MODEL DIAGNOSTICS AND VALIDATION FOR LINEAR MODEL FITTING USING HIGHER-ORDER STATISTICS

Ergang Liu & Jitendra K. Tugnait
Dept. of Electrical Engineering
Auburn University, Auburn, Alabama 36849, USA
tugnait@eng.auburn.edu

ABSTRACT

Given a linear stationary non-Gaussian signal, suppose that we fit a linear model using higher-order statistics and one of several existing methods. The model is fitted under certain assumptions on the data and the underlying (true) model. Having obtained a model, how do we know if the fitted model is "good?" This paper is devoted to the problem of model diagnostics and validation. We propose some simple frequency-domain tests that are applicable to both third-order and fourth-order statistics-based model fitting unlike existing tests. A computer simulation example is presented to illustrate the proposed tests.

1 Introduction

The area of parametric modeling via higher-order cumulant functions has attracted considerable attention in recent years [3],[4]. Use of higher order statistics allows one to identify noncausal as well as nonminimum phase finite-dimensional parametric models from system output measurements alone (blind identification). Most of the published papers thus far have concentrated upon various aspects of parameter estimation and model order selection including algorithm development and analysis. It appears that only [1] has addressed the problem of model validation. This paper is devoted to the problem of model validation using higher order statistics which is appropriate when the model has been fitted using higher order statistics. Model validation involves testing to see if the fitted model is an appropriate representation of the underlying (true) system. It involves devising appropriate statistical tools to test the validity of the assumptions made in obtaining the fitted model.

Given the data and an appropriately fitted linear model, in order to validate the model, we first inverse filter the data using the fitted model. Then the linear model validation problem is cast into a classical binary hypothesis testing problem. Under the null hypothesis that fitted model generates the data, the inverse filtered data is an i.i.d. non-Gaussian sequence, possibly contaminated with a colored Gaussian noise sequence. Under the alternative hypothesis that the fitted model does not describe the given data, the inverse filtered data is a colored non-Gaussian sequence, possibly contaminated with a colored Gaussian noise sequence. The model validation test considered in [1] is based upon testing for constancy of the bispectrum of the inverse filtered data. It is restricted to third-order statistics-based model fitting. In this paper we use some simple frequency-domain tests using integrated polyspectra [5] that are applicable to both third-order and fourth-order statistics-based model fitting.

This work was supported by the National Science Foundation under Grant MIP-9312559.

2 Model Assumptions

Let $\{s(t)\}$ denote a stationary ARMA(p,q) signal given by

$$\sum_{i=0}^{p} a_i\, s(t-i) \;=\; \sum_{i=0}^{q} b_i\, w(t-i), \quad a_0 := 1, \;\; b_0 := 1. \tag{2-1}$$

The measurements of the signal are noisy

$$x(t) \;=\; s(t) \;+\; v(t). \tag{2-2}$$

The input $\{w(t)\}$ is not observed. The following conditions are assumed to hold.

- (H1) $A(z) = \sum_{i=0}^{p} a_i\, z^{-i} \neq 0$ for $|z| \geq 1$.
- (H2) $B(z) = \sum_{i=0}^{q} b_i\, z^{-i} \neq 0$ for $|z| = 1$.
- (H3) The random sequence $\{w(t)\}$ is i.i.d., zero-mean and non-Gaussian such that its $r-$th cumulant γ_{rw} is nonzero for either $r = 3$ and 4 or $r = 4$ and 6. Moreover its moments up to order twelve are assumed to be bounded.
- (H4) The zero-mean noise $\{v(t)\}$ is independent of $\{w(t)\}$ and is colored Gaussian such that $|\mathrm{cov}\{v(t_1), v(t_2)\}| \;\leq\; M\beta^{|t_1-t_2|}$ for some $0 < M < \infty$ and $0 < \beta < 1$, and for all t_1 and t_2.

Condition (H1) can be relaxed to (H1′): $A(z) \neq 0$ for $|z| = 1$. Several schemes are available in the literature to estimate the unknown parameter vector $\theta = (a_1, \cdots, a_p,\, b_1, \cdots, b_q)$ given a sample sequence of observations $X_N = \{x(t),\; 1 \leq t \leq N\}$ [1],[3],[4].

