Available online at www.sciencedirect.com
SCIENCE DIRECT®
Systems & Control Letters 55 (2006) 494–500
SYSTEMS & CONTROL LETTERS
www.elsevier.com/locate/sysconle

The asymptotic model quality assessment for instrumental variable identification revisited

S. Garatti[a], M.C. Campi[b,*], S. Bittanti[a]

[a]Department of Electronics and Information, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[b]Department of Electrical Engineering for Automation, University of Brescia, Via Branze 38, 25123 Brescia, Italy

Received 9 December 2004; received in revised form 27 September 2005; accepted 24 October 2005
Available online 9 December 2005

Abstract

In this paper the problem of computing uncertainty regions for models identified through an instrumental variable technique is considered. Recently, it has been pointed out that, in certain operating conditions, the asymptotic theory of system identification (the most widely used method for model quality assessment) may deliver unreliable confidence regions. The aim of this paper is to show that, in an instrumental variable setting, the asymptotic theory exhibits a certain "robustness" that makes it reliable even with a moderate number of data samples. Reasons for this are highlighted in the paper through a theoretical analysis and simulation examples.

Keywords: System identification; Model quality assessment; Asymptotic theory; Instrumental variable

1. Introduction

Model quality assessment is an important (and also challenging) problem in system identification. In fact, it has been widely recognized that an identified model is of little use in practical applications if an estimate of its reliability is not provided together with the model itself. In other words, if S is the data-generating system and $\widehat{S}$ is the identified model, it is fundamental to characterize the system-model mismatch, i.e. the distance between S and $\widehat{S}$ (see [11,9,5,1]).

One of the best-known tools for model quality assessment is the asymptotic theory of system identification [10,13]. The asymptotic theory works in a probabilistic framework and returns asymptotic ellipsoidal confidence regions for S—namely, regions in the parameter space to which the data-generating system parameter belongs with a pre-assigned probability when the number of data grows unbounded.

In real applications, the major drawback with the use of the asymptotic theory is that only a finite number of data points is available. Consequently, the asymptotic theory applies only approximately, and it is a common experience that it returns sensible results in many cases, but not always. As a matter of fact, it has been recently shown that—in condition of *poor excitation* and depending on the underlying identification setting—the ellipsoid obtained through the asymptotic theory may even be completely unreliable (see [3,6]).

This limitation of the asymptotic theory is quite severe because lack of excitation is common in many applications, particularly when the identification has to be performed in closed-loop with restricted bandwidth. This happens, for example, at the first iterations of iterative controller design schemes (see [2,4,7,8,14]). Moreover, at a more general level, one can argue that the model quality assessment is even more important when the system is poorly excited as this means that the system-model mismatch is significant.

Our previous contribution [6] focuses on prediction error minimization (PEM) identification techniques and shows the problems which may arise if the model structure is not appropriately selected relative to the identification setup. Herein, we consider the instrumental variable (IV) identification methods and we investigate the applicability of the asymptotic theory for the assessment of the model quality in situations where poor information may occur. The good news conveyed by this paper is

* Corresponding author. Tel.: +39 030 3715458; fax: +39 030 380014.
E-mail addresses: sgaratti@elet.polimi.it (S. Garatti), campi@ing.unibs.it (M.C. Campi), bittanti@elet.polimi.it (S. Bittanti).
URL: http://bsing.ing.unibs.it/∼campi/ (M.C. Campi).

0167-6911/$ - see front matter
doi:10.1016/j.sysconle.2005.10.003

that in IV settings the asymptotic theory exhibits a “robustness” property so that it can be safely used in real applications even in case of poor excitation and for moderate data samples. The reasons for such a “robustness” are highlighted through theoretical arguments.

1.1. Structure of the paper

In Section 2 the IV identification setting is presented and a brief summary of the standard asymptotic theory is given. Moreover, the problems that may arise when using the asymptotic theory in presence of poor excitation are pointed out. Section 3 delivers a new asymptotic result, also valid in “singular” conditions, precisely defined in Section 3. This result makes it possible to show in Section 4 that the asymptotic theory for IV methods can be safely used even when data are poorly exciting. Some simulation results are given in Section 5.

