RESEARCH REPORT SERIES
*(Statistics #2006-12)*

# A Spectral Approach for Locally Assessing Time Series Model Misspecification

Tucker McElroy
Scott Holan[1]

Statistical Research Division
U.S. Census Bureau
Washington, DC 20233

University of Missouri-Columbia[1]

Report Issued: November 15, 2006

# A Spectral Approach for Locally Assessing Time Series Model Misspecification

Tucker McElroy*and Scott Holan†
U.S. Census Bureau and University of Missouri-Columbia

**Abstract**

Peaks in the spectrum of a stationary process are indicative of periodic phenomena, such as seasonality or business cycles. Hence one important aspect of developing parametric models for periodic processes is proper characterization of spectral peaks. By using diagnostics constructed from the average ratio of two spectral densities this work proposes to test whether a hypothesized model is locally supported in the frequency domain by the data. The local fit of a model is assessed by considering a subset of the whole frequency band, focused on the locality of the spectral peak. This technique can therefore be used to test the goodness-of-fit of a model with respect to local frequency domain properties of the data. For example, one can test for the appropriateness of a hypothesized seasonal or cyclical spectral peak in the model for the data. In the development of these diagnostics we provide a result of independent interest, the asymptotic distribution of general polynomial functionals of the periodogram. We present theoretical properties for several new diagnostics, and also explore their finite-sample properties through simulation and application to several economic time series.

**Keywords.** Cycle estimation, Goodness-of-fit, Model Misspecification, Peak detection, Seasonal Adjustment, Spectral Density.



## 1 Introduction

Although there is an abundance of literature on time domain methods for detecting model misspecification for a stationary time series (see Li (2004) for a comprehensive discussion), frequency

*Statistical Research Division, U.S. Census Bureau, 4700 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

†Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO, 65211-6100, holans@missouri.edu

domain approaches have received less attention. Perhaps this is due in part to the wide-spread association of modelling in the frequency domain with spectral density estimation via kernel-smoothing methods. Nevertheless, model misspecification can be assessed in the frequency domain very easily, namely by comparing a postulated model spectral density (perhaps the maximum likelihood estimate from a particular model class) with some non-model-based spectral estimate, over a suitable range of frequencies. Although general frequency domain goodness-of-fit tests have been proposed (see Paparoditis (2000) and Chen and Deo (2004), among others), few have been formulated with the intent of being limited to a suitable frequency band. One exception, although not fully developed, is given by Beran (1994). In this paper we propose a wide class of measures, formulated in the frequency domain, that can be implemented in a band-limited fashion. These measures can be applied to a number of problems, such as spectral peak detection, significance of unobserved components, and local model misspecification. We first discuss some of these applications, in order to provide motivation for our general method.

McElroy and Holan (2005) discusses the problem of spectral peak detection, with applications to cycle estimation in econometrics, signal detection in engineering, and seasonal adjustment in federal statistics. If a postulated model fails to adequately capture a prominent spectral peak really present in the data, then certain periodic phenomena will be completely absent from our model, resulting in a loss of explanatory power. So peak detection is important, and one way of measuring the strength of spectral peaks is via aggregating the spectral density's convexity in a neighborhood of the node. This is the approach of McElroy and Holan (2005), though that work adopts a nonparametric framework.

In the basic Unobserved Component (UC) model, each component of economic phenomenon, e.g., trend, cycle, seasonal, is modelled as a separate time series, and the sum of all components yields the observed process. Given the usual issues of parsimony in statistical modelling, one is interested in knowing whether the addition of another component is compelling with respect to the data. A time domain method of answering this question, which enjoys some popularity, is to determine if the variance of the innovation sequence of an ARIMA representation of a given component differs significantly from zero. In contrast, a frequency domain perspective examines the spectral density of the postulated component model at a range of frequencies, and determines whether the data prefers a model that includes that particular component.

More generally, we may be interested in whether a model fits the data at hand in a particular spectral range. The Gaussian maximum-likelihood algorithm involves finding a model spectral density for the data such that it is close to the periodogram in an average sense, in that the discrepancy is aggregated over all frequencies. Thus, maximum likelihood estimates can be expected

to provide a decent model in a global sense, over all frequencies. However, it is possible that the model is less desirable over certain frequency bands, even though it is optimal globally. This is problematic for certain applications, where it may be more important to correctly model certain frequencies, whereas the behavior at other frequencies is deemed to be less important. For example, consider cycle modelling in econometrics; here the low frequency band, where the cycle periodicities are located, is crucial while the high frequency behavior is less vital.

The spectral measures presented in this paper provide a flexible set of tools that can address a wide class of problems, including those already described. In Section 2 we develop a general theory for such measures, including a discussion of finite sample behavior, asymptotics, and statistical power for relevant hypotheses. Section 3 gives relevant details for several applications, including spectral peak identification, component model significance, and local goodness-of-fit. Moreover, we discuss some possible kernels that can be used for spectral averaging. Section 4 provides the results of a simulation study, which show the efficacy of these methods in practice. Finally, in order to illustrate the various applications several time series are considered. Section 5 concludes, and all proofs are left to the Appendix.

## 2 Theory for Spectral Measures

Suppose that, after suitable transformations and differencing if necessary (and removal of regression effects), we have a mean zero stationary time series $X_1, X_2, \cdots, X_n$, which will sometimes be denoted by the vector $\mathbf{X} = (X_1, X_2, \cdots, X_n)'$. The spectral density $f(\lambda)$ is well-defined so long as the autocovariance function $\gamma_f(h)$ is absolutely summable, and can be defined by

$$f(\lambda) = \sum_{h=-\infty}^{\infty} \gamma_f(h) e^{-ih\lambda} \tag{1}$$

with $i = \sqrt{-1}$ and $\lambda \in [-\pi, \pi]$. It follows that the inverse Fourier transform yields

$$\gamma_f(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{ih\lambda} \, d\lambda,$$

a relation that we will use repeatedly in the sequel. That is, $\gamma_f$ and $f$ are Fourier transform pairs; this relationship holds for any integrable function $f$, not just for spectral densities. Furthermore, denoting the Toeplitz matrix associated with $\gamma_f$ by $\Sigma(f)$, it follows that

$$\Sigma_{jk}(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\lambda) e^{i(j-k)\lambda} d\lambda.$$

Finally, let $\hat{f}(\lambda)$ denote the periodogram, defined on a continuum of frequencies. Then $\hat{f}(\lambda)$ can be expressed as

$$\hat{f}(\lambda) = \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-it\lambda} \right|^2 = \sum_{h=1-n}^{n-1} R(h) e^{-ih\lambda}, \tag{2}$$

with $R(h)$ equal to the sample (uncentered) autocovariance function. We consider two basic types of spectral measures: the first depends on the data alone, whereas the second depends on both data and a prespecified model. The measures are defined by

$$\theta_A(g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)g(\lambda)\,d\lambda \tag{3}$$

$$\phi_A(g,f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)g(\lambda)/f(\lambda)\,d\lambda. \tag{4}$$

Of course, $\phi_A$ is a special case of $\theta_A$, but we will distinguish them due to separate applications and interpretations. Here, $g$ and $f$ are integrable functions that are possibly random, e.g., $g = \hat{f}$. The function $A$ denotes a kernel, and is typically chosen to approximate a step function over a suitable interval of frequencies, thereby affecting aggregation of the spectral densities. Based on the above definitions, it follows that

$$\theta_A(\hat{f}) = \frac{1}{n}\mathbf{X}'\Sigma(A)\mathbf{X}$$
$$\phi_A(\hat{f},f) = \frac{1}{n}\mathbf{X}'\Sigma(A/f)\mathbf{X}.$$

Now the fact that $\theta_A(\hat{f})$ can be represented as a quadratic form is convenient, since the Toeplitz matrix $\Sigma(A)$ can be computed separately, and applied to many different time series. When the kernel $A$ corresponds to a rational function, it is fairly simple to compute autocovariances of $A/f$, given that $f$ corresponds to an $ARMA$ model; for this case, the method of Tunnicliffe-Wilson (1979) can be used. If $A$ cannot be written as a rational function in $e^{-i\lambda}$, one can still proceed if $f$ corresponds to an $AR$ process, as Proposition 1 demonstrates. The result requires a few additional concepts. Let $g$ be a polynomial of degree $d$, and define the $(m-d)\times m$ matrix (where $m > d$ is arbitrary) $\Delta(g)$ as having $jk^{th}$ entry given by $g_{j-k+d}$, the $(j-k+d)^{th}$ coefficient of $g$ (if $j-k+d > d$ or $j-k+d < 0$, simply let $g_{j-k+d} = 0$).

**Proposition 1** *Consider polynomials $\theta_1, \theta_2, \cdots, \theta_J$ with degrees $q_j$, and let $g$ be an integrable function on $[-\pi,\pi]$, and let $h$ be defined by*

$$h(\lambda) = \Pi_{j=1}^{J}|\theta_j(e^{-i\lambda})|^2.$$

*Then $\Sigma(g\cdot h) = B\,\Sigma(g)\,B'$, where*

$$B = \Pi_{j=1}^{J}\Delta(\theta_j).$$

*The matrices in the product can be written in any order, so long as their dimensions match up appropriately. If $\Sigma(g)$ has dimension $m$, then $B$ is $(m - \sum_{j=1}^{J} q_j)\times m$ dimensional, so $m$ must be chosen sufficiently large.*

As an application, we can rewrite $\phi_A(\hat{f}, f)$ when $f$ corresponds to an $AR(p)$ process. In particular, suppose that $f(\lambda) = |h(e^{-i\lambda})|^{-2}$ for a polynomial $h$. Then

$$\phi_A(\hat{f}, f) = \frac{1}{n}\mathbf{X}'\Delta(h)\Sigma(A)\Delta'(h)\mathbf{X},$$

where the dimensions are given as follows: $\Delta(h)$ is $n \times (n+p)$ dimensional and $\Sigma(A)$ is square of dimension $n+p$, where $p$ is the order of $h$. This follows from Proposition 1, since $\Sigma(A/f) = \Sigma(A \cdot |h(e^{-i\cdot})|^2)$. Hence, for each new $AR$ model $f$ that we wish to try, it is only necessary to recompute $\Delta(h)$. If $f$ is an $ARMA$ process, one needs to recompute $\Sigma(A/f)$ for each $f$. Since $A(\lambda)$ will often be zero on whole subintervals of $[-\pi, \pi]$, it does not have an $ARMA$ representation, and hence $\Sigma(A/f)$ may be impossible to compute via the Tunnicliffe-Wilson (1979) algorithm. One solution is to use the approximation $\Sigma(A)\Sigma(1/f) \doteq \Sigma(A/f)$, so that $\Sigma(A)$ need only be computed once (two examples of kernels $A$ and their autocovariance functions can be found in McElroy and Holan (2005)). Alternatively, the entries of $\Sigma(A/f)$ can be determined to any desired precision using numerical integration to compute $\phi_A(e^{\cdot ih}, f)$ for any integer $h$.

The first type of measure, $\theta_A$, is well-suited for measuring whether the data is consistent with a specified model. Letting $\tilde{f}$ denote the true (but unknown) spectral density of the data, we might stipulate the hypothesis

$$H_0 : \tilde{f} = f.$$

With $\hat{f}$ (the periodogram) as a plug-in estimate for $\tilde{f}$, the hypothesis can be tested using $\theta_A(\hat{f}) - \theta_A(f)$. In fact, $\theta_A(g) - \theta_A(f)$ can be viewed as a type of signed measure, giving the discrepancy between two spectral densities $g$ and $f$, weighted according to the kernel $A$. The second measure $\phi_A(g, f)$ assesses the average ratio of two spectral densities $g$ and $f$; recall that the ratio of the two spectral densities can be viewed as a residual spectral density (Parzen, 1983). If this ratio is constant, then $g$ effectively "whitens" $f$. So $\phi_A(\hat{f}, f)$ can be used as a test statistic for the hypothesis $H_0$. Note that ultimately the agreement between $\tilde{f}$ and $f$ is important only at those frequencies contained in the support of $A$. So in proposing a model $f$ for the data, it is primarily of importance that the model be appropriate at frequencies in the support of $A$; if the modelling at frequencies outside the support of $A$ is thought to be poor, this will have no effect asymptotically on the testing procedures considered below (though in finite sample, it is important to know $\tilde{f}$ under the Null hypothesis at frequencies outside the support of $A$). Theorems 1 and 2 summarize the important statistical properties of such measures.