3 Model Diagnostics and Validation

The basic premise of the proposed model diagnostics and validation procedures is just as in [1]. Suppose that we fit a linear model to the noisy data based solely upon the third-order or the fourth-order

cumulant sequence of $\{x(t)\}$. Let $\widehat{\theta}$ denote the parameter vector for a linear model that is to be validated given X_N. Let $\{\widehat{h}(i;\widehat{\theta}),\ i \geq 0\}$ denote the impulse response of the model parametrized by $\widehat{\theta}$. Let $\{\widehat{g}(i),\ -\infty < i < \infty\}$ denote its inverse such that $\sum_{l=-\infty}^{\infty} \widehat{g}(l)\widehat{h}(i-l;\widehat{\theta}) = \delta(i)$. Because we can not resolve the ambiguity concerning the scale factor and time shift of the true impulse response, it follows that $\sum_{l=-\infty}^{\infty} \widehat{g}(l)\widehat{h}(i-l;\theta_0) \approx c\delta(i-i_0)$ where c and i_0 are some constant and integer, respectively, and θ_0 denotes the true parameter vector. Define

$$v'(t) := \sum_{i=-\infty}^{\infty} \widehat{g}(i)v(t-i), \quad s'(t) := \sum_{i=-\infty}^{\infty} \widehat{g}(i)s(t-i),$$

$$x'(t) \ = \ \sum_{i=-\infty}^{\infty} \widehat{g}(i)x(t-i). \qquad (3-1)$$

Under the null hypothesis H_0 that the fitted model $\widehat{\theta}$ is the true underlying model, we have $s'(t) = cw(t-i_0)$ and $v'(t)$ is Gaussian.

Thus, after linear inverse filtering, we have a classical binary hypothesis testing problem:

$$H_0 \ : \ x'(t) \ = \ cw(t-i_0) + v'(t), \quad t = 1, 2, \cdots, N,$$

$$H_1 \ : \ x'(t) \ = \ s'(t) + v'(t), \quad t = 1, 2, \cdots, N, \qquad (3-2)$$

where N is "large" and under the alternative hypothesis H_1, $\{s(t)\}$ is some other linear or nonlinear process, therefore, $\{s'(t)\}$ is also a non-i.i.d. random sequence. The test statistic discussed in [1] is based upon testing for constancy of the bispectrum of the filtered measurements. In this paper we propose to use two slices of higher-order cumulant sequences of the filtered measurements to test for higher-order whiteness of $\{x'(t)\}$ under H_0.

3.1 Third-Order Statistics-Based Fitting

Consider the following two slices of cumulants of $\{x'(t)\}$:

$$C_{3x'}(\tau) \ := \ E\{y_{2x'}(t)x'(t+\tau)\}$$

$$C_{4x'}(\tau) \ := \ E\{y_{3x'}(t)x'(t+\tau)\} \qquad (3-3)$$

where

$$y_{2x'}(t) \ := \ x'^2(t) - E\{x'^2(t)\}$$

$$y_{3x'}(t) \ := \ x'^3(t) - 3E\{x'^2(t)\}x'(t) - E\{x'^3(t)\}. \qquad (3-4)$$

It then follows that

$$C_{3x'}(\tau) \ := \ \mathrm{cum}_3\{x'(t), x'(t), x'(t+\tau)\}, \qquad (3-5)$$

$$C_{4x'}(\tau) \ := \ \mathrm{cum}_4\{x'(t), x'(t), x'(t), x'(t+\tau)\}. \qquad (3-6)$$

It is easy to see that under H_0, $C_{3x'}(\tau) = 0 = C_{4x'}(\tau)\ \forall \tau \neq 0$. An important result is the converse of the preceding statement.