2. Model quality assessment for IV identification

2.1. Mathematical setting

Throughout the paper we suppose that the data are generated by the following dynamical system, which is assumed to be asymptotically stable:

$$y(t) = \varphi(t)'\vartheta^0 + v(t), \tag{1}$$

where

$$\varphi(t) = [y(t-1)\dots y(t-n_a) \quad u(t-1)\dots u(t-n_b)]'$$

is the n-vector ($n = n_a + n_b$) of observations and

$$\vartheta^0 = [-a_1^0 \dots - a_{n_a}^0 \quad b_1^0 \dots b_{n_b}^0]'$$

is the true system parameter vector, supposed to be an interior point of an a priori known compact set Θ.

We will also write system (1) in the operational form

$$A(z^{-1})y(t) = B(z^{-1})u(t) + v(t),$$

where

$$A(z^{-1}) = 1 + a_1^0 z^{-1} + \dots + a_{n_a}^0 z^{-n_a},$$
$$B(z^{-1}) = b_1^0 z^{-1} + \dots + b_{n_b}^0 z^{-n_b},$$

and z^{-1} is the unit-time delay operator.

The input $u(t)$ and the residual process $v(t)$ are generated according to the following scheme which encompasses closed-loop as well as open-loop configurations:

$$u(t) = G(z^{-1})r(t) + H(z^{-1})e(t), \quad v(t) = V(z^{-1})e(t), \tag{2}$$

where $G(z^{-1})$, $H(z^{-1})$, $V(z^{-1})$, $r(t)$ and $e(t)$ satisfy the following assumption.

Assumption 1. The transfer functions $G(z^{-1})$, $H(z^{-1})$ and $V(z^{-1})$ are rational, proper and asymptotically stable. In addition, $V(z^{-1})$ has no zeroes on the unit circle in the complex plane. $e(t)$ is a sequence of independent zero mean random variables with variance $\lambda^2 > 0$ and such that $\mathbb{E}[|e(t)|^{4+\delta}] < \infty$, for some $\delta > 0$. $r(t)$ is a wide sense stationary, stochastic, ergodic, external input sequence. $r(t)$ and $e(t)$ are independent.

Remark 1. For subsequent use we note that both $u(t)$ and $y(t)$ can be seen as the sum of two independent processes, one depending on $r(t)$ and the other one depending on $e(t)$. That is, $u(t) = u_r(t) + u_e(t)$ and $y(t) = y_r(t) + y_e(t)$, where

$$u_r(t) = G(z^{-1})r(t), \quad u_e(t) = H(z^{-1})e(t),$$

$$y_r(t) = \frac{B(z^{-1})}{A(z^{-1})}G(z^{-1})r(t),$$

$$y_e(t) = \frac{B(z^{-1})}{A(z^{-1})}H(z^{-1})e(t) + \frac{1}{A(z^{-1})}V(z^{-1})e(t).$$

According to the IV technique [10,13,12] the estimate $\widehat{\vartheta}_N$ of ϑ^0 is computed as

$$\widehat{\vartheta}_N = \mathrm{sol}_{\vartheta\in\Theta}\left\{\frac{1}{N}\sum_{t=1}^{N}\zeta(t)\varphi(t)'\vartheta = \frac{1}{N}\sum_{t=1}^{N}\zeta(t)y(t)\right\}, \tag{3}$$

where N is the number of data points and $\zeta(t)$, the so-called *instrumental variable*, is a n-dimensional, stationary, stochastic process, uncorrelated with the residual process $v(t)$ and correlated with the observation vector $\varphi(t)$.

Throughout the paper we assume that $\zeta(t)$ is chosen as follows:

Assumption 2. $\zeta(t) = \varphi_r(t)$, where $\varphi_r(t)$ is defined as

$$[y_r(t-1)\dots y_r(t-n_a) \quad u_r(t-1)\dots u_r(t-n_b)]'.$$

In other words, the instrumental vector $\zeta(t)$ is the part of the observation vector depending on the external input sequence $r(t)$.

Remark 2. The choice $\zeta(t) = \varphi_r(t)$ is optimal in that it minimizes the estimation error variance (see [12]). In practice, the typical way of generating $\varphi_r(t)$ is to first identify an initial model (through some identification method) and then by operating this model with the only signal $r(t)$ active. This procedure can be refined in an iterative way.

Let Θ^* be the set of solutions to equation

$$\mathbb{E}[\zeta(t)\varphi(t)']\vartheta = \mathbb{E}[\zeta(t)y(t)]. \tag{4}$$

It can be proved (see [10,12,13]) that, in the present setting, the distance between $\widehat{\vartheta}_N$ and $\Theta^* \cap \Theta$ tends to zero, as $N \to \infty$.

