Journal of Multivariate Analysis 100 (2009) 604–621

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

A local spectral approach for assessing time series model misspecification

Tucker McElroy [a,*], Scott Holan [b]

[a] Statistical Research Division, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233-9100, United States

[b] Department of Statistics, University of Missouri-Columbia, 146 Middlebush Hall, Columbia, MO 65211-6100, United States

ARTICLE INFO

Article history:
Received 30 August 2007
Available online 3 July 2008

AMS subject classifications:
62F05
62F03
62M10

Keywords:
Cycle estimation
Goodness-of-fit
Peak detection
Seasonal adjustment
Spectral density
Unobserved components

ABSTRACT

We consider band-limited frequency-domain goodness-of-fit testing for stationary time series, without smoothing or tapering the periodogram, while taking into account the effects of parameter uncertainty (from maximum-likelihood estimation). We are principally interested in modeling short econometric time series, typically with 100 to 150 observations, for which data-driven bandwidth selection procedures for kernel-smoothed spectral density estimates are unlikely to have adequate levels. Our mathematical results take parameter uncertainty directly into account, allowing us to obtain adequate level properties at small sample sizes. The main theorems provide very general results involving joint normality for linear functionals of powers of the periodogram, while accounting for parameter uncertainty, which can be used to determine the level and power of a wide array of statistics. We discuss several applications, such as spectral peak testing and testing for the inclusion of an Unobserved Component, and illustrate our methods on a time series from the Energy Information Administration.

Published by Elsevier Inc.

1. Introduction

There is an abundance of literature on time-domain methods for detecting model misspecification for a stationary time series (see Li [14] for a comprehensive discussion). However, one can also test for model goodness-of-fit (gof) in the frequency domain, 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. For example, there is the likelihood-ratio test in the form of a Whittle likelihood (see Taniguchi and Kakizawa [23]). Also, general frequency-domain gof tests have been proposed by Paparoditis [18,19] and Chen and Deo [5], though these gof tests are not limited to a specific frequency band. The work of Beran [1] and Eichler [8,9] generalize such gof tests to limited frequency bands, and the latter also considers multivariate modeling. Eichler [9] uses kernel-smoothed and/or tapered periodogram estimates for the spectral density, and it is shown that the parameter uncertainty (say of maximum-likelihood estimates) does not affect the asymptotic results for the test statistics.

In this paper we consider band-limited gof testing without smoothing/tapering the periodogram, and we determine the asymptotic effects of parameter uncertainty. Since the applications that we consider (such as the identification of seasonality) involve short economic time series, typically with 100 to 150 observations, we wish to develop diagnostics that do not depend on a bandwidth that grows with sample size. Although the bandwidth selection problem involved with smoothing the periodogram can be adequately handled through data-driven algorithms, simulations indicate that large

* Corresponding author.
E-mail addresses: tucker.s.mcelroy@census.gov (T. McElroy), holans@missouri.edu (S. Holan).

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2008.06.010

samples (1000 or more observations for the global iterative procedure (ITP) used by Eichler [9]—see Table 2 therein) are typically needed to achieve reasonable levels. Since our bandwidths are essentially held fixed, we must take into account the effect of parameter uncertainty on our asymptotic results. We derive mathematical results that take this parameter uncertainty directly into account, and this allows us to obtain reasonable level properties at small sample sizes. Interestingly, parameter uncertainty *decreases* the overall variability in the test statistic (see Remark 1 below and also McElroy [16]). We next discuss some of the applications of our band-limited gof test.

McElroy and Holan [17] discuss the problem of spectral peak detection, with applications to cycle estimation in econometrics and seasonal adjustment in federal statistics; also see Priestley [21] for a background on this subject. If a postulated model fails to adequately capture a prominent spectral peak really present in the data, then certain stochastic periodic phenomena will be completely absent from our model, resulting in a loss of the model's explanatory power. Moreover, such inadequate models will tend to produce flawed seasonal adjustments, since the model-based filters will be faulty (see McElroy [16] for a discussion).

More generally, we may be interested in whether a model fits the data at hand with respect to a particular range of frequencies. 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 Kullback–Leibler discrepancy is computed by aggregating over all frequencies. Thus, maximum-likelihood estimates can be expected to provide an adequate model in a global sense. In contrast, a band-limited diagnostic test can be constructed so as to focus on a narrower band of frequencies (this strategy is also considered in Eichler [9]). For example, if we are interested in estimating or forecasting an ambient signal, such as a trend or seasonal component, then attention is naturally focused upon the band of frequencies where most of the signal's spectral mass is located (e.g., low frequencies for trend). Then one could consider a gof test focused on the pertinent frequency band.

In the basic Unobserved Component (UC) model – see Harvey [10] – 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 modeling, one is interested in knowing whether the addition of another UC 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.

The band-limited gof statistic that we consider here is very similar to the statistic of Paparoditis [18], the difference being that we do not kernel-smooth the periodogram—and we also consider fixed frequency bands (also our work is encompassed by the general results of Eichler [9], only we do not smooth or taper). Since the bandwidth is not allowed to increase with sample size, our methods must take parameter uncertainty into account. Theorem 1 provides a very general mathematical result involving joint normality for linear functionals of powers of the periodogram; Theorem 2 extends this result by taking parameter uncertainty into account. These two results can be used to determine the level and power of a wide array of statistics, and thus may be of general interest to a much wider audience.

The paper is organized as follows. In Section 2 we develop our notation and state our general mathematical results (Theorems 1 and 2). Section 3 specializes these results to the quadratic band-limited gof test statistic that we consider in this paper and gives several of the applications discussed above—peak testing, gof testing, and UC testing. Section 4 provides the results of a simulation study, which show the efficacy of these methods in practice. We also provide an illustration of our methodology, to the application of peak testing, using a series from the Energy Information Administration. Finally, Section 5 concludes. All proofs are left to the Appendix, which also contains some notes for straightforward computer implementation of the proposed diagnostic.

2. Notation and general results

Suppose that, after suitable transformations if necessary (and removal of regression effects), we have a mean zero stationary time series $X_1, X_2, \dots, X_n$, which will sometimes be denoted by the vector $\mathbf{X} = (X_1, X_2, \dots, X_n)'$. If the original data is homogeneously nonstationary, we suppose that a differencing operator $\delta(B)$ has already been applied to reduce the data to stationarity. The spectral density $f(\lambda)$ is well-defined so long as the autocovariance function is absolutely summable. More generally, for any bounded non-negative function f, let $\{\gamma_f(h)\}_{h\in\mathbb{Z}}$ denote its inverse Fourier Transform:

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

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

with $i = \sqrt{-1}$ and $\lambda \in [-\pi, \pi]$ (this definition provides our convention with the 2π factor, which differs from some authors). Finally, let $I(\lambda)$ denote the periodogram:

$$I(\lambda) = \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-it\lambda} \right|^2 = \sum_{h=1-n}^{n-1} R(h) e^{-ih\lambda} \quad \lambda \in [-\pi, \pi],$$

with $R(h)$ equal to the sample (uncentered) autocovariance function.

In this section we consider the asymptotic properties for statistics $Q_n(f, g, \theta)$ of the form

$$Q_n(f, g, \theta) = \frac{1}{n} \sum_{\lambda} g_\theta(\lambda) f^j(\lambda),$$

where g_θ is some weighting function dependent on a parameter vector θ, and $j \geq 1$ is an integer power of the spectral density. The sum is over the Fourier frequencies in $(-\pi, \pi) \setminus \{0\}$; note that when $j = 1$, the sum is asymptotically equivalent to the integral representation

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} g_\theta(\lambda) f(\lambda) \mathrm{d}\lambda.$$

However, this approximation to the sum is not valid when $j > 1$ and f is the periodogram (Chen and Deo [4]). Typically some estimate $\hat{f}$ of the spectrum f is substituted, and g_θ is determined by the practitioner—although θ is not known and an estimate $\hat{\theta}$ is substituted for it. In this paper we only consider $\hat{f} = I$, the periodogram, although other works consider kernel-smoothed and/or tapered periodograms as spectral estimates (Paparoditis [18,19] and Eichler [8,9]). These latter types of estimates can provide a faster rate of convergence of Q_n to its asymptotic distribution, although the sampling distribution will typically be quite sensitive to the choice of bandwidth (Eichler [8,9]).

For motivation, we briefly describe the gof test statistic of this paper (further discussion is given in Section 3), denoted by ψ_A:

$$\frac{1}{n} \sum_{\lambda} A(\lambda) \left(\frac{I(\lambda)}{f_{\hat{\theta}}(\lambda)} - 1 \right)^2 = \frac{1}{n} \sum_{\lambda} A(\lambda) \frac{I^2(\lambda)}{f^2_{\hat{\theta}}(\lambda)} - \frac{2}{n} \sum_{\lambda} A(\lambda) \frac{I(\lambda)}{f_{\hat{\theta}}(\lambda)} + \frac{1}{n} \sum_{\lambda} A(\lambda). \tag{1}$$

Here A is a non-negative function that may correspond to a band-limited function such as $1_{[-a,a]}$ for $0 < a < \pi$. Such a statistic, intuitively speaking, should be approximately zero if the model is correctly specified (and conversely, if it is asymptotically zero, then the model is correctly specified on the support of A). This statistic is equal to

$$Q_n(I^2, A/f^2_\theta, \hat{\theta}) - 2Q_n(I, A/f_\theta, \hat{\theta}) + \frac{1}{n} \sum_{\lambda} A(\lambda).$$

Hence it is needful to consider joint asymptotics involving functionals of different powers of the periodogram.

Note that the dependence of Q_n on $\hat{\theta}$ has a nontrivial effect on the asymptotics. Indeed, if we fix θ at some deterministic quantity, the asymptotics are much simpler to derive. The methods of Eichler [9] achieve a higher rate of convergence for his gof statistics, and hence the $\sqrt{n}$-order error induced by $\hat{\theta} - \tilde{\theta}$ (where $\tilde{\theta}$ is the parameter value) becomes asymptotically negligible. However, in our case the parameter uncertainty has a substantial effect on the asymptotics. Hence we seek to derive the exact asymptotics for such statistics, taking parameter uncertainty into account, while providing a consistent estimate of the variance as well.

In order to proceed, we first establish some further notation. Consider a set J of L integers written as $J = \{j_1, j_2, \ldots, j_L\}$. These positive integers are the various powers of the periodogram to be considered. The ith Q_n statistic then has the form $Q_n(I^{j_i}, g_i, \hat{\theta})$, where $g_i = g_{\theta,i}$ is a user-defined function which in general depends on the parameter vector θ. We consider the joint asymptotics of these statistics under some assumptions on the g_i, the time series $\{X_t\}$, the parameter space Θ, the estimate $\hat{\theta}$, and the model class $\{f_\theta\}_{\theta \in \Theta}$. Generally, we have a family $\mathcal{F} = \{f_\theta\}_{\theta \in \Theta}$ of spectral densities that parameterize the second-order properties of the time series.

Pseudo-true values minimize a certain distance from the model class to the true spectral density. Consider the following Kullback–Leibler distance function, which corresponds to a quasi-Gaussian likelihood (or Whittle likelihood)—see Dahlhaus and Wefelmeyer [7]:

$$D(k, h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\log k(\lambda) + \frac{h(\lambda)}{k(\lambda)} \right) \mathrm{d}\lambda.$$

Typically, the function k is drawn from a parametric family of spectral densities parameterized by θ, i.e., some class $\mathcal{F}$. Note that the Quasi-Maximum-Likelihood (QML) estimate $\hat{\theta}$ is the minimizer of $D(f_\theta, I)$ with respect to $\theta \in \Theta$, where Θ denotes the parameter space. If the model is mis-specified, then $\tilde{\theta}$ is defined to be the minimizer of $D(f_\theta, \tilde{f})$ over $\theta \in \Theta$. Here $\tilde{f}$ denotes the true spectral density. So in this framework, we have

$$H_0 : \tilde{f} \in \mathcal{F}$$
$$H_a : \tilde{f} \notin \mathcal{F}.$$

Note that if $\tilde{\theta}$ is the unique minimizer, then H_0 implies $\tilde{f} = f_{\tilde{\theta}}$ (this is condition 6 below). We will also consider the following set of additional assumptions:

1. $\{X_t\}$ is mean zero and strictly stationary.
2. $\{X_t\}$ is Gaussian.

