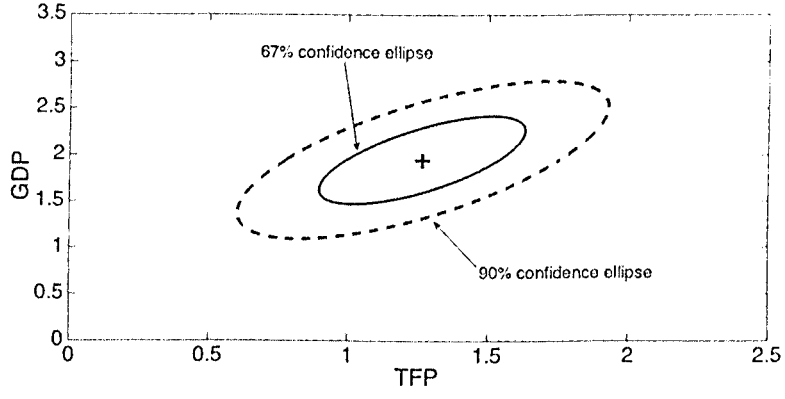

Figure 7 Joint Confidence Sets for Mean GDP and TFP Growth

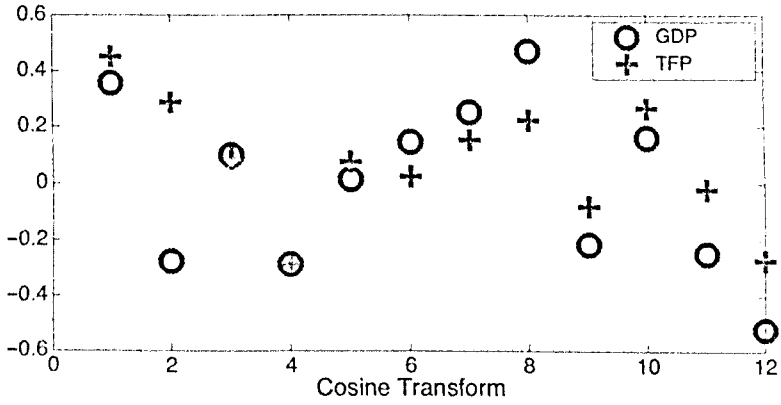

Figure 8 GDP and TFP Growth Cosine Transforms

$T^{-1}\sum_{t=1}^{T}(\hat{z}_t - \bar{z})(\hat{y}_t - \bar{y}) = Z_T'Y_T$. Thus, the variability and covariability of the projection sample paths, $(\hat{y}_t, \hat{z}_t)$, are determined by the cosine transforms (Y_T, Z_T) and the sample means $(\bar{y}, \bar{z})$. Consider the projection sample path of y_t, centered at μ_y and expressed as a fraction s of the sample, $\hat{y}_{\lfloor sT \rfloor} - \mu_y$. Then, from Section 2, $T^{1/2}(\hat{y}_{\lfloor \cdot T \rfloor} - \mu_y) \Rightarrow p_y(\cdot) = Y_0 + Y'\Psi(\cdot)$, and similarly for z_t. Thus, the large-sample variability and covariability of the projections correspond to the variability and covariability of $\tilde{X} = (X_0, X')'$ and $\tilde{Y} = (Y_0, Y')'$. For example, $E\left[\int_0^1 p_y(s)^2 ds\right] = E[\mathrm{tr}(\tilde{Y}\tilde{Y}')] = (q+1)\Omega_{yy}$, $E\left[\int_0^1 p_z(s)^2 ds\right] = (q+1)\Omega_{zz}$, and $E\left[\int_0^1 p_y(s)\, p_z(s)\, ds\right] = (q+1)\Omega_{yz}$.

Thus, $\rho_{yz} = \Omega_{yz}/\sqrt{\Omega_{yy}\Omega_{zz}}$ is alternatively interpreted as the population correlation between the low-frequency trends $p_y(s)$ and $p_z(s)$, averaged over the sample fraction s. Correspondingly, the regression coefficient $\beta = \Omega_{zz}^{-1}\Omega_{zy}$ is the population regression coefficient of a continuous time regression of $p_y(\cdot)$ on $p_z(\cdot)$. The conditionally mean zero error function in this regression, $p_\varepsilon(\cdot) = p_y(\cdot) - \beta p_z(\cdot)$, is the part of the variation of $p_y(\cdot)$ that is independent of $p_z(\cdot)$. Also, the sample correlation coefficient $\hat{\rho}_{yz} = \hat{\Omega}_{yz}/\sqrt{\hat{\Omega}_{yy}\hat{\Omega}_{zz}}$ is recognized as the sample correlation between the projections $\hat{y}_t$ and $\hat{z}_t$, $t = 1, \ldots, T$, and the sample regression coefficient $\hat{\beta}_T$ can alternatively be computed by a linear regression of $\hat{y}_t$ on $\hat{z}_t$ and a constant α.

Table 3 shows the results from the OLS regression of the cosine transforms of the growth rate of GDP onto the growth rate of TFP. The OLS estimate of β is 0.88, suggesting that a trend increase in the TFP growth rate of 1% leads to a 0.88% increase in the trend growth rate of GDP. However, this estimate is based on only 12 observations, and the 90% confidence interval for β ranges from 0.38 to 1.39.[15] The regression R^2 is roughly 50%, and the 90% confidence interval for the correlation of the GDP and TFP trends ranges from 0.31 to 0.86,

[15] To put these values in perspective, recall that the standard neoclassical model, which exhibits balanced growth, implies that a permanent one percentage point increase in TFP leads to a

Table 3 *OLS regression of cosine transform of GDP growth rates on TFP growth rates*

Statistic	
Sample size (q)	12
$\hat{\beta}$	0.88
$se(\hat{\beta})$	0.28
t-statistic	3.16
90% confidence interval for β	[0.38;1.39]
Standard error ($\hat{\sigma}$)	0.22
R^2	0.48
90% confidence interval for ρ^2	[0.09;0.75]

Notes: The 90% confidence interval for β is computed using the Student-t distribution with 11 degrees of freedom. The 90% confidence interval for ρ^2 is based on the 90% confidence interval for ρ, which in turn is computed from the exact distribution of the sample correlation coefficient $\hat{\rho}$ as explained in the text.

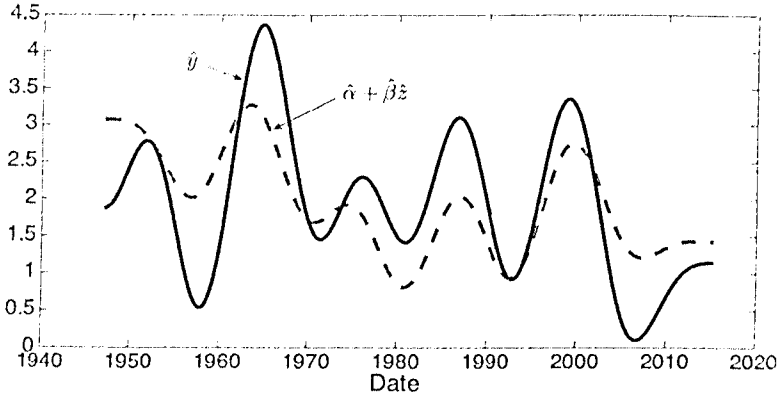

Figure 9 GDP Growth Trend and Predicted Trend from Regression on TFP Growth

suggesting that variations in TFP are important, but not the sole factor, behind variation in trend per-capita GDP growth rates. Figure 9 plots the historical trend growth of GDP, $\hat{y}_t$, along with the predicted values $\hat{\alpha} + \hat{\beta}\hat{z}_t$ from the low-frequency linear regression. TFP explained slightly more than half of the above average trend GDP growth in the 1960s and nearly all of the below and then above average growth in 1990s, but explains little of low growth in the late 1950s. More recently, the plot indicates that TFP growth played only a small role in the slower than average growth in GDP over the past decade.[16]

$\beta = (1 - \zeta)^{-1}$ percentage point long-run increase in GDP, where ζ denotes the elasticity of output with respect to capital. Thus $\beta = 1.5$ if $\zeta = 2/3$.

[16] These empirical results are sensitive to the population series used to construct the per-capita values of real GDP. The data used here are from the US NIPA accounts (Table 7.1) which use the total US population including armed forces overseas and the institutionalized population.

4.2 Cointegration

In the general cointegration model there are n variables with different linear combinations of the variables being integrated of different orders. The notation for the general model can be taxing, but many insights can be gleaned from a bivariate model with $x_t = [y_t, z_t]'$, where y_t and z_t are scalar, $z_t \sim I(1)$ and $v_t = y_t - \beta z_t \sim I(0)$. The partial sum process then satisfies

$$\begin{pmatrix} T^{-1/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} v_t \\ T^{-3/2}\sum_{t=1}^{\lfloor \cdot T \rfloor} z_t \end{pmatrix} \Rightarrow \begin{pmatrix} \omega_1 W_1(\cdot) \\ \omega_2 \int_0^{\lfloor \cdot \rfloor} (\rho W_1(s) + (1-\rho^2)^{1/2} W_2(s))ds \end{pmatrix}$$

where W_1 and W_2 are independent Wiener processes and the parameter ρ denotes the long-run correlation between the $I(0)$ and $I(1)$ components. Thus, letting $V_T = Y_T - \beta Z_T$, $[T^{1/2}V_T, T^{-1/2}Z_T] \Rightarrow [V, Z]$ with $[V', Z']' \sim \mathcal{N}(0, \Sigma)$. Because $v_t \sim I(0)$ and $z_t \sim I(1)$, the covariance matrix Σ has the partitioned form: $\Sigma_{VV} = \omega_1^2 I$, $\Sigma_{ZZ} = \omega_2^2 D$, and $\Sigma_{VZ} = \omega_1\omega_2\rho D^{1/2}$, where D is the $I(1)$ covariance matrix defined in Section 2.

