Monograph:

Zhu, Q.M. and Billings, S.A. (1996) Properties of Higher Order Correlation Function Tests for Nonlinear Model Validation. Research Report. ACSE Research Report 613 . Department of Automatic Control and Systems Engineering

Properties Of Higher Order Correlation Function Tests For Nonlinear Model Validation

Q M Zhu* and S A Billings**

* Department of Mechanical and Manufacturing Engineering, University of Brighton, Brighton, BN2 4GJ, UK

** Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, S1 3JD, UK

Research Report No 613

January 1996

PROPERTIES OF HIGHER ORDER CORRELATION FUNCTION TESTS FOR NONLINEAR MODEL VALIDATION

Q.M. Zhu* and S.A. Billings**

**Department of Mechanical and Manufacturing Engineering, University of Brighton, Brighton BN2 4GJ, UK*

***Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield S1 3JD,UK*

Abstract: Problems related to nonlinear model validation are addressed and properties associated with nonlinear detection, diagnostic power, and asymptotic correction are analysed.

Key words: Model validation; nonlinear systems.

1. INTRODUCTION

An important step in system identification involves validating the estimated model. For parametric models this should include both the parameter estimates and the model structure. Many authors have studied statistical validation procedures based on correlation analysis including Bohlin (1971, 1978), Box and Jenkins (1976), Billings and Voon (1983, 1986), Hjalmarsson (1993), Billings and Zhu (1994, 1995) and these methods have been used to validate both linear and nonlinear parametric models.

There are three problems associated with traditional correlation based model validity tests. The first is the need to check the whiteness of the model residuals. Conventional model based correlation tests can give a false outcome when used to test nonlinear models (Billings and Voon 1983, 1986). The second is the power of diagnosis. It has been noticed that some correlation functions exhibit less power when the noise and input variances are small (Billings and Zhu 1994). Thirdly the tests are based on the assumption that infinite data is available. This has been studied by Hjalmarsson (1993) who showed that this can cause bias in the asymptotic cross-covariance estimate.

The present study presents an extension of the work of Hjalmarsson (1993) and to the tests developed by Billings and Zhu (1994, 1995) and relates to the validation of nonlinear models including neural networks.

2. BACKGROUND

Consider a generalised single input single output (SISO) parametric model

$$\begin{aligned} y(t) &= F(y^{t-1}, u^{t-1}, \varepsilon^{t-1}, \theta) + \varepsilon(t) \\ &= \hat{y}(t) + \varepsilon(t) \end{aligned} \quad (2.1)$$

where t = 1, 2, ... is a time index, F (.) represents either a linear or a nonlinear function and $\hat{y}(t)$ is the one step ahead predicted output. In system identification the delayed output, input, and residual vectors, and parameter vector are defined as

$$\begin{aligned} y^{t-1} &= [y(t-1) \quad \cdots \quad y(t-\tau_y)] \\ u^{t-1} &= [u(t-1) \quad \cdots \quad u(t-\tau_u)] \\ \varepsilon^{t-1} &= [\varepsilon(t-1) \quad \cdots \quad \varepsilon(t-\tau_\varepsilon)] \\ \theta &= [\theta_1 \cdots \theta_m] \end{aligned} \quad (2.2)$$

The parameter vector θ is unknown and is to be estimated from data. Ideally the estimate should be unbiased with finite covariance so that the model residuals can be reduced to a zero mean unpredictable sequence with finite variance and

$$\theta_N \rightarrow \theta \quad as \quad N \rightarrow \infty \qquad (2.3)$$

where θ_N denotes the estimated parameter vector based on N data.