3. $\{X_t\}$ satisfies the Brillinger conditions described in Taniguchi and Kakizawa [23], p. 55.
4. The fourth-order cumulants of $\{X_t\}$ are zero (see Taniguchi and Kakizawa [23], p. 54 for a definition).
5. Θ is compact and convex.
6. $\tilde{\theta}$, the pseudo-true value of the parameter, exists uniquely and lies in the interior of Θ.
7. The spectral density $f_\theta(\lambda)$ is twice continuously differentiable in θ and is continuous in λ.
8. The weighting functions $g_{\theta,i}(\lambda)$ are twice continuously differentiable in θ and are continuous in λ.
9. The matrix $M_f(\theta)$, which is by definition the Hessian of the Kullback–Leibler discrepancy between f_θ and $\tilde{f}$, is nonsingular at $\theta = \tilde{\theta}$.
10. The derivatives of the spectral density are uniformly bounded (in λ) and bounded away from zero.

Most of these conditions are natural and are easily satisfied, though condition 5 assumes the compactness of the parameter space, which would typically include the innovation variance (although from a theoretical standpoint this makes little sense, in practice it matters little). We also note that condition 2 implies conditions 3 and 4.

We utilize the notation ∇_θ and H_θ for the gradient and Hessian matrix operators, which operate on a scalar function of the parameter vector θ. Our starting point is the following joint asymptotic normality result for the fixed parameter case:

Theorem 1. *Under conditions* 1, 3, *and* 8, *for any* $\theta \in \Theta$

$$\left\{\sqrt{n}\left(Q_n(I^{j_i}, g_i, \theta) - j_i! Q_n(\tilde{f}^{j_i}, g_i, \theta)\right)\right\}_{i=1}^{L} \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V(\theta)) \tag{2}$$

as $n \to \infty$, *with* $V(\theta)$ *an* $L \times L$ *variance matrix with klth entry*

$$\begin{aligned} V_{kl}(\theta) &= \frac{(j_k + j_l)! - j_k! j_l!}{4\pi} \int_{-\pi}^{\pi} \left(g_{\theta,k}(\lambda) g_{\theta,l}(-\lambda) + g_{\theta,l}(\lambda) g_{\theta,k}(-\lambda) + 2 g_{\theta,k}(\lambda) g_{\theta,l}(\lambda)\right) \tilde{f}^{j_k + j_l}(\lambda) \mathrm{d}\lambda \\ &\quad + \frac{j_k j_k! j_l j_l!}{2(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} (g_{\theta,k}(\lambda) g_{\theta,l}(\omega) + g_{\theta,l}(\lambda) g_{\theta,k}(\omega)) G^X(\lambda, -\lambda, \omega) \tilde{f}^{j_k - 1}(\lambda) \tilde{f}^{j_l - 1}(\omega) \mathrm{d}\lambda \mathrm{d}\omega. \end{aligned}$$

Note that this result is true whether or not H_0 is true and clearly provides a generalization to Theorem 3 of Chiu [6]. The mean term in (2) is $j_i! Q_n(\tilde{f}^{j_i}, g_i, \theta)$, which under H_0 is equal to $j_i! Q_n(f_{\tilde{\theta}}^{j_i}, g_i, \theta)$. This can be estimated by substituting $\hat{\theta}$ for $\tilde{\theta}$. Then evaluating at $\theta = \hat{\theta}$ yields the centered test statistic

$$\sqrt{n}\left(Q_n(I^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta})\right).$$

The following theorem gives the asymptotics of this centered test statistic. Under H_0 we have $\tilde{f} = f_{\tilde{\theta}}$ and the asymptotic bias in (3) (below) goes away, but under H_a this quantity determines the power of the test.

Theorem 2. *Under conditions* 1, 2, 5, 6, 7, 8, *and* 9 *with* $\hat{\theta}$ *the QML (if* $\hat{\theta}$ *is the MLE, also assume condition* 10*), we have*

$$\begin{aligned} &\left\{\sqrt{n}\left(Q_n(I^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta})\right) + \sqrt{n}\frac{j_i!}{2\pi} \int_{-\pi}^{\pi} g_{\tilde{\theta},i}(\lambda)(f_{\tilde{\theta}}^{j_i}(\lambda) - \tilde{f}^{j_i}(\lambda)) \mathrm{d}\lambda\right\}_{i=1}^{L} \\ &\stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, W(\tilde{\theta})) \end{aligned} \tag{3}$$

as $n \to \infty$, *with* $W(\theta)$ *an* $L \times L$ *variance matrix with klth entry*

$$\begin{aligned} W_{kl}(\theta) &= \frac{(j_k + j_l)! - j_k! j_l!}{4\pi} \int_{-\pi}^{\pi} \left(g_{\theta,k}(\lambda) g_{\theta,l}(-\lambda) + g_{\theta,l}(\lambda) g_{\theta,k}(-\lambda) + 2 g_{\theta,k}(\lambda) g_{\theta,l}(\lambda)\right) \tilde{f}^{j_k + j_l}(\lambda) \mathrm{d}\lambda \\ &\quad + \frac{(j_k + 1)! - j_k!}{4\pi} \int_{-\pi}^{\pi} \left(g_{\theta,k}(\lambda) p_{\theta,l}(-\lambda) + p_{\theta,l}(\lambda) g_{\theta,k}(-\lambda) + 2 g_{\theta,k}(\lambda) p_{\theta,l}(\lambda)\right) \tilde{f}^{j_k + 1}(\lambda) \mathrm{d}\lambda \\ &\quad + \frac{(j_l + 1)! - j_l!}{4\pi} \int_{-\pi}^{\pi} \left(g_{\theta,l}(\lambda) p_{\theta,k}(-\lambda) + p_{\theta,k}(\lambda) g_{\theta,l}(-\lambda) + 2 p_{\theta,k}(\lambda) g_{\theta,l}(\lambda)\right) \tilde{f}^{j_l + 1}(\lambda) \mathrm{d}\lambda \\ &\quad + \frac{1}{4\pi} \int_{-\pi}^{\pi} \left(p_{\theta,k}(\lambda) p_{\theta,l}(-\lambda) + p_{\theta,l}(\lambda) p_{\theta,k}(-\lambda) + 2 p_{\theta,k}(\lambda) p_{\theta,l}(\lambda)\right) \tilde{f}^{2}(\lambda) \mathrm{d}\lambda. \end{aligned}$$

These entries are defined in terms of the following quantities:

$$\begin{aligned} p_{\theta,i}(\lambda) &= -j_i! f_\theta^{-2}(\lambda) b_i'(\theta) M_f^{-1}(\theta) \nabla_\theta f_\theta(\lambda) \\ b_i(\theta) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (\tilde{f}^{j_i}(\lambda) - f_\theta^{j_i}(\lambda)) \nabla_\theta g_{\theta,i}(\lambda) + j_i g_{\theta,i}(\lambda) f_\theta^{j_i - 1}(\lambda) \nabla_\theta f_\theta(\lambda) \mathrm{d}\lambda \\ M_f(\theta) &= \nabla_\theta \nabla_\theta' D(f_\theta, \tilde{f}). \end{aligned}$$

When all the $g_{\theta,i}$ functions are even and H_0 holds, the variance formulas simplify to

$$W_{kl}(\theta) = \frac{(j_k + j_l)! - j_k! j_l!}{\pi} \int_{-\pi}^{\pi} r_{\theta,k}(\lambda) r_{\theta,l}(\lambda) \mathrm{d}\lambda - 2 j_k! j_l! b_k'(\theta) M_f^{-1}(\theta) b_l(\theta)$$

$$h_\theta(\lambda) = \nabla_\theta \log f_\theta(\lambda) = \frac{\nabla_\theta f_\theta(\lambda)}{f_\theta(\lambda)}$$

$$r_{\theta,i}(\lambda) = g_{\theta,i}(\lambda) f_\theta^{j_i}(\lambda).$$

Remark 1. The variance of the kth component of (3) is given by W_{kk}, which is equal to $\pi^{-1}((2j_k)! - j_k!^2) \int_{-\pi}^{\pi} r_{\theta,k}^2(\lambda)\mathrm{d}\lambda - 2j_k!^2 z(\theta)$, where $z(\theta)$ is the quadratic form $b'(\theta) M_f^{-1}(\theta) b(\theta)$. Since $M_f^{-1}(\theta)$ is non-negative definite under H_0 (see Appendix A.2) and parameter uncertainty only affects W_{kk} through $b(\theta)$, we see that the variance is *decreased* by estimating parameters (when all parameters are held fixed, $b(\theta) = 0$ and we reduce to Theorem 1). The results in McElroy [16] treat the case that all parameters are held fixed except for the innovation variance, which is estimated.

Some implications of the various conditions are discussed in the proof. The formula for the asymptotic variance is fairly complicated, but note that when the parameters are all fixed, each p_i function is identically zero and the result just reduces to Theorem 1. In practice, we wish to estimate $W(\tilde{\theta})$ via plugging in $\hat{\theta}$, so that we get the correct size for our test statistic under H_0. Let us first re-express the formula for $M_f(\theta)$:

$$\begin{aligned} \left[M_f(\theta)\right]_{kl} \;=\; & \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial^2}{\partial\theta_k \partial\theta_l} f_\theta(\lambda) \cdot f_\theta^{-1}(\lambda) - \frac{\partial}{\partial\theta_k} f_\theta(\lambda) \cdot \frac{\partial}{\partial\theta_l} f_\theta(\lambda) \cdot f_\theta^{-2}(\lambda) \\ & - \frac{\partial^2}{\partial\theta_k\partial\theta_l} f_\theta(\lambda) \cdot f_\theta^{-2}(\lambda) \cdot \tilde{f}(\lambda) + 2 \frac{\partial}{\partial\theta_k} f_\theta(\lambda) \cdot \frac{\partial}{\partial\theta_l} f_\theta(\lambda) \cdot f_\theta^{-3}(\lambda) \cdot \tilde{f}(\lambda) \mathrm{d}\lambda \\ \left[M_f(\tilde{\theta})\right]_{kl}\Big|_{H_0} = \; & \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_k} f_{\tilde{\theta}}(\lambda) \cdot \frac{\partial}{\partial\theta_l} f_{\tilde{\theta}}(\lambda) \cdot f_{\tilde{\theta}}^{-2}(\lambda) \mathrm{d}\lambda. \end{aligned}$$

Therefore, under H_0, the formula simplifies greatly; so define the Fisher information matrix by

$$M_{f,H_0}(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \nabla_\theta f_\theta(\lambda) \nabla_\theta' f_\theta(\lambda) f_\theta^{-2}(\lambda) \mathrm{d}\lambda.$$

Further, we can re-express the formula for $W(\theta)$ with $M_{f,H_0}(\theta)$ in place of $M_f(\theta)$, so long as H_0 is true—this makes no difference in the value of $W(\tilde{\theta})$, although $W(\theta)$ is altered in general. Call this new variance matrix $W^0(\theta)$. Lastly, we can form the estimate $W^0(\hat{\theta})$, which is shown below to be consistent for $W(\tilde{\theta})$ under H_0.

Proposition 1. *Suppose H_0 is true, and define $W^0(\theta)$ as noted above. Then under conditions* 5, 6, 7, 8*, and* 9 *(for the QML; for the MLE, also assume condition* 10*)*

$$W^0(\hat{\theta}) \xrightarrow{P} W(\tilde{\theta})$$

as $n \to \infty$, in the sense that each matrix entry converges in probability.

Remark 2. As a result, we can compute the variance estimate with only a knowledge of the first derivatives of f_θ. Procedurally, we first compute $\nabla_\theta f_\theta(\lambda)$ (analytically, if possible), and hence obtain $M_{f,H_0}(\hat{\theta})$, $b_i(\hat{\theta})$, and $p_{\hat{\theta},i}(\lambda)$. Then we can compute the entries of $W_{H_0}(\hat{\theta})$, using numerical integration if necessary.

