

Statistical Properties of Signal Extraction Diagnostics

Tucker McElroy
U.S. Census Bureau

Abstract

A model-based diagnostic 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-smoothing” and “under-smoothing.” The new methods are demonstrated on a U.S. Census Bureau time series exhibiting seasonality.

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

Disclaimer This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

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 estimating an ambient signal accurately. 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 composite 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).

A model-based diagnostic for signal extraction was first described in Maravall (2003), and this basic idea was modified and studied in Findley, McElroy, and Wills (2004). The 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. It turns out that the diagnostics proposed in Findley, McElroy, and Wills (2004) can be further modified so that this interpretation of model rejection is mathematically correct.

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:

$$Y_t = S_t + N_t \quad (1)$$

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., they share no common zeroes), such that

$$U_t = \delta^S(B)S_t \quad V_t = \delta^N(B)N_t \quad (2)$$

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 $\delta^N(z) = 1 + z + z^2 + \dots + z^{11}$ 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 $t = 1, 2, \dots, n$. Assumption A states that the initial d values of Y_t , i.e., the variables Y_1, Y_2, \dots, 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

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

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

$$W_t = \delta^N(B)U_t + \delta^S(B)V_t \quad (3)$$

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):

$$W = \underline{\Delta}_N U + \underline{\Delta}_S V \quad (4)$$

We will be interested in estimates of U and V . The minimum mean squared error signal extraction estimate is $\hat{U}_t = \mathbb{E}[U_t|W]$, which is expressed as some linear function of the differenced data vector W when the data

are Gaussian; putting this together for each time t , we obtain the various rows of a matrix F :

$$\hat{U} = FW = \mathbb{E}[U|Y].$$

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):

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

which implicitly defines F .

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 actual 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 orders of the differencing operators d_S and d_N have been correctly ascertained.

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 . The alternative space consists of any other *ARMA* model for W_t , including different polynomial orders, coefficients, and innovation variance. However, the differencing orders d_S and d_N are the same for both the Null and Alternative models. Note that Σ_W could in practice be determined by parameter estimates, which are then treated 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

$$\begin{aligned} H_0 : \tilde{\Sigma}_W &= \Sigma_W \\ H_1 : \tilde{\Sigma}_W &\neq \Sigma_W \end{aligned} \quad (5)$$

In Section 2.2 we consider a modified testing framework that does not make any assumptions about the innovation variance of W_t . Note that the alternative space has no particular directionality that is naturally

associated with it, i.e., conducting a one-sided test is nonsensical. 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 . That work claimed asymptotic normality of the test statistic under H_0 ; in Section 2, this claim is verified under two different scenarios. Section 3 presents a modified test statistic, which arguably offers a ready interpretation to the rejection of H_0 ; also the power of the procedure is discussed.

2 Theoretical Results

The test statistic involves computing the sample second moment of \hat{U}_t and comparing this to its expectation under H_0 ; this quantity is not scale-invariant, so it is then normalized by its standard error under H_0 . We consider two different testing scenarios below; the first is more simplistic, while the second offers an interpretation that is more applicable. Our basic statistic is

$$\hat{T}_n = \frac{\hat{U}'\hat{U}}{n} = \frac{W'\Sigma_W^{-1}\underline{\Delta}_N\Sigma_U\Sigma_U'\underline{\Delta}_N\Sigma_W^{-1}W}{n}. \quad (6)$$

It will be necessary to discuss spectra to an extent, so for any stationary process $\{X_t\}$ we denote its spectral density by $f_X(\lambda)$; this is related to Σ_X by the formula

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

2.1 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 such do, and the mean and variance of (6) will as well. The following theorem summarizes the asymptotics of \hat{T}_n , as well as a normalized version denoted by $\underline{\tau}_n$, which is defined below.