Moreover, thanks to Assumption 2 and Eq. (1), Eq. (4) can be rewritten as

$$\mathbb{E}[\varphi_r(t)\varphi(t)']\vartheta = \mathbb{E}[\varphi_r(t)\varphi(t)']\vartheta^0 + \mathbb{E}[\varphi_r(t)v(t)],$$

and, since $\varphi(t) = \varphi_r(t) + \varphi_e(t)$ and $r(t)$ is independent of $e(t)$, the last equation is equivalent to

$$\mathbb{E}[\varphi_r(t)\varphi_r(t)'](\vartheta - \vartheta^0) = 0. \tag{5}$$

It follows that the cardinality of Θ^* depends on the rank of the matrix $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ and that ϑ^0 always belongs to Θ^*. Thus, if $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ is nonsingular, then Θ^* is the singleton $\{\vartheta^0\}$ and $\widehat{\vartheta}_N \to \vartheta^0$ as $N \to \infty$.

2.2. Asymptotic theory

We turn now to the problem of evaluating the accuracy of a model estimated through the IV method. The asymptotic Theorem 1 can be trivially obtained from the general results presented in [10,12,13]. Before stating the theorem some preliminaries are in order.

Suppose that $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ is nonsingular. Then, let

$$Q_\alpha = \lambda^2 \mathbb{E}[\varphi_r^\alpha(t)\varphi_r^\alpha(t)'], \tag{6}$$

where $\varphi_r^\alpha(t) = \sum_{i=0}^{\infty} \alpha_i \varphi_r(t-i)$ and α_i are the Markov coefficients of $V(z^{-1})$, viz. $V(z^{-1}) = \sum_{i=0}^{\infty} \alpha_i z^{-i}$. In other words, $\varphi_r^\alpha(t)$ is the observation vector $\varphi(t)$ filtered through $V(z^{-1})$.

Further, let

$$P_\alpha = \mathbb{E}[\varphi_r(t)\varphi_r(t)']^{-1} Q_\alpha \mathbb{E}[\varphi_r(t)\varphi_r(t)']^{-1} \tag{7}$$

and consider the following ellipsoid centered in $\widehat{\vartheta}_N$ and intersected with Θ:

$$\mathscr{E}_\alpha(r) = \{\vartheta \in \Theta : (\widehat{\vartheta}_N - \vartheta)' P_\alpha^{-1} (\widehat{\vartheta}_N - \vartheta) \leqslant r\}, \tag{8}$$

where r, the so-called "size" of the ellipsoid, is a real positive number.

Remark 3. It is perhaps worth mentioning that assuming that $V(z^{-1})$ has no zeroes on the unit circle (Assumption 1) serves the purpose of guaranteeing that the definition of $\mathscr{E}_\alpha(r)$ is well posed, i.e. P_α is invertible. See Appendix A.1 for details.

The following theorem suggests how to select r so that $\mathscr{E}_\alpha(r)$ is an ellipsoidal confidence region for ϑ^0 of pre-assigned asymptotic probability.

Theorem 1. *Under the assumptions in this section, we have that*

$$\lim_{N\to\infty} \mathbb{P}\left\{\vartheta^0 \in \mathscr{E}_\alpha\left(\frac{\rho(p)}{N}\right)\right\} = p,$$

where $\rho(p)$ is the inverse of the function $p = \int_0^\rho f_{\chi^2}(x)\,\mathrm{d}x$ and $f_{\chi^2}(x)$ is the probability density of a χ^2 random variable with n degrees of freedom.

In the practical computation of $\mathscr{E}_\alpha(r)$, Q_α and P_α cannot be exactly computed for the following two reasons

(i) The Markov coefficients α_i's and the noise variance λ^2 are not known.
(ii) The expectations cannot be exactly evaluated.

As for (ii), the common way to proceed is to replace $\mathbb{E}$ with $1/N\sum$ i.e. to resort to the sample counterparts of the expectations. This is asymptotically correct since $1/N\sum \to \mathbb{E}$ almost surely as $N \to \infty$.

Point (i) plays a significant role in the analysis to come and is therefore discussed separately in the subsequent Section 2.3.

2.3. Discussion on the practical use of the asymptotic theory

The exact computation of $\mathscr{E}_\alpha(\rho(p)/N)$ requires the knowledge of λ^2 and $V(z^{-1})$ (see Eqs. (6)–(8)). However, both these quantities are unknown in practice and have to be identified from data.

