RESEARCH REPORT SERIES
(Statistics - #2006-6)

Statistical Properties of Model-Based Signal Extraction Diagnostic Tests

Tucker McElroy

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

Report Issued: June 5, 2006

Statistical Properties of Model-Based Signal Extraction Diagnostic Tests

Tucker McElroy
U.S. Census Bureau

Abstract

A model-based diagnostic test for signal extraction was first described in Maravall (2003), and this basic idea was modified and studied in Findley, McElroy, and Wills (2004). The paper at hand improves on the latter work in two ways: central limit theorems for the diagnostics are developed, and two hypothesis-testing paradigms for practical use are explicitly described. A further modified diagnostic provides an interpretation of one-sided rejection of the Null Hypothesis, yielding general notions of "over-modeling" and "under-modeling." The new methods are demonstrated on two U.S. Census Bureau time series exhibiting seasonality.

Keywords. ARIMA model, Seasonal adjustment, Filtering, Central limit theorem.

1 Introduction

The model-based approach to signal extraction, while elegant and optimal under certain conditions, is still in need of a suite of diagnostics capable of identifying the quality of the procedure. Certainly, model inadequacy – assessed for example through Ljung-Box statistics – will imply a poor signal estimate, but the inverse statement need not hold, i.e., goodness of model fit, as indicated through standard ARIMA model diagnostics, need not indicate the goodness of the corresponding signal extraction method. This is the case, because deviations of the data from the fitted ARIMA model, deemed harmless according to standard ARIMA goodness-of-fit measures, may cause serious problems from the perspective of accurately estimating an ambient signal, e.g., a seasonal or a trend. One reason for this phenomenon is that the maximum likelihood procedure for fitting ARIMA models typically fits the best model to the data, without regard to the postulated component models. This is the case even with a structural components approach (Harvey 1989), since poorness of fit of the component models is only assessed in a global sense. Thus, for some series, the quality of the signal extraction may be in doubt and can be assessed through various spectrum diagnostics, assuming a frequency-based characterization of signal and noise – see Findley, Monsell, Bell, Otto, and Chen (1998) and Soukup and Findley (1999).

In the seasonal adjustment program SEATS (Gómez and Maravall, 1997) a series of diagnostics, based on quantifying the variation in estimated signals, have been in use for several years, and have recently been documented in Maravall (2003) and associated with a statistical test. These model-based seasonal adjustment diagnostics for over- and under-smoothing of the seasonal component were later adapted for finite sample signal extraction in Findley, McElroy, and Wills (2004). The basic concept is to measure the variation of an estimated signal – assessed through a variance estimate of the appropriately "differenced" signal extraction – and compare this quantity to what we would expect if our model were true. Thus, extreme values of variation, relative to a benchmark computed from a hypothesized model, would indicate model inadequacy with respect to the component model for the desired signal. For example, an extreme diagnostic computed for the trend would indicate poor modelling of the low frequencies, since these constitute the spectral domain of trends.

Further, one may distinguish between intra- and inter-component variation. The former is concerned with the expected second-order structure of an estimated signal, measured through its auto-covariance function. The latter treats the expected variation across estimated components, measured through the cross-covariance function. When modified slightly, these diagnostics can be interpreted as weighted measures of model fit, placing more weight on discrepancies between model and truth that occur at frequencies pertinent to the signal of interest. With this spectral interpretation the diagnostics of Maravall (2003) seem to be a fairly natural measure.

Mathematically, the diagnostics can typically be viewed as a quadratic form in the differenced data. This simple structure facilitates a finite sample description in terms of mean and variance of the statistic, as well as the analysis of the asymptotic behavior. A full knowledge of the covariance structure of the signal error process is required, which can easily be obtained via a matrix-based approach to filtering (McElroy, 2005). This paper generalizes the results of Findley, McElroy, and Wills (2004), describing the statistical behavior with full rigor. First the background notation for signal extraction is developed, and two hypothesis testing schemas are described. In the next section we discuss the statistical properties of the diagnostics, as well as their applications. These methods are demonstrated on two time series in the following section. Proofs are contained in an Appendix.

1.1 Signal Extraction Notations

Since we wish to consider mean square optimal signal extraction from a finite sample, we follow the approach of McElroy (2005). Consider a nonstationary time series Y_t that can be written as the sum of two possibly nonstationary components S_t and N_t, the signal and the noise:

$$Y_t = S_t + N_t \tag{1}$$

Following Bell (1984), we let Y_t be an integrated process such that $W_t = \delta(B)Y_t$ is stationary, where B is the backshift operator and $\delta(z)$ is a polynomial with all roots located on the unit circle of the complex plane (also, $\delta(0) = 1$ by convention). This $\delta(B)$ is the differencing operator of the series, and we assume it can be factored into relatively prime polynomials $\delta^S(z)$ and $\delta^N(z)$ (i.e., polynomials with no common zero), such that

$$U_t = \delta^S(B)S_t \qquad V_t = \delta^N(B)N_t \tag{2}$$

are stationary time series. Note that included as special cases are $\delta^S = 1$ and/or $\delta^N = 1$, in which case either the signal or the noise or both are stationary. We let d be the order of δ, and d_S and d_N are the orders of δ^S and δ^N; since the latter operators are relatively prime, $\delta = \delta^S \cdot \delta^N$ and $d = d_S + d_N$.

For example, the noise could be a nonstationary seasonal with trend plus irregular signal (or nonseasonal), in which case $\delta^S(z)$ could be $(1-z)^2$, and the noise has differencing operator $\delta^N(z) = 1 + z + z^2 + \cdots z^{11}$ for monthly data. This is the appropriate setup for seasonal adjustment, in which case we are interested in estimating S_t.

As in Bell and Hillmer (1988), we assume Assumption A of Bell (1984) holds on the component decomposition, and we treat the case of a finite sample with $t = 1, 2, \cdots, n$. Assumption A states that the initial d values of Y_t, i.e., the variables $Y_1, Y_2, \cdots, Y_d$, are independent of $\{U_t\}$ and $\{V_t\}$. For a discussion of the implications of this assumption, see Bell (1984) and Bell and Hillmer (1988). A further assumption that we make is that $\{U_t\}$ and $\{V_t\}$ are uncorrelated time series.

Now we can write (2) in a matrix form, as follows. Let Δ be an $n - d \times n$ matrix with entries given by $\Delta_{ij} = \delta_{i-j+d}$ (the convention being that $\delta_k = 0$ if $k < 0$ or $k > d$). The matrices Δ_S and Δ_N have entries given by the coefficients of $\delta^S(z)$ and $\delta^N(z)$, but are $n - d_S \times n$ and $n - d_N \times n$ dimensional respectively. This means that each row of these matrices consists of the coefficients of the corresponding differencing polynomial, horizontally shifted in an appropriate fashion. Hence

$$W = \Delta Y \qquad U = \Delta_S S \qquad V = \Delta_N N$$

where Y is the transpose (denoted by Y') of $(Y_1, Y_2, \cdots, Y_n)$, and W, U, V, S, and N are also column vectors. It follows from the equation

$$W_t = \delta^N(B)U_t + \delta^S(B)V_t \tag{3}$$

that we need to define further differencing matrices $\underline{\Delta}_N$ and $\underline{\Delta}_S$ with row entries given by the coefficients of $\delta^N(z)$ and $\delta^S(z)$ respectively, which are $n-d \times n-d_S$ and $n-d \times n-d_N$ dimensional. Then we can write down the matrix version of (3):

$$W = \underline{\Delta}_N U + \underline{\Delta}_S V \tag{4}$$

We will be interested in estimates of U and V. The minimum mean squared error linear signal extraction estimate is $\hat{U}_t$, which can be expressed as some linear function of the differenced data vector W; putting this together for each time t, we obtain the various rows of a matrix F:

$$\hat{U} = FW.$$

We note that the various rows of F differ (unlike in the bi-infinite filtering case), since only a finite number of Y_t's are available to filter. The last row of F, for example, corresponds to the concurrent filter, i.e., a one-sided filter used to extract a signal at "time present."

For any random vector X, let Σ_X denote its covariance matrix. With these notations in hand, we can now state the signal extraction formulas, which are given in Proposition 1 of McElroy (2005):

$$\hat{U} = \Sigma_U \underline{\Delta}_N' \Sigma_W^{-1} W = FW \tag{5}$$

which implicitly defines F. Later on, it will be necessary to discuss spectra. For a stationary process $\{X_t\}$, we use $f_X(\lambda)$ to denote its spectral density; this is related to the autocovariance matrix Σ_X by the formula

$$[\Sigma_X]_{jk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_X(\lambda) e^{i(j-k)\lambda} \, d\lambda.$$

More generally, for any bounded positive symmetric function $g(\lambda)$, we define

$$[\Sigma(g)]_{jk} = \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) e^{i(j-k)\lambda} \, d\lambda. \tag{6}$$

This will be called the ACF matrix of g, and is always Toeplitz. Note our use of 2π, which differs from some authors.

1.2 Hypothesis Testing Framework

In this paper, we must make a distinction between a specified model for W – whose covariance matrix is denoted Σ_W – and the true covariance matrix for W, based on the true underlying

Data Generating Process (DGP) – denoted by $\tilde{\Sigma}_W$. The perspective is that a specified Σ_W – determined either via ad hoc principles or through maximum likelihood estimation – will differ from $\tilde{\Sigma}_W$. However, we assume that at least the differencing operators δ^S and δ^N have been correctly ascertained. Note that, denoting the true spectral density by $\tilde{f}_W$, $\tilde{\Sigma}_W = \Sigma(\tilde{f}_W)$.