Theorem 1 *Assume that Assumption A holds on the model decomposition (1), and that $\{U_t\}$ and $\{V_t\}$ are*

independent and purely nondeterministic. Also suppose that the third and fourth cumulants of the true DGP of W_t are zero, and that the true autocovariance function $\tilde{\gamma}_W$ satisfies $\sum_h |h| \tilde{\gamma}_W(h) < \infty$. Then the true mean and variance of \hat{T}_n are given by

$$\begin{aligned} \mathbb{E}\hat{T}_n &= \frac{1}{n} \text{tr}(A) \\ \text{Var}\hat{T}_n &= \frac{2}{n^2} \text{tr}(A^2) \end{aligned} \quad (7)$$

where tr denotes the trace of a matrix, and

$$A = \Sigma_U \underline{\Delta}'_N \Sigma_W^{-1} \tilde{\Sigma}_W \Sigma_W^{-1} \underline{\Delta}_N \Sigma_U.$$

Moreover, the mean and variance have limiting values as $n \rightarrow \infty$

$$\begin{aligned} \mathbb{E}\hat{T}_n &\rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda) d\lambda \\ \text{tr}(A^2)/n &\rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} g^2(\lambda) \tilde{f}_W^2(\lambda) d\lambda \end{aligned}$$

where $g(\lambda) = f_U^2(\lambda) |\delta^N(e^{-i\lambda})|^2 / f_W^2(\lambda)$ and \tilde{f}_W denotes the spectral density for the true DGP of W_t . The following Central Limit Theorem holds as $n \rightarrow \infty$:

$$\underline{\tau}_n = \sqrt{n} \frac{(\hat{T}_n - \text{tr}(A)/n)}{\sqrt{2 \text{tr}(A^2)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

In order to compute the mean and standard error, it is necessary to assume something about the DGP, such as 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 fixed (nonrandom) parameters. 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). It makes no sense to think of Σ_W as random in our Null hypothesis, since this amounts to assuming that our m.l.e.'s are without any error. The only sensible path is to treat estimated parameters as fixed and derive the corresponding interpretation.

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) 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. An example is given in Section 2.3 below.

2.2 Estimated Innovation Variance Case

Let σ_a^2 denote the innovation variance of W_t , and hence W/σ_a and U/σ_a do not depend on σ_a . Thus Σ_{W/σ_a} is computed from the *ARMA* model for W_t but assuming a unit innovation variance; Σ_{U/σ_a} is based on the *ARMA* model for U_t with innovation variance given in units of σ_a^2 . Hence we can write $F = \Sigma_{U/\sigma_a} \underline{\Delta}'_N \Sigma_{W/\sigma_a}^{-1}$ which no longer requires estimation of σ_a . Thus, we can compute \hat{T}_n without prior knowledge (or estimation) of σ_a , but the matrix A *does* depend on the innovation variance:

$$\begin{aligned} A &= \Sigma_{U/\sigma_a} \underline{\Delta}'_N \Sigma_{W/\sigma_a}^{-1} \tilde{\Sigma}_W \Sigma_{W/\sigma_a}^{-1} \underline{\Delta}_N \Sigma_{U/\sigma_a} \\ &= \tilde{\sigma}_a^2 \Sigma_{U/\sigma_a} \underline{\Delta}'_N \Sigma_{W/\sigma_a}^{-1} \tilde{\Sigma}_W / \tilde{\sigma}_a \Sigma_{W/\sigma_a}^{-1} \underline{\Delta}_N \Sigma_{U/\sigma_a} \end{aligned}$$

where $\tilde{\sigma}_a^2$ is the innovation variance of the true DGP of W_t . Hence, in order to calculate the expected value of \hat{T}_n , we must know $\tilde{\sigma}_a$; in this section, we consider maximum likelihood estimation, i.e.,

$$\hat{\sigma}_a^2 = \frac{1}{n-d} W' \Sigma_{W/\sigma_a}^{-1} W \quad (8)$$

which depends on the hypothesized unit innovation variance model for W_t . Let us denote by \hat{A} the following “innovation-free” version of A :

$$\hat{A} = \Sigma_{U/\sigma_a} \underline{\Delta}'_N \Sigma_{W/\sigma_a}^{-1} \tilde{\Sigma}_W / \tilde{\sigma}_a \Sigma_{W/\sigma_a}^{-1} \underline{\Delta}_N \Sigma_{U/\sigma_a}$$

which can be calculated under a specified hypothesis on the model for $\tilde{\Sigma}_W / \tilde{\sigma}_a$. In this section we consider a testing framework that is a slightly modified version of (5):

$$\begin{aligned} H'_0 &: \tilde{\Sigma}_W / \tilde{\sigma}_a = \Sigma_{W/\sigma_a} \\ H'_1 &: \tilde{\Sigma}_W / \tilde{\sigma}_a \neq \Sigma_{W/\sigma_a} \end{aligned} \quad (9)$$

