VALIDITY OF THE STANDARD CROSS-CORRELATION TEST FOR MODEL STRUCTURE VALIDATION

Sippe G. Douma * **Xavier Bombois** *
Paul M.J. Van den Hof *,[1]

* *Delft Center for Systems and Control, Delft University of Technology,*
Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract: The standard prediction error framework provides many theoretical results under the assumption that the true system is in the model class. An important example is the expression for the parameter covariance matrix which is used to derive model uncertainty regions. An essential step in a system identification procedure is the (in)validation of this assumption that the model structure is rich enough to contain the true system. The standard test for this purpose is the sample cross-correlation test between the output residuals and the input. It turns out that this standard test itself is valid only under exactly those assumptions it is meant to verify. As a result considerable undermodelling errors can remain undetected. Besides suggesting caution to users of the standard test, methods are presented to adapt the test adequately.

Keywords: model validation, system identification, undermodelling

1. INTRODUCTION

In a standard system identification procedure an uncertainty region is constructed around the (nominal) estimate on the basis of the measurement data and prior assumptions on the data generating process. Particularly in the context of identification for robust control the uncertainty region could be argued to be as or even more important than the nominal model. The set of models induced by the uncertainty region has to be guaranteed (at a certain level of probability) to contain the "true system", i.e. the actual process under consideration, to allow for a robust controller design.

In the system identification theory of the prediction error framework the uncertainty region follows from a parameter covariance matrix associated with a Gaussian distribution corresponding to a stochastic noise assumption (Ljung, 1999*b*)(Söderström and Stoica, 1989). This uncertainty region reflects the effect of noise in the measurements on the estimated parameters. The uncertainty region can be said to contain the actual process (at a certain level of probability) if the model errors are due to the noise in the data only and are not due to a limitation of the model structure. Moreover, the available analytical expressions for the (noise-induced) parameter covariance matrix themselves are valid under the assumption that there are no undermodelling errors. That is, the model uncertainty region is correct under the assumption that the true system is in the model class.

Since the 1990's a number of model uncertainty bounding techniques are available which explicitly take the effect of undermodelling into ac-

[1] Corresponding author; p.m.j.vandenhof@dcsc.tudelft.nl

count (Goodwin *et al.*, 1992)(Hakvoort and den Hof, 1997)(De Vries and Van den Hof, 1998). An alternative approach follows the school of thought which states that the effect of undermodelling can be neglected when its effects are indistinguishable from the effects of measurement noise. A practical identification procedure then consists of a model validation step. The model structure is sequentially enlarged until the model estimate passes the validation test. A best-case scenario is assumed and a model uncertainty region is constructed based on the noise induced errors only.
It is clear that this practice stands or falls with the correctness of the model validation test used to verify the assumption that the system is in the model class. In practice, the test on the sample cross-correlation between the residuals of the model and the input, as implemented in the System Identification toolbox of Matlab (Ljung, 2003), is used extensively and nearly exclusively. This fact not withstanding, an evaluation of this test shows that a refinement of the test is possible and warranted. Amongst other things, it turns out that this test is valid only under exactly that assumption it intends to verify, namely that the system is in the model class. Although the test works well in many situations, the underlying working principle is not the one that is suggested by the theory. Figure 1 depicts a motivating example in which the standard cross-correlation test does not invalidate a model while in fact the model exhibits a large undermodelling error with respect to the true system. Details of the example are provided at the end of this paper. When the conclusion is drawn from the cross-correlation test that the undermodelling is insignificant and that the model uncertainty can be based on variance errors only, the resulting model uncertainty region fails to contain the true system. This paper intends to highlight the problems with the standard test and to supply adequate improvements.