Let us denote a particular choice of model – our Null model – by Σ_W, the model's covariance matrix under the Null Hypothesis. This Null Hypothesis is simply a particular choice of AR and MA polynomials that determine the $ARMA$ model for W_t. We further suppose that a specification of Σ_W in turn determines Σ_U and Σ_V, which will be the case if we obtain the models for S and N via a canonical decomposition (Hillmer and Tiao, 1982) of the model for Y. In this fashion, we can explore model inadequacy directly through the choice of Σ_W, without explicitly accounting for Σ_U and Σ_V; this also covers the approach of SEATS, which uses the canonical decomposition technique.

The alternative space consists of any other $ARMA$ model for W_t, including different polynomial orders, coefficients, and innovation variance. However, the differencing polynomials δ^S and δ^N are the same for both the Null and Alternative models. Note that Σ_W could in practice be determined by $ARMA$ coefficient parameter estimates, which we will then treat as fixed rather than random. This perspective is motivated by the difficulty of stipulating a random quantity for the Null model. So our testing framework is

$$\begin{aligned} H_0 &: \tilde{\Sigma}_W = \Sigma_W \\ H_1 &: \tilde{\Sigma}_W \neq \Sigma_W \end{aligned} \tag{7}$$

As a second testing paradigm, we consider the "innovation-free" version of (7), where the data's covariance matrix is essentially assumed to have unit innovation variance σ_a^2:

$$\begin{aligned} H_0^\dagger &: \tilde{\Sigma}_{W/\tilde{\sigma}_a} = \Sigma_{W/\sigma_a} \\ H_1^\dagger &: \tilde{\Sigma}_{W/\tilde{\sigma}_a} \neq \Sigma_{W/\sigma_a} \end{aligned} \tag{8}$$

This second paradigm is motivated by the observation that the signal extraction filters are determined completely by the $ARMA$ coefficients, and do not depend on the innovation variance. Note that for both testing paradigms, the alternative space has no particular directionality that is naturally associated with it. Thus, there is no basis for one-sided tests with these null and alternative hypotheses. We later argue that the spectrum provides an appropriate tool for determining directionality of rejection of H_0, in the context of estimating signals. The basic idea is, model inadequacy can be assessed in the context of signal extraction by measuring an estimated component's deviance from H_0 in an appropriate spectral range; this will allow for meaningful one-sided tests.

Findley, McElroy, and Wills (2004) presented a test statistic for any differenced signal U based on its sample variance. That work claimed asymptotic normality of the test statistic under H_0; in Section 2, this claim is verified under two different scenarios. We also expand the basic sample variance results to include sample autocovariances and crosscovariances. Section 3 discusses various applications of these asymptotic results to testing, power, and suitable interpretations for rejection and non-rejection of H_0. We consider two examples for which the diagnostic gives interesting results. Proofs are contained in an Appendix.

2 Theoretical Results

We present several test statistics, which can all be written as a symmetric quadratic form in the differenced data $W = (W_1, W_2, \cdots, W_{n-d})^{'}$. First we define two main examples of signal extraction diagnostics, and then we discuss their asymptotic behavior in a theorem. Next, we extend to the case that the data's innovation variance is estimated, and examine the asymptotics of a modified diagnostic for this scenario.

2.1 Autocovariance and Crosscovariance Diagnostics

The basic idea of SEATS's diagnostic for over- and under-adjustment is to measure the second-order properties of estimated signal and noise, and compare to what we should expect if our model is true. Then any gross disparities lead us to rejection of our model, and hence of our filtering as well. As a first step, consider the test statistic given by the sample second moment of $\hat{U}_t$, compared to its expectation under H_0; this quantity is not scale-invariant, so it is then normalized by its standard error under H_0. The formula is given by

$$\hat{A}_n = \frac{\hat{U}'\hat{U}}{n} = \frac{W'\Sigma_W^{-1}\underline{\Delta}_N\Sigma_U\Sigma_U\underline{\Delta}_N^{'}\Sigma_W^{-1}W}{n}, \tag{9}$$

which follows from (5). Note that the normalization of n does not match the length of $\hat{U}$, which is $n - d_S$. Later, we will normalize the diagnostics, so that the choice of n versus $n - d_S$ becomes irrelevant. One way to measure the sample autocovariance is to insert a lag matrix in the inner product above. Let L be an $n - d_S$ dimensional square matrix with $L_{ij} = 0$ unless $i = j + 1$ and $L_{j+1,j} = 1$. Then

$$\hat{A}_n(h) = \frac{\hat{U}'L^h\hat{U}}{n} = \frac{W'\Sigma_W^{-1}\underline{\Delta}_N\Sigma_U L^h\Sigma_U\underline{\Delta}_N^{'}\Sigma_W^{-1}W}{n} \tag{10}$$

for any $0 \leq h < n$ gives the lag h autocovariance estimate. It is necessary to symmetrize this matrix in theoretical formulas, so let

$$L^{sym,\,h} = \frac{1}{2}(L^h + {L'}^h).$$

Then the autocovariance estimate can be rewritten as

$$\hat{A}_n(h) = \frac{\hat{U}' L^{sym,h} \hat{U}}{n} = \frac{W'\Sigma_W^{-1} \underline{\Delta}_N \Sigma_U L^{sym,h} \Sigma_U \underline{\Delta}_N' \Sigma_W^{-1} W}{n}. \tag{11}$$

Notice that this is a symmetric quadratic form in the differenced data W. For the crosscovariance between $\hat{U}$ and $\hat{V}$, which are of length $n - d_S$ and $n - d_N$ respectively, it is necessary to trim the longer vector. Without loss of generality, suppose that $d_N < d_S$ so that $\hat{V}$ is longer. It will be shown later that the *first* values (as opposed to the last values) of the vector should be trimmed; this is achieved via the formula

$$[0 \quad 1_{n-d_S}]\, \hat{V} = [0 \quad 1_{n-d_S}]\, \Sigma_V \underline{\Delta}_S' \Sigma_W^{-1} W. \tag{12}$$

Here 1_{n-d_S} denotes the $n-d_S$ dimensional identity matrix, and 0 denotes a zero matrix of dimension $n - d_S \times d_S - d_N$. We then take the inner product with $\hat{U}$, inserting the lag matrix L^h:

$$\hat{C}_n(h) = \frac{W'\Sigma_W^{-1} \left(\underline{\Delta}_N \Sigma_U L^h [0\,1] \Sigma_V \underline{\Delta}_S' \right)^{sym} \Sigma_W^{-1} W}{n} \tag{13}$$

The $[0\,1]$ matrix has the same dimensions as in (12) above. Note that we have symmetrized the interior matrix. In this manner, the autocovariance and crosscovariance diagnostics are defined.

2.2 Fixed Parameters Case

Suppose that an $ARMA$ model for W_t is completely specified. When following the Hilmer and Tiao (1982) approach, it may be possible to obtain a canonical decomposition and thereby derive $ARMA$ models for U_t and V_t. Or if a structural approach is adopted (Harvey 1989), the models for U_t and V_t could be specified directly. Note that model-based filters do not depend on innovation variance (see McElroy 2005), but the mean squared errors of the estimates they produce do, and the mean and variance of the test statistics will as well. The following theorem summarizes the small sample properties and asymptotics of $\hat{A}_n(h)$ and $\hat{C}_n(h)$, and is appropriate for the testing paradigm described by (7). Since both of these statistics have the form $Q(W) = W' BW/n$ for some symmetric matrix B, we describe the results in terms of $Q(W)$. For the asymptotic results, the various spectral densities need to satisfy a certain smoothness condition. Writing $\gamma_g(h)$ for the hth coefficient of $g(\lambda)$ in the Fourier basis, we say that g is in the space $\mathcal{C}^1$ if

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

It is a standard fact that $\mathcal{C}^1 \subset C^1([-\pi, \pi])$, the space of once continuously differentiable functions on $[-\pi, \pi]$.

Theorem 1 *Assume that Assumption A holds on the component decomposition (1) and that $\{U_t\}$ and $\{V_t\}$ are independent. If the third and fourth cumulants of the true DGP of W_t are zero, then the true mean and variance of $Q(W)$ are given by*

$$\mathbb{E}Q(W) = \frac{1}{n}\, tr(B\tilde{\Sigma}_W) \tag{14}$$
$$VarQ(W) = \frac{2}{n^2}\, tr\left((B\tilde{\Sigma}_W)^2\right)$$

where tr denotes the trace of a matrix. The matrix B is either $\Sigma_W^{-1}\underline{\Delta}_N\Sigma_U L^{sym,h}\Sigma_U\underline{\Delta}_N^{'}\Sigma_W^{-1}$ or $\Sigma_W^{-1}\left(\underline{\Delta}_N\Sigma_U L^h[01]\Sigma_V\underline{\Delta}_S^{'}\right)^{sym}\Sigma_W^{-1}$ depending on whether we are considering autocovariances or crosscovariances. Moreover, if $\tilde{f}_W, f_W, f_U, f_V \in \mathcal{C}^1$, and if f_W is strictly positive with $1/f_W \in \mathcal{C}^1$, then the mean and variance have the following limiting behavior as $n \to \infty$:

$$\mathbb{E}Q(W) \to \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda)\,\tilde{f}_W(\lambda)\,d\lambda$$
$$nVarQ(W) \to \frac{2}{2\pi}\int_{-\pi}^{\pi} g^2(\lambda)\,\tilde{f}_W^2(\lambda)\,d\lambda$$

where $g(\lambda) = f_U^2(\lambda)|\delta^N(e^{-i\lambda})|^2\cos(h\lambda)/f_W^2(\lambda)$ in the autocovariance case, and in the crosscovariance case

$$g(\lambda) = f_U(\lambda)f_V(\lambda)\left(\delta^N(e^{-i\lambda})\delta^S(e^{i\lambda})e^{-ih\lambda} + \delta^N(e^{i\lambda})\delta^S(e^{-i\lambda})e^{ih\lambda}\right)/2f_W^2(\lambda).$$

Also, if the process $\{W_t\}$ satisfies either condition (B) or (HT) referenced in the Appendix, then the following Central Limit Theorem holds as $n \to \infty$:

$$\hat{\tau}_n = \sqrt{n}\,\frac{(Q(W) - \mathbb{E}Q(W))}{\sqrt{nVarQ(W)}} \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,1)$$

Remark 1 The assumptions are common to the literature on convergence of functionals of the periodogram (see Taniguchi and Kakizawa (2000)), and do not seem very stringent. A linear Gaussian process satisfies the cumulant condition as well as condition (B). A non-Gaussian stationary process (with third and fourth order cumulants zero) that is a causal filter of white noise with C^1 spectral density satisfies condition (HT). If we formulate an $ARMA$ model for W_t and apply the canonical decomposition, then necessarily $f_W, f_U, f_V \in \mathcal{C}^1$. If $ARMA$ is the correct specification, then clearly $\tilde{f}_W \in \mathcal{C}^1$ as well. The data need not be Gaussian, but must "look" Gaussian up to fourth order. The programs $X13-AS$ and $TRAMO-SEATS$ provide diagnostics for the presence of skewness and kurtosis; if skewness and kurtosis are present, the asymptotic distribution may not have unit variance.