The asymptotic theory we develop requires some mild conditions on the data. We follow the material in Taniguchi and Kakizawa (2000, Section 3.1.1). The sufficient conditions can be expressed using two different formulations. The first formulation, which is denoted as condition (B) and is due to Brillinger (1981), states that the process is strictly stationary and that condition (B1) of

Taniguchi and Kakizawa (2000, page 55) holds. Alternatively, condition (HT), due to Hosoya and Taniguchi (1982), states that the process has a linear representation, and requires that conditions (H1) through (H6) of Taniguchi and Kakizawa (2000, pages 55 – 56) hold. Neither of these conditions are stringent; for example, a causal $MA(\infty)$ process having finite fourth moments satisfies (HT). Finally, let $\mathcal{C}^1$ denote the space of non-negative definite functions $g$ defined on $[-\pi, \pi]$ such that

$$\sum_h |h||\gamma_g(h)| < \infty.$$

We first discuss the measure $\theta_A$ in Theorem 1:

**Theorem 1** *Let $\hat{f}$ be the periodogram given by (2), and $\tilde{f}$ be the true spectral density of the data. The mean of $\theta_A(\hat{f})$ is given by*

$$\mathbb{E}\theta_A(\hat{f}) = \frac{1}{n} tr(\Sigma(A)\Sigma(\tilde{f})),$$

*where tr denotes the trace. If the third and fourth cumulants of $\{X_t\}$ are zero, then*

$$nVar\theta_A(\hat{f}) = \frac{2}{n} tr(\Sigma(A)\Sigma(\tilde{f})\Sigma(A)\Sigma(\tilde{f})).$$

*If in addition $A, \tilde{f} \in \mathcal{C}^1$, then as $n \to \infty$*

$$\mathbb{E}\theta_A(\hat{f}) \to \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda)\tilde{f}(\lambda)\, d\lambda$$
$$nVar\theta_A(\hat{f}) \to \frac{2}{2\pi} \int_{-\pi}^{\pi} A^2(\lambda)\tilde{f}^2(\lambda)\, d\lambda.$$

*Finally, if $A, \tilde{f} \in \mathcal{C}^1$, the process $\{X_t\}$ satisfies either condition (B) or (HT), and the kernel $A$ is real, continuous, and even, then*

$$\sqrt{n}\left(\theta_A(\hat{f}) - \theta_A(\tilde{f})\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V)$$

*with variance given by*

$$V = \frac{2}{2\pi} \int_{-\pi}^{\pi} A^2(\lambda)\tilde{f}^2(\lambda)\, d\lambda + \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} A(\lambda)A(\omega)G^X(-\lambda, \omega, -\omega)d\lambda d\omega.$$

*Here $G^X$ is the tri-spectral density*

$$G^X(\lambda, \omega, \theta) = \sum_{j,k,l=-\infty}^{\infty} \exp\{-i(\lambda j + \omega k + \theta l)\}c^X(j, k, l).$$

*The function $c^X$ denotes the fourth-order cumulant function.*

Next, we consider the properties of $\phi_A$:

**Theorem 2** *Let $\hat{f}$ be the periodogram given by (2), and $\tilde{f}$ be the true spectral density of the data. The mean of $\phi_A(\hat{f}, f)$ is given by*

$$\mathbb{E}\phi_A(\hat{f}, f) = \frac{1}{n} tr(\Sigma(A/f)\Sigma(\tilde{f})).$$

*If the third and fourth cumulants of $\{X_t\}$ are zero, then*

$$nVar\phi_A(\hat{f}, f) = \frac{2}{n} tr\{\Sigma(A/f)\Sigma(\tilde{f})\Sigma(A/f)\Sigma(\tilde{f})\}.$$

*If in addition $A, \tilde{f} \in \mathcal{C}^1$, then as $n \to \infty$*

$$\mathbb{E}\phi_A(\hat{f}, f) \to \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \frac{\tilde{f}(\lambda)}{f(\lambda)} \, d\lambda$$

$$nVar\phi_A(\hat{f}, f) \to \frac{2}{2\pi} \int_{-\pi}^{\pi} A^2(\lambda) \left( \frac{\tilde{f}(\lambda)}{f(\lambda)} \right)^2 d\lambda.$$

*Finally, if the process $\{X_t\}$ satisfies either condition (B) or (HT), and the kernel $A$ is real, continuous, and even, then*

$$\sqrt{n} \left( \phi_A(\hat{f}, f) - \phi_A(\tilde{f}, f) \right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V)$$

*with variance given by*

$$V = \frac{2}{2\pi} \int_{-\pi}^{\pi} A^2(\lambda) \left( \frac{\tilde{f}(\lambda)}{f(\lambda)} \right)^2 d\lambda + \frac{1}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{A(\lambda)A(\omega)}{f(\lambda)f(\omega)} G^X(-\lambda, \omega, -\omega) d\lambda d\omega.$$

**Remark 1** If the data are Gaussian, then the third and fourth cumulants vanish, and the limiting variances $V$ are straightforward to calculate, given a knowledge of $\tilde{f}$. In the context of hypothesis testing, where some knowledge $\tilde{f}$ is supposed, one can actually compute the limiting distribution in Theorems 1 and 2, as well as the finite sample means and variances. The other conditions of the theorems, such as $\tilde{f} \in \mathcal{C}^1$, are not very restrictive and are satisfied by many processes (for example, all $ARMA$ processes).

Next, we present a much more general central limit theorem for polynomial functionals of the periodogram. These results do not supersede Theorems 1 and 2, because we restrict ourselves to the Brillinger (1975) condition (B). However, under this condition, Theorem 3 below represents a novel generalization of Theorem 3 of Chiu (1988), and thus may be of interest in other applications.

We consider a fairly general "polynomial" functional of the periodogram of the form $\sum_{j=1}^{J} \phi_j(\lambda) \hat{f}^j(\lambda)$, where the deterministic functions $\phi_j$ are real and continuous. Theorem 3 below states a central limit theorem for integrals of such functionals. In Corollary 3 below, we only require the case $J = 2$. However, since this theorem may be of independent interest, we state the results more generally.

**Theorem 3** *Consider a finite collection of real, continuous functions $\phi_j(\lambda)$, for $j = 1, 2, \cdots, J$. If the process $\{X_t\}$ satisfies (B), then*

$$\sqrt{n}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{j=1}^{J}\phi_j(\lambda)\hat{f}^j(\lambda)\,d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{j=1}^{J}\phi_j(\lambda)\,j!\,\tilde{f}^j(\lambda)\,d\lambda\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V)$$

*where the limiting variance $V$ is given by*

$$\begin{aligned} V = &\sum_{j,k=1}^{J}\frac{(j+k)! - j!k!}{2\pi}\int_{-\pi}^{\pi}\left(\phi_j(\lambda)\phi_k(-\lambda) + \phi_j(\lambda)\phi_k(\lambda)\right)\tilde{f}^{j+k}(\lambda)\,d\lambda \\ &+ \frac{jkj!k!}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\phi_j(\lambda)\phi_k(\omega)G^X(\lambda, -\lambda, \omega)\tilde{f}^{j-1}(\lambda)\tilde{f}^{k-1}(\omega)\,d\lambda d\omega. \end{aligned}$$

**Remark 2** Note that when the third and fourth cumulants are zero and the $\phi_j$'s are even, the variance reduces to

$$V = \frac{2}{2\pi}\int_{-\pi}^{\pi}\sum_{j,k=1}^{J}\phi_j(\lambda)\phi_k(\lambda)\left((j+k)! - j!k!\right)\tilde{f}^{j+k}(\lambda)\,d\lambda,$$

which is the limit of the non-negative quantity $\frac{2}{2\pi}\int_{-\pi}^{\pi} Var\left(\sum_{j=1}^{J}\phi_j(\lambda)\hat{f}(\lambda)\right)\,d\lambda$. Compare Remark 1 of Chiu (1988).

# 3 Applications

There are many possible applications of these measures, and we move from specific examples to the more general situation in the subsequent development. We first consider the situation where it is suspected that certain periodic phenomena are present in the data, and it is desired to detect the significance of such phenomena. For example, a seasonal pattern may be present in the time series, which manifests as a peak in the process's spectral density at the so-called "seasonal frequencies." These would be $\pi/6$, $2\pi/6$, $3\pi/6$, $4\pi/6$, $5\pi/6$, and $6\pi/6$ for monthly data. Another example comes from econometrics, where much interest focuses on detection of a business cycle in macro-economic series. A business cycle represents the slowly moving (stationary) oscillations about a smooth trend, and is commonly thought to have a period between 2 and 10 years for most series (Harvey and Trimbur, 2003). Again, the presence of a cycle would be manifested as a peak in the corresponding frequency range of the spectral density.

Now in formulating a model $f$ that attempts to model a peak in the spectrum, one may ask whether the data $\mathbf{X}$ significantly warrants such structure in the model. A crude measure would be given by comparing the model spectral density $f$ and the periodogram $\hat{f}$ at the frequencies of interest. A more thoughtful approach involves first constructing a measure of "peakedness" or

"modality" for any given model $f$; one can then compare the modality measure for both $f$ and $\hat{f}$, and determine if there are significant discrepancies. As suggested by the approach of McElroy and Holan (2005), we consider aggregating the model $f$ over a range of frequencies about the desired frequency of interest, summing against a kernel function that can be interpreted as measuring the second derivative, or convexity, of $f$. In fact, using the second derivative of a smooth aggregation kernel will yield such an interpretation, and so we consider a measure of the form

$$\theta_{\ddot{A}}(g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \ddot{g}(\lambda)\, d\lambda,$$

where the equality follows from integration by parts (along with the boundary conditions that $A$ and its derivative are zero at $\pm\pi$). By letting $g = f$, where $f$ is a hypothesized model, we can compute a modality measure; see McElroy and Holan (2005) for a complete discussion. For computational purposes, one would typically place the derivatives on the known kernel $A$, and the measure would be calculated via the representation

$$\sum_{h \in \mathbb{Z}} \gamma_{\ddot{A}}(-h) \gamma_f(h),$$

which follows from use of the discrete Fourier Transform. In general, such a quantity will require an approximation of the infinite series, or a numerical method for evaluation of the integral. However, since $\theta_{\ddot{A}}(f)$ can be interpreted as the limiting expectation of $\theta_{\ddot{A}}(\hat{f})$ under the assumption that $f = \tilde{f}$, i.e., that the hypothesized model is truth, it follows that another suitable modality measure is given by the exact finite sample expectation, namely

$$\theta_{n,\ddot{A}}(f) = \frac{1}{n} tr(\Sigma(\ddot{A})\Sigma(f)).$$

Note that, even for large samples, this will not take much computer time to calculate. In a similar fashion, we can compute the finite sample variance measure

$$\sigma^2_{n,\ddot{A}}(f) = \frac{2}{n} tr\left(\Sigma(\ddot{A})\Sigma(f)\Sigma(\ddot{A})\Sigma(f)\right).$$

Now Theorem 1 implies that, under the appropriate conditions, $\theta_{n,\ddot{A}}(f) \to \theta_{\ddot{A}}(f)$ and $\sigma_{n,\ddot{A}}(f) \to \sigma_{\ddot{A}}(f)$ as $n \to \infty$, where

$$\sigma^2_{\ddot{A}}(g) = \frac{2}{2\pi} \int_{-\pi}^{\pi} A^2(\lambda) g^2(\lambda)\, d\lambda.$$

Asymptotically the finite sample variance forms a self-normalization for the modality measure, such that $\theta_{\ddot{A}}(f)/\sigma_{\ddot{A}}(f)$ has magnitude bounded by $1/\sqrt{2}$ (this follows from the Cauchy-Schwarz inequality). By comparing $\theta_{\ddot{A}}(\hat{f})$ and $\theta_{n,\ddot{A}}(f)$, we can test the Null Hypothesis that $f = \tilde{f}$ over a relevant spectral region. To see this, we use the following corollary of Theorem 1:

**Corollary 1** *Suppose that the third and fourth cumulants of* $\{X_t\}$ *are zero,* $A$ *is twice continuously differentiable with* $\ddot{A}, \tilde{f} \in \mathcal{C}^1$*, and the process* $\{X_t\}$ *satisfies either condition (B) or (HT). If the kernel* $\ddot{A}$ *is real and even, then*

$$\sqrt{n}\frac{\left(\theta_{\ddot{A}}(\hat{f}) - \theta_{n,\ddot{A}}(\tilde{f})\right)}{\sigma_{n,\ddot{A}}(\tilde{f})} \tag{5}$$

*has mean zero and variance one for all* $n$*, and converges weakly to a standard normal as* $n \to \infty$*.*

**Remark 3** Thus under $H_0$, namely that $f = \tilde{f}$, we compute (5) and use the standard normal quantiles for the test's critical values. In particular, an $\alpha$-level lower one-sided test, with $f_0$ denoting the Null model, has asymptotic critical value

$$x_\alpha = \frac{\sigma_{n,\ddot{A}}(f_0)}{\sqrt{n}} z_\alpha + \theta_{n,\ddot{A}}(f_0)$$

with $z_\alpha = \Phi^{-1}(\alpha)$ and $\Phi$ the standard normal cumulative distribution function. Similar results can be formulated for the 2-sided test. In a similar fashion if we assume $f_0, f_1 \in \mathcal{C}^1$, the asymptotic power of the test can be determined: for each alternative model $f_1$ that differs from the null model $f_0$, the asymptotic power is

$$\Phi\left(\frac{\sigma_{n,\ddot{A}}(f_0)}{\sigma_{n,\ddot{A}}(f_1)} z_\alpha + \frac{\sqrt{n}}{\sigma_{n,\ddot{A}}(f_1)}\left[\theta_{n,\ddot{A}}(f_0) - \theta_{n,\ddot{A}}(f_1)\right]\right).$$

**Example 1:** Suppose that we wish to perform local peak detection; then we would posit a Null model $f_0$ that has a peak at a desired frequency $\lambda_0$. A very simple model is given by the following $AR(2)$:

$$(1 - 2\rho\cos\omega B + \rho^2 B^2) X_t = \epsilon_t \tag{6}$$

where $\epsilon_t$ is a white noise sequence with variance $\tau^2$. The frequency $\omega$ parameterizes the location of the peak, which is at $\cos^{-1}(\cos\omega(1+\rho^2)/2\rho)$. The parameter $\rho$ governs the overall shape of the curve, with $\rho = 1$ corresponding to an infinite peak. The corresponding spectral density is

$$f_0(\lambda) = \frac{\tau^2}{(1 - 2\rho\cos(\lambda+\omega) + \rho^2)(1 - 2\rho\cos(\lambda-\omega) + \rho^2)}. \tag{7}$$

For simplicity of exposition, we will focus on this case, with $\omega$ selected such that the peak occurs at a desired frequency $\lambda_0$.

In order to implement the test, we apply Proposition 1 to compute $\Sigma(\ddot{A}/f_0)$. As discussed above, this is easily computed when $f_0$ corresponds to an $AR$ model. See McElroy and Holan (2005) for more discussion of kernels.