A common routine to test whether two signals $z(t,\theta_N)$, which represents a scalar valued process, and $Z(t,\theta_N)$, which represents a vector valued process, are dependent is to use the sample covariance $\Gamma(\theta_N) = [\gamma_1(\theta_N) \quad \cdots \quad \gamma_\tau(\theta_N)]^T$ which is computed by

$$\Gamma(\theta_N) = \frac{1}{\sqrt{N}} \sum_{t=\tau}^{N} z(t,\theta_N) Z(t,\theta_N) \qquad \tau = 0 \cdots N-1 \qquad (2.4)$$

Typically tests based on two statistics of this test are formulated as follows

$$S(\tau,\theta_N) = \frac{\gamma_\tau(\theta_N)}{\sqrt{p_{\tau\tau}(\theta_N)}} \qquad \tau = 0 \cdots N-1 \qquad (2.5)$$

is called the Gaussian test, where $p_{\tau\tau}(\theta_N)$ is the product of the variance of $z(t,\theta_N)$ and the variance of the τ' th element of $Z(t,\theta_N)$ and

$$T(\theta_N) = \Gamma^T(\theta_N) P^{-1}(\theta_N) \Gamma(\theta_N) \qquad (2.6)$$

is called the χ^2 test, where $P(\theta_N)=\{p_{ij}(\theta_N)\}$ is the covariance matrix of $\Gamma(\theta_N)$ under the null hypothesis (Hjalmarsson 1993). That is

$$P(\theta_N)=\sum_{\tau=0}^{N-1}R_{zz}(\tau)R_{ZZ}(\tau) \quad (2.7)$$

is a consistent estimate of

$$P(\theta)=\sum_{\tau=0}^{\infty}R_{zz}(\tau)R_{ZZ}(\tau) \quad (2.8)$$

where $R_{zz}(\tau)$ and $R_{ZZ}(\tau)$ are the covariance functions of $z(t,\theta)$ and $Z(t,\theta)$ respectively. For linear model validation, the residual auto-correlation test is performed by setting $z(t,\theta_N)=\varepsilon(t,\theta_N)$ (the residuals) and $Z(t)=[\varepsilon(t) \quad \varepsilon(t-1) \quad \cdots]$ and the cross correlation test between the input and residuals can be obtained by setting $z(t,\theta_N)=\varepsilon(t,\theta_N)$ and $Z(t)=[u(t) \quad u(t-1) \quad \cdots]$. Under the hypothesis that the residuals are zero mean white noise (null hypothesis H_0) the statistics introduced will exhibit certain well defined asymptotic distributions, which are given as follows

$$S(\tau,\theta_N)\sim AsN(0,1) \quad T(\theta_N)\sim As\chi^2(n) \quad N\to\infty \quad (2.9)$$

where $N(0,1)$ denotes the standard Gaussian distribution and $\chi^2(n)$ denotes the χ^2 distribution with n degrees of freedom.

The risk of rejecting H_0 when H_0 holds (which is called a type 1 risk) is equal to α, which is the confidence level or equivalently the value of the test threshold and the risk of accepting H_0 when it is not true (which is called a type 2 risk) depends on the properties of the tested model (Soderstrom and Stoica 1989).

3. NONLINEARITY EXAMINATION

In this and the following sections a set of infinite data will be assumed for analytical convenience, but it should be made clear that all the results presented are also applicable to the case with finite data length. An asymptotic correction mechanism is proposed in section four. Throughout all the stochastic processes will be assumed to be ergodic.

Nonlinear components in the residuals may not always be diagnosed when linear model based tests are applied. For example consider a situation where an inadequate model has been estimated to leave the residuals

$$\varepsilon(t)=e(t-2)e(t-5)+e(t) \quad (3.1)$$

where $e(t)$ is a white noise sequence with zero mean and finite variance. Applying traditional linear model validity tests to the residual of (3.1) by setting

$$z(t,\theta_N) = \varepsilon(t) \quad Z(t,\theta_N) = [\varepsilon(t) \quad \varepsilon(t-1) \quad \cdots] \qquad (3.2)$$

and substituting (3.2) into (2.5) gives

$$S_{\varepsilon\varepsilon}(\tau,\theta_N) = \begin{cases} 1 & \tau = 0 \\ 0 & \tau \neq 0 \end{cases} \qquad (3.3)$$

which incorrectly indicates that the residuals are white. To deal with nonlinear elements in the residuals a general approach has been developed by Billings and Zhu (1994), which is formulated by defining