In order to compute the mean and standard error, it is necessary to assume something about the DGP, such as that provided by the hypothesis H_0. In practice, one could use Theorem 1 as follows: estimate a model for W_t, and declare this to be the Null model described by Σ_W, now viewed as having fixed (nonrandom) parameters. If we have a method for uniquely determining f_U and f_V from f_W (such as the canonical decomposition approach of Hillmer and Tiao (1982)), then an α probability of Type I error has the interpretation that an independent replicate, i.e., a series with DGP given by Σ_W, with filters computed *without* re-estimation of model parameters, would have probability α of falsely rejecting H_0. Note that if we were to re-estimate model parameters for the independent replicate, H_0 would no longer be true for that series (since its DGP given by Σ_W would in general differ from maximum likelihood estimates of that DGP).

Hence, this gives the following application for power studies: if a model $\hat{\Sigma}_W$ is fitted to a series, then we set the Null model equal to the estimate, and simulate series (assuming some distribution compatible with our assumptions) from H_0; then, since the Null model is correct for each simulation, we construct model-based filters based on H_0, *without* re-estimating model parameters for each simulation. The quantiles of the test statistic's empirical distribution function form estimates of the testing procedure's critical values. Next, selecting any other choice of Σ_W, we can compute filters and test statistics based on the false Alternative model, and compute the probability of Type II error for a given critical value, thus obtaining a power surface.

2.3 Estimated Innovation Variance Case

We now focus on a scenario where all of the model parameters are specified except for the innovation variance σ_ϵ^2; hence, our results will be appropriate for the second hypothesis testing paradigm (8). For example, if it is thought that the data can be modelled with an airline model, and we wish to test whether the signal extraction using a $(.6, .6)$ airline model is reasonable, i.e.,

$$W_t = (1-B)(1-B^{12})Y_t = (1-.6B)(1-.6B^{12})\epsilon_t,$$

then we should estimate the innovation variance σ_ϵ^2 from the data, assuming that the other two parameters are correct. In this section, we describe how to modify the diagnostics to accommodate this situation.

What is known then is the "innovation free" version of the covariance matrix of W, denoted by $\Sigma_{W/\sigma_\epsilon}$. Likewise $\Sigma_{U/\sigma_\epsilon}$ is based on the model for U_t with innovation variance given in units of σ_ϵ^2. Hence we can write the signal extraction matrix for U as $F = \Sigma_{U/\sigma_\epsilon} \underline{\Delta}_N^{'} \Sigma_{W/\sigma_\epsilon}^{-1}$, which no longer requires knowledge of σ_ϵ. In fact, we can write (for the autocovariance diagnostic)

$$B = \Sigma_{W/\sigma_\epsilon}^{-1} \underline{\Delta}_N \Sigma_{U/\sigma_\epsilon} L^{sym,h} \Sigma_{U/\sigma_\epsilon} \underline{\Delta}_N^{'} \Sigma_{W/\sigma_\epsilon}^{-1}.$$

A similar formula holds for the cross-covariance case. Thus, we can compute $Q(W)$ without prior knowledge (or estimation) of σ_ϵ, but its mean and variance *do* depend on the innovation variance.

An estimate of the innovation variance (which is unbiased under $H_0^\dagger$) is given by the maximum likelihood estimate (Box and Jenkins, 1976):

$$\hat{\sigma_\epsilon}^2 = \frac{1}{n-d} W' \Sigma^{-1}_{W/\sigma_\epsilon} W, \tag{15}$$

which depends on the hypothesized unit innovation variance model for W_t. Now the mean value of the quadratic form is generally given by

$$\mathbb{E}Q(W) = \frac{tr(B\tilde{\Sigma}_W)}{n} = \frac{\tilde{\sigma_\epsilon}^2}{n} tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}).$$

Naturally, we substitute $\hat{\sigma_\epsilon}$ for $\tilde{\sigma_\epsilon}$, and obtain the estimate

$$\widehat{\mathbb{E}Q(W)} = \frac{\hat{\sigma_\epsilon}^2}{n} tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}),$$

which is computable under $H_0^\dagger$. From the form of the mean and variance (shown below), it is seen that the testing paradigm given by (8) is natural for this setting. So

$$P(W) = Q(W) - \widehat{\mathbb{E}Q(W)}$$

forms a mean-corrected, computable diagnostic. The following theorem presents its finite sample and asymptotic properties.

Theorem 2 *Assume that Assumption A holds on the model decomposition (1), and that $\{U_t\}$ and $\{V_t\}$ are independent. If the third and fourth cumulants of the true DGP of W_t are zero, then the true mean and variance of $P(W)$ are given by*

$$\mathbb{E}P(W) = \frac{\tilde{\sigma_\epsilon}^2}{n} tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}) \left(1 - \frac{tr(C)}{n-d}\right)$$

$$VarP(W) = \frac{2\tilde{\sigma_\epsilon}^4}{n} \left(\frac{tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})^2}{n} - 2\frac{tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}) tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}} C)}{n(n-d)} + \frac{tr^2(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}) tr(C^2)}{n(n-d)^2} \right), \tag{16}$$

where $C = \Sigma^{-1}_{W/\sigma_\epsilon} \tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}$. Under $H_0^\dagger$, the mean becomes zero and the variance simplifies to

$$\frac{2\tilde{\sigma_\epsilon}^4}{n^2} \left(tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})^2 - \frac{tr^2(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})}{n-d} \right).$$

Moreover, if $\tilde{f}_W, f_W, f_U, f_V \in \mathcal{C}^1$, and f_W is strictly positive with $1/f_W \in \mathcal{C}^1$, then the mean and variance have the following limiting behavior as $n \to \infty$:

$$\mathbb{E}P(W) \to \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda)\, d\lambda \right) \left(1 - \frac{\sigma_\epsilon^2}{\tilde{\sigma_\epsilon}^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{f}_W(\lambda)}{f_W(\lambda)}\, d\lambda \right)$$

$$\begin{aligned} nVarP(W) \to\ & 2 \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g^2(\lambda) \tilde{f}_W^2(\lambda)\, d\lambda \right) \\ & - 4 \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \frac{\tilde{f}_W^2(\lambda)}{f_W(\lambda)}\, d\lambda \right) \left(\frac{\sigma_\epsilon^2}{\tilde{\sigma_\epsilon}^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda)\, \tilde{f}_W(\lambda)\, d\lambda \right) \\ & + 2 \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{f}_W^2(\lambda)}{f_W^2(\lambda)}\, d\lambda \right) \left(\frac{\sigma_\epsilon^2}{\tilde{\sigma_\epsilon}^2} \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda)\, d\lambda \right)^2 \end{aligned}$$

with g as in Theorem 1. Also, if the process $\{W_t\}$ satisfies either condition (B) or (HT) referenced in the Appendix, then the following Central Limit Theorem holds as $n \to \infty$:

$$\sqrt{n}\,\frac{(P(W) - \mathbb{E}P(W))}{\sqrt{nVarP(W)}} \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0,1) \tag{17}$$

Remark 2 The variance of $P(W)$ under $H_0^\dagger$ must be estimated by the following:

$$n\widehat{VarP(W)} = 2\hat{\sigma_\epsilon}^4 \left(\frac{tr(B\tilde{\Sigma}_{W/\hat{\sigma_\epsilon}})^2}{n} - \frac{tr^2(B\tilde{\Sigma}_{W/\hat{\sigma_\epsilon}})}{n(n-d)} \right)$$

It is easily checked that this converges in probability to the limit of $nVarP(W)$, and so may be substituted into (17) by Slutsky's Theorem. Thus we compute the statistic

$$\tau_n^\dagger = \sqrt{n}\,\frac{P(W)}{\sqrt{n\widehat{VarP(W)}}},$$

which is asymptotically standard normal under $H_0^\dagger$.

Remark 3 If we treat $\hat{\sigma_\epsilon}$ and the $ARMA$ parameters as fixed, then

$$\hat{\sigma_\epsilon}^2 \Sigma_{W/\sigma_\epsilon} = \Sigma_W.$$

It follows that

$$P(W)|_{H_0^\dagger} = Q(W) - \frac{1}{n} tr(B\tilde{\Sigma}_W) = Q(W) - \mathbb{E}Q(W)|_{H_0}$$

$$n\widehat{VarP(W)}|_{H_0^\dagger} = 2\left(\frac{tr(B\tilde{\Sigma}_W)^2}{n} - \frac{tr^2(B\tilde{\Sigma}_W)}{n(n-d)} \right)$$

$$= nVarQ(W)|_{H_0} - 2\frac{tr^2(B\tilde{\Sigma}_W)}{n(n-d)}.$$

Hence in this case, $P(W)$ has less variability than $Q(W)$. Thus it is easier to reject $H_0^\dagger$ than H_0 when the model parameters come from estimates that are treated as fixed. This makes sense, since H_0 implies $H_0^\dagger$.