To estimate λ^2 and $V(z^{-1})$, a common choice is to identify an ARMA model describing the residual error $\varepsilon(t,\widehat{\vartheta}_N) = y(t) - \varphi(t)'\widehat{\vartheta}_N$. This is motivated by the fact that $\varepsilon(t,\widehat{\vartheta}_N) \to v(t) = V(z^{-1})e(t)$ as $N \to \infty$, since, under the assumption of Theorem 1, $\widehat{\vartheta}_N \to \vartheta^0$.

In a practical application, the number of data points is finite so that $\widehat{\vartheta}_N \neq \vartheta^0$ and λ^2 and $V(z^{-1})$ cannot be identified exactly. However, when $\varphi_r(t)$ is well exciting (and therefore $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ is positive definite with all the eigenvalues away from zero) we have $\widehat{\vartheta}_N \approx \vartheta^0$ and the introduced approximation is small.

Consider now the situation of poorly exciting inputs, so that matrix $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ has some eigenvalues close to zero. As long as $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ is not exactly singular, it is still true that the estimate $\widehat{\vartheta}_N$ converges to the true system parameter ϑ^0 as $N \to \infty$. However, such a convergence takes place with a very slow rate and it may happen that $\widehat{\vartheta}_N$ is far from ϑ^0 even for a large N. In this case it is no longer true that $\varepsilon(t,\widehat{\vartheta}_N)$ approximates $v(t)$, so that λ^2 and $V(z^{-1})$ cannot be identified with a good accuracy.

Thus, one could think that, in case of poor excitation, the asymptotic theory may lead to erroneous results. One of the main scopes of this paper is to prove that this is not so and the asymptotic theory can still be safely applied.

To this aim, we first develop in the next section a new asymptotic theory valid for the singular case (lack of excitation) and, then, we show in Section 4 that, in the light of this new theory, the asymptotic results maintain their applicability in case of poor excitation.

3. Asymptotic theory for the singular case

Let us assume now that $\det \mathbb{E}[\varphi_r(t)\varphi_r(t)'] = 0$, i.e. we are in the singular case. The aim of this section is to show that a result similar to Theorem 1 still holds true.

As it has been already noted in Section 2.1, if matrix $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ is singular, then the set of asymptotic estimates Θ^* is not a singleton, but it is an affine subspace whose dimensionality d is equal to the dimension of the kernel of $\mathbb{E}[\varphi_r(t)\varphi_r(t)']$ (see Eq. (5)). Refer the parameter space to a basis having the first d components parallel to Θ^*, and the remaining $n-d$ components orthogonal to Θ^*. Let x be the first d coordinates in this basis, whereas z are the remaining $n-d$ coordinates (see Fig. 1 for a graphical representation in a bi-dimensional space). Thus, $[(x^0)'\ (z^0)']'$ and $[(\widehat{x}_N)'\ (\widehat{z}_N)']'$ represent ϑ^0 and $\widehat{\vartheta}_N$, respectively. Furthermore, in such coordinates, we have that $\Theta^* = \{[x'\ z']' : z = z^0\}$.

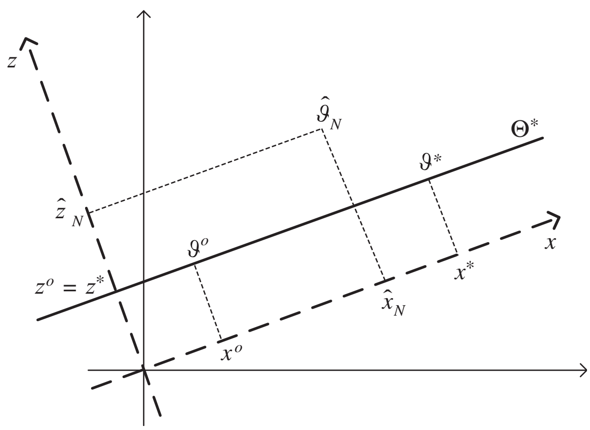

Fig. 1. The parameter space.

In the present singular setting, matrix $(1/N)\sum\zeta(t)\varphi(t)' = (1/N)\sum\varphi_r(t)\varphi(t)'$ in Eq. (3) is singular itself, leaving a degree of freedom in the choice of $\widehat{\vartheta}_N$. In the sequel we assume that $\widehat{\vartheta}_N$ is fixed by a suitable deterministic tie-break rule such that $\widehat{\vartheta}_N$ tends almost surely to a limit estimate $\vartheta^* = [(x^*)'\ (z^*)']'$, where ϑ^* is an interior point of Θ. As an example, if Θ is a sphere, as a tie-break rule one can take the $\widehat{\vartheta}_N$ with smaller distance from the sphere center. Note that, though $\vartheta^* \in \Theta^*$ (and, therefore, $z^* = z^0$), $\vartheta^* \neq \vartheta^0$ in general since $x^* \neq x^0$.