2. BACKGROUND

Let $\mathcal{M}(\theta)$ denote a model class of linear time-invariant models parametrized by a parameter vector θ corresponding to a certain model structure of a plant model $G(q,\theta)$ and noise model $H(q,\theta)$ with q the standard shift operator. In a standard prediction error framework (Ljung, 1999*b*) (Söderström and Stoica, 1989) a model is identified from measurement data $Z^N := \{y,u\}_N$ of data length N according to

$$\hat{\theta}_N = \arg\min_{\theta} V_N\left(\theta, Z^N\right) = \arg\min_{\theta} \frac{1}{N}\sum_{t=1}^{N} \varepsilon^2(t,\theta), \tag{1}$$

where the residuals $\varepsilon(t,\theta)$ are constructed as

$$\varepsilon(t,\theta) = H^{-1}(q,\theta)\left(y(t) - G(q,\theta)\,u(t)\right). \tag{2}$$

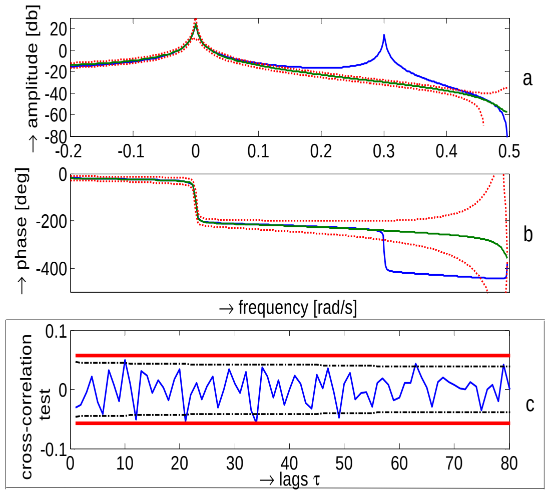

Fig. 1. Bode-plots of a true fourth order model $G_0(q)$ (solid blue), a second order OE-estimate (dashed green) and the standard 99% variance-based confidence bounds of the second order OE-estimate (dotted red). Figure (c) depicts the standard cross-correlation test. The sample cross-correlation function between the residuals of the OE-model and the input is depicted and seen to remain within the standard test-bounds depicted in solid lines. In other words, the standard cross-correlation test does not invalidate the second order OE-model Moreover, the confidence bounds of the OE-model based on variance errors only fail to contain the true system.

Define the parameter θ^* as
$\theta^* = \arg\min_\theta \lim_{N\to\infty} E\left[V_N\left(\theta, Z^N\right)\right]$,with $E[\cdot]$ the expectation operator. Assume that

$$y(t) = G_0(u(t)) + v(t) \tag{3}$$

where G_0 is a possibly nonlinear operator. We further assume that $v(t) = H_0(q)e(t)$ and $e(t)$ a zero-mean white noise sequence and we assume that $v(t)$ is uncorrelated with the input $u(t)$.
Consider a plant and noise model estimate $G(q,\hat{\theta}_N)$ and $H(q,\hat{\theta}_N)$ corresponding to the least-squares estimate $\hat{\theta}_N$ of expression (1). With expression (3) the associated residual sequence $\varepsilon(t,\hat{\theta}_N)$ can be evaluated as

$$\varepsilon(t,\hat{\theta}_N) = \beta_b(t,\hat{\theta}_N,G_0,\theta^*) + \beta_v(t,\hat{\theta}_N,\theta^*) + \varepsilon_v(t,\hat{\theta}_N) \tag{4}$$

with

$$\beta_b(t,\hat{\theta}_N,G_0,\theta^*) = H^{-1}(q,\hat{\theta}_N)\left(G_0(u(t)) - G(q,\theta^*)\,u(t)\right) \tag{5}$$

$$\beta_v(t,\hat{\theta}_N,\theta^*) = H^{-1}(q,\hat{\theta}_N)\left(G(q,\theta^*)\,u(t) - G\left(q,\hat{\theta}_N\right)u(t)\right)$$

$$\varepsilon_v(t,\hat{\theta}_N) = H^{-1}(q,\hat{\theta}_N)v(t) \tag{6}$$

such that

i. $\beta_b(t, \hat{\theta}_N, G_0, \theta^*)$ corresponds to the bias of models estimated within the model class $\mathcal{M}(\theta)$.
ii. $\beta_v(t, \hat{\theta}_N, \theta^*)$ corresponds to the variance error in the model $G\left(q, \hat{\theta}_N\right)$ with respect to $G(q, \theta^*)$.
iii. $\varepsilon_v\left(t, \hat{\theta}_N\right)$ corresponds to the measurement noise $v(t)$.