The application and interpretation of Theorem 2 is similar to that discussed for Theorem 1. For an estimated model, an independent replicate with innovation variance re-estimated would falsely reject $H_0^\dagger$ with probability α, given the appropriate critical value. Note that in this case, the "unit-innovation variance" DGP for the replicate and the model used for the filters exactly coincide, so $H_0^\dagger$ is true. If instead we were to re-estimate the non-innovation variance parameters for the replicate, then we would obtain parameter estimates that would in general be different from the DGP parameters, and thus $H_0^\dagger$ would be false. So only the innovation variance is to be re-estimated in this interpretation.

3 Applications and Extensions

A primary objective of this work is to determine the asymptotics of diagnostics introduced in Findley, McElroy, and Wills (2004). Theorems 1 and 2 provide ways to compute asymptotic power for the procedures, under a two-sided alternative. However, it is desirable to obtain an interpretation for positive versus negative values of the diagnostic, so as to obtain a sensible one-sided test. In addition, we want rejection of the Null Hypothesis to give us meaningful information about *how* our model is incorrect for signal extraction, and how it might be modified.

In this section, we discuss a modification of the autocovariance diagnostic that facilitates such interpretations. Then we demonstrate the properties of the diagnostic on two seasonal time series. Finally, we explore the question of redundancy: how correlated are the various diagnostics with one another when the model is correct? Can all this information be consolidated? We note that even though the original over- and under-adjustment diagnostics of Maravall (2003) and later Findley, McElroy, and Wills (2004) sought to address the issue of over- and under-smoothing of a signal extraction component, in fact these diagnostics assess over- and under-modeling of various components. This distinction is further clarified and explored in this section.

3.1 Over- and Under-Modeling

From the results of the previous section, we know that the autocovariance diagnostic, under some conditions, converges in probability to

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f_U^2(\lambda)|\delta^N(e^{-i\lambda})|^2 \cos(h\lambda)\frac{\tilde{f}_W(\lambda)}{f_W^2(\lambda)}\,d\lambda = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_U(\lambda)\frac{f_S(\lambda)}{f_Y(\lambda)}\cos(h\lambda)\frac{\tilde{f}_W(\lambda)}{f_W(\lambda)}\,d\lambda.$$

So $\hat{A}_n(0)$ is a measure of model discrepancy – $\tilde{f}_W/f_W$ – weighted by the function $f_U f_S/f_Y$. Observe that the function f_S/f_Y is an appealing weighting function, since it is bounded between 0 and 1 for all frequencies, and attains its minimum (zero) at "noise" frequencies and its maximum (one) at "signal" frequencies, i.e., those frequencies that are roots of $\delta^N(e^{-i\lambda})$ and $\delta^S(e^{-i\lambda})$ respectively. Such a weighting function f_S/f_Y fully weights model discrepancy from truth $\tilde{f}_W/f_W$ at signal frequencies, but disregards discrepancies occurring at the noise frequencies. Unfortunately, the asymptotic limit of $\hat{A}_n(0)$ does not provide such a weighting function; it weights the model discrepancy by $f_U f_S/f_Y$, and the function f_U can destroy the interpretation of fully weighting signal frequencies. This is because f_U can be quite general: for example, it can be flat or even have a high-pass form $1+\rho^2-2\rho\cos\lambda$ for ρ close to one (this would be problematic for a trend signal, where low frequencies should receive more weight).

We propose the following modified diagnostic, which will correspond to the desired weighting scheme f_S/f_Y:

$$\tilde{A}_n(h) = \frac{W' \Sigma_W^{-1} \underline{\Delta}_N (\Sigma_U L^h)^{sym} \underline{\Delta}_N' \Sigma_W^{-1} W}{n},$$

where $(\Sigma_U L^h)^{sym} = (\Sigma_U L^h + L'^h \Sigma_U)/2$. This can be written in the form $Q(W)$ as in Section 2, with a matrix $B = \Sigma_W^{-1} \underline{\Delta}_N \Sigma_U L^{h\,sym} \underline{\Delta}_N' \Sigma_W^{-1}$. The same types of theoretical results (Theorems 1 and 2) apply to $\tilde{A}_n$, but now the appropriate function g is defined by

$$\tilde{g}(\lambda) = f_U(\lambda) |\delta^N(e^{-i\lambda})|^2 \cos(h\lambda) f_W^{-2}(\lambda).$$

Thus we have the convergence

$$\tilde{A}_n(h) \xrightarrow{P} \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(h\lambda) \frac{f_S(\lambda)}{f_Y(\lambda)} \frac{\tilde{f}_W(\lambda)}{f_W(\lambda)} \, d\lambda.$$

In addition,

$$\mathbb{E}\tilde{A}_n(h)|_{H_0} = \frac{1}{n} tr(B \tilde{\Sigma}_W)|_{H_0} \to \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(h\lambda) \frac{f_S(\lambda)}{f_Y(\lambda)} \, d\lambda,$$

so we have

$$\tilde{A}_n(h) - \mathbb{E}\tilde{A}_n(h)|_{H_0} \xrightarrow{P} \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos(h\lambda) \frac{f_S(\lambda)}{f_Y(\lambda)} \left(\frac{\tilde{f}_W(\lambda)}{f_W(\lambda)} - 1 \right) d\lambda. \tag{18}$$

This is the numerator of the normalized modified diagnostic. We form a variance normalization along the lines of Theorem 1 (use equation (14) with the above choice of B), and call the normalized quantity $\tilde{\tau}$. That is,

$$\tilde{\tau} = \frac{\tilde{A}_n(h) - \mathbb{E}\tilde{A}_n(h)}{\sqrt{Var(\tilde{A}_n(h))}}.$$

Note that we have developed this modified diagnostic with the testing paradigm H_0 in mind, but a similar treatment can easily be developed for $H_0^\dagger$.

Given these asymptotics, we can offer the following interpretation. Let us refer to the range of frequencies where f_S is high relative to f_Y as the "spectral range" of the signal. Now if $\tilde{f}_W > f_W$ in the spectral range of the signal, then (18) is positive and the diagnostic is too large. But if $\tilde{f}_W < f_W$ in the spectral range, then (18) is negative and the diagnostic is too small. Now when $\tilde{f}_W < f_W$ in a spectral band, the model is too chaotic for those frequencies since it assigns too much variation there. This will be referred to as "over-modeling." Conversely, $\tilde{f}_W > f_W$ indicates the model is too stable, as it assigns too little variation at the signal frequencies. This will be referred to as "under-modeling." Outside the spectral range of the signal these interpretations are less meaningful, since the weighting function will dampen the effect of model discrepancies. These observations can be used to form a meaningful one-sided testing procedure. First we summarize the logic:

$$\begin{aligned} &\tilde{\tau}_n \text{ is significantly negative} \Leftrightarrow \tilde{A}_n(0) \text{ is too small} \\ &\quad \Leftrightarrow \tilde{f}_W << f_W \text{ in the spectral range of the signal} \\ &\quad \text{over-modeling} \\ &\tilde{\tau}_n \text{ is significantly positive} \Leftrightarrow \tilde{A}_n(0) \text{ is too large} \\ &\quad \Leftrightarrow \tilde{f}_W >> f_W \text{ in the spectral range of the signal} \\ &\quad \text{under-modeling} \end{aligned} \tag{19}$$

Now let the functional D be defined, for given f_S and f_Y, by

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

Then the upper one-sided test has hypotheses

$$\begin{aligned} H_0 &: \tilde{f}_W = f_W \\ H_1 &: D(\tilde{f}_W, f_W) > 0 \end{aligned}$$

and H_0 is rejected with confidence $1-\alpha$ if $\tilde{\tau}_n > z_{1-\alpha}$, which indicates significant under-modeling in the signal's spectral band. The lower one-sided test has hypotheses

$$\begin{aligned} H_0 &: \tilde{f}_W = f_W \\ H_1 &: D(\tilde{f}_W, f_W) < 0 \end{aligned}$$

and H_0 is rejected with confidence $1-\alpha$ if $\tilde{\tau}_n < z_{\alpha}$, which indicates significant over-modeling in the signal's spectral band.

We mention here a few other properties of the modified diagnostic $\tilde{A}_n(h)$. When $h = 0$, the symmetrization term is no longer needed. In this case, summing the signal and noise diagnostic produces

$$\frac{W'\Sigma_W^{-1}\underline{\Delta}_N\Sigma_U\underline{\Delta}_N'\Sigma_W^{-1}W}{n} + \frac{W'\Sigma_W^{-1}\underline{\Delta}_S\Sigma_V\underline{\Delta}_S'\Sigma_W^{-1}W}{n} = \frac{W'\Sigma_W^{-1}W}{n}.$$

Under H_0, this has a mean of 1. So, under the correct model case, the signal and noise diagnostics will tend to have opposite signs, e.g., a significantly positive diagnostic for the seasonal will often be accompanied by a significantly negative diagnostic for the seasonally adjusted component. This property does not hold for the original diagnostics $\hat{A}_n(0)$.