Lemma 1. Let $\{x'(t)\}$ be as in (3-1) with $\{\widehat{g}(i)\}$ denoting the impulse response function of a stable rational transfer function. If $C_{3x'}(\tau) = 0 = C_{4x'}(\tau)$ $\forall \tau \neq 0$ where $\{x'(t)\}$ obeys (2-1),(2-2) and (3-1), then $x'(t) = cw(t-i_0) + v'(t)$. •

Proof: Let $\mathcal{S}_r(z)$ denote the Z-transform of the sequence $\{C_{rx'}(\tau)\}$ (r=3 or 4). The hypothesis of the lemma implies that

$$\mathcal{S}_3(z) = c_1 \quad \text{and} \quad \mathcal{S}_4(z) = c_2 \ \ \forall z \qquad (3-7)$$

where c_1 and c_2 are some constants. Let the transfer function of the overall concatenated system comprised of the original system followed by the inverse filter with $w(t)$ as input and $x'(t)$ as output, be denoted by $H(z)$. Then it follows that

$$\mathcal{S}_3(z) \ = \ \gamma_{3w}H_2(z)H(z), \qquad (3-8)$$

$$\mathcal{S}_4(z) \ = \ \gamma_{4w}H_3(z)H(z) \qquad (3-9)$$

where γ_{rw} denotes the $r-$th cumulant of the random variable $w(t)$ and where if $H(z) := \sum_{i=-\infty}^{\infty} h(i)z^{-i}$, then

$$H_2(z) := \sum_{i=-\infty}^{\infty} h^2(i)z^{-i} \qquad (3-10)$$

and

$$H_3(z) := \sum_{i=-\infty}^{\infty} h^3(i)z^{-i}. \qquad (3-11)$$

It also follows from the hypotheses of the lemma that $H(z)$ is a stable, rational transfer function. It further follows from (3-7)-(3-11) that, for some constant $d(\neq 0)$, we have

$$dH_2(z) \ = \ H_3(z) \ \ \forall z$$

$$\Rightarrow \ \sum_{i=-\infty}^{\infty} h^2(i)[d-h(i)]z^{-i} = 0 \ \ \forall z$$

$$\Rightarrow \ h^2(i)[d-h(i)] = 0 \ \ \forall i. \qquad (3-12)$$

Therefore, either $h(i) = 0$ or $h(i) = d$ for any given i. Two cases arise:

Case A. $h(i) = d$ for infinitely many i's. This implies that $H(z)$ is *unstable* – a contradiction.

Case B. $h(i) = d$ for finitely many i's. This implies that $H(z)$ is FIR (finite impulse response). Let i_L and $i_H \geq i_L$ be two integers such that $h(i) = 0$ for any $i < i_L$ and for any $i > i_H$ but $h(i_L) \neq 0$ and $h(i_H) \neq 0$. Then we have

$$C_{3x'}(i_H - i_L) \ = \ \gamma_{3w}\, d^3$$

which violates a hypothesis of the lemma unless $i_H = i_L$, in which case $H(z) = dz^{-i_L}$ yielding the desired result. □

Thus, higher-order whiteness of two cumulant slices of respectively different order cumulants implies that

H_0 is true. [It appears that $C_{3x'}(\tau) = 0\ \forall \tau \neq 0$ alone does not necessarily imply H_0.]

Define ($r = 3$ or 4)

$$S_{rx'}(\omega) := \sum_{\tau=-\infty}^{\infty} C_{rx'}(\tau) e^{-j\omega\tau} = \mathcal{S}(e^{j\omega}). \quad (3-13)$$

The above quantities have been called integrated polyspectra in [5]. Lemma 2 then follows trivially from Lemma 1.

Lemma 2. Let $\{x'(t)\}$ be as in (3-1) with $\{\widehat{g}(i)\}$ denoting the impulse response function of a stable rational transfer function. If $S_{3x'}(\omega) = c_1$ and $S_{4x'}(\omega) = c_2\ \forall\omega$ where c_1 and c_2 are some constants and $\{x'(t)\}$ obeys (2-1),(2-2) and (3-1), then $x'(t) = cw(t - i_0) + v'(t)$. •

Thus, constancy of two integrated polyspectra of respectively different orders implies that H_0 is true. Our proposed test for the binary hypothesis testing problem (3-2) is based upon Lemma 2.