Thus a standardized test statistic in general will be some linear combination (given by constants β_i) of the various Q_n:

$$T_n = \sqrt{n} \frac{\sum_i \beta_i Q_n(I^{j_i}, g_i, \hat{\theta}) - \sum_i \beta_i j_i! Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta})}{\sqrt{\sum_{i,k} \beta_i \beta_k W_{ik}^0(\hat{\theta})}}. \tag{4}$$

Under H_0 and the listed assumptions, this statistic is asymptotically standard normal by Theorem 2 and Proposition 1. Under H_a, the asymptotics are dominated by the quantity

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} g_{\tilde{\theta},i}(\lambda)(f_{\tilde{\theta}}^{j_i}(\lambda) - \tilde{f}^{j_i}(\lambda))\mathrm{d}\lambda, \tag{5}$$

which determines the power of the test. Large discrepancies between $f_{\tilde{\theta}}$ and $\tilde{f}$ in the support of $g_{\tilde{\theta}}$ will increase the asymptotic power; but if the above integral is zero, then there will be little or no power to the procedure. These ideas are further developed in Section 3.

Finally, we note that sometimes the finite-sample properties of our standardized test statistic (4) are inadequate, in that the sampling distribution displays skewness and non-normality (we have observed this in samples of up to $n = 1000$). However, in our experience a logarithmic variance-stabilizing transform has proved beneficial. Letting $\mu = \sum_i \beta_i j_i! Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta})$, we have

$$\sqrt{n}\frac{\log\left(\sum_i \beta_i Q_n(I^{j_i}, g_i, \hat{\theta}) + \epsilon\right) - \log(\mu + \epsilon)}{\sqrt{\sum_{i,k} \beta_i \beta_k W_{ik}^0(\hat{\theta})/(\mu + \epsilon)^2}},$$

where ϵ is any deterministic constant that ensures the arguments of the log function are positive. The above quantity is asymptotically normal by the delta method (Bickel and Doksum [2]).

3. The goodness-of-fit statistic ψ_A

First we consider the statistic $\psi_A(I, f_{\hat{\theta}})$ given by (1). For quadratic statistics ($L = 2$ and $j_1 = 2, j_2 = 1$), using (5) we see that asymptotic power is determined by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{i=1}^{2}\beta_i j_i! g_{\tilde{\theta},i}(\lambda)(f_{\tilde{\theta}}^{j_i}(\lambda) - \tilde{f}^{j_i}(\lambda))d\lambda. \tag{6}$$

For the ψ_A statistic in particular, the band functions are $g_{\theta,i} = A/f_\theta^{j_i}$ and $\beta_1 = 1$ and $\beta_2 = -2$; so (6) yields

$$\frac{1}{\pi}\int_{-\pi}^{\pi} A(\lambda)\frac{\tilde{f}(\lambda)}{f_{\tilde{\theta}}(\lambda)}\left(1 - \frac{\tilde{f}(\lambda)}{f_{\tilde{\theta}}(\lambda)}\right)d\lambda,$$

which could be zero for some alternatives. Note that a quadratic statistic with non-zero asymptotic power is given by letting $\beta_1 = 1/2, \beta_2 = -2$ instead; then we obtain

$$-\frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)\left(1 - \frac{\tilde{f}(\lambda)}{f_{\tilde{\theta}}(\lambda)}\right)^2 d\lambda \le 0,$$

with equality only if $f_{\tilde{\theta}} = \tilde{f}$ on the support of A. This quadratic statistic can therefore be expected to have superior power against a wide class of alternatives. However, we will focus on ψ_A as defined by (1), since this is a more intuitive formulation (and is consistent with the prior work of Paparoditis [18]). This can also be written as

$$\psi_A(I, f_{\hat{\theta}}) = Q_n(I^2, A/f_\theta^2, \hat{\theta}) - 2Q_n(I, A/f_\theta, \hat{\theta}) + \gamma_A(0). \tag{7}$$

The asymptotic mean works out to be $\gamma_A(0)$, while the variance estimate is $W_{11}^0(\hat{\theta}) - 4W_{12}^0(\hat{\theta}) + 4W_{22}^0(\hat{\theta})$, so by (4) our normalized test statistic is

$$\sqrt{n}\frac{Q_n(I^2, A/f_\theta^2, \hat{\theta}) - 2Q_n(I, A/f_\theta, \hat{\theta})}{\sqrt{W_{11}^0(\hat{\theta}) - 4W_{12}^0(\hat{\theta}) + 4W_{22}^0(\hat{\theta})}}.$$

We next discuss some of the properties of ψ_A (such as power and level) as well as potential applications.

The use of band-limited A can produce higher power in certain situations. Suppose that A is the indicator function on an interval of length $\delta < 2\pi$. Then (6) yields

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} A(\lambda)\left(1 - \frac{\tilde{f}(\lambda)}{f_{\tilde{\theta}}(\lambda)}\right)^2 d\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(1 - \frac{\tilde{f}(\lambda)}{f_{\tilde{\theta}}(\lambda)}\right)^2 d\lambda$$

if $f_{\tilde{\theta}} = \tilde{f}$ outside the support of A. Disregarding the effect of parameter uncertainty on the variance for simplicity, we find that the asymptotic variance is $16\gamma_{A^2}(0)$. In analogy with the efficacy of the test (Taniguchi, Puri, and Kondo [24]), we can form a measure related to the asymptotic relative efficiency of Pitman [20] by taking (6) divided by the asymptotic standard deviation, and take the ratio of such quantities for two tests. Then the ratio for the identity kernel compared to the kernel A yields $\sqrt{\gamma_{A^2}(0)} = \sqrt{\delta/2\pi}$ which is less than one. This heuristic argument indicates, in this specialized situation, that taking a band-limited statistic can be more powerful against certain kinds of alternatives. (See Eichler [9] for a related discussion.)

We next illustrate the use of logarithmic variance-stabilizing transforms. Now (7) can be re-written as $n^{-1}\sum_\lambda A(\lambda)(I(\lambda)/f_{\hat{\theta}}(\lambda) - 1)^2$, which is positive. Therefore letting $\epsilon = 0$ and $\mu = \gamma_A(0)$ in the variance-stabilizing method of Section 2, we see that

$$\sqrt{n}\frac{\log\left(Q_n(I^2, A/f_\theta^2, \hat{\theta}) - 2Q_n(I, A/f_\theta, \hat{\theta}) + 2\gamma_A(0)\right) - \log(\gamma_A(0))}{\sqrt{\left(W_{11}^0(\hat{\theta}) - 4W_{12}^0(\hat{\theta}) + 4W_{22}^0(\hat{\theta})\right)/\gamma_A(0)^2}} \quad (8)$$

is asymptotically normal. These modified statistics are just as easy to compute, but tend to have improved symmetry in smaller samples.

We next discuss several applications of the ψ_A statistic mentioned in the Introduction: peak detection, band-limited gof testing, and UC testing. We first consider the situation where it is suspected that certain stochastic periodic phenomena are present in the data, and it is desired to detect the significance of such phenomena. For example, an evolving seasonal pattern may be present in the time series, which manifests itself as peaks in the spectral density of the process at the so-called "seasonal frequencies". These are $\pi/6, 2\pi/6, 3\pi/6, 4\pi/6, 5\pi/6$, and $6\pi/6$ for monthly data. Another example comes from econometrics, where much interest focuses on detection of a business cycle in macroeconomic series. A business cycle represents the slowly moving (stationary) oscillations about a smooth trend, and is commonly thought to have a period between 4 and 10 years for most series (Harvey and Trimbur [11]). Again, the presence of a cycle would be manifested as a peak in the corresponding frequency range of the spectral density.

Suppose that we wish to perform local peak detection; we say a peak is detected if a postulated model f_θ that includes a salient peak cannot be rejected. A very simple model is given by the following $AR(2)$:

$$(1 - 2\rho\cos\omega B + \rho^2 B^2)X_t = \epsilon_t, \quad (9)$$

where ϵ_t is a white noise sequence with variance τ^2. The frequency ω parameterizes the location (maximizer) of the peak, which is at $\cos^{-1}(\cos\omega(1+\rho^2)/2\rho)$; this quantity is close to ω if ρ is close to unity, so for simplicity we will call ω the "peak location" parameter. The parameter ρ governs the overall shape of the curve, with $\rho = 1$ corresponding to the limiting case of an infinite peak. The corresponding spectral density is

$$f_\theta(\lambda) = \frac{\tau^2}{(1 - 2\rho\cos(\lambda+\omega) + \rho^2)(1 - 2\rho\cos(\lambda-\omega) + \rho^2)}, \quad (10)$$

with parameter vector $\theta' = (\omega, \rho, \tau^2)$. To construct the ψ_A statistic, one could use the kernel $1 + \cos\lambda$, appropriately centered at frequencies of interest (see McElroy and Holan [17] for more discussion of specific kernels). For example, we might be interested in detecting seasonal peaks in seasonally adjusted data, in which case the data typically requires at least one trend difference. So we would difference the data, select a kernel with support centered on a particular seasonal frequency (testing one seasonal peak at a time), and perform the test using (8) and the null model f_θ given by (10). Alternatively, perhaps we are interested in detecting a business cycle in raw data; typically the data will require trend differencing, and we follow the same procedure except that now the peak will be located in the low frequency band rather than at a seasonal frequency.

This approach can be generalized to band-limited gof testing by taking kernels A of a more general shape. For example, suppose A consists of several modes, one centered at each spectral region of interest. For monthly data, one could test the fit of a seasonal model by using the kernel $1+\cos(12\lambda)$, which has modes at all six seasonal frequencies. The basic requirement is only that the kernel is non-negative.

Another application involves testing for the inclusion of additional UCs. Suppose that we formulate a model for the time series that consists of two (possibly nonstationary) UCs. We may wish to discern whether the second UC is warranted by the data. That is, we may wish to compare a model having both UCs (say f_θ) to a model containing only one (call this model f_ξ); then the model for the second UC is $f_\theta - f_\xi$ (if the components are stationary). In the spirit of likelihood-ratio tests, we can use a weighting function of the form A/f_ξ, where the denominator corresponds to the specific alternative of only one UC. One then modifies (8) by replacing the weighting functions A/f_θ^j by A/f_ξ^j. The gof statistic is computed by fitting both models – the two UCs (f_θ) and the one UC (f_ξ) – and plugging into the formula for ψ_A. The kernel A would be chosen to weight the difference between the spectra, where the second UC is expected to be concentrated.

For example, 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 (the second UC) is given by the $AR(2)$ model (9), while the first UC just consists of trend— say a generalized random walk. Then the once differenced trend has a spectral density f_ξ, while the (differenced) alternative model consists of differenced trend plus differenced cycle, with spectral density

$$f_\theta(\lambda) = f_\xi(\lambda) + |1 - e^{-i\lambda}|^2 q(\lambda),$$

where $q(\lambda)$ is given by (10). We center the kernel on cycle frequencies (λ's corresponding to 4–10 year periodicities), while excluding trend frequencies (λ's close to zero) completely. Then if the data has a significant spectral peak in a neighborhood of the cycle peak frequency ω and f_θ 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 cycle into our overall model.

More generally, we may wish to test a proposed model class f_θ against a specific alternative family $\mathcal{G} = \{f_\xi\}_{\xi\in\Xi}$. In other words, we suppose that

$$H_a : \tilde{f} \in \mathcal{G}$$

with $\tilde{f} = f_{\tilde{\xi}}$ for the unique pseudo-true value $\tilde{\xi}$. By using f_ξ in lieu of f_θ in the definition of the weighting functions, we may be able to increase our power against the specific alternative H_a. Such test statistics will still be asymptotically normal, since Theorem 2 can easily be adapted: suppose that each g_i in (4) now depends on ξ instead of θ, written as $g_{\xi,i}$. Then so long as $\hat{\xi} - \tilde{\xi} = O_P(n^{-1/2})$ (for MLEs or QMLs), where $\tilde{\xi}$ is the unique pseudo-true value (i.e., it is the minimizer of $D(f_\xi, \tilde{f})$ when H_0 holds), we have

$$\left\{\sqrt{n}\left(Q_n(I^{j_i}, g_i, \hat{\xi}) - j_i! Q_n\left(f_{\hat{\theta}}^{j_i}, g_i, \hat{\xi}\right)\right)\right\}_{i=1}^{L} \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, W(\tilde{\theta}, \tilde{\xi})),$$

where the formulas for $W(\theta, \xi)$ are obtained from $W(\theta)$ by replacing $g_{\theta,k}$ everywhere by $g_{\xi,k}$. (Also note that now $b_k(\theta, \xi) = (2\pi)^{-1} \int_{-\pi}^{\pi} g_{\xi,i}(\lambda) f_\theta^{j_i - 1}(\lambda) \nabla_\theta f_\theta(\lambda) \mathrm{d}\lambda$, and the formula for $p_{\theta,\xi,i}$ is similarly altered.) This assertion is proved by simply adapting the proof of Theorem 2, noting that we can form the Taylor expansion of $g_{\hat{\xi},i}$ about $g_{\tilde{\xi},i}$ without introducing additional error asymptotically.