As a second application, we consider testing a model by considering its ratio to the true model. Our Null hypothesis is still $\tilde{f}(\lambda) = f(\lambda)$. Note that $\tilde{f}/f$ can be interpreted as a residual spectral density; thus our Null hypothesis asserts that this residual spectral density is a constant, the spectrum of unit variance white noise (Parzen, 1983). In general, if we have two models $g$ and $f$ in mind, we might measure their local spectral discrepancy via $\phi_A(g, f)$ (note this is not a true "metric" because it is not symmetric). For example, if $A$ is compactly supported, $\phi_A(g, f)$ aggregates the ratio of the models over the kernel's support. Consider the following approximation to $\phi_A(g, f)$:

$$\phi_{n,A}(g, f) = \frac{1}{n} tr(\Sigma(A)\Sigma(g/f))$$

Similarly, an approximation to the variance in Theorem 2 is given by

$$\sigma^2_{n,A}(g, f) = \frac{2}{n} tr\{\Sigma(A)\Sigma(g/f)\Sigma(A)\Sigma(g/f)\}.$$

Now if $g$ is proportional to $f$, $\phi_{n,A}$ and $\sigma^2_{n,A}$ simplify greatly. In particular, if $g(\lambda) = f(\lambda)$ for all $\lambda$, then

$$\phi_{n,A}(g, f) = \gamma_A(0)$$
$$\sigma^2_{n,A}(g, f) = \frac{2}{n} tr\{\Sigma^2(A)\}.$$

Further, suppose we wish to detect whether the data warrants a spectral model $f$ within a desired spectral range. By choosing the kernel $A$ appropriate to the spectral range of interest, we can appeal to the measure $\phi_A(\hat{f}, f)$. Moreover, adapting the proof of Theorem 2 it can be shown that the mean and variance of $\phi_A(\hat{f}, f)$ are approximately $\phi_{n,A}(\tilde{f}, f)$ and $\sigma^2_{n,A}(\tilde{f}, f)$ respectively. Thus we have the following corollary:

**Corollary 2** *Suppose that the third and fourth cumulants of $\{X_t\}$ are zero, $A, \tilde{f} \in \mathcal{C}^1$, and the process $\{X_t\}$ satisfies either condition (B) or (HT). If the kernel $A$ is real, continuous, and even, then*

$$\sqrt{n}\frac{\Big(\phi_A(\hat{f}, f) - \phi_{n,A}(\tilde{f}, f)\Big)}{\sigma_{n,A}(\tilde{f}, f)}$$

*converges weakly to a standard normal as $n \to \infty$.*

**Remark 4** Now under $H_0$, the convergence result becomes

$$\sqrt{n}\frac{\Big(\phi_A(\hat{f}, f) - \gamma_A(0)\Big)}{\sqrt{n^{-1}2tr\{\Sigma^2(A)\}}} \xRightarrow{\mathcal{L}} \mathcal{N}(0, 1). \tag{8}$$

This quantity can now be used as a test statistic, with standard normal critical values. Note that it does not have mean zero and variance one in finite sample, because we are using approximations to the mean and variance. If desired, the exact quantities from Theorem 2 can be used, but the approximations $\phi_{n,A}(\tilde{f}, f)$ and $\sigma_{n,A}(\tilde{f}, f)$ may be easier to compute.

**Example 2:** Suppose we are interested in peak modelling, which is similar to peak detection. We might postulate a Null model $f_0$ of the $AR(2)$ form given in Example 1. Under $H_0$, the true model is equal to $f_0$; if the peak is not actually present in the spectrum of the data, the division by large values of $f_0$ at the cycle frequencies will not be compensated by the numerator of $\tilde{f}/f_0$, resulting in a low value of $\phi_A$ and rejection of $H_0$. Note that this would result in a negative value of the standardized $\phi_A$, whereas a significantly positive $\phi_A$ indicates that $\tilde{f}/f_0$ is too large at the cycle frequencies. The kernel $A$ should be selected in an approximate band-pass fashion, such that only cycle frequencies are permitted to affect the diagnostic.

Now all the quantities in the test statistic in (8) can be easily calculated; if $\gamma_A(0) = 0$ as well, the statistic will be properly centered automatically. In general though, we conceive of the kernel $A$ as representing aggregation over an interval of frequencies, and thus the assumption that $\gamma_A(0) = 0$ seems incompatible with this concept (since $\gamma_A(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)\, d\lambda$). A more natural example would be a constant function over the interval of desired frequencies (the level of the constant function does not really matter, since this will cancel out of the test statistic). If one wishes to weight middle frequencies more highly, then a kernel of the type

$$A(\lambda) = 1 + \cos\lambda$$

could be used (or powers thereof). Since this choice is appropriate for the interval $[-\pi, \pi]$ with centering at zero, a shifted, scaled and reflected version of the above sinusoidal kernel may be used for other frequency ranges; see McElroy and Holan (2005) for a more detailed discussion. Of course, this method is not limited to peak modelling; we may be interested in the local fit of $f$ to the data at any and all frequencies, not necessarily where peaks or troughs occur. A practitioner should invest some thought into choosing a kernel $A$ appropriate for the local properties of the Null model.

Another application is given by using the measure $\phi_A$ to test for the presence of unobserved components. Suppose that we formulate a model for the time series that consists of two unobserved components, modeled by $g(\lambda)$ and $f(\lambda)$ respectively. We may wish to discern whether the component modeled by $g(\lambda)$ is warranted by the data. That is, we may wish to compare a model having one component to a model containing two components. This suggests using the measure $\phi_A(f, f+g)$; instead we consider $\phi_A(f_0, f_1+g)$, where $f_0$ and $f_1$ have the same form but are allowed to differ in their parameter values (the advantage of not restraining $f_0 = f_1$ is that we can then ensure that both $f_0$ and $f_1 + g$ have approximately equal total power, thereby making it more difficult to reject the Null hypothesis), and with kernel $A$ chosen to weight $g$ more highly. To see that this is the appropriate measure, suppose we wish to assess the data against a hypothesized model $f_1 + g$, which would naturally be done via the measure $\phi_A(\hat{f}, f_1 + g)$. Asymptotically, this quantity is $\phi_A(\tilde{f}, f_1 + g)$. Under the Null hypothesis that no unobserved component is needed,

we have $H_0 : \tilde{f} = f_0$, so that our measure converges to $\phi_A(f_0, f_1 + g)$. Adapting Corollary 2 in a straightforward manner (i.e., assuming $g \in \mathcal{C}^1$) we obtain

$$\sqrt{n}\,\frac{\Big(\phi_A(\hat{f}, f_1 + g) - \phi_{n,A}(\tilde{f}, f_1 + g)\Big)}{\sigma_{n,A}(\tilde{f}, f_1 + g)} \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, 1), \tag{9}$$

where, under $\mathrm{H}_0$, the above quantities are defined by

$$\begin{aligned}
\phi_A(\hat{f}, f_1 + g) &= \frac{1}{n}\mathbf{X}'\Sigma(A/(f_1 + g))\mathbf{X} \\
\phi_{n,A}(f_0, f_1 + g) &= \frac{1}{n}tr\{\Sigma(A)\Sigma(f_0/(f_1 + g))\} \\
\sigma^2_{n,A}(f_0, f_1 + g) &= \frac{2}{n}tr\{(\Sigma(A)\Sigma(f_0/(f_1 + g)))^2\}.
\end{aligned}$$

Thus extreme values of $\phi_A(\hat{f}, f + g)$ indicate that $f_0$ is not a sufficient model for the data at frequencies where the kernel $A$ is centered, and hence that the component model $f_1 + g$ is preferred.

**Example 3:** Suppose that we have a macroeconomic series, and we wish to know whether a cycle should be added as an unobserved component to the overall model. For concreteness, suppose that the cycle is given by the $AR(2)$ model used in Example 1. That is, $g(\lambda)$ is given by (7). The remaining portion $f$ could model the trend (i.e., long-term movements), and might, for example, be given by an $MA$ model. In order to implement the above tests, we will require an explicit form for $f_0$ and $f_1$ so that we can compute $\Sigma(f_0)$, etc. Additionally, the kernel $A$ should be centered at the low frequencies – perhaps at frequency 0. Thus if the data has a significant spectral peak in a neighborhood of the cycle peak frequency $\cos^{-1}(\cos\omega(1 + \rho^2)/2\rho)$ and $f_0$ is unable to capture this behavior by itself, extreme values of the test statistic will tend to be produced, resulting in rejection of $H_0$ and incorporation of the trend plus cycle into our overall model.

Finally, we come to a much more general model comparison procedure. Given a proposed model $f_1$, we know how to assess whether it fits the data locally (or even globally): we can apply Corollary 2 under the Null Hypothesis that $\tilde{f} = f_1$, and choose the kernel $A$ according to the frequency band we are interested in. Of course, this yields a signed measure of goodness-of-fit, and we must have a positive measure if we wish to compare the fit of competing models. This leads us to consider the following measure

$$\psi_A(g, f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)\left(\frac{g(\lambda)}{f(\lambda)} - 1\right)^2 d\lambda = \phi_A(g^2, f^2) - 2\phi_A(g, f) + \gamma_A(0), \tag{10}$$

which follows by expanding the square. Note that $\psi_A(\hat{f}, f)$ yields a positive measure of the fit of a model $f$ to the data, with respect to the kernel $A$. A similar measure is studied in Paparoditis (2000), but that work focuses on a kernel with shrinking bandwidth. If we choose the kernel $A$ to be non-negative everywhere, then the measure $\psi_A(g, f) = 0$ if and only if $g$ and $f$ coincide almost

everywhere. Alternatively, if we have two candidate models $f_1$ and $f_2$, then we can determine which model yields a better fit to the data with respect to $A$, simply by using the measure $\psi_A(\hat{f}, f_1) - \psi_A(\hat{f}, f_2)$; positive values indicate that $f_2$ is superior, whereas negative values indicate that $f_1$ is superior. In order to make these statements statistically significant, we must develop some asymptotic theory for $\psi_A$; it is also necessary to develop a convenient way to compute $\psi_A$ in the spirit of our previous measures.

The second and third terms of $\psi_A(\hat{f}, f)$ in (10) are easily computed by $-2\phi_A(\hat{f}, f) + \gamma_A(0)$, which can be conveniently represented as a quadratic form in the data (plus a constant). The first term involves the square of $\hat{f}$, and hence is more complicated. Following McElroy and Holan (2005), let $R$ denote the column vector of sample autocovariances: $\mathbf{R} = \{R(1-n), \cdots, R(0), \cdots, R(n-1)\}'$, where

$$R(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} X_t X_{t+|h|}.$$

Then, applying Proposition 1 to compute $\Sigma(A/f^2)$ if necessary, we have

$$\mathbf{R}' \Sigma(A/f^2) \mathbf{R} = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \frac{\hat{f}^2(\lambda)}{f^2(\lambda)} \, d\lambda. \tag{11}$$

The derivation of (11) is given in the proof of Corollary 3 below. So our goodness-of-fit measure is calculated via

$$\psi_A(\hat{f}, f) = \frac{1}{2} \mathbf{R}' \Sigma(A/f^2) \mathbf{R} - 2n^{-1} \mathbf{X}' \Sigma(A/f) \mathbf{X} + \gamma_A(0).$$

Note that the factor of $1/2$ associated with the first term is essentially due to the fact that $\mathbb{E}\hat{f}^2(\lambda) \to 2\tilde{f}^2(\lambda)$. We emphasize that $\psi_A(\hat{f}, f)$ is easy to compute given a knowledge of $f$. This measure can be used in two ways: as a stand-alone goodness-of-fit, comparing a proposed model $f$ to the true process $\tilde{f}$, and for comparing two models $f_1$ and $f_2$ to the process $\tilde{f}$. The following result is a corollary of Theorem 3.

**Corollary 3** *Suppose that the third and fourth cumulants of $\{X_t\}$ are zero, $A, \tilde{f} \in \mathcal{C}^1$, and the process $\{X_t\}$ satisfies condition (B). If the kernel $A$ is real, continuous, and even, then*

$$\sqrt{n} \left( \psi_A(\hat{f}, f) - \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \left( \frac{\tilde{f}(\lambda)}{f(\lambda)} - 1 \right)^2 d\lambda \right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V),$$

*where the asymptotic variance is given by*

$$V = \frac{2}{2\pi} \int_{-\pi}^{\pi} A^2(\lambda) \left( \frac{\tilde{f}(\lambda)}{f(\lambda)} \right)^4 d\lambda + \frac{8}{2\pi} \int_{-\pi}^{\pi} \frac{A^2(\lambda) \tilde{f}^2(\lambda)}{f^2(\lambda)} \left( \frac{\tilde{f}(\lambda)}{f(\lambda)} - 1 \right)^2 d\lambda.$$

*In addition, if $f_1, f_2 \in \mathcal{C}^1$, then*

$$\sqrt{n} \left( \psi_A(\hat{f}, f_1) - \psi_A(\hat{f}, f_2) - \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \left( \frac{\tilde{f}(\lambda)}{f_1(\lambda)} - \frac{\tilde{f}(\lambda)}{f_2(\lambda)} \right) \left( \frac{\tilde{f}(\lambda)}{f_1(\lambda)} + \frac{\tilde{f}(\lambda)}{f_2(\lambda)} - 2 \right) d\lambda \right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, W),$$

*where the asymptotic variance is given by*

$$W = \frac{2}{2\pi}\int_{-\pi}^{\pi} A^2(\lambda)\left(\frac{\tilde{f}^2(\lambda)}{f_1^2(\lambda)} - \frac{\tilde{f}^2(\lambda)}{f_2^2(\lambda)}\right)^2 d\lambda + \frac{8}{2\pi}\int_{-\pi}^{\pi} A^2(\lambda)\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} - \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right)^2\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} + \frac{\tilde{f}(\lambda)}{f_2(\lambda)} - 1\right)^2 d\lambda.$$

**Remark 5** Note that $W$ reduces to $V$ when $f_1 = \tilde{f}$ or $f_2 = \tilde{f}$. In the basic goodness-of-fit testing scenario, we formulate the Null hypothesis that $\tilde{f} = f$, so that the asymptotic mean is zero and $V$ reduces to $\frac{2}{2\pi}\int_{-\pi}^{\pi} A^2(\lambda)\, d\lambda$. However, if we are comparing both $f_1$ and $f_2$ to the data, then our Null hypothesis should state that $f_1$ and $f_2$ compare equally favorably to the data, i.e.,

$$\left(\frac{\tilde{f}}{f_1} - 1\right)^2 = \left(\frac{\tilde{f}}{f_2} - 1\right)^2.$$

Assuming that $f_1 \neq f_2$ (if $f_1 = f_2$, the theorem becomes trivial), the above condition is equivalent to the statement that

$$2 = \frac{\tilde{f}}{f_1} + \frac{\tilde{f}}{f_2}.$$

From this we can deduce that $\tilde{f} = 2\frac{f_1 f_2}{f_1 + f_2}$. So under this $H_0$, the latter goodness-of-fit measure has asymptotic mean zero, and the limiting variance $W$ simplifies to

$$\frac{64}{2\pi}\int_{-\pi}^{\pi} A^2(\lambda)\left(\frac{f_1(\lambda) - f_2(\lambda)}{f_1(\lambda) + f_2(\lambda)}\right)^2 d\lambda.$$

Using these facts, one can easily set up two-sided tests with approximately normal critical values.