There are a variety of inference questions that arise in the cointegration model, and several of these are straightforward to address using the low-frequency transforms of the data.[17] For example, one question asks whether the cointegrating coefficient takes on a specific value, that is whether $\beta = \beta_0$. Müller and Watson (2013) take up this question using the low-frequency framework. The idea is straightforward. Write $Y_T - \beta_0 Z_T = V_T - (\beta - \beta_0)Z_T$. If $\beta = \beta_0$, then the term involving Z_T is absent, but otherwise this component is present. Thus, $[T^{1/2}(Y_T - \beta_0 Z_T), T^{-1/2}Z_T] \Rightarrow [V + b_0 Z, Z]$ where $b_0 = T(\beta_0 - \beta)$. The covariance matrix of $[(V + b_0 Z)', Z']'$ is readily computed and depends on b_0. Thus, a hypothesis about the value of the cointegrating coefficient, say $\beta = \beta_0$ or equivalently $b_0 = 0$, can be tested as a restriction on the covariance matrix. After imposing invariance restrictions, Müller and Watson (2013) show that the scales ω_1 and ω_2 can be set to unity, so that ρ is the only remaining parameter, and develop an optimal test for $\beta = \beta_0$ using a numerical approximation to the LFD as discussed in the last section. Thus, as in the other inference examples in the low frequency setting, optimal inference about cointegrating coefficients becomes a standard problem involving the covariance matrix of a normal random vector.

Perhaps a more interesting question involves the cointegration model's assumption that z_t follows an $I(1)$ process. What if z_t followed another persistent process, perhaps one of the parametric processes described in the last section? In the context of "efficient" regression inferences about β, Elliott (1998) showed that the $I(1)$ assumption was crucial in the sense that large

We have also carried out the analysis using the non-institutionalized civilian population over the age of 16. This alternative population series produces per-capita values of GDP growth more closely related to long-run movements in TFP; the regression yields $\hat{\beta} = 1.2$, $R^2 = 0.6$, and 90 percent confidence interval that includes values of β between 0.7 to 1.6.

[17] Also see Bierens (1997) and Phillips (2005a, 2014) for related approaches.

size distortions can arise if z_t follows a LTUM instead of an exact $I(1)$ process. The low-frequency analysis outlined in the last paragraph is not immune to Elliott's critique: it uses the full covariance matrix for $[(V+b_0Z)', Z']'$, and therefore utilizes the $I(1)$ property of z_t (through the $I(1)$ covariance matrix D that appears in Σ_{ZZ} and Σ_{VZ}). Müller and Watson (2013) study optimal tests for $\beta = \beta_0$ under alternative assumptions about the trend process. In particular, they show that if the z_t process is unrestricted, then the (essentially) most powerful test simply involves testing whether $y_t - \beta_0 z_t$ is $I(0)$ using the LFST test in (4), a solution to the Elliott critique originally proposed by Wright (2000), although not in the low-frequency setting.[18]

5 RELATIONSHIP TO SPECTRAL ANALYSIS

This section relates our method to spectral analysis. A first subsection discusses why one cannot simply use traditional spectral analysis inference tools to answer questions about low-frequency variability. A second subsection derives the limiting covariance matrix of the cosine transforms in terms of the spectral density of the underlying time series. A calculation shows that the covariance matrix is fully determined by the shape of the spectral density close to the origin, a function we call the "local-to-zero spectrum."[19] In the third subsection, we present a central limit theorem for low frequency transforms under general forms of the local-to-zero spectrum.

5.1 Scarcity of Low-Frequency Information and its Extraction Using Cosine Transforms

Scarcity of Low-Frequency Information. In traditional spectral analysis (see, for instance, Priestley, 1981, or Brockwell and Davis, 1991), the periodogram is the starting point to learn about the spectral properties of a time series. Figure 10 plots the periodogram for quarterly GDP growth rates. The shaded portion of the figure shows the frequencies corresponding to periods of 11 years or longer. A mere six periodogram ordinates fall into this low-frequency region, and this count would remain unaltered if GDP growth was instead sampled at the monthly or even weekly frequency. Intuitively, 67 years of data contain only limited information about components with periods of 11 years or longer.

Traditional inference about the shape and value of the spectral density is based on averaging periodogram ordinates. The asymptotics are driven by an assumption that the spectrum is locally flat (that is, the spectrum is fixed

[18] The Müller and Watson (2013) result on the near-optimality of the LFST test holds in the cointegration model with a single cointegrating vector. If there are multiple cointegrating vectors then it is possible to obtain more powerful tests even absent assumptions on the process for the common stochastic trends.

[19] This subsection largely follows the development in Müller and Watson (2016).

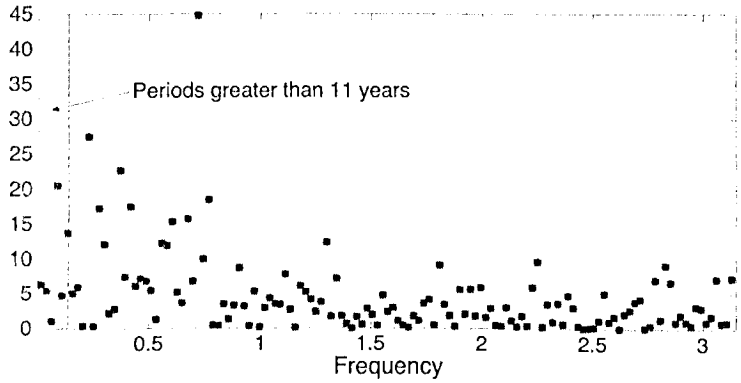

Figure 10 Periodogram of GDP Growth Rates

and continuous), so that as the sample size grows, eventually there will be an increasing number of relevant periodogram ordinates to estimate the value of the spectral density at any given point. Under such asymptotics, Laws of Large Numbers and Central Limit Theorems are applicable and enable asymptotically normal inference.

But with the relevant number of periodogram ordinates as small as six, such asymptotics do not provide good approximations. What is more, in any non-$I(0)$ model, the spectrum has non-negligible curvature even within a $O(T^{-1})$ band around zero, and this band contains a fixed number of periodogram ordinates. For these reasons, the traditional inference tools of spectral analysis are not directly applicable.

Extracting Low-Frequency Information Using Cosine Transforms. Do a small number of cosine transforms do a good job at extracting the low-frequency information that *is* contained in the sample data? One useful way to answer this question is to consider a perfectly periodic time series of frequency ϕ, $x_t = \sin(\phi t + \delta)$ with phase shift $\delta \in [0, \pi)$. Ideally, the low-frequency cosine transforms X_T would capture the entire variation of x_t for $0 \leq \phi \leq \bar{\phi}$ for small values of $\bar{\phi}$ and none of the variation if $\phi > \bar{\phi}$. In large samples, with $\bar{\phi} = \bar{\lambda}/T$ for fixed $\bar{\lambda}$, the corresponding R_T^2 of a regression of x_t on Ψ_T is well approximated by the R^2 of a continuous time regression of $\sin(\lambda s + \phi)$ on the $q \times 1$ cosine functions $\Psi(s)$ and a constant for $s \in [0, 1]$. Figure 11 plots this continuous time R^2 as a function of λ for $\bar{\lambda} = 12\pi$ and q = 12. While the R^2 plot does not follow the ideal step function form, it still provides evidence that the low-frequency cosine transforms do a reasonably good job at extracting the low-frequency information of interest.[20]

[20] Low-frequency Fourier transforms perform similarly well for extracting low-frequency variations. See Müller and Watson (2008) for the corresponding R^2 plot.

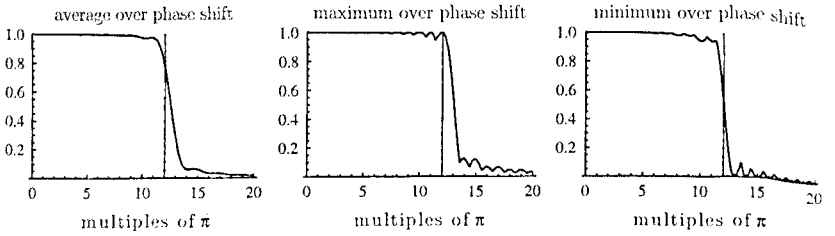

Figure 11 R^2 Plot of Generic Trigonometric Series on Cosine Transforms

5.2 Local-to-Zero Spectra

Because Σ is the limiting covariance matrix of the cosine transforms, which in turn are weighted averages of the underlying series x_t, the elements of Σ depend on the autocovariances of the x-process. These, in turn, are linked to spectrum of the process, and this makes it possible to express Σ as a function of the spectrum. Because the cosine transforms extract low-frequency information in x, it should not be surprising that it is the spectrum close to zero that determines Σ. The remainder of this subsection makes this explicit by showing that Σ can be expressed in terms of the spectrum evaluated at frequencies in a $O(T^{-1})$ neighborhood of frequency zero, a function we call the "local-to-zero spectrum".

Local-to-Zero Spectrum for a Stationary Process. Suppose that x_t is a stationary process with spectral density $f_T(\phi)$.[21] Suppose that in the $O(T^{-1})$ neighborhood of the origin, the suitably scaled spectral density converges (in an L_1 sense)

$$T^{1-2\kappa} f_T(\lambda/T) \to S(\lambda). \tag{5}$$

For example, the FRM is traditionally defined by the assumption that $f_T(\phi)$ is proportional to $|\phi|^{-2d}$ for small ϕ, so that (5) holds with $\kappa = \frac{1}{2} - d$ and $S(\lambda) \propto |\lambda|^{-2d}$. More generally, the function $S(\lambda)$ is the large sample limit of the shape of the original spectrum f_T close to the origin, and we correspondingly refer to it as the local-to-zero spectrum.