To summarize, we can simply compute the normalized diagnostic $\tilde{A}_n(0)$, and test for either over- or under-modeling. By using the above interpretation of the one-sided alternative, we can then proceed to modify our model, to correct the problem. For example, consider the airline model

$$W_t = (1-B)(1-B^{12})Y_t = (1-\theta B)(1-\Theta B^{12})\epsilon_t$$

which has spectral density

$$f_W(\lambda) = (1 + \theta^2 - 2\theta \cos \lambda)(1 + \Theta^2 - 2\Theta \cos 12\lambda)\sigma_\epsilon^2.$$

Over-modeling of the seasonal, for example, indicates that $\tilde{f}_W < f_W$ at seasonal λ, say $\lambda = 2\pi j/12$ for $j = 1, 2, \cdots, 6$. Thus the model discrepancy at the seasonal frequencies is

$$\frac{\tilde{f}_W(2\pi j/12)}{f_W(2\pi j/12)} = \frac{(1 + \tilde{\theta}^2 - 2\tilde{\theta}\cos(2\pi j/12))}{(1 + \theta^2 - 2\theta\cos(2\pi j/12))} \frac{(1 - \tilde{\Theta})^2}{(1 - \Theta)^2} \frac{\tilde{\sigma_\epsilon}^2}{\sigma_\epsilon^2}.$$

This ratio is less than 1 for all j if $\theta = \tilde{\theta}$, $\sigma_\epsilon \geq \tilde{\sigma_\epsilon}$, and $|1 - \Theta| > |1 - \tilde{\Theta}|$. Assuming that $\Theta, \tilde{\Theta} \in (0, 1)$, this can be achieved by letting $\Theta < \tilde{\Theta}$. This indicates that increasing Θ may fix the over-modeling of the seasonal; however, since the above analysis holds for $j = 0$ as well, this may induce over-modeling of the trend. This can then be compensated by adjusting θ downwards.

This example shows that even for the simple airline model, adjustment of the model is tricky (and may be undesirable; after all, the original model is globally optimal if chosen through maximum likelihood estimation). For more general models, where the relationship of the parameters to the seasonal behavior is less clear, it will be even harder to adjust the model.

3.2 Examples

We now consider two time series from the Foreign Trade Division of the U.S. Census Bureau, referred to as $m00100$ and $m00110$ respectively. The first series is Imports of Meat products, and the second series is Imports of Dairy Products and Eggs. Both series are for the time period from January 1989 to December 2003. For the first series we chose a $SARIMA$ model using standard identification techniques (Box and Jenkins, 1976), and the signal extraction diagnostics $\tilde{\tau}$ indicate adequacy for the most part. Below we discuss some slight modifications to the model, which "improve" the model in the sense that the diagnostics are no longer significant. The second series is much more problematic, and many of the diagnostics are highly significant. These two examples furnish a contrast between mild and serious problems with the estimated $SARIMA$ model.

In order not to over-burden the exposition, we focus on the autocovariance diagnostic for the "seasonal-irregular" component, computed at lags 0, 1, and 12. For the first series $m00100$ (length 180), the automatic modeling procedure of X-12-$ARIMA$ determined a $SARIMA(1,1,1)(0,1,1)_{12}$ model for the logged data (after certain fixed regression effects, i.e., trading day and outliers, have been removed):

$$(1 - .736B)(1 - B)(1 - B^{12})Y_t = (1 - .929B)(1 - .795B^{12})\epsilon_t, \qquad \sigma_\epsilon^2 = .0052$$

This is decomposed into trend, seasonal, and irregular, with the nonseasonal AR factor allotted to the trend. Table 1 gives the diagnostics along with their one-sided p-values. Note that we only have a sensible interpretation for one-sided tests for the modified diagnostic $\tilde{\tau}$. According to $\tilde{\tau}$, the model is adequate, whereas $\hat{\tau}$ and $\tau^\dagger$ indicate under-modeling at lag 12. It is interesting that the lag 0 diagnostics are all negative, although not significantly so. Thus we might proceed with the diagnosis that there is mild under-modeling. Since Θ controls the seasonal movement, it can be adjusted downwards to generate more variation at the seasonal frequencies, which increases f_W in the signal's spectral band. In Table 2, we present the diagnostics obtained by fixing $\Theta = .6$ (instead of at the MLE value of .795) and keeping the other parameters the same (alternatively, one could re-estimate the innovation variance with these fixed parameter values). Experimentation with other values of Θ yields similar results: the under-modeling of the seasonal-irregular component is abated.

Table 1. Diagnostics with p-values for $m00100$.

Diagnostic	Lag 0	p-value	Lag 1	p-value	Lag 12	p-value
$\hat{\tau}$	-1.01	.155	-.15	.441	2.43	.007
$\tau^\dagger$	-1.57	.058	-.16	.438	2.62	.004
$\tilde{\tau}$	-1.03	.152	1.27	.101	1.07	.143

Table 2. Diagnostics with p-values for $m00100$, Θ fixed at $.6$

Diagnostic	Lag 0	p-value	Lag 1	p-value	Lag 12	p-value
$\hat{\tau}$	.08	.468	.19	.426	1.55	.061
$\tau^\dagger$	.10	.459	.20	.420	1.55	.061
$\tilde{\tau}$	-.70	.241	1.09	.137	.35	.363

The second series $m00110$ (length 180) in logs is modeled with a $SARIMA(2,0,1)(0,1,1)_{12}$ model (after removing a temporary change and outlier regression effect):

$$(1 - 1.065B + 0.209B^2)(1 - B^{12})Y_t = (1 - 0.528B)(1 - 0.982B^{12})\epsilon_t, \qquad \sigma^2_\epsilon = .0085$$

The model coefficients were estimated using X-12-$ARIMA$. We note that convergence of the maximum likelihood parameter estimation procedure was somewhat slow, and the diagnostics for the seasonal-irregular are highly significant, as shown in Table 3. In the lag 0 case, the sign of $\hat{\tau}$ and $\tau^\dagger$ actually differ from $\tilde{\tau}$. Restricting our attention to $\tilde{\tau}$, we conclude from the significantly negative lag zero diagnostic, together with (19), that there is over-modeling of the seasonal present. However, the high value of Θ indicates that little can be done to increase Θ, and instead we may consider lowering θ. Some experimentation show that this ameliorates the situation, and Table 4 displays results with $\theta = .2$. Even with this low value of θ, some of the $\hat{\tau}$ and $\tau^\dagger$ diagnostics are significant (although less so); however, the $\tilde{\tau}$ is no longer significant, which is the most important measure. Of course, another approach might be to adjust the model specification, perhaps to a $SARIMA$ $(1,1,1)(0,1,1)_{12}$ or $(0,1,1)(0,1,1)_{12}$.

Table 3. **Diagnostics with p-values for $m00110$.**						
Diagnostic	Lag 0	p-value	Lag 1	p-value	Lag 12	p-value
$\hat{\tau}$	9.05	0	11.02	0	17.17	0
$\tau^{\dagger}$	11.36	0	11.69	0	17.86	0
$\tilde{\tau}$	-3.77	0	1.98	.024	.15	.441

Table 4. **Diagnostics with p-values for $m00110$, θ fixed at $.2$**						
Diagnostic	Lag 0	p-value	Lag 1	p-value	Lag 12	p-value
$\hat{\tau}$	.78	.217	2.46	.007	7.02	0
$\tau^{\dagger}$	1.33	.092	2.47	.007	7.84	0
$\tilde{\tau}$	-1.52	.065	.24	.467	.27	.393

It is interesting that the sign of $\hat{\tau}$ and $\tau^{\dagger}$ can be different from $\tilde{\tau}$; this is explained by the theory of Section 3.1. Hence, one should only apply a 2-sided test for the $\hat{\tau}$ and $\tau^{\dagger}$ diagnostics, since the sign can be misleading. If we had increased θ instead, which would have been the indication of the significant positive values of $\hat{\tau}$ and $\tau^{\dagger}$, we would be making the problem worse; the diagnostics become even more extreme with $\theta = .6$ (results not shown here).

Of course, this analysis merely scratches the surface, and is intended only to demonstrate some of the properties of the three diagnostics on real seasonal time series. An extensive empirical investigation of the various diagnostics on dozens of series is beyond the scope of this work.

3.3 Redundancy in the Correct Model Case

All the various diagnostics presented in this work can be written as either a quadratic form in the data or a ratio of such. In the development of Section 2.2, our diagnostic takes the form $Q(W)$, which is normalized by constants that depend on our model choice. In Section 2.3, we examine $P(W)$, which is also a quadratic form in the data. Following Remark 2, the normalized diagnostic $\tau_n^{\dagger}$ can be written as

$$\sqrt{n}\frac{P(W)}{\sqrt{\widehat{VarP(W)}}} = \sqrt{n}\left(\frac{W'BW}{W'\Sigma^{-1}_{W/\sigma_\epsilon}W} - a_0\right)/a_1$$

with $a_0 = tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})/n$ and $a_1 = \sqrt{2\left(tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})^2/n - tr^2(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})/(n(n-d))\right)}$; both a_0 and a_1 are non-random. So $\tau^{\dagger}$ is a ratio of quadratic forms in the data, whereas $\hat{\tau}$ and $\tilde{\tau}$ are simply quadratic forms.