We define the model structure validation problem as the problem of evaluating the hypothesis

$$\Upsilon_0 : G_0 = G(q, \theta^*) \in \mathcal{M}(\theta).$$

In the context of model structure validation, the essential term in (4) is therefore the term $\beta_b(t, \hat{\theta}_N, G_0, \theta^*)$ which reflects the unmodelled dynamics, i.e. that part of the output $y(t)$ that cannot be captured within the model structure of the model class $\mathcal{M}(\theta)$.

3. THE STANDARD CROSS-CORRELATION TEST

Expression (4) shows that if the model $G(q, \hat{\theta}_N)$ is 'good', the contribution of $u(t)$ in $\varepsilon(t, \hat{\theta}_N)$ will be 'small'. A test on the cross-correlation between $u(t)$ and $\varepsilon(t, \hat{\theta}_N)$, therefore, allows for an indication of the quality of the model $G(q, \hat{\theta}_N)$. The cross-correlation test as implemented in the Identification Toolbox in Matlab will be referred to as the *standard cross-correlation test* (Ljung, 2003). This test is based on the sample cross-correlation $\hat{R}_{\varepsilon u}(\tau)$ between $u(t)$ and $\varepsilon(t, \hat{\theta}_N)$, i.e. for $|\tau| \leq n_\tau - 1$

$$\hat{R}_{\varepsilon u}(\tau) := \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t, \hat{\theta}_N) u(t - \tau). \tag{7}$$

Algorithm 1. (The standard cross-correlation test). The model $G(q, \hat{\theta}_N)$ is not invalidated if the sequence $\hat{R}_{\varepsilon u}(\tau)$ satisfies, for $|\tau| \leq n_\tau - 1$,

$$\left|\hat{R}_{\varepsilon u}(\tau)\right| < \gamma(\alpha), \text{ with } \gamma(\alpha) = c_{\mathcal{N}}(\alpha)\sqrt{\frac{P_1}{N}} \tag{8}$$

$$P_1 = \hat{R}_{uu}(0)\hat{R}_{\varepsilon\varepsilon}(0) + 2\sum_{\kappa=1}^{n_\tau - 1} \hat{R}_{uu}(\kappa)\hat{R}_{\varepsilon\varepsilon}(\kappa) \tag{9}$$

where, for $0 \leq \kappa \leq n_\tau - 1$,

$$\hat{R}_{\varepsilon\varepsilon}(\kappa) = \frac{1}{N} \sum_{t=1}^{N} \varepsilon(t, \hat{\theta}_N)\varepsilon(t - \kappa, \hat{\theta}_N) \tag{10}$$

and $\hat{R}_{uu}(\tau)$ defined similarly and where $c_{\mathcal{N}}(\alpha)$ corresponds to the Gaussian distribution $\mathcal{N}(0,1)$ such that for $x \in \mathcal{N}(0,1) \implies \mathrm{p}\left(x \leq c_{\mathcal{N}}(\alpha)\right) = \alpha$ and where the probability level α and the number of considered lags n_τ are a user choice. □

The theoretical motivation of the test is based on an evaluation of the residuals of the model $G(q, \theta^*)$. For this ideal (variance free) model it holds with expression (4) that

$$\hat{R}_{\varepsilon u}(\tau) = \hat{R}_{\beta_b u}(\tau) + \hat{R}_{\varepsilon_v u}(\tau), \tag{11}$$