$$\begin{aligned} z(t,\theta_N) &= \varepsilon^2(t) - E[\varepsilon^2(t)] \\ Z(t,\theta_N) &= [\varepsilon^2(t) - E[\varepsilon^2(t)] \quad \varepsilon^2(t-1) - E[\varepsilon^2(t)] \quad \cdots] \end{aligned} \qquad (3.4)$$

which represent higher order correlation functions. Applying the test to the example above by substituting (3.4) into (2.5) gives

$$S_{\varepsilon^2\varepsilon^2}(\tau,\theta_N) = \begin{cases} 1 & \tau = 0 \\ c & \tau = 2 \\ c & \tau = 5 \\ 0 & otherwise \end{cases} \qquad (3.5)$$

where

$$\begin{aligned} c &= \frac{\lambda\rho - \rho^3}{\lambda^2 + 4\rho^3 + \lambda - \rho^4 - \rho^2} \\ \lambda &= E[e^4(t)] \quad (fourth \quad order \quad moment) \\ \rho &= E[e^2(t)] \quad (second \quad order \quad moment) \end{aligned} \qquad (3.6)$$

and this now correctly indicates that the residual has not been reduced to a white noise sequence.

In summary the traditional first order correlation functions can only be used to test linear models. Higher order correlation functions can test both linear and nonlinear models.

4. DIAGNOSTIC POWER

The nonlinear model validity test discussed in the last section works well when the noise and input have
large variances. However it can sometimes exhibit less power when the noise and input variances are small because the fourth and higher order moments become small (Billings and Zhu 1994).

To explain this problem, consider again the example given in the last section, when the variance of the noise sequence $e(t)$ is much smaller and the constant c in (3.6) becomes

$$
\begin{aligned}
c &= \frac{O(\lambda\rho)-O(\rho^3)}{O(\lambda^2)+O(4\rho^3)+\lambda-O(\rho^4)-\rho^2} \\
&\approx \frac{0}{\lambda-\rho^2} \\
&= 0
\end{aligned}
\qquad (4.1)
$$

where $O(.)$ is the notation of infinitesimals (Soderstrom and Stoica 1989). The higher order correlation function discussed in the last section has less power to deal with this case. Accordingly Billings and Zhu (1994, 1995) suggested an alternative solution by introducing output terms combined with the input and residuals to form more powerful higher order correlation functions defined as

$$
\begin{aligned}
z(t,\theta_N) &= y(t)\varepsilon(t)-E[y(t)\varepsilon(t)] \\
Z(t,\theta_N) &= [\varepsilon^2(t)-E[\varepsilon^2(t)] \quad \varepsilon^2(t-1)-E[\varepsilon^2(t)] \quad \cdots]
\end{aligned}
\qquad (4.2)
$$

The corresponding test of (2.5) therefore becomes

$$
S_{(y\varepsilon)\varepsilon^2}(\tau,\theta_N) = S_{(\hat{y}\varepsilon)\varepsilon^2}(\tau,\theta_N) + S_{\varepsilon^2\varepsilon^2}(\tau,\theta_N) \qquad (4.3)
$$

Inspection of (4.3) shows that $S_{(\hat{y}\varepsilon)\varepsilon^2}(\tau,\theta_N)$ should still contribute even when $S_{\varepsilon^2\varepsilon^2}(\tau,\theta_N)$ is small due to a smaller variance of noise. $S_{(\hat{y}\varepsilon)\varepsilon^2}(\tau,\theta_N)$ will go to zero when the residual is reduced to a zero mean white noise, so that $S_{(y\varepsilon)\varepsilon^2}(\tau,\theta_N) = S_{\varepsilon^2\varepsilon^2}(\tau,\theta_N)$.

In summary the introduction of output terms can enhance the power of model validity tests while maintaining the computational simplicity.

5. ASYMPTOTIC CORRECTION

The asymptotic problem can be addressed by following Hjalmarsson (1993) who showed that for finite length of data

i) the cross-covariance may not be a consistent estimate,

ii) the auto-covariance is a consistent estimate, Therefore both Guassian and χ^2 tests may not give asymptotic outcomes and may cause an incorrect diagnosis of the residuals.