What is the benefit of weighting by a particular alternative? Computing the asymptotic power, we find that (6) for the modified ψ_A statistic (i.e., $\beta_1 = 1/2$ instead of 1) yields (under the specific alternative that $\tilde{f} = f_{\tilde{\xi}}$)

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \left(\frac{f_{\tilde{\theta}}(\lambda)}{f_{\tilde{\xi}}(\lambda)} - 1\right)^2 \mathrm{d}\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \left(\frac{k(\lambda)}{f_{\tilde{\xi}}(\lambda)} - 1\right)^2 \mathrm{d}\lambda,$$

where $k = f_{\tilde{\theta}} - f_{\tilde{\xi}}$ (in the UC testing case, this corresponds to the second UC). In the conventional weighting scheme, the corresponding power quantity is

$$-\frac{1}{2\pi} \int_{-\pi}^{\pi} A(\lambda) \left(\frac{k(\lambda)}{f_{\tilde{\xi}}(\lambda) + k(\lambda)} - 1\right)^2 \mathrm{d}\lambda.$$

The latter integrand will tend to be smaller, being close to unity on frequencies where k is large and $f_{\tilde{\xi}}$ is small. Conversely, in the case where we weight by the alternative, the integrand can be quite large, especially if k is large on a set of frequencies where $f_{\tilde{\xi}}$ is small (in the extreme case that k and $f_{\tilde{\xi}}$ have disjoint support and A is constructed to have support contained within that of k, we have $A(k/f_{\tilde{\xi}})^2 = \infty$ and $A(k/f_{\tilde{\theta}})^2 = 1$). This heuristic argument shows how weighting by a specific alternative model can generate more power against that alternative hypothesis (if A is chosen appropriately).

4. Empirical study

In this section we explore the level and power of ψ_A in small samples via Monte Carlo simulation. Additionally, we illustrate the utility of our approach by considering one of the stated applications in the context of a real time series from the Energy Information Administration.

4.1. Simulation study

A key motivation for this work has been to achieve reasonable levels at small samples ($n \leq 150$) typical of seasonally adjusted data at the US Census Bureau (Gaussianity is a safe assumption for such time series once outlier and calendar effects have been removed). Of course we also want decent power in small samples; both levels and power are explored in a small simulation study. Initial simulations indicated substantial asymmetry in ψ_A (a long right tail), with the bell-shape only appearing for large ($n \geq 1000$) sample sizes; the log transform method of Sections 2 and 3 greatly ameliorated this problem, providing decent levels even for small samples.

The simulation we perform to evaluate level assumes that, under the null hypothesis, the data are Gaussian and come from either an $AR(1)$ model with $\phi = .6$ or an $ARMA(1, 1)$ with $\phi = \theta = .6$; both models have unit innovation variance. Further, we conducted the simulation using two different kernels A under various sample sizes. The first kernel is a band-limited version of the Tukey–Hanning kernel (TH): a kernel of the form $1 + \cos(\lambda)$, restricted to the frequencies $(0, \pi/3)$ and centered at $\pi/6$. The second kernel is the identity kernel on the interval $(-\pi, \pi)$. A complete discussion regarding kernels can be found in [17]. Finally, this simulation consisted of 10,000 repetitions at the nominal α-level $\alpha = .05$.

The simulation we perform to evaluate power assumes that the data are Gaussian and come from the $AR(2)$ cycle model given in (9) with $\rho = .75$ and unit innovation variance. Additionally, the null model was chosen such that a spectral mode is present at frequency $\omega = .3069\pi$. This choice comes from considerations surrounding the model used for the *Oil Data* (analyzed in Section 4.2) where the spectral peak is present in the low frequency band. The simulation was conducted under two separate alternative specifications. In the first scenario the model under the alternative was chosen to be from the class of $AR(1)$ models while in the second case the model under the alternative was chosen to be from the class of $ARMA(1, 1)$ models. Thus both models are mis-specified in a manner that precludes the estimates from achieving the correct spectral shape. Moreover, the simulations were all conducted using a band-limited Tukey–Hanning kernel with centering frequency $\omega_0 = .3069\pi$ and bandwidths $\delta = .6\pi, .4\pi$; see [17] for a complete justification of these choices. Finally, this simulation consisted of 10,000 repetitions at the nominal α-level $\alpha = .05$.

The results of the simulation study demonstrated good finite-sample performance for the limited cases investigated (see Table 1). In general, both test statistics ψ_A and $\log(\psi_A)$ produced α-levels close to the nominal α-level .05 for each of the

Table 1
Results of a Monte Carlo simulation to evaluate the level of the gof statistic ψ_A and its log-transformed counterpart

Level $AR(1) - \phi = .6$						
A = TH kernel; $\omega_0 = \pi/6$; $\delta = \pi/3$	$n = 100$	$n = 150$	$n = 250$	$n = 500$	$n = 1000$	$n = 5000$
mean - ψ_A	−.085	−.061	−.045	−.018	−.015	−.009
sd - ψ_A	.790	.838	.864	.907	.940	.927
level - ψ_A	.025	.032	.032	.034	.036	.034
mean - $\log(\psi_A)$	−.292	−.232	−.195	−.124	−.098	−.048
sd - $\log(\psi_A)$	.823	.847	.880	.933	.974	.993
level - $\log(\psi_A)$	.033	.032	.032	.039	.046	.048
$A = 1_{(-\pi,\pi)}$	$n = 100$	$n = 150$	$n = 250$	$n = 500$	$n = 1000$	$n = 5000$
mean - ψ_1	−.153	−.154	−.130	−.083	−.049	−.027
sd - ψ_1	.917	.938	.944	.971	.995	.992
level - ψ_1	.027	.032	.034	.042	.048	.047
mean - $\log(\psi_1)$	−.006	−.037	−.040	−.023	−.006	−.008
sd - $\log(\psi_1)$	.839	.883	.912	.952	.984	.990
level - $\log(\psi_1)$	.023	.030	.032	.040	.046	.048
Level $ARMA(1, 1) - \phi = \theta = .6$						
A = TH kernel; $\omega_0 = \pi/6$; $\delta = \pi/3$	$n = 100$	$n = 150$	$n = 250$	$n = 500$	$n = 1000$	$n = 5000$
mean - ψ_A	−.058	−.050	−.035	−.035	−.017	−.009
sd - ψ_A	.827	.846	.860	.901	.920	.928
level - ψ_A	.031	.031	.032	.033	.034	.036
mean - $\log(\psi_A)$	−.285	−.238	−.185	−.143	−.106	−.047
sd - $\log(\psi_A)$	.837	.852	.887	.934	.973	.990
level - $\log(\psi_A)$	.031	.030	.033	.040	.044	.047
$A = 1_{(-\pi,\pi)}$	$n = 100$	$n = 150$	$n = 250$	$n = 500$	$n = 1000$	$n = 5000$
mean - ψ_A	.334	.150	.100	.040	.033	.020
sd - ψ_A	1.918	1.352	1.154	1.053	1.031	1.018
level - ψ_A	.101	.078	.068	.058	.054	.053
mean - $\log(\psi_A)$	.046	−.059	−.066	−.080	−.059	−.021
sd - $\log(\psi_A)$	1.062	.965	.964	.971	.989	1.009
level - $\log(\psi_A)$	.043	.037	.038	.043	.049	.052

The simulations consisted of 10,000 replications using the nominal α-level $\alpha = .05$. The top panel, under each model specification, illustrates the use of a band-limited kernel A of the form $1 + \cos(\lambda)$, the Tukey–Hanning kernel (TH), centered at $\omega_0 = \pi/6$ having bandwidth $\delta = \pi/3$. The bottom panel, under each model specification, uses an identity kernel on the interval $(-\pi, \pi)$. Note the innovation variance is taken equal to 1.

sample sizes under consideration, though, in most cases, they were slightly undersized. Furthermore, the distribution of both test statistics under the null hypothesis approached the distribution of a standard normal random variable, confirming our theoretical asymptotic results. However, in small sample sizes (i.e., $n < 1000$) the distribution of the ψ_A test statistic was skewed right, even though the means and standard deviations were at the approximate $(0, 1)$ level; see Table 1. Under these circumstances we found that the log transformation helped to alleviate this problem.

In general, for the cases we investigated, the power was excellent (Table 2). The one exception was for $\log(\psi_A)$ under the $ARMA(1, 1)$ alternative for sample sizes less than or equal to 150. These results are to be expected, since the class of models chosen under the alternative cannot achieve the spectral shape experienced under the null hypothesis. If instead of choosing an $AR(1)$ or ARMA$(1, 1)$ alternative we chose an $AR(3)$ alternative, then the power would be very low. Again this is reasonable since any $AR(2)$ model can be perfectly fit by the class of $AR(3)$ models by taking the last AR coefficient equal to zero. Essentially this would yield a case where the local spectral mass is in close agreement between the models under the null and alternative hypotheses, and thus the test would have diminished power.

Further, it is important to bear in mind that the power performance of our diagnostics is linked to the choice of bandwidth. Specifically, it is possible for the practitioner to take too local/global a perspective and thus exclude/include spectral frequencies of interest resulting in a loss of power. In summary, even for sample size as small as $n = 100$, the power of the tests (using a nominal α-level of .05) were excellent under the gof statistic ψ_A, with less favorable results under $\log(\psi_A)$ depending on the alternative. Although this simulation study is limited, it clearly demonstrates the efficacy of our approach even in small samples.

4.2. Application: Spectral peak identification

For reasons of space we only consider the first application of spectral peak identification. We consider the time series of Annual Crude Oil Prices from 1861–1999 measured in money of 1999 (see http://www.eia.doe.gov/emeu/international/contents.html for more information); this series will be referred to as the Oil series.

Now the data do not exhibit a nonstationary trend, and ACF and $PACF$ plots indicate that a low-order AR model may be adequate. We considered various $AR(p)$ models with $1 \le p \le 9$, fitted using the maximum-likelihood method, and assessed

Table 2
Results of a Monte Carlo simulation to evaluate the power of the gof statistic ψ_A and its log-transformed counterpart

Power $AR(2)$ cycle model				
$\mathrm{H}_a = AR(1)$	$\omega_0 = .3069\pi\,;\,\delta = .6\pi$		$\omega_0 = .3069\pi\,;\,\delta = .4\pi$	
Sample size	ψ_A	$\log(\psi_A)$	ψ_A	$\log(\psi_A)$
$n = 100$	.985	.780	.979	.790
$n = 150$	.999	.907	.997	.900
$n = 250$	1	.983	1	.979
$n = 500$	1	1	1	1
$\mathrm{H}_a = ARMA(1, 1)$	$\omega_0 = .3069\pi\,;\,\delta = .6\pi$		$\omega_0 = .3069\pi\,;\,\delta = .4\pi$	
Sample size	ψ_A	$\log(\psi_A)$	ψ_A	$\log(\psi_A)$
$n = 100$	.657	.359	.694	.370
$n = 150$	.823	.477	.857	.493
$n = 250$	.958	.657	.968	.673
$n = 500$	1	.884	1	.892

The simulations consisted of 10,000 replications using the nominal α-level $\alpha = .05$. Note that we used band-limited kernel A of the form $1 + \cos(\lambda)$, the Tukey–Hanning kernel (TH). Additionally, in this simulation the cycle model parameters are $\rho = .75$, $\omega = .3069\pi$ and $\tau^2 = 1$.