This only assesses the model parameters, apart from the innovation variance; hence it is an “innovation-free” version of (5). Thus under H'_0 ,

$$\hat{A}|_{H'_0} = \Sigma_{U/\sigma_a} \underline{\Delta}'_N \Sigma_{W/\sigma_a}^{-1} \underline{\Delta}_N \Sigma_{U/\sigma_a}$$

by (9). Then, under H'_0 , the computable quantity $\hat{\sigma}_a^2 \text{tr}(\hat{A})/n$ is an estimate of the Null hypothesis mean of \hat{T}_n . It turns out that \hat{T}_n and $\hat{\sigma}_a^2$ are highly correlated, and each is asymptotically normal. Therefore, it is necessary to compute a new standard error for $\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})$; simply using $\sqrt{2\hat{\sigma}_a^4 \text{tr}(\hat{A}^2)/n}$ will not be correct. The mean and variance of $\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})$ are given by

$$\begin{aligned} \mathbb{E}[\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})/n] & \\ &= \frac{\tilde{\sigma}_a^2}{n} \text{tr}(\hat{A}) \left(1 - \frac{\text{tr}(\Sigma_{W/\sigma_a}^{-1} \tilde{\Sigma}_W / \tilde{\sigma}_a)}{n-d} \right) \end{aligned} \quad (10)$$

$$\begin{aligned} \text{Var}[\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})] & \\ &= \frac{2\tilde{\sigma}_a^4}{n} \left(\frac{\text{tr}(\hat{A}^2)}{n} - 2 \frac{\text{tr}(\hat{B}) \text{tr}(\hat{A})}{n(n-d)} + \frac{\text{tr}(\hat{C}) (\text{tr}(\hat{A}))^2}{n(n-d)^2} \right) \end{aligned}$$

where the matrices \hat{C} and \hat{B} are defined by

$$\begin{aligned} \hat{C} &= \Sigma_{W/\sigma_a}^{-1} \tilde{\Sigma}_W / \tilde{\sigma}_a \Sigma_{W/\sigma_a}^{-1} \tilde{\Sigma}_W / \tilde{\sigma}_a \\ \hat{B} &= \Sigma_{U/\sigma_a} \underline{\Delta}'_N \hat{C} \Sigma_{W/\sigma_a}^{-1} \underline{\Delta}_N \Sigma_{U/\sigma_a}. \end{aligned}$$

Under H'_0 , the mean becomes zero and the variance simplifies greatly to

$$\frac{2\tilde{\sigma}_a^4}{n} \left(\frac{\text{tr}((\hat{A}|_{H'_0})^2)}{n} - \frac{(\text{tr}(\hat{A}|_{H'_0}))^2}{n(n-d)} \right) \quad (11)$$

which can be computed by substituting $\hat{\sigma}_a$ for $\tilde{\sigma}_a$. The asymptotics of such a standardized statistic, which we denote by $\bar{\tau}_n$, are presented below.

Theorem 2 *Make the same assumptions as Theorem 1. Then the following limit theorem holds:*

$$\bar{\tau}_n = \frac{(\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A}) - \mathbb{E}[\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})/n])}{\sqrt{\text{Var}[\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})]}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

as $n \rightarrow \infty$. The limits of the mean and variance are given as follows:

$$\begin{aligned} \mathbb{E}[\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})/n] &\rightarrow \\ &\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda) d\lambda \right) \left(1 - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{f}_W(\lambda)}{f_W(\lambda)} d\lambda \right) \\ n \text{Var}[\hat{T}_n - \hat{\sigma}_a^2 \text{tr}(\hat{A})] &\rightarrow \\ &2 \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g^2(\lambda) \tilde{f}_W^2(\lambda) d\lambda \right) \\ &- 4 \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \frac{\tilde{f}_W^2(\lambda)}{f_W(\lambda)} d\lambda \right) \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda) d\lambda \right) \\ &+ 2 \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\tilde{f}_W^2(\lambda)}{f_W^2(\lambda)} d\lambda \right) \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda) d\lambda \right)^2 \end{aligned}$$

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 the H'_0 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 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 would be false. So only the innovation variance is to be re-estimated in this interpretation.