Now consider a weighted average of x_t,

$$\eta_T = \int_0^1 g(s) x_{\lfloor sT \rfloor + 1} ds = T^{-1} \sum_{t=1}^{T} \tilde{g}_{T,t} x_t$$

for some Riemann integrable function g and $\tilde{g}_{T,t} = T \int_{(t-1)/T}^{t/T} g(s) ds$. In light of (1), the cosine transforms are an example of η_T. Recalling that the jth autocovariance of x_t is given by $\int_{-\pi}^{\pi} f_T(\phi) e^{-\mathrm{i}\phi j} d\phi$, where $\mathrm{i} = \sqrt{-1}$, we obtain for the covariance between two such weighted averages, η_T^1 and η_T^2:

[21] The T subscript on f_T accommodates "double-array" processes such as the LTUM in which the AR(1) coefficient depends on T. To ease notation, we omit the corresponding T subscript on $x_{T,t} = x_t$ in this subsection.

$$
\begin{aligned}
T^{2(1-\kappa)}E[\eta_T^1\eta_T^2] &= T^{-2\kappa}E\left[\left(\sum_{s=1}^{T}\tilde{g}_{T,s}^1 x_s\right)\left(\sum_{t=1}^{T}\tilde{g}_{T,t}^2 x_t\right)\right] \\
&= T^{-2\kappa}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{g}_{T,s}^1\tilde{g}_{T,t}^2 E[x_s x_t] \\
&= T^{-2\kappa}\sum_{s=1}^{T}\sum_{t=1}^{T}\tilde{g}_{T,s}^1\tilde{g}_{T,t}^2\int_{-\pi}^{\pi} f_T(\phi)e^{-i\phi(s-t)}d\phi \\
&= T^{-2\kappa}\int_{-\pi}^{\pi} f_T(\phi)\left(\sum_{s=1}^{T}\tilde{g}_{T,s}^1 e^{i\phi s}\right)\left(\sum_{t=1}^{T}\tilde{g}_{T,t}^2 e^{-i\phi t}\right)d\phi \\
&= T^{1-2\kappa}\int_{-\pi T}^{\pi T} f_T(\lambda/T)\left(T^{-1}\sum_{s=1}^{T}\tilde{g}_{T,s}^1 e^{i\lambda s/T}\right) \\
&\quad \left(T^{-1}\sum_{t=1}^{T}\tilde{g}_{T,t}^2 e^{-i\lambda t/T}\right)d\lambda \\
&\to \int_{-\infty}^{\infty} S(\lambda)\left(\int_0^1 g^1(s)e^{i\lambda s}ds\right)\left(\int_0^1 g^2(s)e^{-i\lambda s}ds\right)d\lambda.
\end{aligned}
\tag{6}
$$

Thus the limiting covariance matrix Σ depends on f_T only through S (albeit for $-\infty < \lambda < \infty$, a point we return to below).

Local-to-Zero Pseudo-Spectrum. This calculation requires the spectral density of x_t to exist. For some models, such as the $I(1)$ model, only the spectral density of Δx_t is well defined, so a generalization of (6) to this more general case is required. This is possible when the g-weights add to zero, that is when $\int_0^1 g(s)ds = 0$ (otherwise, if $\bar{x}$ doesn't have a finite second moment, then neither does η_T).

Let $f_{\Delta,T}(\phi)$ denote the spectral density of Δx_t, and assume it has a local-to-zero spectrum defined as

$$
T^{3-2\kappa}f_{\Delta,T}(\lambda/T) \to S_\Delta(\lambda). \tag{7}
$$

With $\tilde{g}_{T,t} = T\int_{(t-1)/T}^{t/T} g(s)ds$ and $\tilde{G}_{T,t} = T^{-1}\sum_{s=1}^{t-1}\tilde{g}_{T,t}$, summation by parts yields $T^{-1}\sum_{t=1}^{T}\tilde{g}_{T,t}x_{T,t} = -\sum_{t=1}^{T}\tilde{G}_{T,t}\Delta x_{T,t}$, since $\tilde{G}_{T,T+1} = 0$. Thus

$$
\begin{aligned}
T^{2(1-\kappa)}E[\eta_T^1\eta_T^2] &= T^{2(1-\kappa)}E\left[\left(\sum_{s=1}^{T}\tilde{G}_{T,s}^1\Delta x_s\right)\left(\sum_{t=1}^{T}\tilde{G}_{T,t}^2\Delta x_t\right)\right] \\
&= T^{3-2\kappa}\int_{-\pi T}^{\pi T} f_{\Delta,T}(\lambda/T)\left(T^{-1}\sum_{s=1}^{T}\tilde{G}_{T,s}^1 e^{i\lambda s/T}\right)\left(T^{-1}\sum_{t=1}^{T}\tilde{G}_{T,t}^2 e^{-i\lambda t/T}\right)d\lambda
\end{aligned}
$$

$$\to \int_{-\infty}^{\infty} S_\Delta(\lambda) \left(\int_0^1 G^1(s) e^{i\lambda s} ds \right) \left(\int_0^1 G^2(s) e^{-i\lambda s} ds \right) d\lambda = \gamma^{12}$$

where $G(s) = \int_0^s g(r)dr$. Furthermore, by integration by parts and $G(1) = 0$, $\int_0^1 G(s)e^{i\lambda s}ds = -\int_0^1 g(s)e^{i\lambda s}ds/(i\lambda)$, so that

$$\gamma^{12} = \int_{-\infty}^{\infty} \frac{S_\Delta(\lambda)}{\lambda^2} \left(\int_0^1 g^1(s) e^{i\lambda s} ds \right) \left(\int_0^1 g^2(s) e^{-i\lambda s} ds \right) d\lambda. \quad (8)$$

Thus, when $\int_0^1 g^1(s)ds = \int_0^1 g^2(s)ds = 0$, equation (8) generalizes (6): if the spectral density of x_t is well defined, then $f_T(\phi)|1 - e^{-i\phi}|^2 = f_{T,\Delta}(\phi)$, and (5) and (7) are equivalent, since $T^2|1 - e^{-i\lambda/T}|^2 \to \lambda^2$, so that $S_\Delta(\lambda) = S(\lambda)\lambda^2$. On the other hand, if the spectral density of x_t does not exist, then $f_T(\phi) = f_{T,\Delta}(\phi)/|1 - e^{-i\phi}|^2$ suitably defines a "pseudo spectrum" of x_t with local-to-zero limit $S(\lambda) = S_\Delta(\lambda)/\lambda^2$, so (8) still applies.

Σ as a Weighted Average of $S(\lambda)$. Several features of Σ follow from the representation (8). Since S is an even function and g^1 and g^2 are real valued, γ^{12} can be rewritten as $\gamma^{12} = 2\int_0^\infty S(\lambda)w^{12}(\lambda)d\lambda$, where $w^{12}(\lambda) = \mathrm{Re}[\left(\int_0^1 g^1(s)e^{i\lambda s}ds\right)\left(\int_0^1 g^2(s)e^{-i\lambda s}ds\right)]$. Elements of Σ are thus equal to a weighted average of the local-to-zero spectrum S. With $g^1(s) = \sqrt{2}\cos(\pi j s)$ and $g^2(s) = \sqrt{2}\cos(\pi k s)$, a calculation shows that $w^{12}(\lambda) = w_{j,k}(\lambda) = 0$ for $1 \le j, k \le q$ and $j + k$ odd, so that $E[X_j X_k] = 0$ for all odd $j + k$, independent of the local-to-zero spectrum S. With X_0 the suitably scaled limit of the sample mean $\bar{x}$ (so that $g(s) = 1$), this holds for all $j, k = 0, 1, 2 \ldots$.

Figure 12 plots $w_{j,k}(\cdot)$ for selected values of j and k. The figure displays the weights $w_{j,k}$ corresponding to the covariance matrix of the vector $(X_0, X_1, X_2, X_3, X_{10}, X_{11}, X_{12})'$, organized into a symmetric matrix of nine panels corresponding to X_0, $(X_1, X_2, X_3)'$ and $(X_{10}, X_{11}, X_{12})'$. The weights for the variances $w_{j,j}$ are shown in bold and the weights for the covariances $w_{j,k}$, $j \neq k$ are shown as thin curves. A calculation shows that covariance weights $w_{j,k}$, $j \neq k$, integrate to zero, which implies that for a flat local-to-zero spectrum S, Σ is diagonal. As can be seen from the panels on the diagonal, the variance of X_j is mostly determined by the values of S in the interval $\pi j \pm 2\pi$. Further, as long as S is somewhat smooth, the correlation between X_j and X_k is very close to zero for $|j - k|$ large. For X_j, $j > 0$, the weights $w_{j,j}$ put zero mass on $\lambda = 0$; it is this feature that makes it possible for Σ to be well defined even if $S(\lambda)$ is not integrable, as is the case if f_T is only a "pseudo-spectrum." In contrast, because $\int_0^1 g(s)ds \neq 0$ for X_0, there is positive weight placed on $\lambda = 0$, so the variance of X_0 will not exist for processes such as the $I(1)$ model.

Since the local-to-zero spectrum determines Σ, S is the key property of the underlying process x_t for the purposes of "low-frequency econometrics." It is instructive to revisit the three benchmark models LTUM, LLM and FRM introduced in Section 3.2 from this perspective. As already noted above, the

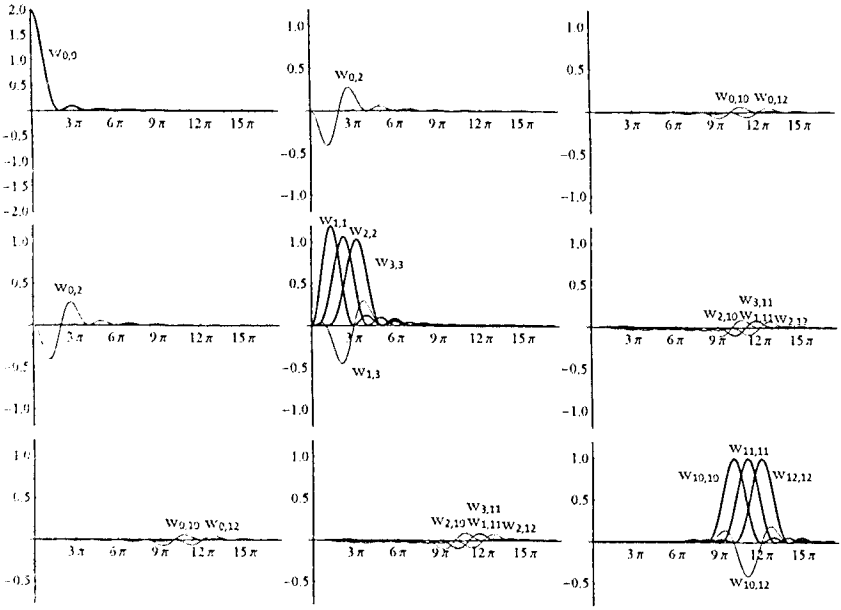

Figure 12 Weights on Local-to-Zero Spectrum for Covariance Matrix of $(X_0, X_1, X_2, X_3, X_{10}, X_{11}, X_{12})'$

FRM model has a (pseudo-) spectrum proportional to $|\phi|^{-2d}$ for ϕ close to zero, so that $S(\lambda) \propto |\lambda|^{-2d}$, $-1/2 < d < 3/2$. In particular S is constant for the $I(0)$ model and it is proportional to λ^{-2} for the $I(1)$ model. This implies that the local-to-zero spectrum of the LLM satisfies $S(\lambda) \propto b^2 + \lambda^{-2}$. Finally, in the LTUM, a straightforward calculation shows that $S(\lambda) \propto 1/(\lambda^2 + c^2)$. (Note that $(b, c) \to (\infty, \infty)$ and $(b, c) \to (0, 0)$ recover the $I(0)$ and $I(1)$ model, respectively). Figure 13 plots the logarithm of the local-to-zero spectra of the three benchmark models for selected values of b, c and d, and scale normalised so that $S(6\pi) = 1$.