# 4 Empirical Study

The spectral measures we provide can be used in many different settings. The examples described in Section 3 demonstrate the utility of our approach for several applications. However, the examples we provide do not constitute an exhaustive list of problems for which these diagnostics are potentially useful, and we restrict our attention to just a few case studies. In this section we present the results of a simulation study which gives insight into the finite sample performance for several of the proposed measures. Specifically, using the AR(2) cycle model presented in Example 3 as our Null hypothesis, we investigate the size and power of our diagnostics under several alternatives from the class of ARMA models. Finally, we illustrate our methodology using two separate examples from macro-economic time series. The first example is concerned with assessing the appropriateness of a business cycle model and is therefore concerned with local spectral goodness-of-fit in the low frequency band. The second example considers the addition of extra components when fitting UC models.

### 4.1 Simulation Study

We simulated data from the AR(2) model using (6). This model satisfies all of the assumptions of the previous theorems and corollaries as well as providing a Null model capable of illustrating several different hypothesis tests. For example, this model can be used when the goal is to assess local goodness of fit or, alternatively, to carry out model based peak detection. It is worth noting that although these problems are not unrelated they can be tested using different diagnostics.

For simplicity of generation and exposition we restrict ourselves to one Null model and several alternative models from the class of ARMA processes (including white noise). These models were chosen to be representative of different peak and trough behavior in the neighborhood of the spectral node under consideration; see Figure 1. It should be noted that all of the models investigated were chosen to have identical innovation variance. With these models in hand, we consider several of the diagnostics simultaneously. Moreover, we consider several sample sizes and generate one thousand repetitions for each sample size. Specifically, we consider the hypothesis $H_0 : \tilde{f} = f$; this hypothesis is tested using test statistics (5) and (8). Additionally, the Null model was chosen such that a spectral node is present at frequency .353. This choice comes from considerations of cycle modelling where spectral peaks are present in the low frequency band. Finally, the simulations were all conducted using a sinusoidal kernel with centering frequency $\mu = .353$ and bandwidth $\beta = .706$; see McElroy and Holan (2005) for a complete justification of these choices.

The results of the simulation study demonstrated good finite sample performance for the cases investigated. Both test statistics (5) and (8) produced $\alpha$-levels close to the nominal $\alpha$-level .05 for each of the sample sizes under consideration. Additionally, the power of the test (under a nominal $\alpha$-level of .05) ranged from adequate to excellent depending on the alternative hypothesis being considered, even for sample sizes as small as $n = 120$. Furthermore, the power increased with increased sample size, as expected, and increased along with the departure from the Null hypothesis in the frequency domain over the bandwidth being considered; see Figure 1 and Table 1.

The distribution of both test statistics under the Null hypothesis approached the distribution of a standard Normal random variable, confirming our theoretical asymptotic results. Moreover, the test statistics achieved approximate Normality even in small sample sizes and the distribution means and standard deviations were at the approximate (0,1) level; see Table 2 and Figures 2 and 3.

The speed of convergence to Normality of the distribution of the test statistic $\phi_A$ under $H_0$ is sensitive to our choice of Null model. As seen in Figure 3 the distribution has a fairly long right tail for this choice of Null process, while for the $\theta_{\ddot{A}}$ diagnostic Normality is achieved at a faster rate.

This property of the diagnostics suggest that the choice of test statistic should be made according to the intended application and that for small sample sizes p-values based on Monte Carlo simulation may be preferred. Finally, simulations were conducted for the other diagnostics yielding similar results.

### 4.2 Example 1: Cycle Model Identification

As an illustration we consider the annual series of U.S. GDP. Figure 4 displays the logarithm of the data from 1870 to 1998; a single differencing seems sufficient to render the series stationary. Figure 5 plots an AR(9)-spectrum of the differenced logged data. The model was obtained using maximum likelihood estimation and AIC model selection. The left-hand peak may indicate the presence of a cycle in the differenced data. Thus, we apply our data analytic method to discern if the peak is significant as well as to assess the local (low frequency band) fit of the AR(9) model.

Based on the analysis in Harvey and Trimbur (2003) of quarterly post-World War II GDP and the nonparametric results of McElroy and Holan (2005), a reasonable frequency for the cycle is $2\pi/17.8 = .353$, which we take as the frequency for centering our kernel, $\mu$, in the diagnostics. Additionally, we choose our window width $\beta$ as wide as possible (in this case, the constraint is that the left-hand boundary of the kernel's support is at the origin). Hence we set $\beta = .706$, which places the support of the kernel on the interval $[0, .706]$.

Using the sinusoidal kernel with the $\mu = .353$ and $\beta = .706$ we test the hypothesis $\mathrm{H}_0 : \widetilde{f} = f$, where $f$ is our hypothesized AR(9) model. This hypothesis is tested using both (5) and (8) respectively. In both hypothesis tests our alternative is chosen to be two sided. The p-value for the $\theta_{\ddot{A}}$ diagnostic is .48 while the p-value for the $\phi_A$ diagnostic is .75. Therefore these tests indicate that there is insufficient evidence to reject the AR(9) model on the frequency band $[0, .706]$. Moreover, the $\theta_{\ddot{A}}$ test can be interpreted as rejection of flatness in the direction of negative convexity. That is, with significance, there is insufficient evidence to conclude local flatness or trough behavior of the spectral density. This corroborates the findings in Harvey and Trimbur (2003) as well as McElroy and Holan (2005).

### 4.3 Example 2: Unobserved Component Modelling

As an illustration of our diagnostics in the setting of unobserved component models we first consider the United States Consumer Price Index Series (CPI) from January 1913 through August 2005. Specifically we analyzed the logarithm of the CPI (all items; 1982-84=100; NSA), see Figure 6. Complete information regarding this series can be found on the Bureau of Labor Statistics website (http://www.bls.gov).

Using the Ox software program several models were fit to the logarithm of the CPI series. Among the models under consideration a component model having only an IMA(2,2) trend was determined best according to AICc and log-likelihood value. Another model under consideration was a model having an ARMA(2,1) cycle component with IMA(2,2) trend, and by the same model selection criteria was deemed a competitive model ($AICc = -6.5617$ vs. $AICc = -6.46479$); see Harvey and Trimbur (2003) for a detailed description of the ARMA(2,1) cycle model under consideration. In light of the underlying data generating process and the fact that the model fitting algorithm is sensitive to the starting values supplied in the numerical algorithm, we conducted the hypothesis test $\mathrm{H}_0 : \tilde{f} = f_0$ (trend) vs. $\mathrm{H}_a : \tilde{f} = f_1 + g$ (trend/irregular plus cycle component). Using the sinusoidal kernel, this hypothesis test was evaluated using the test statistic in (9) with $f_0, f_1$ equal to the spectral density of the trend-irregular component and $g$ equal to the cycle spectral density[1]. Note that these time series (and their spectra) were second differenced in order to achieve stationarity.

The business cycle, if it exists, is of unknown period. Since it was of interest to determine whether the data supported a cycle component in the model, we performed four separate tests centering the domain of support for the kernel diagnostic at different frequencies. The four frequencies chosen were $\pi/6$, $\pi/12$, $\pi/24$ and .62 (the hypothesized cycle frequency); the kernel bandwidths were set equal to the centering frequencies. The four tests produced p-values all less than .01. Thus, we conclude that the data supports an IMA(2,2) trend-irregular component plus an ARMA(2,1) cycle component in favor of a model having only an IMA(2,2) trend-irregular component (the fact that the procedure indicates the presence of cyclical power at all four choices of $\mu$ demonstrates that the practitioner need only have a vague notion of the location of the potential spectral peak; it does not indicate the presence of four cycles!).

As a final illustration we consider the Producer Price Index(PPI) from January 1913 through August 2005. Again we analyzed the logarithm of the data (PPI, all commodities; NSA), see Figure 7. Complete information regarding this index can be found at http://www.bls.gov. In exactly the same manner as the CPI, we compared a model having only an IMA(2,2) trend component with a model having both an IMA(2,2) trend and an ARMA(2,1) cycle. For this series the model containing a cycle component was found best according AICc ($AICc = -5.158$ vs. $AICc = -5.14049$). Finally this series was tested using centering frequencies, $\mu$, equal to $\pi/6$, $\pi/12$,, $\pi/24$ and .34 (the hypothesized cycle frequency); again the kernel bandwidths were set equal to the centering frequencies. With the exception of $\mu = \pi/6$ (p-value=.02) all the p-values were less than .01. Therefore based on all the tests conducted we conclude, at any reasonable level $\alpha$, there is sufficient

[1]That is, $f_0$ and $f_1$ are trend-irregular models chosen such that $f_0$ and $f_1 + g$ are both feasible models of the data, in the sense that their total power is roughly comparable.

evidence to reject the Null Hypothesis of a model having only a trend component in favor of a model with both a trend and a cycle component. Our procedure was really useful in this case, because the difference in $AICc$ values was so minute.

Although the AICc model selection criteria slightly favored a model having only trend component one must keep in mind that this model selection criteria is a global measure. Given that in both cases CPI and PPI produced values of AICc that were extremely similar for both trend and trend plus cycle models, it is not surprising that when the models are compared locally (in the low frequency band) that a model having cycle component is preferred.

# 5 Conclusion

In this paper we employ a model based approach for assessing time series model misspecification. In particular, we appeal to the frequency domain and develop asymptotic theory for several spectral measures. These measures provide a flexible set of tools that can be used in a wide class of problems including spectral peak identification, component significance, and local goodness of fit. To this end, this article extends the work of McElroy and Holan (2005) as well as providing several results of independent interest.

Further we provide the results of a simulation study that detail the finite sample performance of our diagnostics as well as the level and power for several relevant hypotheses. The simulations show the improvement gained by adopting a parametric framework, versus the nonparametric approach of McElroy and Holan (2005). Additionally, the methodology is illustrated through the analysis of several economic time series. Although several spectral domain diagnostics are currently available for assessing goodness-of-fit, they are usually concerned with a global measure. The framework we develop considers local assessment of goodness-of-fit. Finally, as we have demonstrated, these local measures may be more appropriate for several of the applications encountered in practice.

One limitation of the current approach is the requirement that the input series be stationary. Since economic data is typically nonstationary, it must be suitably differenced before applying these measures. However, the very operation of differencing affects the shape of the spectral density, and this affect is not just felt at frequencies corresponding to the unit roots. Future work must focus on addressing this important topic.

**Acknowledgements** Holan's research was supported by an ASA/NSF/BLS research fellowship.

## Appendix: Proofs

**Proof of Proposition 1.** We first show that if $g$ is a polynomial, then $\Delta(g)\Sigma(f)\Delta'(g) = \Sigma(f \cdot |g(e^{-i\cdot})|^2)$. Here $\Sigma(f)$ is $m \times m$ dimensional, and $\Delta(g)$ is $(m-q) \times m$ for any $m > q$, the order of $g$. Suppose $\{X_t\}$ is a weakly stationary time series with spectral density $f$, and let $Y_t = g(B)X_t$ so that $\{Y_t\}$ has spectral density $f\,|g(e^{-i\cdot})|^2$. Then we have the following matrix relation for any $m > q$:

$$\begin{bmatrix} Y_{q+1} \\ \vdots \\ Y_m \end{bmatrix} = \Delta(g) \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}.$$

Computing autocovariances now, we obtain

$$\Sigma_Y = \Delta(g)\Sigma_X\Delta'(g)$$

where $\Sigma_Y = \Sigma(f|g(e^{-i\cdot})|^2)$ and is $m-q$ dimensional, and $\Sigma_X = \Sigma(f)$ and is $m$-dimensional. □