To compute $\bar{\tau}_n$ under H'_0 , our estimate of the standard error is

$$\frac{2\hat{\sigma}_a^4}{n} \left(\frac{\text{tr}((\hat{A}|_{H'_0})^2)}{n} - \frac{(\text{tr}(\hat{A}|_{H'_0}))^2}{n(n-d)} \right)$$

which has the same asymptotics as (11) by Slutsky’s Theorem. In the case that estimated parameters are actually used to compute the H'_0 filters, the estimated standard error will simplify to

$$\frac{2}{n} \left(\frac{\text{tr}((A|_{H'_0})^2)}{n} - \frac{(\text{tr}(A|_{H'_0}))^2}{n(n-d)} \right)$$

which is less than the H_0 standard error of $\underline{\tau}_n$:

$$\frac{2}{n} \left(\frac{\text{tr}((A|_{H'_0})^2)}{n} \right)$$

This reduced variation is due to the fact that the variability in estimating the innovation variance has been accounted for. Hence it is easier to reject H'_0 than H_0 for the same series. The trade-off is that rejection of H'_0 provides less information than rejection of H_0 does.

To summarize, letting $\dot{A} = \hat{\sigma}_a^2 \hat{A}|_{H'_0}$,

$$\bar{\tau}_n = \sqrt{n} \frac{\hat{T}_n - \dot{A}}{\sqrt{\frac{\text{tr}(\dot{A}^2)}{n} - \frac{\text{tr}^2(\dot{A})}{n(n-d)}}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

as $n \rightarrow \infty$, when H'_0 is true.

2.3 Example

Consider the following application of Theorems 1 and 2 to the series of U.S. Exports of Other Agricultural Materials (non-Manufactured), from January 1989 – December 2001 ($N = 156$). Suppose that we wish to

estimate a trend from the logged series, fitted to the Box-Jenkins airline model:

$$W_t = (1 - B)^2 U(B) Y_t = (1 - \theta B)(1 - \Theta B^{12}) a_t$$

where $U(z) = 1 + z + z^2 + \dots + z^{11}$. So S_t is a trend, and N_t represents seasonal plus irregular. The maximum likelihood estimates are $\hat{\theta} = .799$, $\hat{\Theta} = .801$, and $\hat{\sigma}_a^2 = .011$. Based on a sample size of $n = 156$, we find the values

$$\begin{aligned} \hat{T}_{156} &= 0.00000535 \\ \underline{\tau}_{156} &= 1.459 \\ \bar{\tau}_{156} &= 2.101 \end{aligned}$$

which indicate a rejection of H_0 and H'_0 with p -values .144 and .036 respectively (using two-sided alternatives). The interpretation of the rejection of H_0 is that an independent replicate of the DGP ($\theta = .799$, $\Theta = .801$, $\sigma_a^2 = .011$) using that model for a filter, would only have such an extreme $\underline{\tau}_{156}$ statistic with probability .144. But the rejection of H'_0 implies that an independent replicate of the DGP ($\theta = .799$, $\Theta = .801$, $\sigma_a^2 = .011$), with filter derived from this model with innovation variance re-estimated (but assuming this model to compute (8)), would only have such an extreme $\bar{\tau}_{156}$ statistic with probability .036. The rejection of H'_0 indicates model inadequacy. Since the average sum of squares of the differenced trend estimate is too large, this may indicate that insufficient smoothing in the low frequencies has taken place, possibly pointing us to adjusting θ downward. However, there is little theoretical basis for this conclusion at this point, because there is no natural directionality to the alternative hypothesis for these statistics.

3 Extensions

By modifying \hat{T}_n slightly, we can improve the interpretability of the signal extraction diagnostics, which will allow us to determine a form of directionality for rejection of H_0 .

3.1 Modifying the Diagnostics

The covariance matrix Σ_U has a Cholesky factorization $\Sigma_U = \sqrt{\Sigma_U'} \sqrt{\Sigma_U}$; consider a modified estimate of U :

$$\check{U} = \sqrt{\Sigma_U'}^{-1} \hat{U} = \sqrt{\Sigma_U} \Delta'_N \Sigma_W^{-1} W$$

So \hat{T}_n defined in (6) is modified to

$$\check{T}_n = \frac{\check{U}' \check{U}}{n} = \frac{W' \Sigma_W^{-1} \Delta'_N \Sigma_U \Delta'_N \Sigma_W^{-1} W}{n}$$

Essentially the same results carry through for this modified statistic, but with $\Sigma_U \Sigma_U$ replaced everywhere by Σ_U . Note that under H_0 , the mean and variance of \tilde{T}_n are innovation-free, so there is no need to estimate σ_a . Hence, we will focus on the analogue of Theorem 1.