Table 3
This table contains the models along with the associated p-values for the Oil data analysis described in Section 4.2

Oil data analysis								
Test statistic	ψ_A				$\log(\psi_A)$			
Model	$AR(1)$	$AR(2)$	$AR(3)$	$AR(4)$	$AR(1)$	$AR(2)$	$AR(3)$	$AR(4)$
$A = 1_{(-\pi,\pi)}$	.030	.128	.564	.827	.067	.232	.715	.622
$A =$ TH kernel; $\delta = .6\pi$	<.001	<.001	.148	.700	.004	.025	.3527	.986
$A =$ TH kernel; $\delta = .4\pi$	<.001	<.001	.056	.764	.003	.027	.330	.939
$A =$ TH kernel; $\delta = .2\pi$	<.001	<.001	.006	.997	.004	.042	.518	.284

Note that we used band-limited kernel A of the form $1 + \cos(\lambda)$, the Tukey–Hanning kernel (TH). Additionally, in this analysis the kernel was centered at $\omega_0 = .3069\pi$ in all cases where a band-limited kernel was used.

their goodness-of-fit via *AICc* and Ljung–Box [15] statistics ($p = 9$ was chosen as a threshold, since the *AICc* (see Hurvich and Tsai [13]) values were much higher for *AR* models having more than this number of parameters). Note that although the exact model specifications are not provided here they are available upon request from the first author. The $AR(4)$ was the preferred model according to *AICc* (among models not deemed inadequate according to the Ljung–Box statistics). In fact, the Ljung–Box statistics rejected all models but the $AR(3)$ and $AR(4)$. Both of these models contain a minor peak in the spectrum in the appropriate "cycle band". Given that stochastic cycles have a period between 4 and 10 years (see the discussion in [11]), the cycle band for annual data consists of those frequencies between $\pi/2$ and $\pi/5$. In both the $AR(3)$ and $AR(4)$ models, there is a pair of complex conjugate roots with frequency $.443\pi$ and $.358\pi$ respectively; these frequencies give approximate locations for the spectral peaks, and are in the right region for a cycle. The $AR(1)$ and $AR(2)$ models fail to capture the cycle dynamics in the data, as they do not have any complex conjugate roots.

Hence in running our diagnostics, we center the kernel A in the cycle band $(\pi/5, \pi/2)$ using centering frequency $\omega_0 = .3069\pi$ with bandwidths $.6\pi$, $.4\pi$ and $.2\pi$. For this analysis we used a TH kernel ([17]) as well as the identity kernel on $(-\pi, \pi)$. As expected, if we take a more refined perspective (a smaller bandwidth) the local properties in the neighborhood of the spectral mode become more salient (Table 3). In fact, the only model that is deemed appropriate, using the untransformed statistic under the identity kernel is the $AR(1)$ model. In contrast, if we take a weighting kernel centered at the peak frequency with narrow bandwidth (i.e., $.2\pi$) then the only models deemed acceptable, using the log-transformed diagnostic, are the $AR(3)$ and $AR(4)$, while using the ψ_A diagnostics only the $AR(4)$ model is deemed acceptable. These results should be contrasted with the Ljung–Box diagnostics, which consist of a series of p-values at various lags (results available upon request). For the latter two models ($AR(3)$ and $AR(4)$), all the Ljung–Box statistics are well above the .05 level, indicating adequacy. Thus, the proper diagnostic in this case is one that focuses on the locality of the postulated peak.

5. Conclusion

This paper treats band-limited gof testing using a quadratic functional of the periodogram. Theorems 1 and 2 together provide a complete asymptotic theory (under both null and alternative hypotheses) for these types of statistics, taking parameter uncertainty into account in the asymptotic variance. We develop several applications, such as peak testing and UC testing, and provide simulations documenting the level and power properties. The performance in small samples is fairly good in comparison to similar procedures (cf. Eichler [9]).

A limitation of our method is the regularity conditions required by the theory, in particular near-Gaussianity of the data. We would not expect these techniques to work well in the context of a high degree of non-normality. Moreover, the practitioner must choose a band-limited kernel A, upon which the power of the test will be sensitive. Using the identity kernel weights all frequencies equally, while restricting to a frequency band may generate increased power against certain alternatives.

Our simulation studies and data analysis were necessarily limited, and future studies will focus on expanding these empirical results to testing for UCs and weighting by a specific alternative model. Additionally, many other types of statistics may be considered by applying Theorem 2—for example, the signal extraction diagnostics of McElroy [16] easily fall under this scope. Another potential application is to compare gof tests for two fitted models in the spirit of Rivers and Vuong [22].

In summary, our gof statistic ψ_A is a flexible addition to more traditional time-domain diagnostics, such as the Ljung–Box [15] statistics. Specifically, our approach allows the modeler the ability to focus on particular frequency bands of interest, and so the theory and methods of this paper can be fruitfully adapted to many different applications.

Acknowledgments

This work has greatly benefited from comments from the editor and an anonymous referee, as well as discussions with David Findley and Agustin Maravall. Holan's research was supported by an ASA/NSF/BLS research fellowship.

Disclaimer

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical, methodological, technical, and operational issues are those of the authors and not necessarily those of the US Census Bureau.

Appendix

A.1. Proofs

Proof of Theorem 1. The technique of proof is a simple adaption of the proof of Theorem 3 of Chiu [6], using the material from Brillinger [3]. We use the Cramér–Wold device: take scalars $\alpha_1, \ldots, \alpha_L$ and consider

$$\frac{1}{\sqrt{n}}\left(\sum_\lambda \sum_{i=1}^L \alpha_i g_{\theta,i}(\lambda) I^{j_i}(\lambda) - \sum_\lambda \sum_{i=1}^L j_i! \alpha_i g_{\theta,i}(\lambda) \tilde{f}^{j_i}(\lambda)\right).$$

For simplicity let $\phi_i(\lambda) = \alpha_i g_{\theta,i}(\lambda)$, so that we consider

$$\frac{1}{n}\sum_\lambda \sum_{i=1}^L \phi_i(\lambda) I^{j_i}(\lambda), \tag{A.1}$$

appropriately centered. From the proof of Theorem 2 of Chiu [6], this centering is $\frac{1}{n}\sum_\lambda \sum_{i=1}^L j_i! \phi_i(\lambda)\tilde{f}^{j_i}(\lambda)$. We will generalize Theorem 5.10.2 of Brillinger [3] to higher powers of the periodogram. Define the discrete Fourier transform of the data at Fourier frequencies λ 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.1) is given by

$$n^{-1}\sum_{\lambda_1}\sum_{\lambda_2}\sum_{i,k=1}^L \phi_i(\lambda_1)\phi_k(\lambda_2)\mathrm{cum}(I^{j_i}(\lambda_1), I^{j_k}(\lambda_2)).$$

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

$$n^{-(j_i+j_k+1)}\sum_\nu \sum_{\lambda_1}\sum_{\lambda_2} \phi_i(\lambda_1)\phi_k(\lambda_2)\mathrm{cum}\{d(\omega_{lm}); lm \in \nu_1\}\cdots \mathrm{cum}\{d(\omega_{lm}); lm \in \nu_q\},$$

where $\omega_{lm} = (-1)^m \lambda_l$ and the summation in ν is over all indecomposable partitions of the following table (see Brillinger [3])

$$(1,1)\cdots(1,2j_k)\cdots(1,2j_i)$$
$$(2,1)\cdots(2,2j_k).$$

This result is obtained by applying Theorem 2.3.2 of Brillinger [3] to

$$I^{j_i}(\lambda_1) = n^{-j_i}\,(d(-\lambda_1)\cdot d(\lambda_1))^{j_i}$$
$$I^{j_k}(\lambda_2) = n^{-j_k}\,(d(-\lambda_2)\cdot d(\lambda_2))^{j_k}.$$

Now our task is to determine which indecomposable partitions ν 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 ν 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 ν 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), $j_k - 1$ non-straddling 2-sets (contained in the first row) and another $j_i - 1$ non-straddling 2-sets (contained in the second row); or there are $j_k + j_i$ 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 [3], which requires condition 3. 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 $j_i + j_k - 1$ powers of n—then the overall exponent of n will be -2, which will counteract the growth in the double sum over λ_1 and λ_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 $j_i + j_k$ powers of n. Now according to (4.3.15) of Brillinger [3], for a particular p-set in a partition ν, the function Δ is evaluated at the sum of the ω_{lm}'s such that (l, m) are in that p-set. Moreover, Δ 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 ω_{lm}'s sum to zero. We refer to $\sum_{(l,m)\in B} \omega_{lm}$ as the ω-sum of the p-set B. For visualization, it is helpful to write out the table of ω_{lm}'s corresponding to the table given above:

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

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

4-set, 2-set partitions. Now for this type of partition, the ω-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 λ_1, λ_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 $j_i + j_k$, so that the overall exponent of n is -2; since a single summation in λ is only order n, it is asymptotically negligible. Hence the ω-sum for each of the 2-sets is only zero if they do not straddle, i.e., they are contained in a row. For those 2-sets in the first row, they consist of exactly one choice of λ_1 and one choice of $-\lambda_1$; for 2-sets in the second row, they consist of exactly one choice of λ_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 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 λ_1 in position $(1, 2)$ in the table. There are j_i choices of the element $-\lambda_1$ that it can form a 2-set with, such that the ω-sum is zero. Moving on to the second such element in position $(1, 4)$, there are now $j_i - 1$ such choices. Proceeding in this fashion, we obtain $j_i!$ such 2-set configurations. Independently, we pair up λ_2 with elements $-\lambda_2$ in the second row, and obtain $j_k!$ 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 $j_i j_k$ ways of doing this (j_i 2-set choices for the first row, and j_k 2-set choices for the second row). Therefore, the number of 4-set, 2-set partitions is $j_i j_k j_i! j_k!$.

Next, we see from (4.3.15) of Brillinger [3] that each of these partitions yields the same contribution to the variance, namely

$$G^X(\lambda_1, -\lambda_1, \lambda_2)\widetilde{f}^{j_i-1}(\lambda_1)\widetilde{f}^{j_k-1}(\lambda_2).$$

Note that the 2π factors do not appear, since we define our cumulant spectral densities without this normalization, which differs from Brillinger [3]. Combining with the ϕ_j's and replacing the Riemann sum by an integral (which is valid asymptotically, because these integrands are deterministic) yields

$$\frac{j_i j_k j_i! j_k!}{(2\pi)^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi_i(\lambda)\phi_k(\omega) G^X(\lambda, -\lambda, \omega)\widetilde{f}^{j_i-1}(\lambda)\widetilde{f}^{j_k-1}(\omega) d\lambda d\omega.$$

2-set partitions. Again we must have the ω-sum of each 2-set to be zero, but since there are $j_i + j_k$ 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 λ_1 is paired with a λ_2 or a $-\lambda_1$ in every 2-set, or λ_1 is paired with a $-\lambda_2$ or a $-\lambda_1$ in every 2-set. In other words, the first case stipulates that no λ_1 and $-\lambda_2$ are together in a 2-set. Focusing on this case, if a given 2-set contains a λ_1 and a $-\lambda_1$, then the corresponding ω-sum is zero; likewise for 2-sets containing a λ_2 and a $-\lambda_2$. However if the elements are λ_1 and λ_2, the ω-sum is only zero if $\lambda_1 = -\lambda_2$, which essentially stipulates a condition on the double sum. Since all the ω-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 $j_i + j_k$ 2-sets, the overall exponent of n is -1, which balances the single sum. The number of $\widetilde{f}(\lambda_1)$ and $\widetilde{f}(\lambda_2)$ terms is difficult in principal to determine, but since $\lambda = -\lambda_2$ and $\widetilde{f}$ is even, we are only concerned with the total number of such terms, which is $j_i + j_k$.