**Proof of Theorem 1.** The formula for the expectation is immediate from the definition of $\theta_A(\hat{f})$. For the variance, rewrite as

$$\theta_A(\hat{f}) = \frac{1}{n}\mathbf{Z}'\Sigma^{1/2}(\tilde{f})\Sigma(A)\Sigma^{1/2}(\tilde{f})\mathbf{Z},$$

where $\Sigma^{1/2}(\tilde{f})$ is the symmetric matrix square root of $\Sigma(\tilde{f})$, which exists because the matrix is positive definite (see Golub and Van Loan (1996)). Here $\mathbf{Z}$ is a vector of uncorrelated random variables defined by $\mathbf{Z} = \Sigma^{-1/2}(\tilde{f})\mathbf{X}$. Then if the third and fourth order cumulants vanish, the variance of the quadratic form (McCullagh, 1987) is given by

$$\frac{2}{n^2}tr\left((\Sigma^{1/2}(\tilde{f})\Sigma(A)\Sigma^{1/2}(\tilde{f}))^2\right) \tag{A.1}$$

from which the result follows. For the convergence of the mean and variance, we use several lemmas from McElroy (2006). A discrete approximation to $\gamma_g(j-k)$ is given by

$$\overline{\gamma}_g(h) = \frac{1}{n}\sum_{k=1}^{n} g(\lambda_k)e^{ih\lambda_k}$$

with $\lambda_k = 2\pi k/n$. In this manner the Toeplitz approximation to $\Sigma(g)$ is defined by

$$\overline{\Sigma}_{jk}(g) = \overline{\gamma}_g(j-k).$$

See Taniguchi and Kakizawa (2000) for related results, which we cite here for easy reference. It can be shown that

$$\overline{\Sigma}(g) = H^*DH$$

with $H_{jk} = n^{-1/2} e^{i2\pi jk/n}$ and $D = diag\{g(\lambda_1), \cdots, g(\lambda_n)\}$. The $*$ denotes conjugate transpose. Now Lemmas 2 and 3 of McElroy (2006) can be adapted to deterministic vectors $\mathbf{Z}$ and $\mathbf{X}$. For example, let $\mathbf{e}$ denote the vector of ones; then

$$tr(\Sigma(A)\Sigma(\tilde{f})) = \mathbf{e}'\Sigma(A)\Sigma(\tilde{f})\mathbf{e} = O(1) + \mathbf{e}'\overline{\Sigma}(A)\overline{\Sigma}(\tilde{f})\mathbf{e}.$$

This follows by applying Lemma 2 and 3 of McElroy (2006). Then we use $\overline{\Sigma}(A)\overline{\Sigma}(\tilde{f}) = \overline{\Sigma}(A\tilde{f})$, and apply Lemma 2 once more:

$$\mathbf{e}'\overline{\Sigma}(A\tilde{f})\mathbf{e} = O(1) + \mathbf{e}'\Sigma(A\tilde{f})\mathbf{e} = O(1) + tr(A\tilde{f}).$$

The order statements are as $n \to \infty$, and are based on a result of Wahba (1968). Finally,

$$\frac{1}{n} tr(\Sigma(A)\Sigma(\tilde{f})) = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda)\tilde{f}(\lambda)\, I_e(\lambda)\, d\lambda$$

where $I_e(\lambda) = n^{-1} |\sum_{t=1}^n e^{-i\lambda t}|^2 = \sum_{|h|<n} (1 - |h|/n) e^{-i\lambda h}$. Thus

$$\begin{aligned} &\frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda)\tilde{f}(\lambda) \left( \sum_{h\in\mathbb{Z}} e^{-ih\lambda} - I_e(\lambda) \right) d\lambda \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda)\tilde{f}(\lambda) \left( \sum_{|h|\geq n} e^{-ih\lambda} + n^{-1} \sum_{|h|<n} |h| e^{-ih\lambda} \right) d\lambda. \end{aligned}$$

Let $g(\lambda) = A(\lambda)\tilde{f}(\lambda)$, so that the above error can be written

$$\sum_{|h|\geq n} \gamma_g(h) + n^{-1} \sum_{|h|<n} |h|\gamma_g(h).$$

It is a simple exercise to show that $g \in \mathcal{C}^1$. As a result, the second term above is $O(1/n)$. Also, it is necessary that $|\gamma_g(h)| \leq Ch^{-2}$, from which it follows that the first term is $O(1/n)$ as well. This establishes that

$$\frac{1}{n} tr(\Sigma(A)\Sigma(\tilde{f})) = O(1/n) + \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda)\tilde{f}(\lambda)\, d\lambda$$

using $\sum_{h\in\mathbb{Z}} \gamma_g(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda)\, d\lambda$. So the mean converges to its limit at rate $1/n$; the proof for the variance is similar. The asymptotic normality result, it follows from Lemma 3.1.1 of Taniguchi and Kakizawa (2000) that

$$\sqrt{n}(\theta_A(\hat{f}) - \theta_A(\tilde{f})) = \sqrt{n} \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda)(\hat{f}(\lambda) - \tilde{f}(\lambda))\, d\lambda \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V)$$

with $V$ as stated in the theorem. $\square$

**Proof of Theorem 2.** This proof is similar to that of Theorem 1; the calculations for the mean and variance are straightforward. The limiting behavior of the mean and variance are proved similarly to Theorem 1. For the asymptotic normality we apply Lemma 3.1.1 of Taniguchi and Kakizawa (2000):

$$\sqrt{n}\left(\phi_A(\hat{f},f)-\phi_A(\tilde{f},f)\right)=\sqrt{n}\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{A(\lambda)}{f(\lambda)}(\hat{f}(\lambda)-\tilde{f}(\lambda))\,d\lambda \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,V)$$

with $V$ as stated in the theorem. This proves the theorem. □

**Proof of Theorem 3.** The technique of proof is a routine adaption of the proof of Theorem 3 of Chiu (1988), using material from Brillinger (1975). However, due to the terseness of Chiu (1988)'s treatment, we provide considerably more details here. The expression for the asymptotic mean follows from the proof of Theorem 2 of Chiu (1988). Now generalizing Theorem 5.10.2 of Brillinger (1975) to higher powers of the periodogram, we can approximate the integral

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{j=1}^{J}\phi_j(\lambda)\hat{f}^j(\lambda)\,d\lambda \tag{A.2}$$

by a Riemann sum, with error $O_P(1/n)$ as follows:

$$\frac{1}{n}\sum_{\lambda}\sum_{j=1}^{J}\phi_j(\lambda)I^j(\lambda), \tag{A.3}$$

where the sum is over the Fourier frequencies in $(-\pi,\pi)\setminus\{0\}$ as in Chiu (1988), and $I(\lambda)$ denotes the periodogram $\hat{f}(\lambda)$ restricted to these Fourier frequencies. Note that the factor of $2\pi$ that occurs in our integral expression, in Chiu (1988) is instead incorporated into the definition of the periodogram. Define the discrete Fourier transform of the data at Fourier frequencies $\lambda$ by

$$d(\lambda)=\sum_{t=1}^{n}X_t\,e^{-i\lambda t}.$$

Hence $I(\lambda)=d(-\lambda)d(\lambda)/n$. Now the variance of $\sqrt{n}$ times (A.3) is given by

$$n^{-1}\sum_{\lambda_1}\sum_{\lambda_2}\sum_{j,k=1}^{J}\phi_j(\lambda_1)\phi_k(\lambda_2)\,cum(I^j(\lambda_1),I^k(\lambda_2)).$$

Fix $j$ and $k$ for the moment, and without loss of generality suppose that $j\geq k$. Then the corresponding term in the variance is given by

$$n^{-(j+k+1)}\sum_{\nu}\sum_{\lambda_1}\sum_{\lambda_2}\phi_j(\lambda_1)\phi_k(\lambda_2)\,cum\{d(\omega_{lm});lm\in\nu_1\}\cdots cum\{d(\omega_{lm});lm\in\nu_q\},$$

where $\omega_{lm}=(-1)^m\lambda_l$ and the summation in $\nu$ is over all indecomposable partitions of the following table (see Brillinger (1975, p. 20))

$$\begin{array}{c}(1,1)\cdots(1,2k)\cdots(1,2j)\\(2,1)\cdots(2,2k).\end{array}$$

This result is obtained by applying Theorem 2.3.2 of Brillinger (1975) to

$$I^j(\lambda_1) = n^{-j}(d(-\lambda_1) \cdot d(\lambda_1))^j$$
$$I^k(\lambda_2) = n^{-k}(d(-\lambda_2) \cdot d(\lambda_2))^k.$$

Now our task is to determine which indecomposable partitions $\nu$ will yield asymptotically non-negligible contributions to the variance. In order to do this, we introduce some terminology. Let a $p$-set be any subset of a given table with exactly $p$ elements. We will say that a $p$-set straddles the table if it has at least one element in each row. Now a partition $\nu$ consists of a disjoint collection of $p$-sets (for various $p$), such that the union yields the whole table. Note that these $p$-sets need not be connected. Below, we will show that the only partitions $\nu$ that we need to consider are of two types: either they contain exactly one 4-set (which straddles, with 2 elements in the top row and 2 in the bottom), $k-1$ non-straddling 2-sets (contained in the first row) and another $j-1$ non-straddling 2-sets (contained in the second row); or there are $k+j$ 2-sets, where at least one 2-set straddles. There are additional conditions on these partitions as well, which are discussed below.

In the following analysis, we use Theorem 4.3.2 of Brillinger (1975), which requires condition (B1). Specifically, we use (4.3.15), which is a special case of the above theorem. Asymptotically, the term $\Delta(\lambda) = \sum_{t=1}^{n} e^{-i\lambda t}$ tends to zero unless $\lambda = 0$, in which case the sum is $n$. Now in order for a particular partition to contribute to the variance asymptotically, the corresponding cumulants must together produce $j+k-1$ powers of $n$ – then the overall exponent of $n$ will be $-2$, which will counteract the growth in the double sum over $\lambda_1$ and $\lambda_2$. However, it is possible for the double sum to collapse into a single sum (e.g., when $\lambda_1 = \lambda_2$), in which case we require $j+k$ powers of $n$. Now according to (4.3.15) of Brillinger (1975), for a particular $p$-set in a partition $\nu$, the function $\Delta$ is evaluated at the sum of the $\omega_{lm}$'s such that $(l,m)$ are in that $p$-set. Moreover, $\Delta$ evaluated at this sum is asymptotically negligible unless the sum is zero; hence, we can only supply powers of $n$ by considering $p$-sets such that all the $\omega_{lm}$'s sum to zero. We refer to $\sum_{(l,m)\in B} \omega_{lm}$ as the $\omega$-sum of the $p$-set $B$. For visualization, it is helpful to write out the table of $\omega_{lm}$'s corresponding to the table given above:

$$\begin{array}{cccccc} -\lambda_1 & & \lambda_1 \cdots & -\lambda_1 & \lambda_1 \cdots & -\lambda_1 \quad \lambda_1 \\ -\lambda_2 & & \lambda_2 \cdots & -\lambda_2 & \lambda_2. & \end{array}$$

Clearly, the 2-set given by $\{(1,1),(1,2)\}$ has corresponding $\omega$-sum of zero. Now it follows that if $p$ is odd, the $\omega$-sum of that $p$-set cannot be zero. Since we always need to generate $j+k-1$ powers of $n$ (and possibly $j+k$ powers of $n$), we must have at least $j+k-1$ $p$-sets (but for different $p$, possibly) in a partition $\nu$. Since the total size of the table is $2j+2k$, this excludes $p \geq 6$ outright. Also, having more that one 4-set is excluded as well. Hence, the only possible partitions would

have a single 4-set and $j+k-1$ 2-sets, or simply $j+k$ 2-sets. Let us consider the former type in more detail.

**4-set, 2-set partitions** Now for this type of partition, the $\omega$-sum over the 4-set and over each of the 2-sets must be zero. Note that we can effectively ignore the "diagonal" aspect of the double sum over $\lambda_1, \lambda_2$, i.e., the cases that $\lambda_1 = \lambda_2$ or $\lambda_1 = -\lambda_2$. This is because the total number of sets in this partition is $j+k$, so that the overall exponent of $n$ is $-2$; since a single summation in $\lambda$ is only order $n$, it is asymptotically negligible. Hence the $\omega$-sum for each of the 2-sets is only zero if they don't straddle, i.e., they are contained in a row. For those 2-sets in the first row, they consist of exactly one choice of $\lambda_1$ and one choice of $-\lambda_1$; for 2-sets in the second row, they consist of exactly one choice of $\lambda_2$ and one choice of $-\lambda_2$. In order for the partition to be indecomposable, the 4-set must straddle (essentially, the condition of indecomposability for a two row table amounts to the condition that at least one $p$-set in the partition straddles). It is easy to see that the 4-set must contain the elements $\lambda_1, -\lambda_1, \lambda_2, -\lambda_2$ in some order (it is not possible to draw three elements from one row and one from another). This gives a fairly precise description of the $p$-sets in this type of partition; it is sufficient to count up the number of such partitions using elementary combinatorics.

Ignore for a moment the 4-set and consider the first element $\lambda_1$ in position $(1,2)$ in the table. There are $j$ choices of the element $-\lambda_1$ that it can form a 2-set with, such that the $\omega$-sum is zero. Moving on to the second such element in position $(1,4)$, there are now $j-1$ such choices. Proceeding in this fashion, we obtain $j!$ such 2-set configurations. Independently, we pair up $\lambda_2$ with elements $-\lambda_2$ in the second row, and obtain $k!$ configurations there. Now we wish to pick one of the first row 2-sets and one of the second row 2-sets, and combine them into a 4-set: there are $j\,k$ ways of doing this ($j$ 2-set choices for the first row, and $k$ 2-set choices for the second row). Therefore, the number of 4-set, 2-set partitions is $jkj!k!$.