It follows that if the hypothesis Υ_0 is true, the term $\beta_b(t, \theta^*, G_0, \theta^*) = 0$ and the sample cross correlation term $\hat{R}_{\varepsilon u}(\tau)$ equals the sample cross-correlation $\hat{R}_{\varepsilon_v u}(\tau)$ between the input $u(t)$ and the prefiltered noise $\varepsilon_v(t, \theta^*) = H^{-1}(q, \theta^*)v(t)$. The cross-correlation test, therefore, intends to verify whether the data provides evidence that it is unlikely that the sample cross-correlation $\hat{R}_{\varepsilon u}(\tau)$ is a realization of sample cross-correlations $\hat{R}_{\varepsilon_v u}(\tau)$. The bound $\gamma(\alpha)$ of expression (8) reflects (an estimate of) the $\alpha \cdot 100\%$ confidence region for a realization of the sample cross-correlation term $\hat{R}_{\varepsilon_v u}(\tau)$. If the cross-correlation term $\hat{R}_{\varepsilon u}(\tau)$ exceeds the bound $\gamma(\alpha)$, it is unlikely, against a false alarm rate of $100(1-\alpha)\%$, that the term $\hat{R}_{\varepsilon u}(\tau)$ has the same statistical properties as $\hat{R}_{\varepsilon_v u}(\tau)$.

In particular, it holds that

$$\begin{aligned}&\Upsilon_0 \text{ is true and } \hat{\theta}_N = \theta^* \\ &\implies \hat{R}_{\varepsilon u}(\tau) = \hat{R}_{\varepsilon_v u}(\tau) \in \mathcal{N}(0, P_{\theta^*}(\tau)),\end{aligned}$$

with

$$P_{\theta^*}(1) = \hat{R}_{uu}(0)R_{\varepsilon_v\varepsilon_v}(0) + 2\sum_{\kappa=1}^{N-1} \hat{R}_{uu}(\kappa)R_{\varepsilon_v\varepsilon_v}(\kappa)$$

with $R_{\varepsilon_v\varepsilon_v}(\kappa) = E\left[\varepsilon_v(t, \theta^*)\varepsilon_v(t+\kappa, \theta^*)\right]$. We see that P_1 in the standard cross-correlation test is an estimate of $P_{\theta^*}(\tau)$ where it is assumed that $P_{\theta^*}(\tau) = P_{\theta^*}(1)$ at each lag.

While the theoretical motivation is sound so far, a crucial problem with the standard test appears when expression (10) is used to estimate the auto-correlation $R_{\varepsilon_v\varepsilon_v}(\tau)$ which is required to formulate the bounds based on P_{θ^*}. The line of reasoning is followed that if Υ_0 is true, $\varepsilon(t, \theta^*) = \varepsilon_v(t, \theta^*)$ and the residuals $\varepsilon(t, \theta^*)$ can be used to estimate the properties of $\varepsilon_v(t, \theta^*)$. That is,

$$\begin{aligned}&\Upsilon_0 \text{ is true and } \hat{\theta}_N = \theta^* \\ &\implies \Lambda_{\varepsilon_v} = \Lambda_\varepsilon = E\left[\varepsilon(\theta^*)\varepsilon^T(\theta^*)\right]\end{aligned}$$

However, this cannot be extended to the situation in which Υ_0 is not true. When undermodelling is present, the residuals $\varepsilon(t, \theta^*)$ are not equal to $\varepsilon_v(t, \theta^*)$. The next section will show that expression (10) is not a good estimator of the noise properties and that its use is detrimental in the context of model validation. Two alternative approaches to estimating the noise properties are presented, based on a repeated input sequence and on an auxiliary high-order model, respectively.

4. EFFECT OF UNDERMODELLING ON NOISE MODELS

4.1 Difficulties when estimating noise properties

Expression (4) immediately reveals the problem which occurs when using expression (10) to estimate the required noise properties. For low model orders n (with respect to G_0) the residual sequence $\varepsilon(t,\hat{\theta}_N)$ contains a large undermodelling contribution $\beta_b(t,\hat{\theta}_N,G_0,\theta^*)$. On the other hand, while for large model orders n the undermodelling contribution will be small, a part of the measurement noise realization $v(t)$ is fitted by the parameter $\hat{\theta}_N$ (becoming its variance error) and as such does not appear in the residuals as reflected in the term $\beta_v(t,\hat{\theta}_N,\theta^*)$. In other words, the residual error $\varepsilon(t,\hat{\theta}_N)$ cannot be considered to be representative for $\varepsilon_v(t,\hat{\theta}_N)$ without due care.