Given $\{x'(t),\ 1 \leq t \leq N\}$, calculate $\{y_{2x'}(t),\ 1 \leq t \leq N\}$ and $\{y_{3x'}(t),\ 1 \leq t \leq N\}$ replacing expectations in (3-4) with appropriate sample averages. Let $X'(\omega)$ denote the DFT of $\{x'(t),\ \ 1 \leq t \leq N\}$ given by

$$X'(\omega_k) \ = \ \sum_{t=0}^{N-1} x'(t+1)\exp(-j\omega_k t), \quad (3-14)$$

$$\omega_k \ = \ \frac{2\pi}{N}k, \quad k = 0, 1, \cdots, N-1. \quad (3-15)$$

Similarly define $Y_{2x'}(\omega_k)$ and $Y_{3x'}(\omega_k)$. Given the above DFT's, define the cross-spectrum (integrated polyspectrum [5]) estimators as

$$\widehat{S}_{3x'}(k) := \frac{1}{N(2m_N+1)} \sum_{i=-m_N}^{m_N} X'(\omega_{k-i}) Y^*_{2x'}(\omega_{k-i}), \quad (3-16)$$

$$\widehat{S}_{4x'}(k) := \frac{1}{N(2m_N+1)} \sum_{i=-m_N}^{m_N} X'(\omega_{k-i}) Y^*_{3x'}(\omega_{k-i}). \quad (3-17)$$

Let us choose m_N to be such that as $N \to \infty$, we have $m_N N^{-1} \to 0$ and $m_N \to \infty$. In light of (3-16) define a coarser frequency grid:

$$\omega_l = \frac{2\pi l}{L_N} = \frac{2\pi l(2m_N+1)}{N} \quad (3-18)$$

with $l = 0, 1, \cdots, L_N - 1$ where $L_N = \lfloor \frac{N}{2m_N+1} \rfloor$. It then follows from [7] that asymptotically,

$$\widehat{S}_{3x'}(k) \sim \mathcal{N}^C\left(S_{3x'}(\omega_k), \Delta_N^{-1} S_{x'x'}(\omega_k) S_{22}(\omega_k)\right), \quad (3-19)$$

$$\widehat{S}_{4x'}(k) \sim \mathcal{N}^C\left(S_{4x'}(\omega_k), \Delta_N^{-1} S_{x'x'}(\omega_k) S_{33}(\omega_k)\right) \quad (3-20)$$

where $\Delta_N = 2m_N + 1$, $\mathcal{N}^C(\mu, \sigma^2)$ denotes a complex (circularly symmetric) Gaussian [7] distribution with mean μ and variance σ^2, and $S_{22}(\omega)$ and $S_{33}(\omega)$ denote the power spectra of $y_{2x'}(t)$ and $y_{3x'}(t)$, respectively. Moreover, the estimators $\widehat{S}_{3x'}(k)$ for various k's on the coarse grid (3-18) are asymptotically mutually independent. The same is true for $\widehat{S}_{4x'}(k)$.

Pick P points on the coarse grid (3-18) in the interval $(0, \pi)$; call this set Ω_P. Define the vectors/matrices

$$\widehat{\mathbf{R}}_{3x'} \ := \ \left[\ \widehat{S}_{3x'}(l),\ l \in \Omega_P\ \right]^T, \quad (3-21)$$

$$\widehat{\mathbf{R}}_{4x'} \ := \ \left[\ \widehat{S}_{4x'}(l),\ l \in \Omega_P\ \right]^T, \quad (3-22)$$

$$\Sigma_3 \ := \ \mathrm{cov}\left\{\widehat{\mathbf{R}}_{3x'}\right\} = \ \text{a diagonal matrix}, \quad (3-23)$$

$$\Sigma_4 \ := \ \mathrm{cov}\left\{\widehat{\mathbf{R}}_{4x'}\right\} = \ \text{a diagonal matrix}. \quad (3-24)$$

Consider a $(P-1) \times P$ matrix $\mathbf{D}$ defined as

$$\mathbf{D} := \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{bmatrix}. \quad (3-25)$$