Next, we see from (4.3.15) of Brillinger (1975) that each of these partitions yields the same contribution to the variance, namely

$$G^X(\lambda_1, -\lambda_1, \lambda_2)\tilde{f}^{j-1}(\lambda_1)\tilde{f}^{k-1}(\lambda_2).$$

Note that the $2\pi$ factors do not appear, since we define our cumulant spectral densities without this normalization, which differs from Brillinger (1975). Combining with the $\phi_j$'s and replacing the Riemann sum by an integral yields

$$\frac{jkj!k!}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\phi_j(\lambda)\phi_k(\omega)G^X(\lambda,-\lambda,\omega)\tilde{f}^{j-1}(\lambda)\tilde{f}^{k-1}(\omega)\,d\lambda d\omega.$$

**2-set partitions** Again we must have the $\omega$-sum of each 2-set to be zero, but since there are $j+k$ sets, the contribution to the variance from these types of partitions will still be negligible unless the double sum collapses. Hence we will have two classes of partitions: either $\lambda_1$ is paired with a $\lambda_2$ or a $-\lambda_1$ in every 2-set, or $\lambda_1$ is paired with a $-\lambda_2$ or a $-\lambda_1$ in every 2-set. In other words, the first case stipulates that no $\lambda_1$ and $-\lambda_2$ are together in a 2-set. Focusing on this case, if a given 2-set contains a $\lambda_1$ and a $-\lambda_1$, then the corresponding $\omega$-sum is zero; likewise for 2-sets containing a $\lambda_2$ and a $-\lambda_2$. However if the elements are $\lambda_1$ and $\lambda_2$, the $\omega$-sum is only zero if $\lambda_1 = -\lambda_2$, which essentially stipulates a condition on the double sum. Since all the $\omega$-sums must be zero for the partition to make a non-negligible contribution, we see that we must have $\lambda_1 = -\lambda_2$. Since there are $j + k$ 2-sets, the overall exponent of $n$ is $-1$, which balances the single sum. The number of $\tilde{f}(\lambda_1)$ and $\tilde{f}(\lambda_2)$ terms is difficult in principal to determine, but since $\lambda = -\lambda_2$ and $\tilde{f}$ is even, we are only concerned with the total number of such terms, which is $j + k$.

On the other hand, if no $\lambda_1$ and $\lambda_2$ can be in the same 2-set, we obtain a zero $\omega$-sum for 2-sets containing a $\lambda_1$ and a $-\lambda_2$ only if $\lambda_1 = \lambda_2$. Hence the double sum collapses to a single sum here as well. The contribution to the variance will then be

$$n^{-1} \sum_{\lambda_1} \left( \phi_j(\lambda_1)\phi_k(-\lambda_1)\tilde{f}^{j+k}(\lambda_1) + \phi_j(\lambda_1)\phi_k(\lambda_1)\tilde{f}^{j+k}(\lambda_1) \right).$$

It remains to count how many such partitions exist; we count the number of partitions yielding the first case, and the same argument can be applied to the second case. First consider including decomposable partitions in the count. Taking the first $\lambda_1$ element in the $(1, 2)$ location of the table, there are $j$ choices of $-\lambda_1$ to pair with, and $k$ choices of $\lambda_2$, so $j + k$ choices total. For the second $\lambda_1$, there is one less $-\lambda_1$ or one less $\lambda_2$, for a total $j+k-1$ remaining choices. All together, we find mates for the $\lambda_1$ elements in $(j+k)(j+k-1)\cdots(k+1)$ ways. Now consider the first $-\lambda_2$ element (none of the $-\lambda_2$ elements have yet been paired). It may only pair with $\lambda_2$ or $-\lambda_1$, of which in total there are only $k$ remaining choices. Proceeding, we obtain $k!$ choices of mates for the various $-\lambda_2$, and so have $(j+k)!$ configurations of 2-sets satisfying our conditions. However, some of these partitions are decomposable, so we must subtract off their contribution. As discussed above, there are $j!k!$ such decomposable partitions, thus our summary count is $(j + k)! - j!k!$. Now replacing the Riemann sum by an integral, we have a variance contribution of

$$\frac{(j + k)! - j!k!}{2\pi} \int_{-\pi}^{\pi} \left( \phi_j(\lambda)\phi_k(-\lambda) + \phi_j(\lambda)\phi_k(\lambda) \right) \tilde{f}^{j+k}(\lambda)\, d\lambda.$$

All together, the asymptotic variance of $\sqrt{n}$ times (A.2) yields $V$, as stated in Theorem 3. Finally, we must consider the higher order cumulants, and show that they always tend to zero ad $n \to \infty$.

Consider the $h^{th}$ cumulant of (A.3), which yields

$$n^{-h}\sum_{\lambda_1}\cdot\sum_{\lambda_h}\sum_{j_1,\cdots,j_h}^{J}\phi_{j_1}(\lambda_1)\cdots\phi_{j_h}(\lambda_h)\,cum\{I^{j_1}(\lambda_1),\cdots,I^{j_h}(\lambda_h)\}$$
$$=n^{-h-r}\sum_{\lambda_1}\cdot\sum_{\lambda_h}\sum_{j_1,\cdots,j_h}^{J}\phi_{j_1}(\lambda_1)\cdots\phi_{j_h}(\lambda_h)\sum_{\nu}cum\{d(\omega_{lm});lm\in\nu_1\}\cdots cum\{d(\omega_{lm});lm\in\nu_q\},$$

where $r=j_1+\cdots+j_h$ and the summation in $\nu$ is over all indecomposable partitions of the table

$$\begin{array}{l}(1,1)\cdots(1,2j_1)\\(2,1)\cdots(2,2j_2)\\\vdots\\(h,1)\cdots(h,2j_h).\end{array}$$

We seek the dominant term in the above cumulant. If we consider an indecomposable partition $\nu$ of the above table, many of the same principles apply from our variance analysis. In particular, we need not consider $p$-sets in $\nu$ with $p$ odd. And the greatest number of factors of $n$ are produced from $\Delta$ evaluated at $\omega$-sums, if we were to take $\nu$ to be a partition consisting solely of 2-sets, where each $\lambda_k$ is paired with a $-\lambda_k$. This would produce $r$ factors of $n$. However, this approach leads to a decomposable partition, since no 2-set straddles. By joining two such 2-sets into a 4-set, we decrease our exponent of $n$ by one. In order to maximize the powers of $n$ contributed by the partition, and at the same time obtain an indecomposable partition, we need $h-1$ 4-sets that straddle such that no row consists purely of 2-sets. Then the rest are row-contained 2-sets, and the total powers of $n$ contributed will be $r-h+1$. This is the most that can be contributed; note that partitions that require a collapsing of $\lambda$-sums actually lower the order (which can be compensated by choosing the partition appropriately). So the maximum exponent of $n$ will be $-2h+1$. Now $h$ of these factors will go towards offsetting the growth due to the $\lambda$ sums. This leaves an overall order for (A.3) of $n^{-h+1}$.

Finally, we multiply by $n^{h/2}$ and obtain the order $n^{-h/2+1}$. This is negative if $h>2$, and hence all cumulants of order $h\geq 3$ tend to zero. Hence the characteristic function tends to that of a Gaussian with mean zero and variance $V$, which proves the theorem. $\square$

**Proof of Corollary 3.** First we derive (11). The quadratic form is seen at once to equal

$$\begin{aligned}
&\sum_{j=1}^{2n-1}\sum_{k=1}^{2n-1} R(j-n)R(n-k)\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{A(\lambda)}{f^2(\lambda)}e^{i\lambda(j-k)}\,d\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{A(\lambda)}{f^2(\lambda)}\sum_{j=1-n}^{n-1} R(j)e^{i\lambda j}\sum_{k=1-n}^{n-1} R(-k)e^{-i\lambda k}\,d\lambda \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{A(\lambda)}{f^2(\lambda)}\hat{f}^2(\lambda)\,d\lambda.
\end{aligned}$$

Hence the measure can be rewritten as

$$\psi_A(\hat{f},f) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{A(\lambda)}{2f^2(\lambda)}\hat{f}^2(\lambda) + \frac{-2A(\lambda)}{f(\lambda)}\hat{f}(\lambda)\,d\lambda + \gamma_A(0).$$

So to obtain the first convergence we simply apply Theorem 3, using $\phi_1 = -2A/f$ and $\phi_2 = A/2f^2$. These functions are even, so the variance is

$$\begin{aligned}
V &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(2\phi_1^2(\lambda)\tilde{f}^2(\lambda) + 16\phi_1(\lambda)\phi_2(\lambda)\tilde{f}^3(\lambda) + 40\phi_2^2(\lambda)\tilde{f}^4(\lambda)\right)d\lambda \\
&= \frac{2}{2\pi}\int_{-\pi}^{\pi} 4A^2(\lambda)\frac{\tilde{f}^2(\lambda)}{f^2(\lambda)} - 8A^2(\lambda)\frac{\tilde{f}^3(\lambda)}{f^3(\lambda)} + 5A^2(\lambda)\frac{\tilde{f}^4(\lambda)}{f^4(\lambda)}\,d\lambda,
\end{aligned}$$

which can be re-expressed in the stated form. The calculation for the mean is standard. Next for the two-model comparison we have $\phi_1 = -2A(1/f_1 - 1/f_2)$ and $\phi_2 = A/2\cdot(1/f_1^2 - 1/f_2^2)$. So the asymptotic mean of $\psi_A(\hat{f},f_1) - \psi_A(\hat{f},f_2)$ is

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)\tilde{f}^2(\lambda)\left(\frac{1}{f_1^2(\lambda)} - \frac{1}{f_2^2(\lambda)}\right) - 2A(\lambda)\tilde{f}(\lambda)\left(\frac{1}{f_1(\lambda)} - \frac{1}{f_2(\lambda)}\right)d\lambda + \gamma_A(0) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} - \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right)\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} + \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right) - 2A(\lambda)\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} - \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right)d\lambda + \gamma_A(0),
\end{aligned}$$

which simplifies to the stated expression. Next, the variance is

$$\begin{aligned}
W &= \frac{2}{2\pi}\int_{-\pi}^{\pi} 4A^2(\lambda)\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} - \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right)^2 - 8A^2(\lambda)\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} - \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right)\left(\frac{\tilde{f}^2(\lambda)}{f_1^2(\lambda)} - \frac{\tilde{f}^2(\lambda)}{f_2^2(\lambda)}\right) \\
&\quad + 5A^2(\lambda)\left(\frac{\tilde{f}^2(\lambda)}{f_1^2(\lambda)} - \frac{\tilde{f}^2(\lambda)}{f_2^2(\lambda)}\right)^2 d\lambda \\
&= \frac{8}{2\pi}\int_{-\pi}^{\pi} A^2(\lambda)\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} - \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right)^2\left(1 - 2\left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} + \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right) + \left(\frac{\tilde{f}(\lambda)}{f_1(\lambda)} + \frac{\tilde{f}(\lambda)}{f_2(\lambda)}\right)^2\right)d\lambda \\
&\quad + \frac{2}{2\pi}\int_{-\pi}^{\pi} A^2(\lambda)\left(\frac{\tilde{f}^2(\lambda)}{f_1^2(\lambda)} - \frac{\tilde{f}^2(\lambda)}{f_2^2(\lambda)}\right)^2 d\lambda
\end{aligned}$$

as desired. This completes the proof. □

## References


[1] Brockwell, P. and Davis, R. (1996) *Introduction to Time Series and Forecasting.* New York: Springer.

[2] Beran, J. (1994) *Statistics for Long Memory Processes.* New York: Chapman and Hall.

[3] Brillinger, D. (1981) *Time Series Data Analysis and Theory,* San Francisco: Holden-Day.

[4] Chen, W. and Deo, R. (2004) A generalized portmanteau goodness-of-fit test for time series models. *Econometric Theory* **20**, 382 – 416. 5

[5] Chiu, S. (1988) Weighted least squares estimators on the frequency domain for the parameters of a time series. *The Annals of Statistics* **16**, 1315–1326.

[6] Golub, G. and Van Loan, C. (1996) *Matrix Computations.* Baltimore: The Johns Hopkins University Press.

[7] Harvey, A. and Trimbur, T. (2003) General Model-Based Filters for Extracting Cycles in Economic Time Series. *Review of Economics and Statistics* **85**, 244 – 255.

[8] Hosoya, Y. and Taniguchi, M. (1982) A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics* **10**, 132–153.

[9] Li, W. (2004) *Diagnostic Checks in Time Series.* CRC Press.

[10] McCullagh, P. (1987) *Tensor Methods in Statistics.* London: Chapman and Hall.

[11] McElroy, T. (2006) Statistical Properties of Model-Based Signal Extraction Diagnostics. SRD Research Report No. $RRS2006-06$, U.S. Census Bureau.

[12] McElroy, T. and Holan, S. (2005) A Nonparametric Test for Assessing Spectral Peaks. SRD Research Report No. 2005/10, Bureau of the Census.

[13] Paparoditis, E. (2000) Spectral density based goodness-of-fit tests for time series models. *Scandinavian Journal of Statistics* **27**, 143 – 176.

[14] Parzen, E. (1983) "Autoregressive spectral estimation" in *Handbook of Statistics,* eds. D. Brillinger and P. Krishnaiah.

[15] Taniguchi, M. and Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series,* New York City, New York: Springer-Verlag.

[16] Tunnicliffe-Wilson, G. (1979) Some efficient computational procedures for high order $ARMA$ models. *Journal of Statistical Computation and Simulation* **8**, 301 – 309.

[17] Wahba, G. (1968) On the distribution of some statistics useful in the analysis of jointly stationary time series. *Annals of Mathematical Statistics* **39**, 1849 – 1862.

| power | | | | | | |
|---|---|---|---|---|---|---|
| $\theta_{\ddot{A}}$ | Level | AR(2) | AR(9) | MA(3) | ARMA(1,1) | White Noise |
| n=120 | .053 | .440 | .885 | .712 | .891 | .946 |
| 144 | .052 | .469 | .895 | .744 | .907 | .954 |
| 180 | .043 | .501 | .899 | .752 | .902 | .945 |
| 244 | .045 | .542 | .919 | .779 | .916 | .954 |
| 280 | .049 | .571 | .921 | .808 | .927 | .957 |
| 360 | .046 | .594 | .920 | .824 | .940 | .971 |
| $\phi_A$ | | | | | | |
| n=120 | .039 | .683 | 1.0 | .996 | 1.0 | 1.0 |
| 144 | .036 | .763 | 1.0 | .999 | 1.0 | 1.0 |
| 180 | .054 | .826 | 1.0 | 1.0 | 1.0 | 1.0 |
| 244 | .068 | .913 | 1.0 | 1.0 | 1.0 | 1.0 |
| 280 | .051 | .947 | 1.0 | 1.0 | 1.0 | 1.0 |
| 360 | .055 | .974 | 1.0 | 1.0 | 1.0 | 1.0 |

Table 1: The results of 1000 replications of a simulation study to evaluate the level and power. Presented results include the level of the test and the power under different alternatives from the class of ARMA processes. The nominal $\alpha$-level in all cases was $\alpha = .05$.

| | $n = 120$ | $n = 144$ | $n = 180$ | $n = 240$ | $n = 288$ | $n = 360$ |
|---|---|---|---|---|---|---|
| mean - $\theta_{\ddot{A}}$ | -.168 | -.168 | -.175 | -.154 | -.094 | -.060 |
| sd - $\theta_{\ddot{A}}$ | .935 | .961 | .933 | .966 | .974 | .959 |
| mean - $\phi_A$ | -.060 | -.046 | -.022 | -.009 | .014 | .068 |
| sd - $\phi_A$ | 1.006 | .956 | 1.000 | 1.031 | .984 | 1.035 |

Table 2: The sample mean and standard deviation for the distribution of test statistics (5) and (8). The simulations consisted of 1000 repetitions at the nominal $\alpha$-level of $\alpha = .05$.