On the other hand, if no λ_1 and λ_2 can be in the same 2-set, we obtain a zero ω-sum for 2-sets containing a λ_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

$$n^{-1} \sum_{\lambda_1} \left(\phi_i(\lambda_1)\phi_k(-\lambda_1)\widetilde{f}^{j_i+j_k}(\lambda_1) + \phi_i(\lambda_1)\phi_k(\lambda_1)\widetilde{f}^{j_i+j_k}(\lambda_1) \right).$$

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 λ_1 element in the $(1, 2)$ location of the table, there are j_i choices of $-\lambda_1$ to pair with, and j_k choices of λ_2, so $j_i + j_k$ choices total. For the second λ_1, there is one less $-\lambda_1$ or one less λ_2, for a total $j_i + j_k - 1$ remaining choices. All together, we find mates for the λ_1 elements in $(j_i+j_k)(j_i+j_k-1)\cdots(j_k+1)$ ways. Now consider the first $-\lambda_2$ element (none of the λ_2 elements have yet been paired). It may only pair with λ_2 or $-\lambda_1$, of which in total there are only j_k remaining choices. Proceeding, we obtain $j_k!$ choices of mates for the various $-\lambda_2$, and so have $(j_i + j_k)!$ 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 $j_i!j_k!$ such decomposable partitions, thus our summary count is $(j_i + j_k)! - j_i!j_k!$. Now replacing the Riemann sum by an integral, we have a variance contribution of

$$\frac{(j_i + j_k)! - j_i!j_k!}{2\pi} \int_{-\pi}^{\pi} (\phi_i(\lambda)\phi_k(-\lambda) + \phi_i(\lambda)\phi_k(\lambda))\widetilde{f}^{j_i+j_k}(\lambda)\mathrm{d}\lambda.$$

All together, the asymptotic variance of $\sqrt{n}$ times (A.1) yields V given by

$$\begin{aligned} V = \sum_{k,l=1}^{L} & \frac{(j_k + j_l)! - j_k!j_l!}{2\pi} \int_{-\pi}^{\pi} (\phi_k(\lambda)\phi_l(-\lambda) + \phi_k(\lambda)\phi_l(\lambda))\tilde{f}^{j_k+j_l}(\lambda)\mathrm{d}\lambda \\ & + \frac{j_k j_k! j_l j_l!}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \phi_k(\lambda)\phi_l(\omega)G^X(\lambda, -\lambda, \omega)\tilde{f}^{j_k-1}(\lambda)\tilde{f}^{j_l-1}(\omega)\mathrm{d}\lambda\mathrm{d}\omega. \end{aligned} \tag{A.2}$$

Finally, we must consider the higher-order cumulants, and show that they always tend to zero as $n \to \infty$. Consider the hth cumulant of (A.1), which yields

$$\begin{aligned} & n^{-h} \sum_{\lambda_1} \cdot \sum_{\lambda_h} \sum_{i_1,\ldots,i_h}^{L} \phi_{i_1}(\lambda_1)\cdots\phi_{i_h}(\lambda_h)\mathrm{cum}\{I^{j_{i_1}}(\lambda_1), \ldots, I^{j_{i_h}}(\lambda_h)\} \\ & = n^{-h-r} \sum_{\lambda_1} \cdot \sum_{\lambda_h} \sum_{i_1,\ldots,i_h}^{L} \phi_{i_1}(\lambda_1)\cdots\phi_{i_h}(\lambda_h) \sum_{\nu} \mathrm{cum}\{d(\omega_{lm}); lm \in \nu_1\}\cdots\mathrm{cum}\{d(\omega_{lm}); lm \in \nu_q\}, \end{aligned}$$

where $r = j_{i_1} + \cdots + j_{i_h}$ and the summation in ν is over all indecomposable partitions of the table

$$\begin{array}{ccc} (1, 1) & \cdots & (1, 2j_{i_1}) \\ (2, 1) & \cdots & (2, 2j_{i_2}) \\ & \vdots & \\ (h, 1) & \cdots & (h, 2j_{i_h}). \end{array}$$

We seek the dominant term in the above cumulant. If we consider an indecomposable partition ν of the above table, many of the same principles apply from our variance analysis. In particular, we need not consider p-sets in ν with p odd. And the greatest number of factors of n are produced from Δ evaluated at ω-sums, if we were to take ν to be a partition consisting solely of 2-sets, where each λ_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 λ-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 λ sums. This leaves an overall order for (A.1) 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 (A.2). This establishes the joint asymptotic normality, and the asymptotic covariance matrix is obtained as follows. With e_j denoting the jth unit vector in $\mathbb{R}^L$ and $\alpha' = (\alpha_1, \ldots, \alpha_L)$, we find that the variance $V_{kl}(\theta)$ in the statement of Theorem 1 is given by $(V(e_k + e_l) - V(e_k) - V(e_l))/2$, where $V(\alpha)$ is given by (A.2), recalling that $\phi_i = \alpha_i g_{\theta,i}$. $\square$

Proof of Theorem 2. For each i we have

$$Q_n(I^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta}) = \left(Q_n(I^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(\tilde{f}^{j_i}, g_i, \hat{\theta})\right) + \left(j_i! Q_n(\tilde{f}^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta})\right). \tag{A.3}$$

The first term on the right-hand side above is asymptotically normal under certain conditions on g_i. The second term in (A.3) is also asymptotically normal (under H_0), and correlated with the first term. The first term can be written as

$$Q_n(I^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(\tilde{f}^{j_i}, g_i, \hat{\theta}) = \frac{1}{n} \sum_{\lambda} g_{\hat{\theta},i}(\lambda) \left(I^{j_i}(\lambda) - j_i! \tilde{f}^{j_i}(\lambda)\right).$$

Expanding $g_{\hat{\theta},i}$ about $\tilde{\theta}$ (use condition 8) yields

$$\begin{aligned} g_{\hat{\theta},i}(\lambda) &= g_{\tilde{\theta},i}(\lambda) + \nabla'_\theta g_{\tilde{\theta},i}(\lambda)(\hat{\theta} - \tilde{\theta}) + \frac{1}{2}(\hat{\theta} - \tilde{\theta})' H_\theta g_{\dot{\theta},i}(\lambda)(\hat{\theta} - \tilde{\theta}) \\ \nabla_\theta g_{\tilde{\theta},i}(\lambda) &= \nabla_\theta g_{\theta,i}(\lambda)|_{\theta=\tilde{\theta}} \\ [H_\theta g_{\dot{\theta},i}(\lambda)]_{kl} &= \frac{\partial}{\partial \theta_k} \frac{\partial}{\partial \theta_l} g_{\theta,i}(\lambda)|_{\theta=\dot{\theta}}, \end{aligned}$$

where each component of $\dot{\theta}$ lies in-between (condition 5) the respective components of $\hat{\theta}$ and $\tilde{\theta}$. Since $\hat{\theta} - \tilde{\theta} = O_P(n^{-1/2})$ (Theorem 3.1.2 of Taniguchi and Kakizawa [23]; use conditions 6 and 9, and also condition 2 which implies the needed Hosoya–Taniguchi conditions [12]) and $H_\theta g_{\dot{\theta},i}(\lambda) \xrightarrow{P} H_\theta g_{\tilde{\theta},i}(\lambda)$ by the smoothness of $g_{\theta,i}$ (condition 8),

$$g_{\hat{\theta},i}(\lambda) = g_{\tilde{\theta},i}(\lambda) + (\hat{\theta} - \tilde{\theta})' \nabla_\theta g_{\tilde{\theta},i}(\lambda) + O_P(n^{-1})$$

uniformly in λ. Hence

$$\begin{aligned} Q_n(I^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(\tilde{f}^{j_i}, g_i, \hat{\theta}) &= \frac{1}{n} \sum_{\lambda} g_{\tilde{\theta},i}(\lambda)(I^{j_i}(\lambda) - j_i! \tilde{f}^{j_i}(\lambda)) \\ &\quad + (\hat{\theta} - \tilde{\theta})' \frac{1}{n} \sum_{\lambda} \nabla_\theta g_{\tilde{\theta},i}(\lambda)(I^{j_i}(\lambda) - j_i! \tilde{f}^{j_i}(\lambda)) + O_P(n^{-1}). \end{aligned}$$

Now applying Theorem 1 (conditions 1, 3, 8), we see that the second term on the right-hand side is $O_P(n^{-1})$ as well. Thus

$$\sqrt{n}\left(Q_n(I^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(\tilde{f}^{j_i}, g_i, \hat{\theta})\right) = \frac{1}{\sqrt{n}} \sum_{\lambda} g_{\tilde{\theta},i}(\lambda)(I^{j_i}(\lambda) - j_i! \tilde{f}^{j_i}(\lambda)) + O_P(n^{-1/2}).$$

Now for the second term of (A.3), which can be written as

$$j_i! Q_n(\tilde{f}^{j_i}, g_i, \hat{\theta}) - j_i! Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta}) = -\frac{j_i!}{n} \sum_{\lambda} g_{\hat{\theta},i}(\lambda) \left[\left(f_{\tilde{\theta}}^{j_i}(\lambda) - \tilde{f}^{j_i}(\lambda)\right) + \left(f_{\hat{\theta}}^{j_i}(\lambda) - f_{\tilde{\theta}}^{j_i}(\lambda)\right)\right].$$

We expand $g_{\hat{\theta},i}$ as before, and also we have (condition 7)

$$f_{\hat{\theta}}^{j_i}(\lambda) = f_{\tilde{\theta}}^{j_i}(\lambda) + j_i f_{\tilde{\theta}}^{j_i - 1}(\lambda)(\hat{\theta} - \tilde{\theta})' \nabla_\theta f_{\tilde{\theta}}(\lambda) + O_P(n^{-1})$$

by the smoothness of f_θ. Hence we obtain

$$\begin{aligned} \sqrt{n} j_i! \left(Q_n(\tilde{f}^{j_i}, g_i, \hat{\theta}) - Q_n(f_{\hat{\theta}}^{j_i}, g_i, \hat{\theta})\right) &= -\frac{j_i!}{\sqrt{n}} \sum_{\lambda} \left(g_{\tilde{\theta},i}(\lambda) + (\hat{\theta} - \tilde{\theta})' \nabla_\theta g_{\tilde{\theta},i}(\lambda)\right) \\ &\quad \times \left(f_{\tilde{\theta}}^{j_i}(\lambda) - \tilde{f}^{j_i}(\lambda) + j_i f_{\tilde{\theta}}^{j_i - 1}(\lambda)(\hat{\theta} - \tilde{\theta})' \nabla_\theta f_{\tilde{\theta}}(\lambda)\right) + o_P(1) \\ &= -j_i! \sqrt{n}(\hat{\theta} - \tilde{\theta})' \frac{1}{n} \sum_{\lambda} \left[(f_{\tilde{\theta}}^{j_i}(\lambda) - \tilde{f}^{j_i}(\lambda)) \nabla_\theta g_{\tilde{\theta},i}(\lambda) + j_i g_{\tilde{\theta},i}(\lambda) f_{\tilde{\theta}}^{j_i - 1}(\lambda) \nabla_\theta f_{\tilde{\theta}}(\lambda)\right] \\ &\quad - \sqrt{n} \frac{j_i!}{n} \sum_{\lambda} g_{\tilde{\theta},i}(\lambda)(f_{\tilde{\theta}}^{j_i}(\lambda) - \tilde{f}^{j_i}(\lambda)) + o_P(1). \end{aligned}$$

The second term is a deterministic bias, which is zero under H_0; therefore it must be subtracted off in order to obtain asymptotic normality. This quantity is just (5), up to the factor $\sqrt{n} j_i!$. The quantity multiplying the parameter estimation error is deterministic, and is a Riemann sum approximation of $b_i(\tilde{\theta})$. So by utilizing the arguments in the proof of Theorem 3.1.2 of Taniguchi and Kakizawa [23], we have

$$\sqrt{n}\left(\hat{\theta}-\tilde{\theta}\right) = M_f^{-1}(\tilde{\theta})\frac{1}{\sqrt{n}}\sum_{\lambda}\nabla_\theta f_{\tilde{\theta}}(\lambda)(I(\lambda)-\tilde{f}(\lambda))f_{\tilde{\theta}}^{-2}(\lambda)+o_P(1).$$

Therefore we obtain

$$\begin{aligned}\sqrt{n}j_i!\left(Q_n(\tilde{f}^{j_i},g_i,\hat{\theta})-Q_n(f_{\hat{\theta}}^{j_i},g_i,\hat{\theta})\right) &= -j_i!b_i'(\tilde{\theta})M_f^{-1}(\tilde{\theta})\frac{1}{\sqrt{n}}\sum_{\lambda}\nabla_\theta f_{\tilde{\theta}}(\lambda)(I(\lambda)-\tilde{f}(\lambda))f_{\tilde{\theta}}^{-2}(\lambda)\\ &\quad -j_i!\sqrt{n}\frac{1}{2\pi}\int_{-\pi}^{\pi}g_{\tilde{\theta},i}(\lambda)(f_{\tilde{\theta}}^{j_i}(\lambda)-\tilde{f}^{j_i}(\lambda))\mathrm{d}\lambda+o_P(1).\end{aligned}$$