Choice of q. The x-axes in Figures 11–13 are usefully thought of zooming in on the shaded area in the periodogram plot of Figure 10: in terms of the original time series, local-to-zero frequencies of, say, $|\lambda| < 12\pi$ correspond to cycles of periodicity $T/6$. Thus, with 66 years worth of data (of any sampling frequency), the shape of the spectrum for frequencies below 11 year cycles more or less determines the properties of the cosine transforms for $q = 12$. In the context of a specific model of low-frequency variability, such as the $I(0)$ or $I(1)$ model, or parametric families such as the FRM, LLM or LTUM, the choice of q then governs the frequencies for which the model's implications are exploited for inference. The choice of q is thus a trade-off between robustness and efficiency: a large q enables more powerful inference, as more cosine transforms are used to learn about model parameters of interest, but a large q also increases the danger of relying on a misspecified model for this

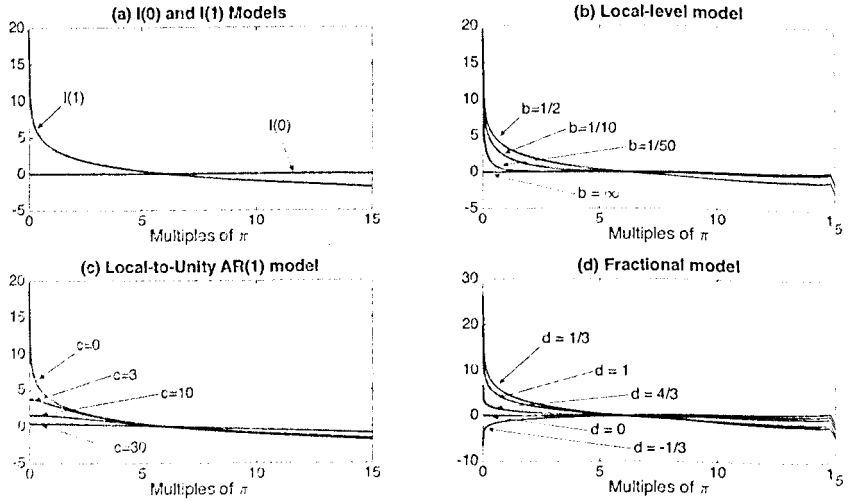

Figure 13 Local-to-Zero Log-Spectra of some Benchmark Models of Low-Frequency Variability

inference, as the low frequency model must accurately describe the spectrum over a wider frequency band. Roughly speaking, the low-frequency model is fit to the q observations X_T, which suggests that in general, the degree of misspecification is a function of q, without any obvious relationship to the underlying sample size or time span. The choice of q then amounts to a regularity assumption about the properties of the underlying data (relative to the model of low-frequency variability) that makes inference possible.[22] For instance, conducting inference about μ in the context of the $I(0)$ model with $q = 12$ and 66 years of data implicitly assumes that the spectrum is approximately flat over all frequencies corresponding to cycles of 11 years or more.

Multivariate Local-to-Zero Spectra. As a final generalization of the connection between the second moments of weighted averages and the spectrum of the underlying series, consider now the case where x_t is a $n \times 1$ vector. Its spectral density then is a $n \times n$ Hermitian matrix valued function $f_T(\phi)$. Assume that for some $n \times n$ matrix Υ_T, $T\Upsilon_T f_T(\lambda/T)\Upsilon_T' \to S(\lambda)$. Let $\eta_T^i = T\Upsilon_T \int_0^1 g^i(s)' x_{\lfloor sT \rfloor + 1} ds = \Upsilon_T \sum_{t=1}^T \tilde{g}_{T,t}^{i\prime} x_t$ for $i = 1, 2$, where $g^i : [0, 1] \mapsto \mathbb{R}^n$ and $\tilde{g}_{T,t}^i = T \int_{(t-1)/T}^{t/T} g^i(s) ds$. Then proceeding as for (6) yields

$$E[\eta_T^1 \eta_T^{2\prime}] \to \int_{-\infty}^{\infty} \left(\int_0^1 g^1(s) e^{i\lambda s} ds \right)' S(\lambda) \left(\int_0^1 g^2(s) e^{-i\lambda s} ds \right) d\lambda,$$

[22] Without additional strong cross-frequencies restrictions, it is not possible to systematically select the appropriate q from data: if one insists on correct inference whenever the crucial central limit theorem (2) holds for some conservatively small q_0, then no data dependent rule for selecting $q > q_0$ can improve on inference that simply uses $q = q_0$. See Müller (2011), especially page 414, for a formal discussion.

and if $T^3\Upsilon_T f_{\Delta,T}(\lambda/T)\Upsilon_T' \to S_\Delta(\lambda)$ and $\int_0^1 g^1(s)ds = \int_0^1 g^2(s)ds = 0$, then also

$$E[\eta_T^1\eta_T^2] \to \int_{-\infty}^{\infty}\left(\int_0^1 g^1(s)e^{i\lambda s}ds\right)'\frac{S_\Delta(\lambda)}{\lambda^2}\left(\int_0^1 g^2(s)e^{-i\lambda s}ds\right)d\lambda.$$

5.3 A Central Limit Theorem

The inference results presented in Sections 3 and 4 all depend crucially on the large sample *Gaussianity* of the suitably scaled cosine transform X_T, and not just on the value of the limiting covariance matrix Σ. In Section 2, we provided an argument for this asymptotic normality based on existing results about the large sample behavior of the partial sum process for some benchmark models. The last subsection identified the local-to-zero spectrum S as the key feature of x_t that determines the asymptotic covariance matrix. We now present a corresponding vector central limit theorem that takes the local-to-zero behavior of the spectral density function as its starting point and is applicable to general weighted averages $T^{1/2}\int_0^1 g(s)'x_{T,\lfloor sT\rfloor+1}ds$. Müller and Watson (2016) derive a corresponding scalar CLT for weighted averages with weights that add to zero, $\int_0^1 g(s)ds = 0$.

Theorem 1 *Let $x_{T,t} = \sum_{s=-\infty}^{\infty} c_{T,s}\varepsilon_{t-s}$, where $c_{T,s}$ are $n\times n$ and ε_t is $n\times 1$. Suppose that*

(i) $\{\varepsilon_t, \mathcal{F}_t\}$ is a martingale difference sequence with $E[\varepsilon_t\varepsilon_t'] = \Sigma_\varepsilon$, Σ_ε invertible, $\sup_t E[||\varepsilon_t||^{2+\delta}] < \infty$ for some $\delta > 0$, and

$$E||E[\varepsilon_t\varepsilon_t' - \Sigma_\varepsilon|\mathcal{F}_{t-m}]|| \le \xi_m \tag{9}$$

for some sequence $\xi_m \to 0$.

(ii) For every $\epsilon > 0$ the exists an integer $L_\epsilon > 0$ such that $\limsup_{T\to\infty} T^{-1}\sum_{l=L_\epsilon T+1}^{\infty}\left(T\sup_{|s|\ge l}||c_{T,s}||\right)^2 < \epsilon$.

(iii) $\sum_{s=-\infty}^{\infty}||c_{T,s}|| < \infty$ (but not necessarily uniformly in T). The spectral density of $x_{T,t}$ thus exists; denote it by $f_T : [-\pi,\pi] \mapsto \mathbb{H}$, where $\mathbb{H}$ is the space of $n\times n$ Hermitian matrices.

(iv.a) Assume that there exists a function $S : \mathbb{R} \mapsto \mathbb{H}$ such that $\int_0^1 ||S(\lambda)||d\lambda < \infty$, $\int_1^\infty ||S(\lambda)||\lambda^{-2}d\lambda < \infty$ and for all fixed K,

$$\int_0^K ||f_T(\frac{\lambda}{T}) - S(\lambda)||d\lambda \to 0. \tag{10}$$

(iv.b) For every diverging sequence $K_T \to \infty$

$$T^{-1}\int_{K_T/T}^{\pi}||f_T(\phi)||\phi^{-2}d\phi = \int_{K_T}^{\pi T}||f_T(\lambda/T)||\lambda^{-2}d\lambda \to 0. \tag{11}$$

(iv.c)

$$T^{-1/2}\int_{1/T}^{\pi} ||f_T(\phi)||^{1/2}\phi^{-1}d\phi = T^{-1/2}\int_{1}^{\pi T} ||f_T(\lambda/T)||^{1/2}\lambda^{-1}d\lambda \to 0. \tag{12}$$

(v) Each component of the function $g : [0, 1] \mapsto \mathbb{R}^n$ *is of bounded variation. Then*

$$T^{1/2}\int_0^1 g(s)'x_{T,\lfloor sT\rfloor+1}ds \Rightarrow \mathcal{N}\left(0, \int_{-\infty}^{\infty}\left(\int_0^1 e^{i\lambda s}g(s)ds\right)' \times S(\lambda)\left(\int_0^1 e^{-i\lambda s}g(s)ds\right)d\lambda\right). \tag{13}$$

The proof of Theorem 1 is in Appendix A.