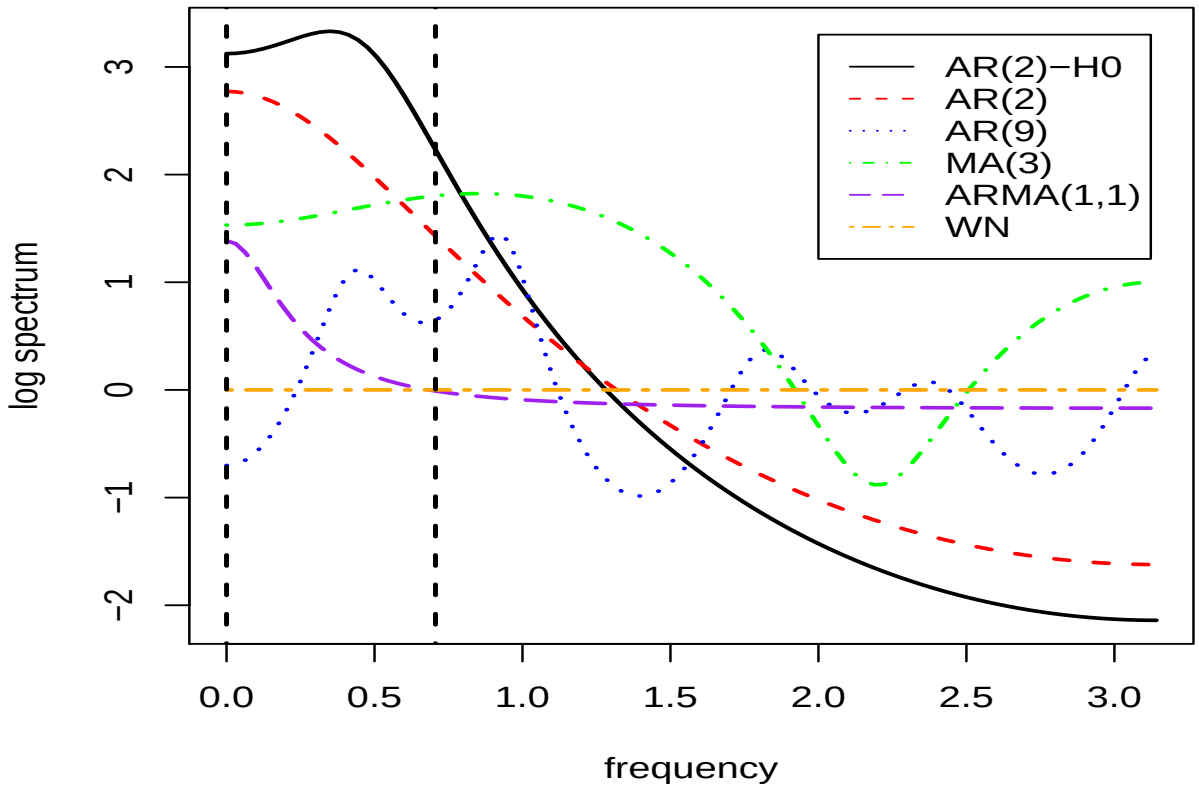


Figure 1: This figure contains plots of the spectral densities used in the simulation study. The dashed vertical lines denote the support of the Kernel (bandwidth) for this investigation.

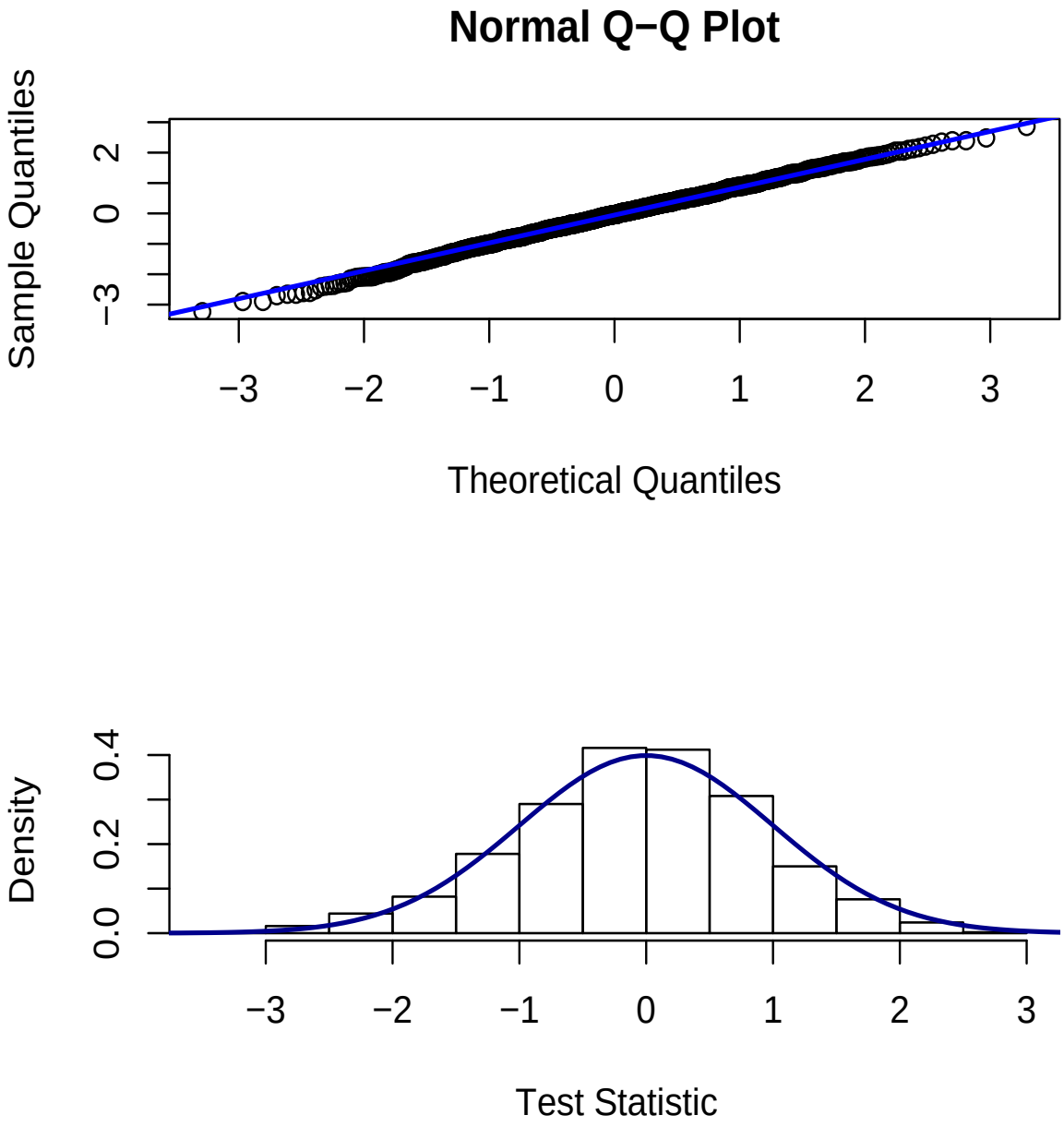


Figure 2: This figure contains a qq plot and histogram for test statistic (5) involving $\theta_{\ddot{A}}$ when $n = 360$. For convenience the theoretical distribution of the standard Normal distribution is superimposed on the histogram.

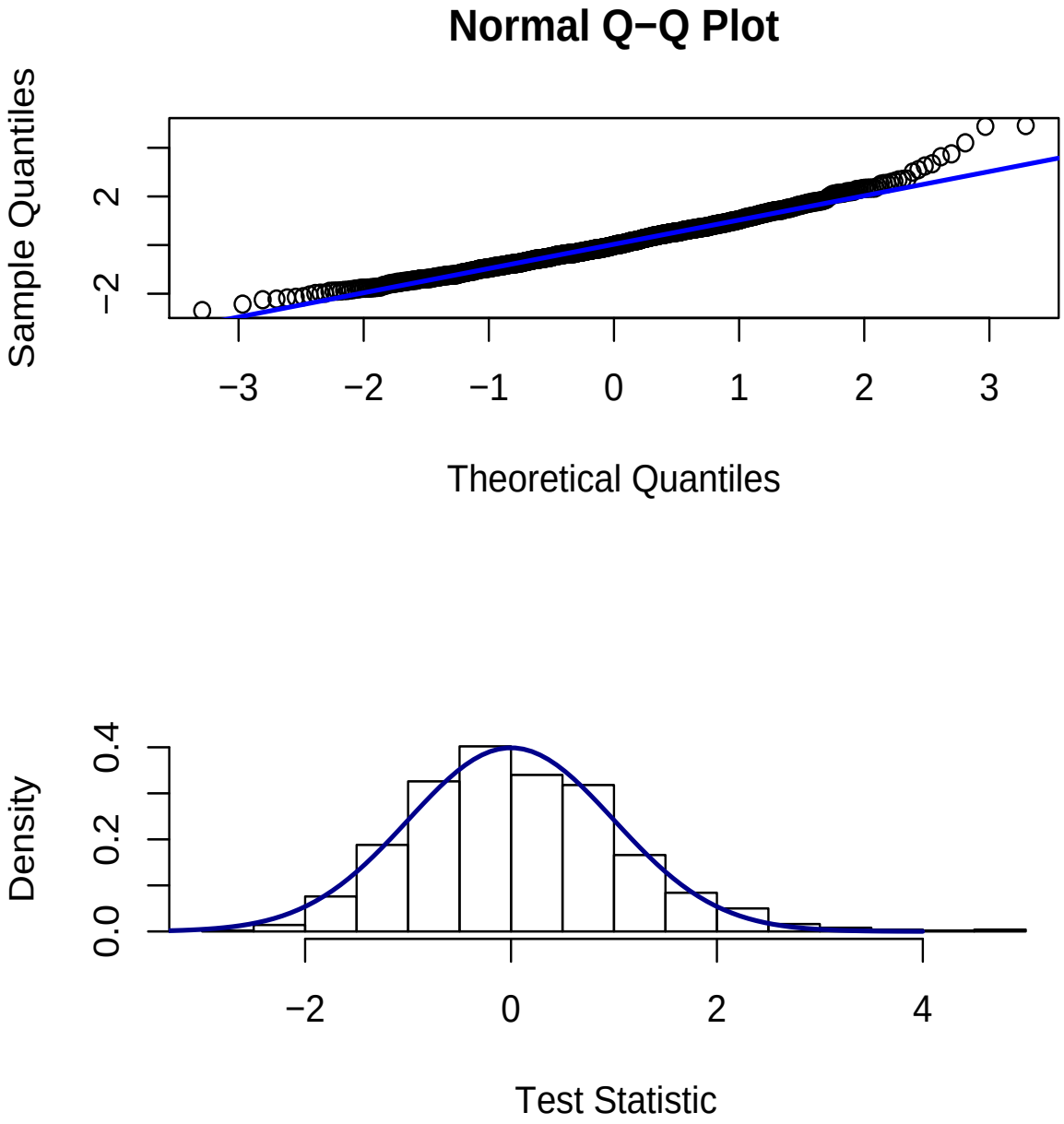


Figure 3: This figure contains a qq plot and histogram for test statistic (8) involving $\phi_A$ when $n = 360$. For convenience the theoretical distribution of the standard Normal distribution is superimposed on the histogram.