For diagnostics of the form $Q(W)$, we can easily compute correlations. Given two such normalized diagnostics for two signals S and $S^\sharp$, the correlation is given by

$$\frac{Cov(Q(W), Q^\sharp(W))}{\sqrt{VarQ(W)\, VarQ^\sharp(W)}} = \frac{2tr(B\tilde{\Sigma}_W B^\sharp \tilde{\Sigma}_W)}{2\sqrt{tr(B\tilde{\Sigma}_W)^2 tr(B^\sharp \tilde{\Sigma}_W)^2}}.$$

Thus, under H_0 we can use this formula to compute the correlation for either $\hat{\tau}$ or $\tilde{\tau}$. We work out some details for the latter case. The numerator works out to be

$$2tr\left(\underline{\Delta}_N \Sigma_U^{1/2} L^{sym,h} \Sigma_U^{1/2'} \underline{\Delta}_N' \Sigma_W^{-1} \tilde{\Sigma}_W \Sigma_W^{-1} \underline{\Delta}_N^\sharp {\Sigma_U^\sharp}^{1/2} L^{sym,k} {\Sigma_U^\sharp}^{1/2'} {\underline{\Delta}_N^\sharp}' \Sigma_W^{-1} \tilde{\Sigma}_W \Sigma_W^{-1}\right)/n^2.$$

Multiplying by n, it converges to

$$\frac{2}{2\pi}\int_{-\pi}^{\pi} |\delta^N(e^{-i\lambda})|^2 f_U(\lambda)\cos(h\lambda)\frac{\tilde{f}_W(\lambda)}{f_W^2(\lambda)} |\delta_\sharp^N(e^{-i\lambda})|^2 f_U^\sharp(\lambda)\cos(k\lambda)\frac{\tilde{f}_W(\lambda)}{f_W^2(\lambda)} d\lambda.$$

This can be rewritten as

$$\frac{2}{2\pi}\int_{-\pi}^{\pi} \frac{f_S(\lambda)}{f_Y(\lambda)}\frac{f_S^\sharp(\lambda)}{f_Y(\lambda)}\cos(h\lambda)\cos(k\lambda)\frac{\tilde{f}_W^2(\lambda)}{f_W^2(\lambda)} d\lambda.$$

This combines the weighting functions f_S/f_Y and $f_S^\sharp/f_Y$. Taking $h = k = 0$, we see that the correlation is always non-negative, but is closer to zero when S and $S^\sharp$ occupy separate frequency bands. For example, taking $S^\sharp = N$, the weighting function is $f_S f_N / f_Y^2$.

Note that the positive correlation effect does not imply that a particular diagnostic computed for signal and noise should have opposite signs. It means that if one diagnostic is higher on another data set, the other diagnostic will be higher too. This effect is reduced by using the modified diagnostic, and by taking complementary signals.

4 Conclusion

This paper defines several diagnostics for goodness of signal extraction, and describes their statistical behavior. The original diagnostics $\hat{\tau}$ and $\tau^\dagger$ were proposed in Findley, McElroy, and Wills (2004), although essentially only for the lag zero auto-covariance case. Here we expand the class of such diagnostics to include other lags and cross-covariance as well, and in addition present a modified $\tilde{\tau}$ that allows for interpretability of the sign. This allows us to conduct meaningful one-sided tests, and provides some direction for how the model should be changed in order to correct over- or under-modeling in particular signal frequency bands. We do not claim that it is easy to know how a given $SARIMA$ model should be altered; but we have tried to demonstrate that $\hat{\tau}$ and $\tau^\dagger$ cannot be used reliably to alter a model.

Of course, a given model chosen by maximum likelihood should be the best fit for the data, since it is preferred to other $SARIMA$ models (at least with the same differencing orders) by AIC or another model comparison criterion. However, the Gaussian maximum likelihood procedure looks for a model f_W that is close to the truth $\tilde{f}_W$ in an average sense, over all the frequencies. This is because the quantity $W^{'}\Sigma^{-1}_{W/\sigma_\epsilon}W/(n-d)$ occurring in the likelihood function converges in probability to

$$\sigma^2_\epsilon \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\tilde{f}_W(\lambda)}{f_W(\lambda)}\,d\lambda.$$

One way to view Gaussian maximum likelihood estimation, is that it is a procedure that seeks out f_W from a model class such that the above quantity is close to σ^2_ϵ. Of course, this formula only holds asymptotically. The diagnostics $\tilde{\tau}$ examine model discrepancy $\tilde{f}_W/f_W$ over a range of "signal frequencies," which is achieved by integrating against the weighting function f_S/f_Y; in contrast, maximum likelihood estimation uses the weight $f_Y/f_Y = 1$. This can be seen as addressing the issue of global versus local fit, in a spectral sense. That is, a model f_W may be good as an overall fit to $\tilde{f}_W$, considering all frequencies at once, but may be a poor fit at some particular band of frequencies. By sacrificing the global fit, a model can arguably be improved locally, i.e., at the signal and noise frequency bands.

In practice, a plethora of signal extraction diagnostics can be produced, and it is reasonable to think that there may be some redundancy in them. For $\hat{\tau}$ and $\tilde{\tau}$, correlations between various auto-covariance diagnostics can be calculated for each example, and in this fashion any redundancy can be assessed. Future work will focus on large-scale empirical investigations of these statistics.

5 Appendix: Proofs

We first delineate a proposition that establishes asymptotic normality of a certain special class of quadratic forms. Being of interest in its own right, this result can be applied to an approximation of the diagnostics considered in this paper. We consider a stationary process $\{W_t\}$, whose spectral density is denoted by $\tilde{f}_W(\lambda)$. Given a sample $W = \{W_1, W_2, \cdots, W_n\}^{'}$, we denote the periodogram by $\hat{f}_W(\lambda)$:

$$\hat{f}_W(\lambda) = \frac{1}{n}\left|\sum_{t=1}^{n} W_t e^{-i\lambda t}\right|^2$$

Then it follows at once that

$$\frac{1}{n}W'\Sigma(g)W = \frac{1}{2\pi}\int_{-\pi}^{\pi} g(\lambda)\,\hat{f}_W(\lambda)\,d\lambda.$$

The following proposition describes the asymptotics of this quadratic form. Some mild conditions on the data are required for the asymptotic theory; we follow the material in Taniguchi and Kakizawa

(2000, Section 3.1.1). Condition (B), due to Brillinger (1981), states that the process is strictly stationary and condition (B1) of Taniguchi and Kakizawa (2000, page 55) holds. Condition (HT), due to Hosoya and Taniguchi (1982), states that the process has a linear representation, and conditions (H1) through (H6) of Taniguchi and Kakizawa (2000, pages 55 – 56) hold. Neither of these conditions are stringent; for example, a causal linear filter of a white noise process with fourth moments satisfies (HT).

Proposition 1 *Suppose that the stationary process $\{W_t\}$ satisfies either condition (B) or (HT). Let g be any real, even, continuous function on $[-\pi, \pi]$. Then as $n \to \infty$ the following assertions hold:*

$$\int_{-\pi}^{\pi} g(\lambda)\, \hat{f}_W(\lambda)\, d\lambda \xrightarrow{P} \int_{-\pi}^{\pi} g(\lambda)\, \tilde{f}_W(\lambda)\, d\lambda$$
$$\sqrt{n} \int_{-\pi}^{\pi} g(\lambda) \left(\hat{f}_W(\lambda) - \tilde{f}_W(\lambda) \right) d\lambda \overset{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, V)$$

where the limiting variance V is given by

$$V = 4\pi \int_{-\pi}^{\pi} g^2(\lambda)\, \tilde{f}_W^2(\lambda)\, d\lambda + \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} g(\lambda)\, g(\omega)\, Q^W(-\lambda, \omega, -\omega)\, d\lambda\, d\omega$$

and Q^W is the tri-spectral density of $\{W_t\}$, i.e.,

$$Q^W(\lambda_1, \lambda_2, \lambda_3) = \sum_{t_1, t_2, t_3 \in \mathbb{Z}} \exp\{-i(\lambda_1 t_1 + \lambda_2 t_2 + \lambda_3 t_3)\} C^W(t_1, t_2, t_3)$$

and C^W is the fourth-order cumulant function of the process.

Proof of Proposition 1. Since g is real, even, and continuous, the first convergence result follow at once from Lemma 3.1.1 of Taniguchi and Kakizawa (2000). Note that they define the periodogram with a 2π factor. □

We next state and prove several approximation lemmas for the analysis of quadratic forms, which are of interest in their own right and are used in the proof of Theorem 1.

Lemma 1 *Let Δ be the $m - d \times m$ dimensional matrix associated to an order d polynomial $\delta(z)$, given by*

$$\Delta_{ij} = \delta_{j+d-i}$$

where we let $\delta_k = 0$ if $k < 0$ or $k > d$. Then, for m-dimensional $\Sigma(g)$,

$$\Delta \Sigma(g) \Delta' = \Sigma(g \cdot f)$$

where $f(\lambda) = |\delta(e^{-i\lambda})|^2$.

Proof of Lemma 1. Based on formula (6), we have

$$\begin{aligned}\left(\Delta\Sigma(g)\Delta^{'}\right)_{jk} &= \sum_{l,t}\Delta_{jl}\Sigma_{lt}(g)\Delta_{kt} \\ &= \frac{1}{2\pi}\sum_{l,t}\delta_{j+d-l}\delta_{k+d-t}\int_{-\pi}^{\pi}g(\lambda)e^{i(l-t)\lambda}\,d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}g(\lambda)e^{i(j-k)\lambda}\sum_{l}\delta_{j+d-l}e^{-i(j+d-l)\lambda}\sum_{t}\delta_{k+d-t}e^{i(k+d-t)\lambda}\,d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}g(\lambda)\delta(e^{-i\lambda})\delta(e^{i\lambda})e^{i(j-k)\lambda}\,d\lambda,\end{aligned}$$

which is the jkth entry of the Toeplitz matrix associated to $g\cdot f$, as desired. □

For the next three lemmas, we need a concept from Taniguchi and Kakizawa (2000). If we let $\gamma_g(j-k)=\Sigma_{jk}(g)$, then a discrete approximation is given by the Riemann sum

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

with $\lambda_k=2\pi k/n$. In this manner we define the Toeplitz approximation to $\Sigma(g)$ via

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

See Taniguchi and Kakizawa (p. 490, 2000) for related results, which we cite here for easy reference. It can be shown that

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

with $H_{jk}=n^{-1/2}e^{i2\pi jk/n}$ and $D=diag\{g(\lambda_1),\cdots,g(\lambda_n)\}$. The $*$ denotes conjugate transpose.