4.2 Effect for the validation test

In the standard cross-correlation test the noise properties are estimated from the residuals of the model to be evaluated. When a model is undermodelled the residual sequence $\varepsilon(t,\hat{\theta}_N)$ contains the undermodelling contribution $\beta_b(t,\hat{\theta}_N,G_0,\theta^*)$. Therefore, the standard test is based on bounds which has the undermodelling incorporated which it intends to detect. In other words, the bound $\gamma(a)$ becomes too large and the validation test too lenient. From a practical point of view it follows that many models with significant undermodelling can pass the standard test (cf. the example in Figure 1).

From a theoretical point of view it is awkward that the test is valid only under the assumptions it is meant to evaluated. Moreover, the theoretical principle underlying the test is to detect undermodelling with respect to the statistical behaviour of the noise. In the standard test, however, the reason that in many cases undermodelling is detected is different. Undermodelling is now detected if its contribution to the sample cross-correlation $\hat{R}_{\varepsilon u}(\tau)$ (left-hand side of inequality (8)) is larger than its contribution to the estimation of the noise properties (appearing in the right-hand side of inequality (8)). That is, from expressions (10), (11) and (8) it holds (approximately) that a model can only be invalidated if

$$\left|\hat{R}_{\beta_b u}(\tau)\right| > \frac{c_{\mathcal{N}}(\alpha)}{N}\sqrt{\sum_{t=1}^{N}\beta_b^2(t,\hat{\theta}_N,G_0,\theta^*)}.$$

This explains why the standard test still works well in many cases as the bound $\gamma(\alpha)$ is based on an average behaviour of the undermodelling errors while the actual undermodelling contribution to $\hat{R}_{\varepsilon u}(\tau)$ will not be evenly distributed over the lags τ. For example, if the undermodelling $\bar{G}_0(q) := G_0 - G(q,\theta^*)$ is linear and the input $u(t)$ is white, the term on the left corresponds to the pulse response of $H^{-1}(q,\hat{\theta}_N)\bar{G}_0(q)$. Since the pulse response of $\hat{H}^{-1}(q)\bar{G}_0(q)$ decays with τ, for the first lags τ the left hand side will often be larger than its "average over τ" on the right hand side.

4.3 Alternatives for estimating noise properties

In the following we focus on estimating the properties of the measurement noise $v(t)$, from which the required properties of $\varepsilon_v(t,\hat{\theta}_N) = H^{-1}(q,\hat{\theta}_N)v(t)$ are readily derived. The problem of separating the noise contributions from modelling errors in the residuals $\varepsilon(t,\hat{\theta}_N)$ can conveniently be circumvented if use can be made of a repeated input signal. Consider two sets of measurement data $\{u,y_1\}$ and $\{u,y_2\}$, both generated with the same input u. With expression (3) it holds that the difference $\varepsilon_d(t)$ between the two measured outputs satisfies

$$\varepsilon_d(t) = y_1(t) - y_2(t) = \{v_1(t) - v_2(t)\}\,. \qquad (12)$$

The scaled difference signal $\sqrt{2}\varepsilon_d(t)$ qualifies for estimating the noise properties. Indeed, the difference signal $\varepsilon_d(t)$ does not contain terms related to the input u. Only some care has to be taken with the initial conditions, i.e. the effect of past input values. Note that in a subsequent plant model identification step use can be made of all the measurements.

In case the input signal is not a user's choice, we suggest to estimate an auxiliary plant model $G_a(q,\hat{\theta}_a)$ of a model order high enough to be reasonably certain that the effects of undermodelling are small. Construct the residual signal $\varepsilon(t,\hat{\theta}_a)$ as in expression (2) with $G(q,\theta) = G(q,\hat{\theta}_a)$ and $H(q,\theta) = 1$. Since it holds with expression (4) that $\varepsilon(t,\hat{\theta}_a) \approx \beta_v(t,\hat{\theta}_N,\theta^*) + v(t)$, $\varepsilon(t,\hat{\theta}_a)$ can be used to estimate a noise model $H(q,\hat{\theta}_v)$. We suggest to use a parametric model structure and proper model order selection (cf. the ARMASA toolbox of (Broersen, 2003)(Broersen, 2002)). Finally, estimate the noise variance $R_v(0) = E[v(t)]^2$ as