The scaling of $x_{T,t}$ in the theorem is such that $f_T(\lambda/T)$ converges to the local-to-zero spectrum without any additional scaling by a power of T (so relative to the discussion in Section 5.2, $x_{T,t}$ in Theorem 1 is multiplied by $T^{1/2-\kappa}$, or $T^{1/2}\Upsilon_T$ in the vector case).

To better understand the role of assumptions (ii) and (iii), consider some leading examples for scalar series, $n = 1$. Suppose first that $x_{T,t}$ is causal and weakly dependent with exponentially decaying $c_{T,s}$, $|c_{T,s}| \le C_0 e^{-C_1 s}$ for some $C_0, C_1 > 0$, as would arise in causal and invertible ARMA models of any fixed and finite order. Then $T^{-1}\sum_{l=LT+1}^{\infty}\left(T\sup_{|s|\ge l}|c_{T,s}|\right)^2 \to 0$ for any $L > 0$, $S(\lambda)$ is constant and equal to the long-run variance of $x_{T,t}$ divided by 2π, and (11) and (12) hold, since f_T is bounded, $\int_{K_T}^{\infty}\lambda^{-2}d\lambda \to 0$ for any $K_T \to \infty$ and $T^{-1/2}\int_1^{\pi T}\lambda^{-1}d\lambda = T^{-1/2}\ln(\pi T) \to 0$.

Second, suppose $x_{T,t}$ is fractionally integrated with parameter $d \in (-1/2, 1/2)$. With $x_{T,t}$ scaled by T^{-d}, $c_{T,s} \approx C_0 T^{-d}s^{d-1}$, so that $T^{-1}\sum_{l=LT+1}^{\infty}\left(T\sup_{|s|\ge l}|c_{T,s}|\right)^2 \to C_0^2\int_L^{\infty}s^{2d-2}ds$, which can be made arbitrarily small by choosing L large. Further, for ϕ close to zero, $f_T(\phi) \approx C_0^2(\phi T)^{-2d}$, so that $S(\lambda) = (2\pi)^{-1}C_0^2\lambda^{-2d}$. Under suitable assumptions about higher frequency properties of $x_{T,t}$, (11) and (12) hold, since $T^{-1}\int_{K_T/T}^{\pi}(\phi T)^{-2d}\phi^{-2}d\phi = \int_{K_T}^{\pi T}\lambda^{-2d-2}d\lambda \to 0$ and $T^{-1/2}\int_{1/T}^{\pi}(\phi T)^{-d}\phi^{-1}d\phi = T^{-1/2}\int_1^{\pi T}\lambda^{-d-1}d\lambda = T^{-1/2}d^{-1}(1 - (\pi T)^{-d}) \to 0$. For instance, even integrable poles in f_T at frequencies other than zero can be accommodated.

Third, suppose $x_{T,t}$ is an AR(1) process with local-to-unity coefficient $\rho_T = 1 - c/T$ and innovation variance equal to T^{-1}. Then $c_{T,s} = T^{-1}\rho_T^s$, $s \ge 0$. Thus $T^{-1}\sum_{l=LT+1}^{\infty}\left(T\sup_{|s|\ge l}|c_{T,s}|\right)^2 \to \int_L^{\infty}e^{-2cs}ds$, which can be made arbitrarily small by choosing L large. Further, $f_T(\phi) = (2\pi)^{-1}T^{-2}/|1 - \rho_T e^{-i\phi}|^2$, which is seen to satisfy (10) with $S(\lambda) = (2\pi)^{-1}(\lambda^2 + c^2)^{-1}$. Conditions (11) and (12) also hold in this example, since $f_T(\phi) \le (2\pi)^{-1}$.

Corresponding central limit theorems for a vector of weighted averages of one or multiple time series follow readily from Theorem 1 by invoking the Cramer–Wold device.

For completeness, we also state the corresponding result for weights that add up to zero, which only requires the existence of the spectral density of the first differences.[23]

Corollary 1 *Let $\Delta y_{T,t} = x_{T,t}$, where $x_{T,t}$ is as in Theorem 1. Assume that $g : [0, 1] \mapsto \mathbb{R}^n$ satisfies $\int_0^1 g(s)ds = 0$. Suppose that in addition to assumptions (i)–(iii) and (v) of Theorem 1*

(iv.a)′ There exists a function $S : \mathbb{R} \mapsto \mathbb{H}$ such that $\int_0^1 ||S(\lambda)||d\lambda < \infty$, $\int_1^\infty ||S(\lambda)||\lambda^{-4}d\lambda < \infty$ and for all fixed K,

$$\int_0^K ||f_T(\frac{\lambda}{T}) - S(\lambda)||d\lambda \to 0. \tag{14}$$

(iv.b)′ For every diverging sequence $K_T \to \infty$

$$T^{-3}\int_{K_T/T}^{\pi} ||f_T(\phi)||\phi^{-4}d\phi = \int_{K_T}^{\pi T} ||f_T(\lambda/T)||\lambda^{-4}d\lambda \to 0.$$

(iv.c)′

$$T^{-3/2}\int_{1/T}^{\pi} ||f_T(\phi)||^{1/2}\phi^{-2}d\phi = T^{-1/2}\int_1^{\pi T} ||f_T(\lambda/T)||^{1/2}\lambda^{-2}d\lambda \to 0.$$

Then

$$T^{-1/2}\int_0^1 g(s)'y_{T,\lfloor sT\rfloor+1}ds \Rightarrow \mathcal{N}\left(0, \int_{-\infty}^{\infty}\left(\int_0^1 e^{i\lambda s}g(s)ds\right)' \times \frac{S(\lambda)}{\lambda^2}\left(\int_0^1 e^{-i\lambda s}g(s)ds\right)d\lambda\right).$$

With $f_{T,y}(\phi) = f_T(\phi)/|1 - e^{-i\phi}|^2$ the pseudo-spectrum of $y_{T,t}$ and $S_y(\lambda) = S(\lambda)/\lambda^2$, the convergence in (14) is equivalent to

$$\int_1^K ||f_{T,y}(\frac{\lambda}{T}) - S_y(\lambda)||d\lambda \to 0$$

for all fixed K, so that the corollary proves a CLT with limiting variance $\int_{-\infty}^{\infty}\left(\int_0^1 e^{i\lambda s}g(s)ds\right)' S_y(\lambda)\left(\int_0^1 e^{-i\lambda s}g(s)ds\right)d\lambda$, a function of the weights

[23] The corollary corrects two inaccuracies in the statement of Theorem 1 of Müller and Watson (2016): $\sum_{s=-\infty}^{\infty}||c_{T,s}|| < \infty$ (and not $\sum_{s=-\infty}^{\infty}||c_{T,s}||^2 < \infty$) is necessary for the spectral density of $\Delta y_{T,t}$ to exist, and the local-to-zero spectral density S of $\Delta y_{T,t}$ (corresponding to $\omega^2 S$ in Theorem 1 of Müller and Watson, 2016) does not need to be integrable on $\mathbb{R}$, but the assumption (iv.a) of Theorem 1 above suffices.

g and the local-to-zero pseudo-spectrum S_y. The relatively weaker assumptions in (iv)′ accommodate potential overdifferencing when y_t is, say, $I(d)$ with $d \in (-1/2, 1/2)$, so that for a scalar series, the local-to-zero spectrum of $x_t = \Delta y_t$, $S(\lambda) \propto \lambda^{2(1-d)}$, is increasing in λ.

6 CONCLUDING REMARKS

Inference about low-frequency phenomena is challenging because of the scarcity of corresponding sampling information. We suggest extracting this information by computing q trigonometrically weighted averages of the time series of interest. We then apply an asymptotic framework that explicitly accounts for the scarcity by keeping the number q fixed, so that in the limit, the inference problem becomes a small sample problem involving q Gaussian variables. In many instances, this small sample problem is readily solved by classic results about inference in small Gaussian samples. In other cases, one can apply analytical or numerical approaches to derive powerful inference from first principles.

The results presented here did not allow for a deterministic time trend in the x_t process. One approach to dealing with deterministic trends is to use inference methods that are unaffected by their presence. Alternatively, for example in the context of measuring the covariability of multiple time series, one might model the presence of common time trends. In either case, it is straightforward to isolate the deterministic trend component by using a suitable set of q weight functions, where one is equal to the (demeaned) trend (and whose value is potentially ignored in the subsequent analysis), and the other $q - 1$ trigonometric weights are orthogonal to both a constant and a time trend. One such set of weights is derived in Müller and Watson (2008).

The results presented here were also based on the large sample Gaussianity of the q weighted averages due to a central limit theorem. The key mechanisms of the approach, however, can be used to deduce a non-Gaussian limit distribution after suitably adjusting the solution of the now non-Gaussian limit problem. For example, consider a stochastic volatility model for the scalar series x_t, where the volatility process is, say, local-to-unity. Conditional on the volatility path, the cosine transforms have a Gaussian limit, with a covariance matrix that depends on the realization of the volatility path. The unconditional distribution thus becomes a mixture of normals, and the corresponding small sample problem becomes inference about the parameters of this normal mixture.

More generally, we believe our approach to low-frequency econometrics to be useful for issues beyond those discussed here. For instance, in Müller and Watson (2016), we use this framework to derive predictive sets for very long-run predictions that are valid for a wide range of persistent processes. We are currently working on inference about the degree of covariability of potentially non-$I(0)$ series. And it would also be of great interest to connect the

low-frequency econometrics approach more directly with structural economic models, such as asset pricing models that stress long-run uncertainty.

A PROOF OF THEOREM 1 AND COROLLARY 1

The proof of Theorem 1 is based on the following Lemmas. Without loss of generality, let $||\cdot||$ denote the spectral norm.

Lemma 1 *(i) Under assumption (v) of Theorem 1, there exists* $C < \infty$ *such that (i.a)* $\sup_{0\le\phi\le\pi,T}(||\sum_{t=1}^{T} e^{i\phi t}\tilde{g}_{T,t}|| - C\phi^{-1}) \le 0$ *and (i.b)* $\sup_{\lambda}(||\int_0^1 e^{i\lambda s}g(s)ds|| - C\lambda^{-1}) \le 0$.