Under H_0 (cf. Lemma 2), asymptotically $\mathbf{D}\widehat{\mathbf{R}}_{3x'} \sim \mathcal{N}^C(0, \mathbf{D}\Sigma_3\mathbf{D}^T)$ and $\mathbf{D}\widehat{\mathbf{R}}_{4x'} \sim \mathcal{N}^C(0, \mathbf{D}\Sigma_4\mathbf{D}^T)$. Let $\widehat{\Sigma}_r$ denote a consistent estimate of Σ_r ($r = 3$ or 4) obtained by using estimators similar to (3-16) in (3-19) and (3-20). Then by [2, Lemma B.4], under H_0, asymptotically

$$(\mathbf{D}\widehat{\mathbf{R}}_{3x'})^{\mathcal{H}}(\mathbf{D}\widehat{\Sigma}_3\mathbf{D}^T)^{-1}(\mathbf{D}\widehat{\mathbf{R}}_{3x'}) \ \sim \ \chi^2(2P-2), \quad (3-26)$$

$$(\mathbf{D}\widehat{\mathbf{R}}_{4x'})^{\mathcal{H}}(\mathbf{D}\widehat{\Sigma}_4\mathbf{D}^T)^{-1}(\mathbf{D}\widehat{\mathbf{R}}_{4x'}) \ \sim \ \chi^2(2P-2). \quad (3-27)$$

The preceding discussion suggests the following simplified (sub-optimal) test procedure for higher-order whiteness. Accept H_0 if the following two inequalities hold true:

$$(\mathbf{D}\widehat{\mathbf{R}}_{3x'})^{\mathcal{H}}(\mathbf{D}\widehat{\Sigma}_3\mathbf{D}^T)^{-1}(\mathbf{D}\widehat{\mathbf{R}}_{3x'}) \leq T_\alpha, \quad (3-28)$$

$$(\mathbf{D}\widehat{\mathbf{R}}_{4x'})^{\mathcal{H}}(\mathbf{D}\widehat{\Sigma}_4\mathbf{D}^T)^{-1}(\mathbf{D}\widehat{\mathbf{R}}_{4x'}) \leq T_\alpha \quad (3-29)$$

else reject it, where T_α is the threshold corresponding to a significance level α, i.e., $Pr\{Y > T_\alpha\} = \alpha$ where $Y \sim \chi^2(2P-2)$.

3.2 Fourth-Order Statistics-Based Fitting

Now consider the cumulants slices $C_{4x'}(\tau)$ and $C_{6x'}(\tau)$ of $\{x'(t)\}$:

$$C_{6x'}(\tau) \ := \ E\{y_{5x'}(t)x'(t+\tau)\}$$

$$= \mathrm{cum}_6\{x'(t), \cdots, x'(t), x'(t+\tau)\} \quad (3-30)$$

where

$$y_{5x'}(t) := y'_{5x'}(t) - E\{y'_{5x'}(t)\}, \qquad (3-31)$$

$$y'_{5x'}(t) := x'^5(t) - 10E\{x'^2(t)\}x'^3(t) + \left[30(E\{x'^2(t)\})^2 - 5E\{x'^4(t)\}\right] x'(t). \qquad (3-32)$$

Under H_0, $C_{4x'}(\tau) = 0 = C_{6x'}(\tau)\ \forall \tau \neq 0$. A converse just as in Lemma 1 also holds true. Using (3-13) the counterpart to Lemma 2 is

Lemma 3. Let $\{x'(t)\}$ be as in (3-1) with $\{\widehat{g}(i)\}$ denoting the impulse response function of a stable rational transfer function. If $S_{4x'}(\omega) = c_1$ and $S_{6x'}(\omega) = c_2\ \forall \omega$ where c_1 and c_2 are some constants and $\{x'(t)\}$ obeys (2-1),(2-2) and (3-1), then $x'(t) = cw(t - i_0) + v'(t)$. •

Proof of Lemma 3 mimics that for Lemma 2; it is omitted.

With obvious notation, define as in (3-17),

$$\widehat{S}_{6x'}(k) := \frac{1}{N(2m_N+1)} \sum_{i=-m_N}^{m_N} X'(\omega_{k-i}) Y^*_{5x'}(\omega_{k-i}). \qquad (3-33)$$