We now prove the normality result using the Cramér–Wold device; first we dispense with the asymptotic bias term by subtracting it off—essentially we suppose that this term is zero (as if H_0 were true) so as not to burden the formulas. Letting $\alpha'=(\alpha_1,\ldots,\alpha_L)$ be a sequence of constants, we have

$$\begin{aligned}&\sum_{i=1}^{L}\alpha_i\sqrt{n}\left(Q_n(I^{j_i},g_i,\hat{\theta})-j_i!Q_n(f_{\hat{\theta}}^{j_i},g_i,\hat{\theta})\right)\\ &\quad=\frac{1}{\sqrt{n}}\sum_{\lambda}\left(\sum_{i=1}^{L}\alpha_i g_{\tilde{\theta},i}(\lambda)(I^{j_i}(\lambda)-j_i!\tilde{f}^{j_i}(\lambda))+\sum_{i=1}^{L}\alpha_i p_{\tilde{\theta},i}(\lambda)(I(\lambda)-\tilde{f}(\lambda))\right)+o_P(1)\\ &\quad=\frac{1}{\sqrt{n}}\sum_{\lambda}\sum_{i=1}^{L+1}\phi_i(\lambda)(I^{j_i}(\lambda)-j_i!\tilde{f}^{j_i}(\lambda))+o_P(1)\end{aligned}$$

with $j_{L+1}=1$ and $\phi_i(\lambda)=\alpha_i g_{\tilde{\theta},i}(\lambda)$ for $i=1,2,\ldots,L$, and $\phi_{L+1}(\lambda)=\sum_{i=1}^{L}\alpha_i p_{\tilde{\theta},i}(\lambda)$. We can now apply Theorem 1 to the final expression, and obtain asymptotic normality with variance $V(\alpha)$ given by

$$V(\alpha)=\sum_{k,l=1}^{L}\frac{(j_k+j_l)!-j_k!j_l!}{4\pi}\int_{-\pi}^{\pi}(\phi_k(\lambda)\phi_l(-\lambda)+\phi_k(-\lambda)\phi_l(\lambda)+2\phi_k(\lambda)\phi_l(\lambda))\tilde{f}^{j_k+j_l}(\lambda)\mathrm{d}\lambda.$$

Note the variance expression is simplified because condition 4 follows from condition 2. Hence the joint asymptotic normality result for $\sqrt{n}(Q_n(I^{j_i},g_i,\hat{\theta})-j_i!Q_n(f_{\hat{\theta}}^{j_i},g_i,\hat{\theta}))$ is proved, with asymptotic covariance matrix $W(\tilde{\theta})$ with entries

$$W_{k,l}(\tilde{\theta})=\frac{1}{2}\left(V(e_k+e_l)-V(e_k)-V(e_l)\right). \tag{A.4}$$

We compute these quantities next. If $\alpha=e_k$, then ϕ_j is zero unless $j=k$ or $j=L+1$, in which case it is $g_{\tilde{\theta},l}$ or $p_{\tilde{\theta},l}$ respectively. It follows that

$$\begin{aligned}V(e_k) &= \frac{(2j_k)!-j_k!^2}{2\pi}\int_{-\pi}^{\pi}(g_k(\lambda)g_k(-\lambda)+g_k^2(\lambda))\tilde{f}^{2j_k}(\lambda)\mathrm{d}\lambda+\frac{1}{2\pi}\int_{-\pi}^{\pi}(p_k(\lambda)p_k(-\lambda)+p_k^2(\lambda))\tilde{f}^2(\lambda)\mathrm{d}\lambda\\ &\quad+\frac{(j_k+1)!-j_k!}{2\pi}\int_{-\pi}^{\pi}(g_k(\lambda)p_k(-\lambda)+p_k(\lambda)g_k(-\lambda)+2g_k(\lambda)p_k(\lambda))\tilde{f}^{j_k+1}(\lambda)\mathrm{d}\lambda.\end{aligned}$$

Next (say $k\neq l$) if $\alpha=e_k+e_l$, then ϕ_j is zero unless $j=k,l,L+1$, in which case it equals $g_{\tilde{\theta},k}$, $g_{\tilde{\theta},l}$, or $p_{\tilde{\theta},k}+p_{\tilde{\theta},l}$ respectively. Then

$$\begin{aligned}V(e_k+e_l) &= \frac{(2j_k)!-j_k!^2}{2\pi}\int_{-\pi}^{\pi}(g_k(\lambda)g_k(-\lambda)+g_k^2(\lambda))\tilde{f}^{2j_k}(\lambda)\mathrm{d}\lambda\\ &\quad+\frac{(2j_l)!-j_l!^2}{2\pi}\int_{-\pi}^{\pi}(g_l(\lambda)g_l(-\lambda)+g_l^2(\lambda))\tilde{f}^{2j_l}(\lambda)\mathrm{d}\lambda\\ &\quad+\frac{1}{2\pi}\int_{-\pi}^{\pi}(p_k(\lambda)+p_l(\lambda))(p_k(-\lambda)+p_l(-\lambda))+(p_k(\lambda)+p_l(\lambda))^2\tilde{f}^2(\lambda)\mathrm{d}\lambda\\ &\quad+\frac{(j_k+j_l)!-j_k!j_l!}{2\pi}\int_{-\pi}^{\pi}(g_k(\lambda)g_l(-\lambda)+g_k(-\lambda)g_l(\lambda)+2g_k(\lambda)g_l(\lambda))\tilde{f}^{j_k+j_l}(\lambda)\mathrm{d}\lambda\\ &\quad+\frac{(j_k+1)!-j_k!}{2\pi}\int_{-\pi}^{\pi}(g_k(\lambda)(p_k(-\lambda)+p_l(-\lambda))+g_k(-\lambda)(p_k(\lambda)+p_l(\lambda))\\ &\quad+2g_k(\lambda)(p_k(\lambda)+p_l(\lambda)))\tilde{f}^{j_k+1}(\lambda)\mathrm{d}\lambda\end{aligned}$$

$$+\frac{(j_l+1)!-j_l!}{2\pi}\int_{-\pi}^{\pi}(g_l(\lambda)(p_k(-\lambda)+p_l(-\lambda))+g_l(-\lambda)(p_k(\lambda)+p_l(\lambda))$$
$$+2g_l(\lambda)(p_k(\lambda)+p_l(\lambda)))\tilde{f}^{j_l+1}(\lambda)\mathrm{d}\lambda.$$

Now applying (A.4) we obtain the stated formula for $W_{kl}(\tilde{\theta})$ when $k \neq l$. Of course $W_{kk}(\tilde{\theta}) = V(e_k)$, but the same formula covers this case too. Finally, when the weighting functions are even, we have $b_k(\theta) = (2\pi)^{-1}\int_{-\pi}^{\pi} r_{\theta,k}(\lambda)\nabla_\theta h_\theta(\lambda)\mathrm{d}\lambda$ and

$$\begin{aligned}W_{kl}(\tilde{\theta}) &= \frac{(j_k+j_l)!-j_k!j_l!}{\pi}\int_{-\pi}^{\pi} r_{\tilde{\theta},k}(\lambda)r_{\tilde{\theta},l}(\lambda)\mathrm{d}\lambda + \frac{(j_k+1)!-j_k!}{\pi}\int_{-\pi}^{\pi} r_{\tilde{\theta},k}(\lambda)\left(-j_lj_l!b_l'M_f^{-1}(\tilde{\theta})\nabla_\theta h_{\tilde{\theta}}(\lambda)\right)\mathrm{d}\lambda\\ &\quad+\frac{(j_l+1)!-j_l!}{\pi}\int_{-\pi}^{\pi} r_{\tilde{\theta},l}(\lambda)\left(-j_kj_k!b_k'M_f^{-1}(\tilde{\theta})\nabla_\theta h_{\tilde{\theta}}(\lambda)\right)\mathrm{d}\lambda\\ &\quad+\frac{1}{\pi}\int_{-\pi}^{\pi}\left(-j_lj_l!b_l'M_f^{-1}(\tilde{\theta})\nabla_\theta h_{\tilde{\theta}}(\lambda)\right)\left(-j_kj_k!b_k'M_f^{-1}(\tilde{\theta})\nabla_\theta h_{\tilde{\theta}}(\lambda)\right)\mathrm{d}\lambda.\end{aligned}$$

This simplifies to the stated expression, recognizing that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\nabla_\theta' h_{\tilde{\theta}}(\lambda)\nabla_\theta h_{\tilde{\theta}}(\lambda)\mathrm{d}\lambda = M_f(\tilde{\theta})$$

under H_0. This proves the theorem for QMLs; for the MLE case, we need only show that the difference between parameter estimates is $o_P(1/\sqrt{n})$—but this follows from Theorems 3.2 and 3.3 of Dahlhaus and Wefelmeyer [7] under the additional condition 10. Finally, we make some comments on the assumptions. Condition 4 is not really necessary, but for the purposes of variance estimation (and stating the result more simply), we have decided to leave off the contribution of fourth-order cumulants. Since a Gaussian assumption is needed for the MLE case anyway, and this also implies the Hosoya–Taniguchi conditions needed for Theorem 3.1.2 of Taniguchi and Kakizawa [23], we have proved the theorem under the more restrictive condition 2. It seems likely that the asymptotic results in the proof of Theorem 3.1.2 could be proved under condition 3 only, and then condition 2 could be relaxed to condition 4 in the QML case. Conditions 5 through 9 are fairly standard and are often satisfied in practice. □

Proof of Proposition 1. The Fisher information matrix is a continuous function of θ by our assumptions on f_θ, and likewise each entry of the inverse matrix is continuous in θ. The vector functions $b_k(\theta)$ are also continuous (by our assumptions on $g_{\theta,k}$), and so the result follows immediately. □

A.2. Implementation notes for ARMA models

First consider the case where the parametric family is an $AR(p)$, where p is fixed throughout. Then $\theta' = (\phi_1, \phi_2, \ldots, \phi_p, \theta_q)$, where $q = p+1$ and θ_q is the innovation variance. As usual we write $\Phi(B) = 1-\phi_1B-\phi_2B-\cdots-\phi_pB^p$ for the autoregressive polynomial. Then the spectral density for this $AR(p)$ is

$$f_\theta(\lambda) = \left|1-\phi_1\mathrm{e}^{-\mathrm{i}\lambda}-\phi_2\mathrm{e}^{-\mathrm{i}2\lambda}-\cdots-\phi_p\mathrm{e}^{-\mathrm{i}p\lambda}\right|^{-2}\theta_q.$$

We need to compute the gradient with respect to θ. Now the last derivative is just the innovation-free spectrum. In general,

$$\begin{aligned}\frac{\partial}{\partial\theta_j}f_\theta(\lambda) &= \frac{\mathrm{e}^{-\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})}f_\theta(\lambda)+\frac{\mathrm{e}^{\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{\mathrm{i}\lambda})}f_\theta(\lambda)\quad j=1,2,\ldots,p\\ \frac{\partial}{\partial\theta_q}f_\theta(\lambda) &= \frac{1}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})\Phi(\mathrm{e}^{\mathrm{i}\lambda})}.\end{aligned}$$

Starting with the calculation of $M_{f,H_0}(\theta)$, we divide the gradient by f_θ:

$$\begin{aligned}h_{\theta,j}(\lambda) &= \frac{\mathrm{e}^{-\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})}+\frac{\mathrm{e}^{\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{\mathrm{i}\lambda})}\quad j=1,2,\ldots,p\\ h_{\theta,q}(\lambda) &= \theta_q^{-1}.\end{aligned}$$

Now we observe that $(2\pi)^{-1}\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}\lambda h}/\Phi(\mathrm{e}^{-\mathrm{i}\lambda})\mathrm{d}\lambda$ is zero if $h<0$. A similar result holds for $\Phi(\mathrm{e}^{\mathrm{i}\lambda})$. Hence we can compute the entries of $M_{f,H_0}(\theta)$ as follows: the klth entry is the integral of the product of h_k and h_l. Using the above observations, we find that the qth row and column of $M_{f,H_0}(\theta)$ are both zero, except for the diagonal entry, which is θ_q^{-2}. As for the other entries, suppose $1 \le k, l \le p$. Then the matrix entry is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{\mathrm{e}^{-\mathrm{i}\lambda(k+l)}}{\Phi^2(\mathrm{e}^{-\mathrm{i}\lambda})}+\frac{\mathrm{e}^{\mathrm{i}\lambda(k+l)}}{\Phi^2(\mathrm{e}^{\mathrm{i}\lambda})}+\frac{\mathrm{e}^{\mathrm{i}\lambda(k-l)}}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})\Phi(\mathrm{e}^{\mathrm{i}\lambda})}+\frac{\mathrm{e}^{\mathrm{i}\lambda(l-k)}}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})\Phi(\mathrm{e}^{\mathrm{i}\lambda})}\right)\mathrm{d}\lambda = 2\gamma_{\tilde{f}_\theta}(k-l).$$

Here $\overline{f}_\theta = f_\theta/\theta_q$, the innovation-free spectral density. We note that software exists to easily compute these quantities rapidly from a knowledge of θ.