(ii) Under assumptions of Corollary 1, with $G(s) = \int_0^s g(r)dr$, *there exists* $C < \infty$ *such that (i.a)* $\sup_{0<\phi\le\pi,T}(||\sum_{t=1}^{T} e^{i\phi t}G(\frac{t-1}{T})|| - CT^{-1}\phi^{-2}) \le 0$ *and (i.b)* $\sup_{\lambda>0}(||\int_0^1 e^{i\lambda s}G(s)ds|| - C\lambda^{-2}) \le 0$.

Proof (i.a) Let $r_t = \sum_{s=1}^{t} e^{i\phi s} = e^{i\phi}(e^{i\phi t} - 1)/(1 - e^{i\phi})$. By summation by parts,

$$\sum_{t=1}^{T} e^{i\phi t}\tilde{g}_{T,t} = r_T\tilde{g}_{T,T} - \sum_{t=1}^{T} r_{t-1}(\tilde{g}_{T,t} - \tilde{g}_{T,t-1})$$

$$= \frac{e^{i\phi}}{1 - e^{i\phi}}\left((e^{i\phi T} - 1)\tilde{g}_{T,T} - \sum_{t=1}^{T}(e^{i\phi(t-1)} - 1)(\tilde{g}_{T,t} - \tilde{g}_{T,t-1})\right)$$

with $r_0 = \tilde{g}_{T,0} = 0$. For $0 \le \phi \le \pi$, $|e^{i\phi}(1-e^{i\phi})^{-1}| = (2\sin(\phi/2))^{-1} \le 1/\phi$. Furthermore, $\limsup_{T\to\infty}\sum_{t=1}^{T}||\tilde{g}_{T,t} - \tilde{g}_{T,t-1}|| < \infty$, since g is of bounded variation, and $\sup_{\phi,t}|e^{i\phi t} - 1| \le 2$.

(i.b) Let $R(s) = \int_0^s e^{i\lambda u}du = i(1 - e^{i\lambda s})/\lambda$. By Riemann–Stieltjes integration by parts,

$$\int_0^1 e^{i\lambda s}g(s)ds = R(1)g(1) - \int_0^1 R(s)dg(s)$$

$$= \lambda^{-1}i[(1 - e^{i\lambda})g(1) - \int_0^1 (1 - e^{i\lambda s})dg(s)]$$

and the Riemann–Stieltjes integral with respect to dg exists, since g is of bounded variation. The result now follows from $\sup_{\lambda,0\le s\le 1}|1 - e^{i\lambda s}| \le 2$.

(ii.a) By summation by parts and $G(1) = G(0) = 0$,

$$\sum_{t=1}^{T} e^{i\phi t}G(\frac{t-1}{T}) = -T^{-1}\sum_{t=1}^{T} r_t\tilde{g}_{T,t}$$

$$= \frac{e^{i\phi}}{1 - e^{i\phi}}T^{-1}\sum_{t=1}^{T} e^{i\phi t}\tilde{g}_{T,t}$$

and the result follows from part (i.a) and $|e^{i\phi}(1-e^{i\phi})^{-1}| = (2\sin(\phi/2))^{-1} \leq 1/\phi$.

(ii.b) By integration by parts, $i\lambda \int_0^1 e^{i\lambda s} G(s)ds = -\int_0^1 e^{i\lambda s} g(s)ds$, and the result follows from part (i.b). □

Lemma 2 $\sigma_T^2 = \mathrm{Var}[T^{-1/2}\sum_{t=1}^T \tilde{g}'_{T,t} x_{T,t}] \to \sigma^2 = \int_{-\infty}^{\infty} \left(\int_0^1 e^{i\lambda s} g(s)ds\right)' S(\lambda) \left(\int_0^1 e^{-i\lambda s} g(s)ds\right) d\lambda$.

Proof We first show that σ^2 exists. For $K > 0$, define

$$\sigma_\infty^2(K) = \int_{-K}^{K} \left(\int_0^1 e^{i\lambda s} g(s)ds\right)' S(\lambda) \left(\int_0^1 e^{-i\lambda s} g(s)ds\right) d\lambda$$

so that we can write $\sigma^2 = \sigma_\infty^2(1) + (\sigma^2 - \sigma_\infty^2(1))$. Then $\sigma_\infty^2(1)$ exists by assumption (iv.a), since $\|\int_0^1 e^{i\lambda s} g(s)ds\| \leq \sup_{0\leq s\leq 1} \|g(s)\| < \infty$ (bounded variation implies boundedness), and

$$\begin{aligned} \tfrac{1}{2}|\sigma^2 - \sigma_\infty^2(1)| &\leq \int_1^\infty \left\| \int_0^1 e^{i\lambda s} g(s)ds \right\|^2 \|S(\lambda)\| d\lambda \\ &\leq C^2 \int_1^\infty \lambda^{-2} \|S(\lambda)\| d\lambda < \infty \end{aligned} \tag{15}$$

by assumption (iv.a), where the second inequality follows from Lemma 1 (i.b).

Let $\Gamma_T(j) = E[x_{T,t} x'_{T,t-j}] = \int_{-\pi}^{\pi} e^{i\phi j} f_T(\phi) d\phi$. We find

$$\begin{aligned} \sigma_T^2 &= T^{-1} \sum_{j,l=1}^T \tilde{g}'_{T,l} \Gamma_T(l-j) \tilde{g}_{T,j} \\ &= T^{-1} \sum_{j,l=1}^T \tilde{g}'_{T,l} \left(\int_{-\pi}^{\pi} e^{i\phi(l-j)} f_T(\phi) d\phi\right) \tilde{g}_{T,j} \\ &= T^{-1} \int_{-\pi}^{\pi} \left(\sum_{t=1}^T e^{i\phi t} \tilde{g}_{T,t}\right)' f_T(\phi) \left(\sum_{t=1}^T e^{-i\phi t} \tilde{g}_{T,t}\right) d\phi \\ &= \int_{-\pi T}^{\pi T} \left(T^{-1} \sum_{t=1}^T e^{i\lambda t/T} \tilde{g}_{T,t}\right)' f_T\left(\frac{\lambda}{T}\right) \left(T^{-1} \sum_{t=1}^T e^{-i\lambda t/T} \tilde{g}_{T,t}\right) d\lambda. \end{aligned}$$

Now for any fixed K,

$$\sigma_T^2(K) = \int_{-K}^{K} \left(T^{-1} \sum_{t=1}^T e^{i\lambda t/T} \tilde{g}_{T,t}\right)' f_T\left(\frac{\lambda}{T}\right) \left(T^{-1} \sum_{t=1}^T e^{-i\lambda t/T} \tilde{g}_{T,t}\right) d\lambda \tag{16}$$

$$\to \sigma_\infty^2(K) = \int_{-K}^{K} \left(\int_0^1 e^{i\lambda s} g(s) ds \right)' S(\lambda) \left(\int_0^1 e^{-i\lambda s} g(s) ds \right) d\lambda$$

where the convergence follows from assumption (iv.a), and

$$\left\| T^{-1} \sum_{t=1}^{T} e^{i\lambda t/T} \tilde{g}_{T,t} - \int_0^1 e^{i\lambda s} g(s) ds \right\| \leq \int_0^1 ||e^{i\lambda s} g(s) - e^{i\lambda \lfloor sT \rfloor / T} g(s)|| ds$$

$$\leq \sup_{0 \leq s \leq 1} ||g(s)|| \int_0^1 |1 - e^{i\lambda(\lfloor sT \rfloor / T - s)}| ds$$

along with $|1 - e^{ia}| < |a|$ for any real a, so that $\sup_{|\lambda| \leq K} |1 - e^{i\lambda(\lfloor sT \rfloor / T - s)}| \leq K/T \to 0$.

Thus, for any fixed K,

$$\delta_K(T) = \sigma_T^2(K) - \sigma_\infty^2(K) \to 0.$$

Now for each T, define K_T as the largest integer $K \leq T$ for which $\sup_{T' \geq T} |\delta_K(T')| \leq 1/K$ (and zero if no such K exists). Note that $\delta_K(T) \to 0$ for all fixed K implies that $K_T \to \infty$, and by construction, also $\delta_{K_T}(T) \to 0$. Thus

$$\sigma_T^2 - \sigma^2 = \delta_{K_T}(T) + (\sigma_T^2 - \sigma_T^2(K_T)) - (\sigma^2 - \sigma_\infty^2(K_T)).$$

But

$$\tfrac{1}{2} |\sigma_T^2 - \sigma_T^2(K_T)| \leq T^{-1} \int_{K_T/T}^{\pi} f_T(\phi) \cdot \left\| \sum_{t=1}^{T} e^{i\phi t} \tilde{g}_{T,t} \right\|^2 d\phi$$

$$\leq C^2 T^{-1} \int_{K_T/T}^{\pi} ||f_T(\phi)|| \phi^{-2} d\lambda \to 0 \qquad (17)$$

by assumption (iv.b) and Lemma 1 (i.a), and similar to (15),

$$\tfrac{1}{2} |\sigma^2 - \sigma_\infty^2(K_T)| \leq C^2 \int_{K_T}^{\infty} \lambda^{-2} ||S(\lambda)|| d\lambda \qquad (18)$$

so that from $\int_1^\infty \lambda^{-2} ||S(\lambda)|| d\lambda < \infty$ in assumption (iv.a), $\int_{K_T}^\infty \lambda^{-2} ||S(\lambda)|| d\lambda \to 0$ by dominated convergence. □

Lemma 3 $T^{-1/2} \sup_{t,T} || \sum_{j=1}^{T} \tilde{g}'_{T,j} c_{T,j-t} || \to 0.$