Hjalmarsson's solution (1993) is the same as that presented in section two, comparing the two studies Billings and Zhu (1994) emphasised the possibility of diagnosing nonlinear elements in the residuals and Hjalmarsson (1993) emphasised asymptotically correct correlation tests for linear model validation. Therefore the approach can be extended for asymptotically correct correlation tests for nonlinear model validation.

The condition for asymptotically correct correlation tests was derived by Hjalmarsson (1993) as below

$$E\left[\frac{d\Gamma(\theta_N)}{d(\theta)}\right]=0 \qquad (5.1)$$

The test with output terms introduced in (4.3) can be proved to satisfy the above condition when substituting (4.3) into (5.1)

$$E\left[\frac{d\Gamma(\theta_N)}{d\theta}\right]=E\left[\frac{dE[z(t,\theta_N)Z(t,\theta_N)]}{d\theta}\right]$$
$$=E\left[z'Z\frac{dz}{d\theta}+zZ'\frac{dZ}{d\theta}\right]$$
$$=2E\left[\left[\varepsilon^2(t,\theta_N)\frac{d\varepsilon(t,\theta_N)}{d\theta}-E\left[\varepsilon^2(t,\theta_N)\frac{d\varepsilon(t,\theta_N)}{d\theta}\right]\right][x1(.)]\right] \qquad (5.2)$$
$$+E\left[\left[\varepsilon^2(t,\theta_N)-E\left[\varepsilon^2(t,\theta_N)\right]\right][x2(.)]\right]$$
$$=[c1 \quad 0 \quad \cdots \quad 0]^T$$

where all the elements are zero except the first element (constant $c1$) corresponding to $\tau = 0$ and $x1(.)$ and $x2(.)$ are remaining terms. In the proof $E[\varepsilon(t-i,\theta_N)]$was used to replace $E[\varepsilon(t,\theta_N)]$ at $\tau = i$.

Following the above statement the higher order correlation functions including delayed output terms as presented in section three also produce asymptotically correct correlation tests for nonlinear model validation with higher diagnostic power. For cross correlation tests between input and residual and all forms of corresponding χ^2 tests see Billings and Zhu (1994, 1995).

6. CONCLUSIONS

The properties of higher order correlation function tests for nonlinear model validation have been analysed to provide a comprehensive understanding of nonlinear model validation to show that the nonlinearity detection and asymptotic correction can be implemented by introducing higher order statistic tests and that the diagnostic power of the tests can be improved by introducing delayed output terms.

7. REFERENCES

Billings, S.A. and Voon, W.S.F., 1983, Structure detection and model validity tests in the identification of nonlinear systems, IEE Proceedings, Pt D, Vol. 130, 193-199.

Billings, S.A. and Voon, W.S.F., 1986, Correlation based model validity tests for nonlinear models, Int. J. Control, Vol. 44, 235-244.

Billings, S.A. and Zhu, Q.M., 1994, Nonlinear model validation using correlation tests, Int. J. Control, Vol. 60, 1107-1120.

Billings, S.A. and Zhu, Q.M., 1995, Model validation tests for multivariable nonlinear models including neural networks, Int, J. Control, Vol. 62, 749-766.

Bohlin, T., 1971, On the problem of ambiguities in maximum likelihood identification, Automatica, Vol. 7, 199-210.

Bohlin, T., 1978, Maximum power validation of models without higher order fitting, Automatica, Vol. 14, 137-146.

Box, G.E.P. and Jenkins, G.M., 1976, Time series analysis forecasting and control, San Francisco, Holden-Day.

Hjalmarsson, H., 1993, Asymptotic correction tests in model validation, Proceedings of the 32nd Conference on Decision and Control, San Antonio, Texas, USA, 2058-2059.

Soderstrom, T. and Stoica, P., 1989, System identification, Prentice Hall International (UK) Ltd, Cambridge.