Next, consider $b_1(\theta)$ and $b_2(\theta)$. Since the index $i = 1$ corresponds to the squared periodogram ($j_1 = 2$), it follows that $g_{\theta,1} = A/f_\theta^2$. Also, $i = 2$ corresponds to $j_2 = 1$, so $g_{\theta,2} = A/f_\theta$. Thus $b_1(\theta) = (2\pi)^{-1}\int_{-\pi}^{\pi} A(\lambda)h_\theta(\lambda)\mathrm{d}\lambda = b_2(\theta)$, where h_θ is a q-vector function with components $h_{\theta,j}$. Since b_1 and b_2 are identical in this case, call the vector b instead. Hence we can compute $b'(\theta)M_{f,H_0}^{-1}(\theta)$, plugging in MLEs for θ. Next, we can write simplified expressions for the entries of $W_{H_0}(\theta)$ under the assumption that A is even:

$$W_{11}(\theta) = 40\gamma_{A^2}(0) - 32b'(\tilde{\theta})M_f^{-1}(\tilde{\theta})b(\tilde{\theta})$$
$$W_{12}(\theta) = 8\gamma_{A^2}(0) - 8b'(\tilde{\theta})M_f^{-1}(\tilde{\theta})b(\tilde{\theta})$$
$$W_{22}(\theta) = 2\gamma_{A^2}(0) - 2b'(\tilde{\theta})M_f^{-1}(\tilde{\theta})b(\tilde{\theta}).$$

So computation of the quadratic form $b'(\theta)M_{f,H_0}^{-1}(\theta)b(\theta)$, together with $\gamma_{A^2}(0)$, produces these values. The variance of the ψ_A statistic is then

$$W_{11} - 4W_{12} + 4W_{22} = 16\gamma_{A^2}(0) - 8b'(\tilde{\theta})M_f^{-1}(\tilde{\theta})b(\tilde{\theta}).$$

Now consider the case where the parametric family is an $MA(r)$, where r is fixed throughout. Then $\theta' = (\theta_1, \theta_2, \ldots, \theta_r, \theta_q)$, where $q = r + 1$ and θ_q is the innovation variance. As usual we write $\Theta(B) = 1 + \theta_1 B + \theta_2 B + \cdots + \theta_r B^r$ for the moving average polynomial. Then the spectral density for this $MA(r)$ is

$$f_\theta(\lambda) = \Theta(\mathrm{e}^{-\mathrm{i}\lambda})\Theta(\mathrm{e}^{\mathrm{i}\lambda})\theta_q.$$

We need to compute the gradient with respect to θ. Now the last derivative is just the innovation-free spectrum. In general,

$$\frac{\partial}{\partial\theta_j}f_\theta(\lambda) = \mathrm{e}^{-\mathrm{i}\lambda j}\Theta(\mathrm{e}^{\mathrm{i}\lambda})\theta_q + \mathrm{e}^{\mathrm{i}\lambda j}\Theta(\mathrm{e}^{-\mathrm{i}\lambda})\theta_q \quad j = 1, 2, \ldots, r$$
$$\frac{\partial}{\partial\theta_q}f_\theta(\lambda) = \Theta(\mathrm{e}^{-\mathrm{i}\lambda})\Theta(\mathrm{e}^{\mathrm{i}\lambda}).$$

As with the $AR(p)$, we have

$$h_{\theta,j}(\lambda) = \frac{\mathrm{e}^{-\mathrm{i}\lambda j}}{\Theta(\mathrm{e}^{-\mathrm{i}\lambda})} + \frac{\mathrm{e}^{\mathrm{i}\lambda j}}{\Theta(\mathrm{e}^{\mathrm{i}\lambda})} \quad j = 1, 2, \ldots, r$$
$$h_{\theta,q}(\lambda) = \theta_q^{-1}.$$

This has the exact same form as the $AR(r)$ case, only with $\Theta(B)$ substituted for $\Phi(B)$. Therefore all the rest of the formulas are the identical, essentially substituting θ_k for $-\phi_k$ everywhere with $1 \le k \le r$.

Finally, consider the case where the parametric family is an $ARMA(p, r)$, where p and r are fixed throughout. Then $\theta' = (\phi_1, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_r, \theta_q)$, where $q = p + r + 1$ and θ_q is the innovation variance. We employ the notations from the AR and MA discussions. The spectral density is

$$f_\theta(\lambda) = \frac{\Theta(\mathrm{e}^{-\mathrm{i}\lambda})\Theta(\mathrm{e}^{\mathrm{i}\lambda})}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})\Phi(\mathrm{e}^{\mathrm{i}\lambda})}\theta_q.$$

We need to compute the gradient with respect to θ. Now the last derivative is just the innovation-free spectrum. For the other derivatives, we can combine the AR and MA results:

$$\frac{\partial}{\partial\theta_j}f_\theta(\lambda) = \frac{\mathrm{e}^{-\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})}f_\theta(\lambda) + \frac{\mathrm{e}^{\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{\mathrm{i}\lambda})}f_\theta(\lambda) \quad j = 1, 2, \ldots, p$$
$$\frac{\partial}{\partial\theta_j}f_\theta(\lambda) = \left(\mathrm{e}^{-\mathrm{i}\lambda(j-p)}\Theta(\mathrm{e}^{\mathrm{i}\lambda}) + \mathrm{e}^{\mathrm{i}\lambda(j-p)}\Theta(\mathrm{e}^{-\mathrm{i}\lambda})\right)\frac{\theta_q}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})\Phi(\mathrm{e}^{\mathrm{i}\lambda})} \quad j = p+1, \ldots, p+r$$
$$\frac{\partial}{\partial\theta_q}f_\theta(\lambda) = f_\theta(\lambda).$$

Then we obtain

$$h_{\theta,j}(\lambda) = \frac{\mathrm{e}^{-\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{-\mathrm{i}\lambda})} + \frac{\mathrm{e}^{\mathrm{i}\lambda j}}{\Phi(\mathrm{e}^{\mathrm{i}\lambda})} \quad j = 1, 2, \ldots, p$$
$$h_{\theta,j}(\lambda) = \frac{\mathrm{e}^{-\mathrm{i}\lambda(j-p)}}{\Theta(\mathrm{e}^{-\mathrm{i}\lambda})} + \frac{\mathrm{e}^{\mathrm{i}\lambda(j-p)}}{\Theta(\mathrm{e}^{\mathrm{i}\lambda})} \quad j = p+1, \ldots, p+r$$
$$h_{\theta,q}(\lambda) = \theta_q^{-1}.$$

Next we compute the entries of $M_{f,H_0}(\theta)$ as follows: the qth row and column of $M_{f,H_0}(\theta)$ are both zero, except for the diagonal entry, which is θ_q^{-2}. As for the other entries, we can divide into blocks. For the k, lth entry, if $1 \leq k, l \leq p$ then the entries correspond to the pure $AR(p)$ case matrix coefficients. But if $p+1 \leq k, l \leq p+r$ then the entries correspond to the pure $MA(r)$ case. Finally, if $1 \leq k \leq p$ and $p+1 \leq l \leq p+r$ (the same result holds with k and l swapped) then the entry is

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{e^{-i\lambda(k+l-p)}}{\Theta(e^{-i\lambda})\Phi(e^{-i\lambda})}+\frac{e^{i\lambda(k+l-p)}}{\Theta(e^{i\lambda})\Phi(e^{i\lambda})}+\frac{e^{i\lambda(k-l+p)}}{\Theta(e^{-i\lambda})\Phi(e^{i\lambda})}+\frac{e^{i\lambda(l-k-p)}}{\Phi(e^{-i\lambda})\Theta(e^{i\lambda})}\right)d\lambda = 2\gamma_v(k-l)$$

with $v(\lambda) = 1/(\Theta(e^{-i\lambda})\Phi(e^{i\lambda}))$. Now the calculations for $b(\theta)$ are the same as the pure AR and MA cases, only we now substitute the h_θ given above.

References

[1] J. Beran, Statistics for Long Memory Processes, Chapman and Hall, New York, 1994.
[2] P. Bickel, K. Doksum, Mathematical Statistics, Prentice Hall, New Jersey, 1977.
[3] D. Brillinger, Time Series Data Analysis and Theory, Holden-Day, San Francisco, 1981.
[4] W. Chen, R. Deo, On the integral of the squared periodogram, Stochastic Processes and their Applications 85 (2000) 159–176.
[5] W. Chen, R. Deo, A generalized portmanteau goodness-of-fit test for time series models, Econometric Theory 20 (2004) 382–416.
[6] S. Chiu, Weighted least squares estimators on the frequency domain for the parameters of a time series, The Annals of Statistics 16 (1988) 1315–1326.
[7] R. Dahlhaus, W. Wefelmeyer, Asymptotically optimal estimation in misspecified time series models, Annals of Statistics 24 (1996) 952–974.
[8] M. Eichler, A frequency-domain based test for non-correlation between stationary time series, Metrika 65 (2007) 133–157.
[9] M. Eichler, Testing nonparametric and semiparametric hypotheses in vector stationary processes, Journal of Multivariate Analysis 99 (2008) 968–1009.
[10] A. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1989.
[11] A. Harvey, T. Trimbur, General model-based filters for extracting cycles in economic time series, Review of Economics and Statistics 85 (2003) 244–255.
[12] Y. Hosoya, M. Taniguchi, A central limit theorem for stationary processes and the parameter estimation of linear processes, The Annals of Statistics 10 (1982) 132–153.
[13] C. Hurvich, C. Tsai, Regression and time series model selection in small samples, Biometrika 76 (1989) 297–307.
[14] W. Li, Diagnostic Checks in Time Series, CRC Press, 2004.
[15] G. Ljung, G. Box, On a measure of lack of fit in time series models, Biometrika 65 (1978) 297–303.
[16] T. McElroy, Statistical properties of model-based signal extraction diagnostic tests, Communications in Statistics, Theory and Methods 37 (2008) 591–616.
[17] T. McElroy, S. Holan, A nonparametric test for assessing spectral peaks, SRD Research Report No. 2005/10, US Census Bureau. http://www.census.gov/srd/papers/pdf/rrs2005-10.pdf. 2005.
[18] E. Paparoditis, Spectral density based goodness-of-fit tests for time series models, Scandinavian Journal of Statistics 27 (2000) 143–176.
[19] E. Paparoditis, Testing the fit of a vector autoregressive moving average model, Journal of Time Series Analysis 26 (2005) 543–568.
[20] J. Pitman, Some Basic Theory for Statistical Inference, Chapman and Hall, London, 1979.
[21] M. Priestley, Spectral Analysis and Time Series, Academic Press, London, 1981.
[22] D. Rivers, Q. Vuong, Model selection tests for nonlinear dynamic models, Econometrics Journal 5 (2002) 1–39.
[23] M. Taniguchi, Y. Kakizawa, Asymptotic Theory of Statistical Inference for Time Series, Springer-Verlag, New York City, New York, 2000.
[24] M. Taniguchi, M. Puri, M. Kondo, Nonparametric approach for non-Gaussian vector stationary processes, Journal of Multivariate Analysis 56 (1996) 259–283.