Proof Recall that for any $\mathbb{R}^n$ valued sequence $\{a_j\}_{j=-\infty}^{\infty}$ and $n \times n$ valued sequence $\{b_j\}_{j=-\infty}^{\infty}$ with $\sum_{j=-\infty}^{\infty} ||a_j||^2 < \infty$ and $\sum_{j=-\infty}^{\infty} ||b_j||^2 < \infty$, $\sum_{j=-\infty}^{\infty} a_j' b_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{A}(\phi)^* \hat{B}(\phi) d\phi$, where $\hat{A}(\phi) = \sum_{j=-\infty}^{\infty} a_j e^{-i\phi j}$ and $\hat{B}(\phi) = \sum_{j=-\infty}^{\infty} b_j e^{-i\phi j}$, and "*" denotes the conjugate transpose. Thus

$$\sum_{j=1}^{T} \tilde{g}'_{T,j} c_{T,j-t} = \sum_{j=1-t}^{T-t} \tilde{g}'_{T,t+j} c_{T,j} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\phi t} \hat{G}_T(\phi)^* \hat{C}_T(\phi) d\phi$$

where $\hat{C}_T(\phi) = \sum_{j=-\infty}^{\infty} c_{T,j} e^{-i\phi j}$ and $\hat{G}_T(\phi) = \sum_{j=1}^{T} e^{-i\phi j} \tilde{g}_{T,j}$, so that $e^{i\phi t}\hat{G}_T(\phi) = \sum_{j=1}^{T} e^{-i\phi(j-t)} \tilde{g}_{T,t} = \sum_{j=1-t}^{T-t} e^{-i\phi j} \tilde{g}_{T,t+j}$. Now from $f_T(\phi) = \frac{1}{2\pi}\hat{C}_T(\phi)\Sigma_\varepsilon \hat{C}_T(\phi)^*$ (cf. Theorem 11.8.3 in Brockwell and Davis, 1991), it follows that $||\hat{C}_T(\phi)|| \leq \sqrt{2\pi}||f_T(\phi)|| \cdot ||\Sigma_\varepsilon^{-1}||$. We thus find

$$2\pi||\sum_{j=1}^{T} \tilde{g}'_{T,j} c_{T,j-t}|| = ||\int_{-\pi}^{\pi} e^{i\phi t}\hat{G}_T(\phi)^*\hat{C}_T(\phi)d\phi||$$
$$\leq \sqrt{2\pi}||\Sigma_\varepsilon^{-1}|| \int_{-\pi}^{\pi} ||\hat{G}_T(\phi)|| \cdot ||f_T(\phi)||^{1/2} d\phi$$

so that it suffices to show that

$$T^{-1/2} \int_0^{\pi} ||\hat{G}_T(\phi)|| \cdot ||f_T(\phi)||^{1/2} d\phi \to 0.$$

Using $||\hat{G}_T(\phi)|| \leq \sum_{j=1}^{T} ||\tilde{g}_{T,j}||$, we have

$$\int_0^{1/T} ||\hat{G}_T(\phi)|| \cdot ||f_T(\phi)||^{1/2} d\phi \leq T^{-1} \sum_{j=1}^{T} ||\tilde{g}_{T,j}|| \cdot \int_0^1 ||f_T(\lambda/T)||^{1/2} d\lambda$$
$$\leq T^{-1} \sum_{j=1}^{T} ||\tilde{g}_{T,j}|| \cdot \left(\int_0^1 ||f_T(\lambda/T)|| d\lambda\right)^{1/2}$$
$$\to \int_0^1 ||g(s)|| ds \cdot \left(\int_0^1 ||S(\lambda)|| d\lambda\right)^{1/2} < \infty$$

where the convergence follows from assumption (iv.a) and straightforward arguments. Furthermore, by Lemma 1 (i.a),

$$T^{-1/2} \int_{1/T}^{\pi} ||\hat{G}_T(\phi)|| \cdot ||f_T(\phi)||^{1/2} d\phi \leq C T^{-1/2} \int_{1/T}^{\pi} ||f_T(\phi)||^{1/2} \phi^{-1} d\phi \to 0 \tag{19}$$

where the convergence follows from assumption (iv.c). □

Lemma 4 *For every $\epsilon > 0$ there exists a $M > 0$ such that*

$$\limsup_{T\to\infty} \mathrm{Var}[T^{-1/2} \sum_{t=1}^{T} \tilde{g}'_{T,t} x_{T,t} - T^{-1/2} \sum_{t=-MT}^{MT} \left(\sum_{j=1}^{T} \tilde{g}'_{T,j} c_{T,j-t}\right) \varepsilon_t] < \epsilon.$$

For this M, $\sigma^2_{M,T} = \mathrm{Var}[T^{-1/2} \sum_{t=-MT}^{MT} \left(\sum_{j=1}^{T} \tilde{g}'_{T,j} c_{T,j-t}\right) \varepsilon_t] = T^{-1} \sum_{t=-MT}^{MT} \left(\sum_{j=1}^{T} \tilde{g}'_{T,j} c_{T,j-t} \Sigma_\varepsilon c'_{T,j-t} \tilde{g}_{T,j}\right)$ satisfies $\limsup_{T\to\infty} |\sigma^2_{M,T} - \sigma^2| < \epsilon$.

Proof We have

$$T^{-1/2}\sum_{t=1}^{T}\tilde{g}'_{T,t}x_{T,t} = \sum_{t=-\infty}^{\infty}\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right)\varepsilon_t$$

so that, with $\bar{g} = \sup_{0\le s\le 1}||g(s)|| < \infty$ (bounded variation implies boundedness),

$$\begin{aligned}
&\mathrm{Var}[T^{-1/2}\sum_{t=1}^{T}\tilde{g}'_{T,t}x_{T,t} - T^{-1/2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right)\varepsilon_t] \\
&= T^{-1}\sum_{t=-\infty}^{-MT-1}\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j} + T^{-1}\right. \\
&\quad \left.\sum_{t=MT+1}^{\infty}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j}\right) \\
&\le \bar{g}^2||\Sigma_\varepsilon||T^{-1}\sum_{t=MT+1}^{\infty}\left(\sum_{j=1}^{T}(||c_{T,j-t}|| + ||c_{T,j+t}||)\right)^2 \\
&\le 4\bar{g}^2||\Sigma_\varepsilon||T^{-1}\sum_{t=MT+1}^{\infty}\left(T\sup_{|s|\ge t-T}||c_{T,s}||\right)^2 \\
&= 4\bar{g}^2||\Sigma_\varepsilon||T^{-1}\sum_{t=(M-1)T+1}^{\infty}\left(T\sup_{|s|\ge t}||c_{T,s}||\right)^2
\end{aligned}$$

which can be made smaller than ϵ uniformly in T by choosing M large enough via assumption (ii).

The second claim follows directly follows from this result and Lemma 2. □

Lemma 5 *For any large enough integer* $M > 0$, $\sigma_{M,T}^{-1}T^{-1/2}\sum_{t=-MT}^{MT}\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right)\varepsilon_t \Rightarrow \mathcal{N}(0,1)$.

Proof By the second claim in Lemma 4 and Lemma 2, $\sigma_{M,T} = O(1)$ and $\sigma_{M,T}^{-1} = O(1)$. By Theorem 24.3 in Davidson (1994), it thus suffices to show (a) $T^{-1/2}\sup_{1\le t\le T}|\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\varepsilon_t|^p\to 0$ and (b) $T^{-1}\sum_{t=-MT}^{MT}\left[\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\varepsilon_t\right)^2 - \sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j}\right]\xrightarrow{p}0$.

(a) is implied by the Lyapunov condition via Davidson's (1994) Theorems 23.16 and 23.11. Thus, it suffices to show that

$$\sum_{t=-MT}^{MT} E\left[\left\|T^{-1/2}\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right)\varepsilon_t\right\|^{2+\delta}\right] \to 0.$$

Now

$$\sum_{t=-MT}^{MT} E\left[\left\|T^{-1/2}\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right)\varepsilon_t\right\|^{2+\delta}\right]$$

$$\leq T^{-1-\delta/2}\sum_{t=-MT}^{MT}\left\|\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right\|^{2+\delta} E[\|\varepsilon_t\|^{2+\delta}]$$

$$\leq (\sup_t E[\|\varepsilon_t\|^{2+\delta}])\cdot T^{-\delta/2}\sup_t\left\|\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right\|^{\delta}$$

$$\cdot T^{-1}\sum_{t=-MT}^{MT}\left\|\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right\|^{2}$$

$$\leq \|\Sigma_\varepsilon^{-1}\|\cdot(\sup_t E[\|\varepsilon_t\|^{2+\delta}])\cdot\left(T^{-1/2}\sup_t\left\|\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right\|\right)^{\delta}\cdot\sigma^2_{M,T}\to 0$$

where the last line uses $\left\|\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right\|^2 \leq \|\Sigma_\varepsilon^{-1}\|\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j}$ and the convergence follows from Lemma 3.

For (b), we apply Theorem 19.11 of Davidson (1994) with Davidson's X and c chosen as $X^D_{T,t} = T^{-1}\left(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\varepsilon_t\right)^2 - c^D_{T,t}$ and $c^D_{T,t} = T^{-1}\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j}$. Then $X^D_{T,t}/c^D_{T,t}$ is uniformly integrable, since

$$\frac{X^D_{T,t}}{c^D_{T,t}} + 1 = \frac{(\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\varepsilon_t)^2}{\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j}}$$

$$\leq \|\Sigma_\varepsilon^{-1}\|\frac{\sum_{j=1}^{T}\|\tilde{g}'_{T,j}c_{T,j-t}\|^2\|\varepsilon_t\|^2}{\sum_{j=1}^{T}\|\tilde{g}'_{T,j}c_{T,j-t}\|^2} = \|\Sigma_\varepsilon^{-1}\|\cdot\|\varepsilon_t\|^2$$

almost surely, so that $\sup_t E[(X^D_{T,t}/c^D_{T,t}+1)^{1+\delta/2}] \leq \|\Sigma_\varepsilon^{-1}\|^{1+\delta/2}\sup_t E[\|\varepsilon_t\|^{2+\delta}] < \infty$ by assumption (i). Further,

$$\sum_{t=-MT}^{MT} c^D_{T,t} = \sigma^2_{M,T} = O(1)$$

by Lemma 4, and

$$\sum_{t=-MT}^{MT}(c_{T,t}^{D})^2 = T^{-2}\sum_{t=-MT}^{MT}\left[\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j}\right]^2$$

$$\leq ||\Sigma_\varepsilon|| \sup_{-MT\leq t\leq MT}\left\|T^{-1/2}\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right\|^2$$

$$\cdot T^{-1}\sum_{t=-MT}^{MT}\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\Sigma_\varepsilon c'_{T,j-t}\tilde{g}_{T,j}$$

$$= ||\Sigma_\varepsilon|| \sup_{-MT\leq t\leq MT}\left\|T^{-1/2}\sum_{j=1}^{T}\tilde{g}'_{T,j}c_{T,j-t}\right\|^2 \cdot \sigma^2_{M,T} \to 0$$

where the convergence follows from Lemma 3. □

Proof of Theorem 1: By Lemmas 2, 4 and 5, for large enough M,

$$\sigma^{-1}T^{-1/2}\sum_{t=1}^{T}\tilde{g}'_{T,t}x_{T,t} = \frac{\sigma_{M,T}}{\sigma}A_T + \frac{\sigma_{M,T}}{\sigma}B_T$$

where $A_T \Rightarrow \mathcal{N}(0,1)$, $\limsup_{T\to\infty} E(B_T^2) < \epsilon$ and $\limsup_{T\to\infty}|\sigma^2_{M,T} - \sigma^2| < \epsilon$. Thus, by Slutzky's Theorem, as $\epsilon \to 0$, the desired convergence in distribution follows. But ϵ was arbitrary, which proves the Theorem. □