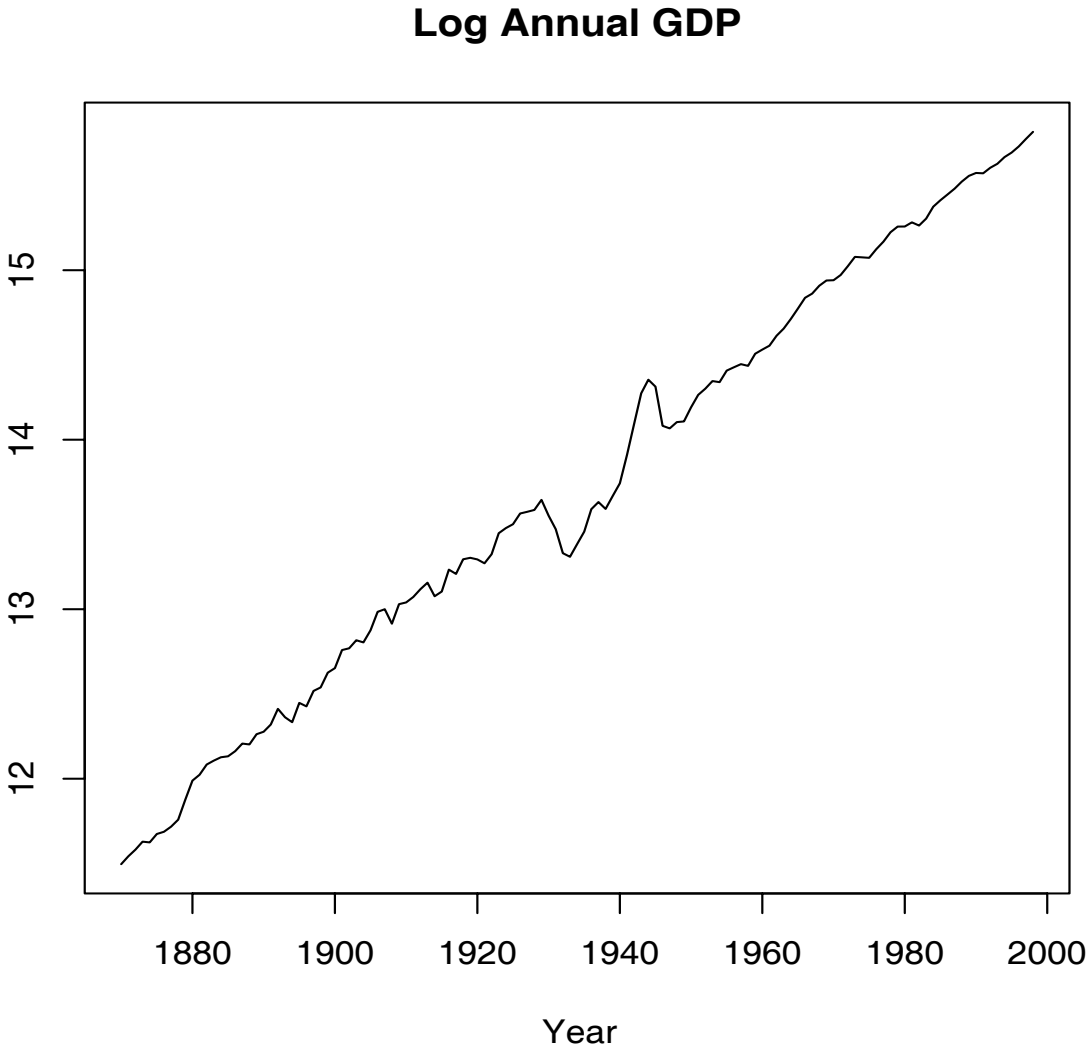


Figure 4: Logarithm of U.S. GDP, 1870-1998.

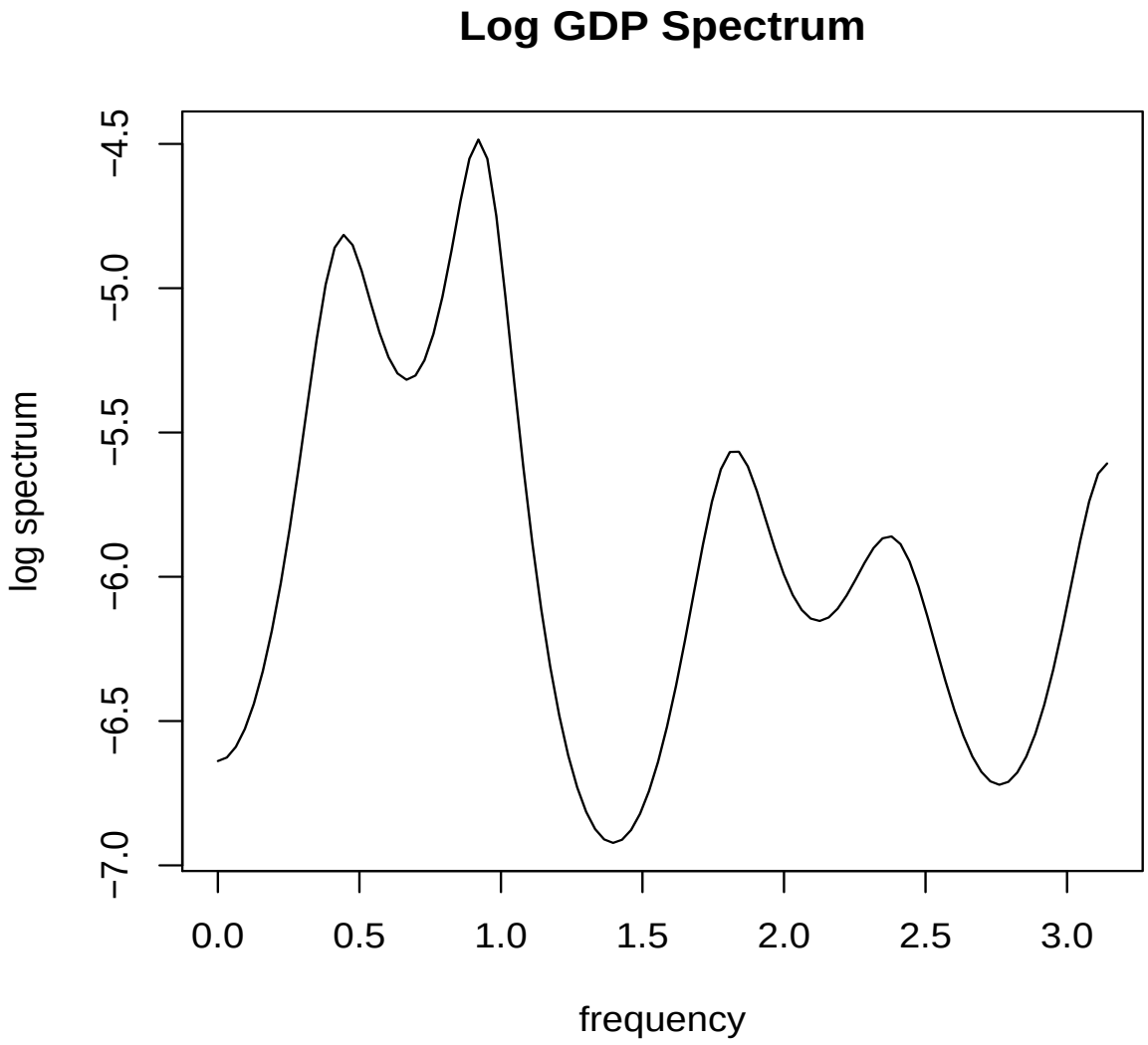


Figure 5: AR(9) Spectrum of differenced U.S. GDP. Left-hand peak may indicate presence of a cycle in the differenced logged data.

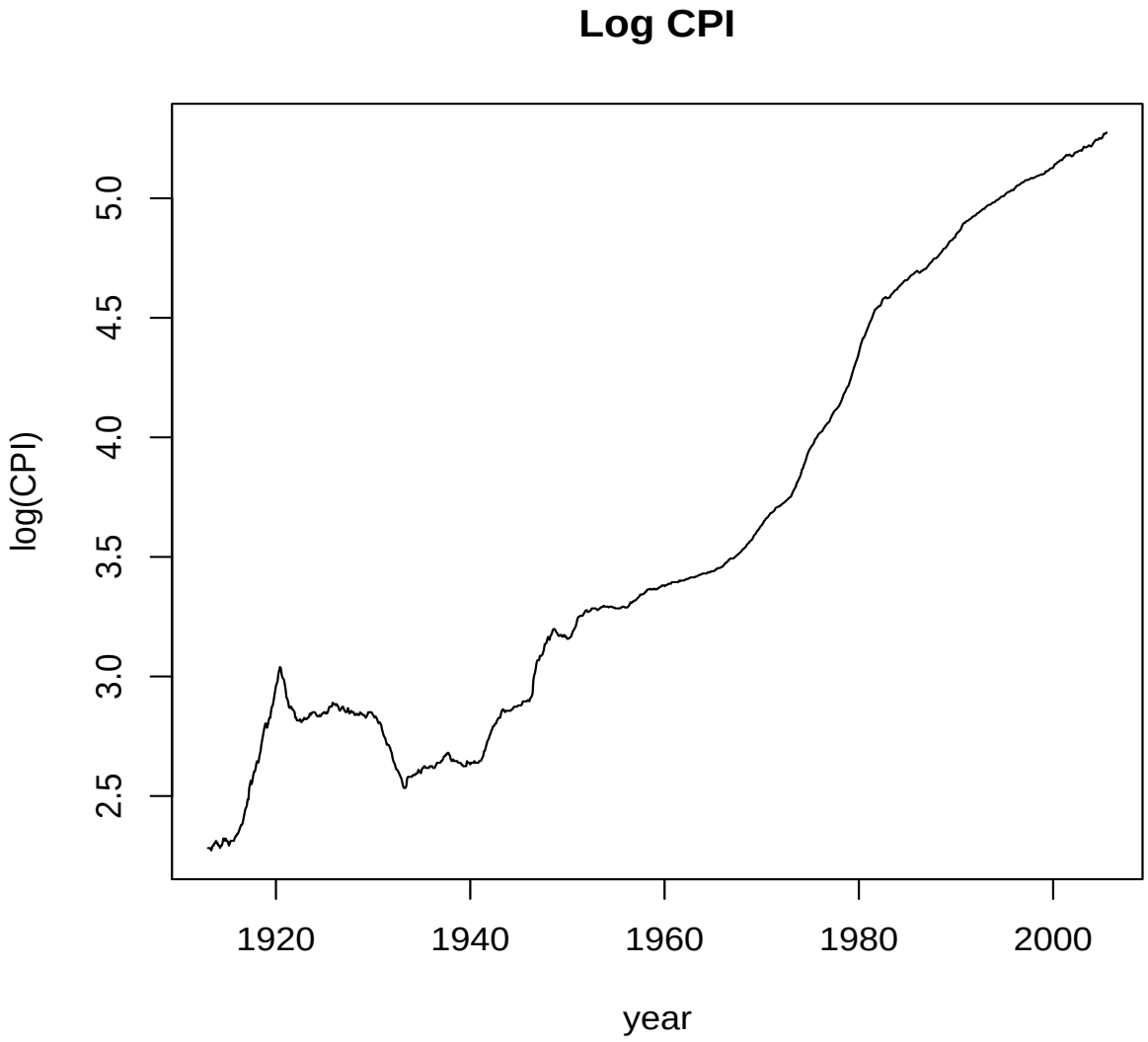


Figure 6: Logarithm of U.S. CPI, 1913-2005.

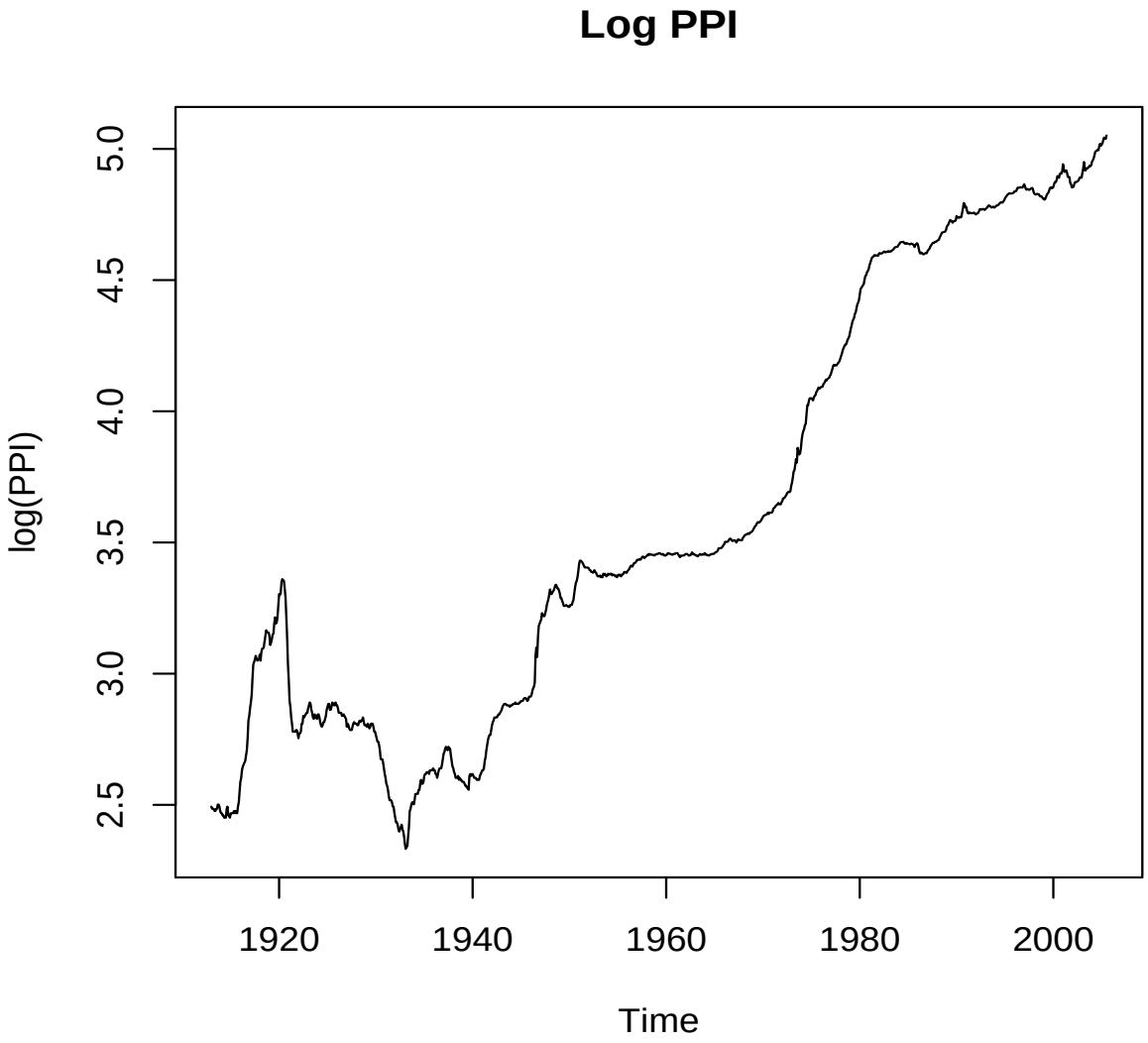


Figure 7: Logarithm of U.S. PPI, 1913-2005.