Lemma 2 *Let f be a continuous function on $[-\pi,\pi]$ in $\mathcal{C}^1$, and let Z_t and X_t be two sequences of (possibly correlated) random variables with uniformly bounded second moments. Let $Z=(Z_1,Z_2,\cdots,Z_n)^{'}$ and $X=(X_1,X_2,\cdots,X_n)^{'}$, and let $\Sigma(f)$ be n-dimensional. Then*

$$Z^{'}\Sigma(f)X=O_P(1)+Z^{'}\overline{\Sigma}(f)X$$

as $n\to\infty$.

Proof of Lemma 2.

$$\begin{aligned}&\mathbb{E}\left|Z^{'}\Sigma(f)X-Z^{'}\overline{\Sigma}(f)X\right| \\ &\leq\sum_{j,k}\mathbb{E}|Z_jX_k|\left|\Sigma_{jk}(f)-\overline{\Sigma}_{jk}(f)\right|\leq\sup_j\sqrt{\mathbb{E}[Z_j^2]}\sup_k\sqrt{\mathbb{E}[X_k^2]}\sum_{j,k}\left|\Sigma_{jk}(f)-\overline{\Sigma}_{jk}(f)\right|\end{aligned}$$

using the Cauchy-Schwarz inequality. Now apply Lemma 7.2.9 of Taniguchi and Kakizawa (2000), which is due to Wahba (1968), noting that boundedness in absolute mean implies that the expression is bounded in probability. □

Lemma 3 *Let g and f be continuous functions on $[-\pi, \pi]$ satisfying the same conditions as in Lemma 2, and suppose that the random variables Z_t have uniformly bounded second moments. Then*

$$Z'\Sigma(f)\Sigma(g)\Sigma(f)Z = O_P(1) + Z'\Sigma(f^2 \cdot g)Z$$

as $n \to \infty$.

Proof of Lemma 3.

$$\begin{aligned} Z'\Sigma(f)\Sigma(g)\Sigma(f)Z - Z'\Sigma(f^2 \cdot g)Z = Z' \left(\Sigma(f)\Sigma(g) - \Sigma(f \cdot g)\right)\Sigma(f)Z \\ + Z' \left(\Sigma(f \cdot g)\Sigma(f) - \Sigma(f^2 \cdot g)\right) Z \end{aligned} \tag{20}$$

Let $X = \Sigma(f)Z$; a fact that we use repeatedly is that X will satisfy the same boundedness properties as Z, so long as γ_f is absolutely summable:

$$\begin{aligned} X_j &= \sum_{k=1}^{n} \Sigma_{jk}(f)Z_k = \sum_{k=1}^{n} \gamma_f(j-k)Z_k \\ X_j^2 &= \sum_{k,l=1}^{n} \gamma_f(j-k)\gamma_f(j-l)Z_kZ_l \\ \mathbb{E}X_j^2 &\leq \left(\sup_k \sqrt{\mathbb{E}Z_k^2}\right)^2 \sup_j\{\sum_{k,l=-\infty}^{\infty} |\gamma_f(j-k)||\gamma_f(j-l)|\} \end{aligned}$$

It is easy to see that the latter supremum is bounded by $(\sum_{k=-\infty}^{\infty} |\gamma_f(k)|)^2$. So in a like manner, $Y = \Sigma(g)X$ has bounded second moments, and we can apply Lemma 2 to $Z'\Sigma(f)Y$:

$$Z'\Sigma(f)Y = O_P(1) + Z'\overline{\Sigma}(f)Y$$

Now $Z'\overline{\Sigma}(f)Y = V'\Sigma(g)X$, where $V = \overline{\Sigma}(f)Z$ also has bounded second moments (this follows from Lemma 7.2.9 of Wahba (1968)). Again we apply Lemma 2, and finally obtain

$$Z'\Sigma(f)\Sigma(g)X = O_P(1) + Z'\overline{\Sigma}(f)\overline{\Sigma}(g)X;$$

the advantage of this representation is that $\overline{\Sigma}(f)\overline{\Sigma}(g) = \overline{\Sigma}(f \cdot g)$ follows from the Hermitian form of these matrices. In a like manner,

$$Z'\Sigma(f \cdot g)X = O_P(1) + Z'\overline{\Sigma}(f \cdot g)X$$

by another application of Lemma 2. This shows that the first term on the right hand side of (20) is $O_P(1)$. An analogous argument holds for the second term. Note that we only need that the Fourier coefficients of f and g are absolutely summable, which follows from the conditions of the Lemma. $\square$

We also wish to consider inverses of Toeplitz matrices. It is clear from the Hermitian form of $\overline{\Sigma}(f)$ that

$$\overline{\Sigma}(f)^{-1} = H^* D^{-1} H = \overline{\Sigma}(1/f)$$

so long as f is nowhere equal to zero. We can extend Lemma 2 to this case as follows:

Lemma 4 *Make the same assumptions as in Lemma 2, and suppose that f is strictly positive. Then*

$$Z^{'}\Sigma(f)^{-1}X = O_P(1) + Z^{'}\overline{\Sigma}(1/f)X$$

as $n \to \infty$.

Proof of Lemma 4. Use the same proof as Lemma 2, but now apply the result of Liggett (1971), also found on page 491 and 492 of Taniguchi and Kakizawa (2000). Note that it is necessary that f be bounded away from zero, so that we avoid division by zero. □

With these lemmas, we can now prove Theorem 1 for the autocovariance diagnostics; for the crosscovariances diagnostics, some additional tinkering with the lemmas will be needed.

Proof of Theorem 1. First consider the formulas given in (14); the expectation is immediate, and the variance formula is standard, assuming that third and fourth order cumulants vanish – see McCullagh (1987, p. 65). Note that the matrix of the quadratic form $Q(W)$ must be symmetric; otherwise this formula for the variance is not true. This explains why we use the "symmetrization" technique.

For the asymptotic results, we first prove the Central Limit Theorem, and note that the convergences of mean and variance can be proved in an analogous manner, but are actually easier since the formulas are deterministic. Consider the autocovariance diagnostic first, and let $Z = \Sigma_W^{-1} W$. Then

$$\hat{A}_n(h) = O_P(1/n) + Z^{'}\underline{\Delta}_N \Sigma(f_U^2 \cdot l)\underline{\Delta}_N^{'} Z/n$$

using Lemmas 2 and 3. Here $l(\lambda) = \cos(\lambda h)$ and $L^{sym,h} = \Sigma(l)$ (note that $l \in \mathcal{C}^1$). Applying Lemma 1,

$$Z^{'}\underline{\Delta}_N \Sigma(f_U^2 \cdot l)\underline{\Delta}_N^{'} Z/n = W^{'}\Sigma_W^{-1}\Sigma\left(f_U^2 \cdot l \cdot |\delta^N(e^{-i\cdot})|^2\right)\Sigma_W^{-1}W/n.$$

Next, apply Lemma 4 to obtain

$$\hat{A}_n(h) = O_P(1) + W^{'}\overline{\Sigma}\left(f_U^2 \cdot l \cdot |\delta^N(e^{-i\cdot})|^2 \cdot f_W^{-2}\right)W.$$

Now since $1/f_W \in \mathcal{C}^1$ and $\mathcal{C}^1$ is closed under multiplication, it follows that $f_U^2 \cdot l \cdot |\delta^N(e^{-i\cdot})|^2 \cdot f_W^{-2} \in \mathcal{C}^1$, and we can apply Lemma 2:

$$\hat{A}_n(h) = O_P(1) + W^{'}\Sigma\left(f_U^2 \cdot l \cdot |\delta^N(e^{-i\cdot})|^2 \cdot f_W^{-2}\right)W.$$

Now under the additional assumption (B) or (HT), we can apply Proposition 1, since the function $f_U^2 \cdot l \cdot |\delta^N(e^{-i\cdot})|^2 \cdot f_W^{-2}$ is continuous and even (continuity of each function follows from its summability, and the product is continuous because f_W is strictly positive). This gives a limit theorem

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

where g is defined in the theorem's statement. Using the convergence for the mean and variance, the limit result is proved.