Proof of Corollary 1: Define $G(s) = \int_0^s g(r)dr$. Then by summation by parts and $\int_0^1 g(s)ds = 0$,

$$T^{-1/2}\int_0^1 g(s)y_{T,\lfloor sT\rfloor+1}ds = T^{-3/2}\sum_{t=1}^{T}\tilde{g}_{T,t}y_{T,t}$$

$$= -T^{-1/2}\sum_{t=1}^{T}G(\frac{t-1}{T})\Delta y_{T,t}$$

$$= -T^{-1/2}\sum_{t=1}^{T}G(\frac{t-1}{T})x_{T,t}$$

since $T^{-1}\sum_{s=1}^{t-1}\tilde{g}_{T,s} = \int_0^{(t-1)/T} g(s)ds = G(\frac{t-1}{T})$. The proof now proceeds analogously to the proof of Theorem 1, with $-G(\frac{t-1}{T})$ replacing $\tilde{g}_{T,t}$ throughout. The weights $G(\frac{t-1}{T})$ are an order of magnitude smoother than $\tilde{g}_{T,t}$, as formalized in the stronger inequalities of part (ii) of Lemma 1 compared to part (i). The analogue of Lemma 2,

$$\sigma_T^2 = \mathrm{Var}[T^{-1/2}\sum_{t=1}^{T} G(\frac{t-1}{T})' x_{T,t}] \to \sigma^2$$
$$= \int_{-\infty}^{\infty} \left(\int_0^1 e^{i\lambda s} G(s)ds\right)' S(\lambda) \left(\int_0^1 e^{-i\lambda s} G(s)ds\right) d\lambda,$$

is proved just like Lemma 2, except that Lemma 1 part (ii) is employed to sharpen the analogue of the inequalities in (15), (17) and (18), so that the weaker assumptions (iv.a)′ and (iv.b)′ are sufficient. Similarly, in the analogue of Lemma 3, apply Lemma 1 part (ii.a) to the analogue of inequality (19), (where now $\hat{G}_T(\phi) = -\sum_{j=1}^{T} e^{-i\phi j} G(\frac{t-1}{T})$) so that assumption (iv.c)′ is sufficient. With these modifications, proceeding as in the proof of Theorem 1 yields

$$T^{-1/2}\sum_{t=1}^{T} G(\frac{t-1}{T}) x_{T,t} \Rightarrow \mathcal{N}\left(0, \int_{-\infty}^{\infty} \left(\int_0^1 e^{i\lambda s} G(s)ds\right)' \times S(\lambda) \left(\int_0^1 e^{-i\lambda s} G(s)ds\right) d\lambda\right)$$

and the result follows from $\int_0^1 e^{i\lambda s} g(s)ds = -i\lambda \int_0^1 e^{i\lambda s} G(s)ds$ using integration by parts. □

References

Akdi, Y. and D. A. Dickey (1998): "Periodograms of Unit Root Time Series: Distributions and Tests," *Communications in Statistics: Theory and Methods*, 27, 69–87.

Anderson, T. W. (1984): *An Introduction to Multivariate Statistics*, New York: Wiley.

Andrews, D. W. K. and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

Baillie, R. T. (1996): "Long Memory Processes and Frational Integration in Econometrics," *Journal of Econometrics*, 73, 5–59.

Baxter, M. and R. G. King (1999): "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *Review of Economics and Statistics*, 81(4), 575–593.

Bierens, H. J. (1997): "Nonparametric Cointegration Analysis," *Journal of Econometrics*, 77, 379–404.

Brockwell, P. J. and R. A. Davis (1991): *Time Series: Theory and Methods*, 2nd Edn., New York: Springer.

Cavanagh, C. L. (1985): "Roots Local To Unity," *Working Paper*, Harvard University.

Chan, N. H. and C. Z. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

Davidson, J. (1994): *Stochastic Limit Theory*. New York: Oxford University Press.

Dufour, J.-M. and M. L. King (1991): "Optimal Invariant Tests for the Autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors," *Journal of Econometrics*, 47, 115–143.

Elliott, G. (1998): "The Robustness of Cointegration Methods When Regressors Almost Have Unit Roots," *Econometrica*, 66, 149–158.

(1999): "Efficient Tests for a Unit Root When the Initial Observation is Drawn From its Unconditional Distribution," *International Economic Review*, 40, 767–783.

Elliott, G., U. K. Müller and M. W. Watson (2015): "Nearly Optimal Tests When a Nuisance Parameter is Present Under the Null Hypothesis," *Econometrica*, 83, 771–811.

Elliott, G., T. J. Rothenberg and J. H. Stock (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Engle, R. F. (1974): "Band Spectrum Regression," *International Economic Review*, 15, 1–11.

Fernald, J. (2014): "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity," FRBSF *Working Paper 2012-9*.

Gonçalves, S. and T. Vogelsang (2011): "Block Bootstrap and HAC Robust Tests: The Sophistication of the Naive Bootstrap," *Econometric Theory*, 27(4), 745–791.

Harvey, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hodrick, R. J. and E. C. Prescott (1997): "Postwar US Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking*, 29, 1–16.

Hotelling, H. (1931): "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, 2, 360–378.

Jansson, M. (2004): "The Error in Rejection Probability of Simple Autocorrelation Robust Tests," *Econometrica*, 72, 937–946.

Kiefer, N. and T. J. Vogelsang (2002): "Heteroskedasticity-Autocorrelation Robust Testing Using Bandwidth Equal to Sample Size," *Econometric Theory*, 18, 1350–1366.

(2005): "A New Asymptotic Theory for Heteroskedasticity-Autocorrelation Robust Tests," *Econometric Theory*, 21, 1130–1164.

Kiefer, N. M., T. J. Vogelsang, and H. Bunzel (2000): "Simple Robust Testing of Regression Hypotheses," *Econometrica*, 68, 695–714.

King, M. L. (1980): "Robust Tests for Spherical Symmetry and their Application to Least Squares Regression," *The Annals of Statistics*, 8, 1265–1271.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992): "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root," *Journal of Econometrics*, 54, 159–178.

Lehmann, E. L. and J. P. Romano (2005): *Testing Statistical Hypotheses*. New York: Springer.

Müller, U. K. (2004): "A Theory of Robust Long-Run Variance Estimation," *Working paper*, Princeton University.

(2007): "A Theory of Robust Long-Run Variance Estimation," *Journal of Econometrics*, 141, 1331–1352.

(2011): "Efficient Tests under a Weak Convergence Assumption," *Econometrica*, 79, 395–435.

(2014): "HAC Corrections for Strongly Autocorrelated Time Series," *Journal of Business and Economic Statistics*, 32, 311–322.

Müller, U. K. and M. W. Watson (2008): "Testing Models of Low-Frequency Variability," *Econometrica*, 76, 979–1016.

(2013): "Low-Frequency Robust Cointegration Testing," *Journal of Econometrics*, 174, 66–81.

(2016): "Measuring Uncertainty about Long-Run Predictions," *Review of Economic Studies*, 83.

Newey, W. K. and K. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Nyblom, J. (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.

Phillips, P. C. B. (1987): "Towards a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, 535–547.

(1998): "New Tools for Understanding Spurious Regression," *Econometrica*, 66, 1299–1325.

(2005a): "Challenges of Trending Time Series Econometrics," *Mathematics and Computers in Simulation*, 68, 401–416.

(2005b): "HAC Estimation by Automated Regression," *Econometric Theory*, 21, 116–142.

(2014): "Optimal Estimation of Cointegrated Systems with Irrelevant Instruments," *Journal of Econometrics*, 178, 210–224.

Phillips, P. C. B., Y. Sun, and S. Jin (2006): "Spectral Density Estimation and Robust Hypothesis Testing using Steep Origin Kernels without Truncation," *International Economic Review*, 47, 837–894.

(2007): "Long Run Variance Estimation and Robust Regression Testing Using Sharp Origin Kernels with No Truncation," *Journal of Statistical Planning and Inference*, 137, 985–1023.

Pratt, J. W. (1961): "Length of Confidence Intervals," *Journal of the American Statistical Association*, 56, 549–567.

Priestley, M. (1981): *Spectral Analysis and Time Series*. Academic Press.

Robinson, P. M. (2003): "Long-Memory Time Series," in *Time Series with Long Memory*, ed. by P. M. Robinson, pp. 4–32. Oxford: Oxford University Press.

Stock, J. H. (1994): "Unit Roots, Structural Breaks and Trends," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4, pp. 2740–2841. North Holland, New York.

Sun, Y., P. C. B. Phillips, and S. Jin (2008): "Optimal Bandwidth Selection in Heteroskedasticity-Autocorrelation Robust Testing," *Econometrica*, 76, 175–794.

Wright, J. H. (2000): "Confidence Sets for Cointegrating Coefficients Based on Stationarity Tests," *Journal of Business and Economic Statistics*, 18, 211–222.