Then we have

$$\widehat{S}_{6x'}(k) \sim \mathcal{N}^C\left(S_{6x'}(\omega_k), \Delta_N^{-1} S_{x'x'}(\omega_k) S_{55}(\omega_k)\right) \qquad (3-34)$$

where $S_{55}(\omega)$ denotes the power spectrum of $y_{5x'}(t)$. Define the vectors/matrices

$$\widehat{\mathbf{R}}_{6x'} := \left[\ \widehat{S}_{6x'}(l),\ \ l \in \Omega_P\ \right]^T, \qquad (3-35)$$

$$\Sigma_6 := \text{cov}\left\{\widehat{\mathbf{R}}_{6x'}\right\} = \quad \text{a diagonal matrix.} \quad (3-36)$$

Mimicking the developments in Sec. 3, we have: Accept H_0 if the following two inequalities hold true:

$$(\mathbf{D}\widehat{\mathbf{R}}_{4x'})^{\mathcal{H}}(\mathbf{D}\widehat{\Sigma}_4\mathbf{D}^T)^{-1}(\mathbf{D}\widehat{\mathbf{R}}_{4x'}) \leq T_\alpha, \qquad (3-37)$$

$$(\mathbf{D}\widehat{\mathbf{R}}_{6x'})^{\mathcal{H}}(\mathbf{D}\widehat{\Sigma}_6\mathbf{D}^T)^{-1}(\mathbf{D}\widehat{\mathbf{R}}_{6x'}) \leq T_\alpha \qquad (3-38)$$

else reject it, where T_α is the threshold corresponding to a significance level α for a $\chi^2(2P-2)$ distribution.

4 Simulation Example

We will illustrate the proposed model diagnostics and validation approach by using it for model order selection, as in [1]. The signal $s(t)$ is ARMA(2,1) with either zero-mean, i.i.d. binary (± 1 with probability 0.5 each) driving sequence $\{w(t)\}$ or zero-mean, i.i.d. one-sided exponential $\{w(t)\}$, and noise is zero-mean, white Gaussian:

$$s(t) = 0.8s(t-1) - 0.52s(t-2) + w(t) - 1.5w(t-1).$$

We fit AR(p,q) models for $p, q = 0, 1, \cdots, 6$ using cumulant matching (third-order for exponential and fourth-order for binary)[1]. The "smallest" (p,q) for which the fitted model can be validated is declared the correct model order; see [1] for exact details. To apply the tests (3-37)-(3-38) (binary case) or (3-28)-(3-29) (exponential case), we selected $\alpha = 0.01$ and used various record lengths N and SNR's. The smoothing window sizes $2m_N + 1$ were 11, 21, 45 and 89 for N = 1024, 2048, 4096 and 8192, respectively. We used all the points on the coarse grid in the interval $(0, \pi)$ to select P. Results of 100 Monte Carlo runs are shown in Table 1 for exponential input and in Table 2 for binary input.

TABLE 1: Exponential $w(t)$
No. of times the correct order $(p,q) = (2,1)$ is selected out of 100 runs (N = record length)

SNR → / N ↓	30 dB	20 dB	10 dB
1024	96	97	88
2048	98	98	100
4096	99	100	100

TABLE 2: Binary $w(t)$
No. of times the correct order $(p,q) = (2,1)$ is selected out of 100 runs (N = record length)

SNR → / N ↓	30 dB	20 dB	10 dB
2048	83	82	63
4096	91	92	88
8192	96	97	93

5 References

[1] J.K. Tugnait, "Linear model validation and order selection using higher-order statistics," *IEEE Transactions on Signal Processing*, vol. SP-42, pp. 1728-1736, July 1994.

[2] T. Söderström and P. Stoica, *System Identification*. Prentice Hall Intern.: London, 1989.

[3] J.M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proc. IEEE*, vol. 79, pp. 278-305, March 1991.

[4] C.L. Nikias and A. Petropulu, *Higher Order Spectra Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[5] J.K. Tugnait, "Detection of non-Gaussian signals using integrated polyspectrum," *IEEE Transactions on Signal Processing*, vol. SP-42, pp. 3137-3149, Nov. 1994. [Corrections, vol. SP-43, Nov. 1995.]

[6] J.K. Tugnait, "An improved test for linear model validation and order selection using higher-order statistics," *IEEE Signal Processing Letters*, vol. SPL-2, pp. 123-125, June 1995.

[7] D.R. Brillinger, *Time Series Data Analysis and Theory*. Holt, Rhinehart & Winston,: New York, 1975.