Further, let a noise extraction estimate be $\check{V} = \sqrt{\Sigma_V} \underline{\Delta}'_S \Sigma_W^{-1} W$, so that our corresponding diagnostic is

$$\frac{\check{V}' \check{V}}{n} = \frac{W' \Sigma_W^{-1} \underline{\Delta}_S \Sigma_V \underline{\Delta}'_S \Sigma_W^{-1} W}{n}$$

from which it follows that the sum of both diagnostics is

$$\frac{\check{U}' \check{U}}{n} + \frac{\check{V}' \check{V}}{n} = \frac{W' \Sigma_W^{-1} W}{n}$$

which is proportional to the m.l.e. for the innovation variance. Hence, the signal and noise diagnostics are inversely related, since their sum is a constant that does not depend on the component models. Moreover this additivity property, which captures some of the intuitive notions motivating Maravall (2003), does not hold for the original diagnostics of (6). We summarize the asymptotics of \tilde{T}_n below.

Theorem 3 *Make the same assumptions as Theorem 1. The true mean and variance of \tilde{T}_n are given by (7), where A is now defined as*

$$A = \underline{\Delta}_N \Sigma_U \underline{\Delta}'_N \Sigma_W^{-1} \tilde{\Sigma}_W \Sigma_W^{-1}. \quad (12)$$

Moreover, the mean and variance have limiting values as $n \rightarrow \infty$

$$\begin{aligned} \mathbb{E} \tilde{T}_n &\rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} g(\lambda) \tilde{f}_W(\lambda) d\lambda \\ \text{tr}(A^2)/n &\rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} g^2(\lambda) \tilde{f}_W^2(\lambda) d\lambda \end{aligned}$$

where now $g(\lambda) = f_U(\lambda) |\delta^N(e^{-i\lambda})|^2 / f_W^2(\lambda)$. The following Central Limit Theorem holds as $n \rightarrow \infty$:

$$\tau_n = \sqrt{n} \frac{(\tilde{T}_n - \text{tr}(A)/n)}{\sqrt{2 \text{tr}(A^2)/n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

Note that the function g can be written

$$g(\lambda) = \frac{f_U(\lambda) |\delta(e^{-i\lambda})|^2}{f_W^2(\lambda)} = \frac{f_S(\lambda)}{f_Y(\lambda)} \frac{1}{f_W(\lambda)}$$

where f_S and f_Y are the pseudo-spectral densities of S_t and Y_t under H_0 . Hence the limiting expectation of \tilde{T}_n under H_0 is

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_S(\lambda) \tilde{f}_W(\lambda)}{f_Y(\lambda) f_W(\lambda)} d\lambda$$

which is a weighted average of the discrepancy between model spectrum f_W and DGP spectrum \tilde{f}_W , with high weight being given to those frequencies pertaining to the signal spectrum f_S . This observation will allow us to give a directionality to the rejection of H_0 , as discussed below.

3.2 Power and Interpretations of Test Rejection

We can compute asymptotic formulas for the power of the diagnostic based on \tilde{T}_n . Let A_0 denote the matrix A of (12) computed under H_0 , whereas A_1 is computed under H_1 . Using the asymptotic normality of \tilde{T}_n , we reject H_0 at level α if \tilde{T}_n is outside the interval

$$\text{tr}(A_0)/n \pm z_{1-\alpha/2} \frac{\sqrt{2 \text{tr}(A_0^2)/n}}{\sqrt{n}}$$

Hence β , the probability of Type II error, is by Theorem 3 approximately equal to

$$\begin{aligned} &\Phi \left(\frac{z_{1-\alpha/2} \frac{\sqrt{2 \text{tr}(A_0^2)/n}}{\sqrt{n}} + \text{tr}(A_0 - A_1)/n}{\frac{\sqrt{2 \text{tr}(A_1^2)/n}}{\sqrt{n}}} \right) \\ &- \Phi \left(\frac{z_{\alpha/2} \frac{\sqrt{2 \text{tr}(A_0^2)/n}}{\sqrt{n}} + \text{tr}(A_0 - A_1)/n}{\frac{\sqrt{2 \text{tr}(A_1^2)/n}}{\sqrt{n}}} \right) \end{aligned}$$

where Φ denotes the standard normal cumulative distribution function. Now asymptotically,

$$\text{tr}(A_0 - A_1)/n \rightarrow \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f_S(\lambda)}{f_Y(\lambda)} \left(1 - \frac{\tilde{f}_W(\lambda)}{f_W(\lambda)} \right) d\lambda \quad (13)$$