$$\hat{R}_v(0) = \frac{N}{N-n}\frac{1}{N}\left\|H^{-1}(q,\hat{\theta}_v)\varepsilon(t,\hat{\theta}_a)\right\|_2^2, \quad (13)$$

where $\|x(t)\|_2^2$ denotes $\sum_{t=1}^{N}x^2(t)$ and n is the order of the auxiliary model $G_a(q,\hat{\theta}_a)$. This is a consistent estimate of the noise variance (if $\hat{\theta}_v$ is consistent), a proof of which can be found in (Ljung, 1999*b*, p. 471).

Note that the method does suffer from the fact that the noise model $H(q,\hat{\theta}_v)$ is based on the properties of $\varepsilon(t, \hat{\theta}_a)$ rather than $v(t)$ itself. Still, the distortion will be much smaller than the effect of allowing undermodelling to be present in the residuals. Important is the fact that the variance is estimated consistently. Finally, note that the auxiliary model can be a very poor model from a variance point of view; the variance in the model is of no consequence for the purpose of estimating the variance $R_v(0)$ consistently.

Other approaches to the estimation of noise properties in case of undermodelling are suggested in (Tjärnström and Ljung, 2002)(Hjalmarsson and Ljung, 1992). In (Tjärnström and Ljung, 2002) the use of high order models for noise variance estimation is also suggested. The estimation procedure applied there is based on bootstrapping. The effect of modelling part of the noise into the model is not taken into account in that paper.

5. IMPROVED TEST

To improve the standard test three adaptations are suggested. Firstly and most importantly, an improved cross-correlation test is based on a more accurate estimation of the noise properties as suggested in the previous section. Secondly, the standard test evaluates the cross-correlation at each lag separately. However, the presence of an undermodelling term $\beta_v(t, \hat{\theta}_N, \theta^*)$ is much more accurately detected when considering the correlation over lags. In other words, the test should be vector-valued. In particular, it can be shown that

Υ_0 is true and $\hat{\theta}_N = \theta^*$

$$\Longrightarrow \hat{\mathbf{R}}_{\varepsilon u} = \hat{\mathbf{R}}_{\varepsilon_v u} = \Phi_u \varepsilon_v(\theta^*)$$
$$\Longrightarrow \hat{\mathbf{R}}_{\varepsilon u} \in \mathcal{N}(0, \tfrac{1}{N^2} P_{\theta^*}) \text{ with}$$
$$P_{\theta^*} = \Phi_u^T \Lambda_{\varepsilon_v} \Phi_u \text{ and } \Lambda_{\varepsilon_v} = E\left[\varepsilon_v(\theta^*)\varepsilon_v^T(\theta^*)\right],$$

where the boldface indicates a vector, i.e. $\mathbf{x} = \left[\, x(1) \; \cdots \; x(N) \,\right]^T$, and where $\Phi_u \in \mathbb{R}^{n \times N}$ denotes the matrix with columns given by $\phi(t) := \left[\, u(t) \ldots q^{-1}u(t) \ldots q^{-(n_\tau - 1)}u(t) \,\right]^T$ for $t = [1, N]$. The literature (Ljung, 1999*b*)(Söderström and Stoica, 1989) indeed mentions a vector-valued cross-correlation test based on the inequality

$$\hat{\mathbf{R}}_{\varepsilon u}^T P_{\theta^*}^{-1} \hat{\mathbf{R}}_{\varepsilon u} \leq c_\chi(\alpha, n_\tau), \qquad (14)$$