Next, consider the cross-covariance diagnostic. Letting $Z = \Sigma_W^{-1} W$ again, apply Lemma 2 several times to obtain

$$Z' \underline{\Delta}_N \Sigma_U L^h[0\,1]\Sigma_V \underline{\Delta}_S' Z = O_P(1) + Z' \underline{\Delta}_N \overline{\Sigma}(f_U) L^h[0\,1]\overline{\Sigma}(f_V)\underline{\Delta}_S' Z.$$

In forming the approximations $\overline{\Sigma}$, we utilize the dimension $n - d_S$ for both Σ_U and Σ_V. For the latter matrix, this differs from our former definition, since Σ_V is $n - d_N \times n - d_N$ dimensional. Explicitly,

$$\overline{\Sigma}(f_V) = G^* E G$$

with $G_{jk} = (n-d_S)^{-1/2} e^{i2\pi jk/(n-d_S)}$ and $E = diag\{g(\lambda_1), \cdots, g(\lambda_{n-d_S}), 0, \cdots\}$; $\lambda_j = 2\pi j/(n-d_S)$. It is easy to see that the approximation Lemmas are still valid when we use $n-d_S$ instead of $n-d_N$ in the definition of $\overline{\Sigma}(f_V)$. For convenience let $m = n - d_S$ and calculate as follows:

$$\begin{aligned}
&\left(H\,L^h\,[0\,1]\,G^*\right)_{jk} \\
&= \sum_{t=1}^{m} m^{-1/2} e^{-i2\pi j(t+h)/m} m^{-1/2} e^{i2\pi(t+d_S-d_N)k/m} \\
&= m^{-1} e^{-i2\pi jh/m} e^{i2\pi(d_S-d_N)k/m} \sum_{t=1}^{m} e^{-i2\pi t(j-k)/m} \\
&= \delta_{j,k} e^{-i2\pi j(h-d_S+d_N)/m}
\end{aligned}$$

where $\delta_{j,k}$ is the Kronecker delta (not to be confused with the coefficients of $\delta(z)$). This calculation illustrates why we needed to modify the definition of $\overline{\Sigma}(f_V)$. So the above matrix is $n-d_S \times n-d_N$ and "diagonal" in the sense that the only nonzero entries occur when the row and column index are equal. Call this matrix F. Then

$$\overline{\Sigma}(f_U) L^h[0\,1]\overline{\Sigma}(f_V) = H^*\,C\,F\,E\,G$$

and $C\,F\,E$ is another "rectangular diagonal" matrix, with jth diagonal entry given by

$$f_U(\lambda_j) f_V(\lambda_j) e^{-i\lambda_j(h-d_S+d_N)}$$

so long as $j \leq m$. Therefore,

$$
\begin{aligned}
&(H^* \, C \, F \, E \, G)_{jk} \\
&= \sum_{l,t} m^{-1/2} e^{i2\pi jl/m} \delta_{l,t} f_U(\lambda_l) f_V(\lambda_l) e^{-i\lambda_l (h - d_S + d_N)} m^{-1/2} e^{-i2\pi tk/m} \\
&= m^{-1} \sum_{t=1}^{m} e^{-i2\pi t(k-j)/m} f_U(\lambda_t) f_V(\lambda_t) e^{-i\lambda_t (h - d_S + d_N)},
\end{aligned}
$$

which can be viewed as just the jkth entry of a rectangular Toeplitz matrix $\overline{\Sigma}(f_U \cdot f_V \cdot e^{-i(h-d_S+d_N)\cdot})$. Applying the random vectors $Z^{'} \underline{\Delta}_N$ and $\underline{\Delta}_S^{'} Z$, the inner product is $O_P(1)$ away from the form

$$
Z^{'} \underline{\Delta}_N \Sigma(f_U \cdot f_V \cdot e^{-i(h-d_S+d_N)\cdot}) \underline{\Delta}_S^{'} Z,
$$

by applying Lemmas 2 and 3, extended to the rectangular scenario. Finally, we show how the differencing matrices account for the time delay factor $e^{-i(h-d_S+d_N)\lambda}$:

$$
\begin{aligned}
&\left(\underline{\Delta}_N \Sigma(f_U \cdot f_V \cdot e^{-i(h-d_S+d_N)\cdot}) \underline{\Delta}_S^{'} \right)_{jk} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f_U(\lambda) f_V(\lambda) e^{-i\lambda(h-d_S+d_N)} \sum_l \delta^N_{j+d_N-l} \sum_t \delta^S_{k+d_S-t} e^{i\lambda(l-t)} \, d\lambda \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f_U(\lambda) f_V(\lambda) \sum_l \delta^N_{j+d_N-l} e^{-i\lambda(j+d_N-l)} \sum_t \delta^S_{k+d_S-t} e^{i\lambda(k+d_S-t)} e^{-i\lambda(h-j+k)} \, d\lambda \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f_U(\lambda) f_V(\lambda) \delta^N(e^{-i\lambda}) \delta^S(e^{i\lambda}) e^{-i\lambda h} e^{i\lambda(j-k)} \, d\lambda \\
&= \Sigma_{jk} \left(f_U \cdot f_V \cdot \delta^N(e^{-i\cdot}) \cdot \delta^S(e^{i\cdot}) \cdot e^{-ih\cdot} \right),
\end{aligned}
$$

which can have complex entries. Notice that in order for the delay factors induced by the differencing matrices to cancel with $e^{-i(h-d_S+d_N)\lambda}$, it is necessary to trim $\hat{V}$ of its first $d_S - d_N$ entries, as a careful examination of our proof will reveal. At this point, we consider the symmetrization of this form, which amounts to symmetrizing the delay factors $\delta^N(e^{-i\cdot}) \cdot \delta^S(e^{i\cdot}) \cdot e^{-ih\cdot}$. Apply Lemma 4 to obtain

$$
\hat{C}_n(h) = O_P(1/n) + W^{'} \Sigma\,(g)\, W/n,
$$

using the fact that $1/f_W \in \mathcal{C}^1$ as before. From here we apply Proposition 1 and the proof is complete. □

Proof of Theorem 2. We begin with the mean and variance calculations.

$$
\mathbb{E}P(W) = \frac{tr(B\tilde{\Sigma}_W)}{n} - \mathbb{E}\hat{\sigma_\epsilon}^2 \frac{tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})}{n}
$$

and

$$
\mathbb{E}\hat{\sigma_\epsilon}^2 = \tilde{\sigma_\epsilon}^2 \frac{tr(\Sigma^{-1}_{W/\sigma_\epsilon} \tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})}{n-d}
$$

together yield the formula for the mean. For the variance we have

$$
\begin{aligned}
Var P(W) &= Var Q(W) - 2Cov(Q(W), \widehat{\mathbb{E}Q(W)}) + Var \widehat{\mathbb{E}Q(W)} \\
&= \frac{2}{n^2} tr(B\tilde{\Sigma}_W)^2 - \frac{4}{n^2(n-d)} tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}) tr(B\tilde{\Sigma}_W \Sigma^{-1}_{W/\sigma_\epsilon} \tilde{\Sigma}_W) \\
&+ \frac{2}{n^2(n-d)^2} tr^2(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}) tr(\Sigma^{-1}_{W/\sigma_\epsilon} \tilde{\Sigma}_W)^2,
\end{aligned}
$$

which is easily manipulated into the stated form. The limiting mean and variance formulas are justified using the same techniques as in Theorem 1. For the central limit theorem, we note that

$$
P(W) = \frac{1}{n} W' BW - \frac{\hat{\sigma_\epsilon}^2}{n} tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}) = \frac{1}{n} W' \left[B - tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}}) \Sigma^{-1}_{W/\sigma_\epsilon} \right] W.
$$

Thus $P(W)$ is a symmetric quadratic form, and the same asymptotic techniques apply to the matrix $B - tr(B\tilde{\Sigma}_{W/\tilde{\sigma_\epsilon}})\Sigma^{-1}_{W/\sigma_\epsilon}$. Asymptotically, the associated Toeplitz matrix is $\Sigma(g - \beta\sigma^2_\epsilon/f_W)$ with

$$
\beta = \tilde{\sigma_\epsilon}^{-2} \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda) \, d\lambda.
$$

Hence $P(W)$ is asymptotically normal with mean and variance given by

$$
\begin{aligned}
&\frac{1}{2\pi} \int_{-\pi}^{\pi} (g(\lambda) - \beta\sigma^2_\epsilon / f_W(\lambda)) \tilde{f}_W(\lambda) \, d\lambda \\
&n^{-1} \frac{2}{2\pi} \int_{-\pi}^{\pi} (g(\lambda) - \beta\sigma^2_\epsilon / f_W(\lambda))^2 \tilde{f}^2_W(\lambda) \, d\lambda
\end{aligned}
$$

respectively. It is easily checked that these are identical to the stated expressions for the limit of the mean and variance. □

References

[1] Bell, W. (1984) Signal Extraction for Nonstationary Time Series. *The Annals of Statistics* **12**, 646 –664.

[2] Bell, W. and Hillmer, S. (1988) A Matrix Approach to Likelihood Evaluation and Signal Extraction for ARIMA Component Time Series Models. SRD Research Report No. $RR-88/22$, Bureau of the Census.

[3] Box, G. and Jenkins, G. (1976) *Time Series Analysis,* San Francisco, California: Holden-Day.

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

[5] Findley, D., McElroy, T., Wills, K. (2004) "Diagnostics for ARIMA-Model-Based Seasonal Adjustment," *2004 Proceedings of the American Statistical Association* [CD-ROM]: Alexandria, VA.

[6] Findley, D. F., Monsell, B. C., Bell, W. R., Otto, M. C. and Chen, B. C. (1998) New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program. " *Journal of Business and Economic Statistics*, **16** 127–177 (with discussion).

[7] Gómez, V. and Maravall, A. (1997) Program SEATS Signal Extraction in ARIMA Time Series: Instructions for the User (beta version: June 1997), Working Paper 97001, Ministerio de Economía y Hacienda, Dirrectión General de Análysis y Programación Presupuestaria, Madrid.

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

[9] Hillmer, S. and Tiao, G. (1982) An ARIMA-Model-Based Approach to Seasonal Adjustment. *Journal of the American Statistical Association* **77**, 377, 63 – 70.

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

[11] Liggett, W.S. (1971) On the asymptotic optimality of spectral analysis for testing hypotheses about time series. *Annals of Mathematical Statistics* **42**, 1348–1358.

[12] Maravall, A. (2003) A Class of Diagnostics in the ARIMA-Model-Based Decomposition of a Time Series. Memorandum, Bank of Spain.

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

[14] McElroy, T. (2005) Matrix Formulas for Nonstationary Signal Extraction. SRD Research Report No. 2005/04, Bureau of the Census.

[15] Soukup, R. and Findley, D. (1999) On the spectrum diagnostics used by X-12-ARIMA to indicate the presence of trading day effects after modeling or adjustment. *Proceedings of the Business and Economic Statistics Section*, 144–149, American Statistical Association. Also www.census.gov/pub/ts/papers/rr9903s.pdf

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

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