as $n \rightarrow \infty$. Let us refer to the range of frequencies where f_S is high relative to f_Y as the ‘‘spectral range’’ of S_t . Now if $\tilde{f}_W \gg f_W$ in the spectral range of S_t , then (13) is negative and \tilde{T}_n is too large. But if $\tilde{f}_W \ll f_W$ in the spectral range, then (13) is positive with an upper bound of $\text{trace}(A_0)/n$, and \tilde{T}_n is too small. Now when $\tilde{f}_W \ll f_W$ in a spectral band, the model is too chaotic for those frequencies, and hence it oversmooths (or over-estimates) the data. Conversely, $\tilde{f}_W \gg f_W$ indicates the model is too stable, and hence it undersmooths (or under-estimates) the data. Outside the spectral range of S_t these interpretations are not meaningful, since the pseudo-spectrum of S_t will damp discrepancies between model and DGP spectrum.

These observations can be used to form a meaningful one-sided testing procedure. First we summarize the

logic:

$$\begin{aligned} \tau_n \text{ is significantly negative} &\Leftrightarrow \check{T}_n \text{ is too small} \\ &\Leftrightarrow \check{f}_W \ll f_W \\ &\Leftrightarrow \text{over-smoothing} \\ \tau_n \text{ is significantly positive} &\Leftrightarrow \check{T}_n \text{ is too large} \\ &\Leftrightarrow \check{f}_W \gg f_W \\ &\Leftrightarrow \text{under-smoothing} \end{aligned}$$

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

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

Then the upper one-sided test has hypotheses

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

and H_0 is rejected with confidence $1 - \alpha$ if $\tau_n > z_{1-\alpha}$, which indicates significant under-smoothing in the relevant spectral band. The lower one-sided test has hypotheses

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

and H_0 is rejected with confidence $1 - \alpha$ if $\tau_n < z_\alpha$, which indicates significant over-smoothing in the appropriate spectral band.

For example, if S_t is a trend, then the spectral range consists of the low frequencies. In an airline model, θ governs the trend behavior; a significantly high \check{T}_n means that there is undersmoothing, which indicates that the value of θ is too high. A significantly low \check{T}_n indicates oversmoothing, or that θ is too low. Of course, one might also try to fix the problem by adjusting Θ upwards or downwards respectively, since the trend diagnostic is inversely related to the seasonal plus irregular diagnostic. However, the example below shows this may not be a wise approach.

3.3 Example

Continuing the example of subsection 2.3, we now compute \check{T}_{156} and τ_{156} for the trend:

$$\check{T}_{156} = 0.00094 \quad \tau_{156} = -3.537$$

with (two-sided test) p -value 0.0002, which strongly suggests oversmoothing of the trend. By adjusting the value of θ from .799 upwards, we can try to correct the

oversmoothing. For a filter using $\theta = .98$ (and the same Θ), we obtain

$$\check{T}_{156} = 0.00042 \quad \tau_{156} = -1.526$$

with p -value 0.063, which is no longer significant. Adjusting Θ downwards is less effective, since using the model $\theta = .799$, $\Theta = .01$ yields

$$\check{T}_{156} = 0.0014 \quad \tau_{156} = -2.991$$

with p -value 0.0014. The diagnostic tells us that the trend is oversmoothed, and increased smoothing of the seasonal-irregular will not ameliorate this deficiency.

References

- [1] Bell, W. (1984) Signal Extraction for Nonstationary Time Series. *The Annals of Statistics* **12**, 646–664.
- [2] Bell, W. and Hilmer, S. (1988) A Matrix Approach to Likelihood Evaluation and Signal Extraction for ARIMA Component Time Series Models. SRD Research Report No. *RR – 88/22*, U.S. Census Bureau.
- [3] 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.
- [4] 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).
- [5] Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- [6] Hilmer, S. and Tiao, G. (1982) An ARIMA-Model-Based Approach to Seasonal Adjustment. *Journal of the American Statistical Association* **77**, 377, 63–70.
- [7] Maravall, A. (2003) A Class of Diagnostics in the ARIMA-Model-Based Decomposition of a Time Series. Memorandum, Bank of Spain.
- [8] McElroy, T. (2005) Matrix Formulas for Nonstationary Signal Extraction. SRD Research Report No. *RRS2005/04*, U.S. Census Bureau. www.census.gov/srd/papers/pdf/rrs2005-04.pdf