with $c_\chi(\alpha, n_\tau)$ corresponding to the chi-squared distribution such that for $x \in \chi^2(n) \Longrightarrow \mathrm{p}(x \leq c_\chi(\alpha, n)) = \alpha$. That is, the correlation in the sequence $\hat{R}_{\varepsilon u}(\tau)$ is compared to the correlation in the sequence $\hat{R}_{\varepsilon_v u}(\tau)$. Thirdly, the standard test is derived for θ^*. When applying the test to an estimate $\hat{\theta}_N$ the variance error $\beta_v(t, \hat{\theta}_N, \theta^*)$ should be taken into account (Söderström and Stoica, 1989)(Hjalmarsson, 1993). In (Söderström and Stoica, 1989) it was shown that this corresponds to adapting the covariance matrix P_{θ^*} by considering the properties of $\beta_v(t, \hat{\theta}_N, \theta^*) + \varepsilon_v(t, \hat{\theta}_N)$ instead of those of $\varepsilon_v(t, \hat{\theta}_N)$ only. Unfortunately, the last two improvements are not considered as standard yet. More importantly, however, is the fact that these two improvements alone do not result in a better test when the bounds are still based on an estimate of the noise properties derived from the residuals of the model to be evaluated.

Remark 1. The tests considered here are non-parametric. Alternatively, the model error model approach (Ljung, 1999*a*) uses a parametric model to detect undermodelling, which is favourable from a variance point of view. Still, it is emphasized that a model structure validation stands or falls with a proper noise model.

6. ILLUSTRATIVE EXAMPLE

This example corresponds to Figure 1 of the introduction. Consider $N = 256$ measurements $\{y, u\}_N$ generated according to $y(t) = G_0(q)u(t) + v(t)$ with $u(t)$ and $v(t)$ white noise sequences with variances $\sigma_u^2 = 1$ and $\sigma_v^2 = 0.09$, respectively, and with the true system

$$G_0(q) = \frac{0.01293q^{-1} + 0.1062q^{-2} + 0.1058q^{-3} + 0.01279q^{-4}}{1 - 0.2482q^{-1} + 1.091q^{-2} - 0.2441q^{-3} + 0.9822q^{-4}}$$

OE-models of orders 1 to 10 are estimated from this data; in particular, a model structure given by

$$G(q, \theta, n) = \frac{q^{-1}\left(\sum_{k=1}^{n} b(k)q^{-k}\right)}{1 + \sum_{k=1}^{n} f(k)q^{-k}} \text{ for } n = [1, 10] \qquad (15)$$

and noise model $H(q, \theta) = 1$.

Figure 1 clearly shows the undermodelling error in the second order OE-estimate $\hat{G}_{OE}(q, \theta, 2)$ (dotted green). However, the standard cross-correlation test does not invalidate this second OE-estimate $\hat{G}_{OE}(q, \theta, 2)$, while its confidence region based on variance errors only fails to contain the true system. The bound $\gamma(a)$ is overestimating the noise contribution, due to the incorporation of the undermodelling terms itself in the estimation of the variance with expression (10). Using the second method of Section 4.3 by estimating the noise model with the ARMASA-toolbox on the residuals of an OE-model of order $N/5$ (which is extreme to show the potential of the method) and estimating the variance according to (13), new bounds (dashed-dotted black) are obtained. The new bounds are not corrupted by undermodelling effects. They do detect the effect of undermodelling and the model will correctly be invalidated.

Figure 2 depicts the results of vector-valued cross-correlation tests. The curves depict the test value $\hat{\mathbf{R}}_{\epsilon u}^T P^{-1} \hat{\mathbf{R}}_{\epsilon u}$, for $\hat{\mathbf{R}}_{\epsilon u} \in \mathbb{R}^{n_\tau}$ and different P, which should be below the bound $c_\chi(a, n_\tau)$ for

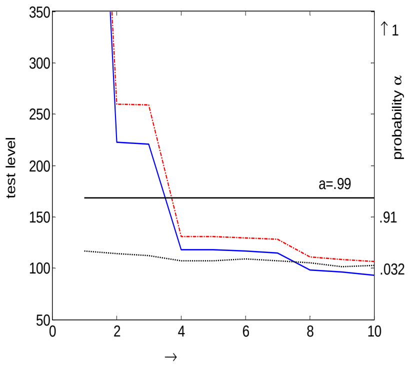

Fig. 2. The vector-valued cross-correlation test for an OE-model of increasing order $n = [1, 10]$. If the test value $\hat{\mathbf{R}}_{\epsilon u}^T P^{-1} \hat{\mathbf{R}}_{\epsilon u}$ is below the bound $c_\chi(.99, 128)$ the model is not invalidated. The matrix P is based on the true noise covariance (solid blue), on the estimate of expression (10) (dotted black) and on the estimate following Section 4.3 (dashed-dotted red).

the model not to be invalidated (cf. expression (14)) against a false-alarm rate of $(1 - \alpha)$. The blue solid line depicts the test value $\hat{\mathbf{R}}_{\epsilon u}^T P_{\theta^*}^{-1} \hat{\mathbf{R}}_{\epsilon u}$ when using the true noise properties to estimate P_{θ^*}. The correct model order of four is easily identified from the graph as the first model order for which the test value becomes smaller than the bound ($\alpha = .99$). When estimating the noise covariance following Section 4.3 to estimate P_{θ^*} the test levels differ from the actual levels, but the model orders are still properly evaluated (red dashed-dotted line). When estimating the noise properties from the residuals of the model to be evaluated (cf. expression (10)) to estimate P_{θ^*} the undermodelling errors are incorporated in the covariance estimate and a vector-valued cross-correlation test would let all model orders (1 to 10) pass (dotted black curve). That is, upgrading the standard point-wise test to a vector-valued test would not be beneficial if the noise properties are still estimated from the residuals of the model to be tested itself.

7. CONCLUDING REMARKS

The standard cross-correlation test works very well in many practical situations. However, significant undermodelling errors can remain undetected. This is due to the fact that the test is based on an estimation of the noise properties derived from residual signals that may contain considerable undermodelling errors. From a theoretical point of view it is important to note that the test actually operates under exactly those assumptions it intends to verify.

In this paper we propose an alternative approach which is based on a vector-valued test for which the bounds are based on an improved estimation of the noise properties allowing undermodelling in the plant model. The former issue has been suggested before in the literature, but is not considered as standard yet. It has been shown in a simulation example, that improved noise modelling is necessary to avoid large undermodelling errors to remain undetected.

REFERENCES

Broersen, P.M.T. (2002). Automatic spectral analysis with time series models. *IEEE Transactions on Instrumentation and Measurement* **51**(2), 211–216.

Broersen, P.M.T. (2003). ARMASA toolbox in matlab. The Mathworks, Inc.

De Vries, D.K. and P.M.J. Van den Hof (1998). Frequency domain identification with generalized orthonormal basis functions. *IEEE Transactions on Automatic Control* **43**(5), 656–669.

Goodwin, G.C., M. Gevers and B.M. Ninness (1992). Quantifying the error in estimated transfer functions with application to model order selection. *IEEE Trans. Autom. Control* **37**, 913–928.

Hakvoort, R.G. and P.M.J. Van den Hof (1997). Identification of probabilistic system uncertainty regions by explicit evaluation of bias and variance errors. *IEEE Trans. Automatic Control* **42**(11), 1516–1528.

Hjalmarsson, H. (1993). Asymptotic correct correlation tests in model validation. Proceedings of the 32nd Conference on Decision and Control. San Antonio, Texas, USA. pp. 2058–2059.

Hjalmarsson, H. and L. Ljung (1992). Estimating model variance in the case of undermodeling. *IEEE Trans. Automatic Control* **37**, 1004–1008.

Ljung, L. (1999*a*). *Model Validation and Model Error Modeling*. pp. 15–42. B. Wittenmark and A. Rantzer, Eds.

Ljung, L. (1999*b*). *System Identification: Theory for the User, 2nd Edition*. Prentice-Hall. Englewood Cliffs, NJ.

Ljung, L. (2003). System identification toolbox in matlab. The Mathworks, Inc.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice-Hall International. Hemel Hempstead, Hertfordshire.

Tjärnström, F. and L Ljung (2002). Using the bootstrap to estimate the variance in the case of undermodeling. *IEEE Transactions on Automatic Control* **47**(2), 395 – 398.