We now turn our attention to the problem of model quality assessment. We concentrate on characterizing how $\widehat{z}_N$ converges to z^0 (Theorem 2). In contrast, the difference $\widehat{x}_N - x^0$ does not go to zero as $\widehat{x}_N \to x^* \neq x^0$ so that $\widehat{x}_N - x^0$ tends to a deterministic number that depends on the introduced tie-break rule.

We need a simple preliminary lemma. Let $\varphi_r^x(t)$ be the projection of $\varphi_r(t)$ along the x-axis, and $\varphi_r^z(t)$ the projection along the z-axis. We have:

Lemma 1. $\varphi_r^x(t) = 0$ *almost surely, while* $\mathbb{E}[\varphi_r^z(t)\varphi_r^z(t)']$ *is nonsingular. Moreover,* $\varepsilon(t,\vartheta^*) = y(t) - \varphi(t)'\vartheta^*$ *only depends on* $e(t)$ *and can be written as* $\sum_{i=0}^{\infty}\beta_i e(t-i)$, *for suitable* β_i*'s.*

Proof. See Appendix A.2. □

Now, let

$$Q_\beta^z = \lambda^2\mathbb{E}\left[\sum_{i=0}^{\infty}\beta_i\varphi_r^z(t-i)\sum_{j=0}^{\infty}\beta_j\varphi_r^z(t-j)'\right],$$

and

$$P_\beta^z = \mathbb{E}[\varphi_r^z(t)\varphi_r^z(t)']^{-1}Q_\beta^z\mathbb{E}[\varphi_r^z(t)\varphi_r^z(t)']^{-1}.$$

Furthermore, consider the following ellipsoid centered in $\widehat{z}_N$:

$$\mathscr{E}_\beta^z(r) = \{z \in \Theta_z : (\widehat{z}_N - z)'(P_\beta^z)^{-1}(\widehat{z}_N - z) \leqslant r\},$$

where Θ_z is the point-wise projection of Θ onto the z-axis and r is again the size of the ellipsoid and is a real positive number.

Remark 4. $\mathbb{E}[\varphi_r^z(t)\varphi_r^z(t)']^{-1}$ exists in view of Lemma 1. Instead, similarly to Q_α in Remark 3, invertibility of Q_β^z requires that $\sum_{i=0}^{\infty}\beta_i z^{-i}$ has no zeroes on the unit circle. Such condition is assumed here for granted.

The following theorem suggests how to select r so that $\mathscr{E}_\beta^z(r)$ is an ellipsoidal confidence region for z^0 of pre-assigned asymptotic probability.

Theorem 2. *We have that*

$$\lim_{N\to\infty}\mathbb{P}\left\{z^0 \in \mathscr{E}_\beta^z\left(\frac{\rho(p)}{N}\right)\right\} = p,$$

where $\rho(p)$ *is the inverse of the function* $p = \int_0^\rho f_{\chi^2}(x)\,\mathrm{d}x$ *and* $f_{\chi^2}(x)$ *is the probability density of a* χ^2 *random variable with* $n-d$ *degrees of freedom.*

Proof. See Appendix A.3. □

Note that $\beta_i \neq \alpha_i$ in general. Thus, if one uses the Markov coefficients α_i's of $V(z^{-1})$ when computing $\mathscr{E}_\beta^z$, the resulting ellipsoid fails to represent a confidence region with the pre-assigned level of confidence. What is remarkable in Theorem 2 is that, in order to compute a correct ellipsoid, one has to use alternative coefficients β_i's and these coefficients can in fact be estimated from the residual error $\varepsilon(t,\widehat{\vartheta}_N)$ since $\widehat{\vartheta}_N \to \vartheta^*$ and $\varepsilon(t,\vartheta^*) = \sum_{i=0}^{\infty}\beta_i e(t-i)$ (see Lemma 1).

Remark 5. It is worth mentioning that Theorem 2 is a generalization of Theorem 1. As a matter of fact, in the nonsingular case, $d=0$ so that $z=\vartheta$ and the statement of Theorem 2 reduces to that of Theorem 1.

In view of the result of Theorem 2, it is possible to determine a confidence region for ϑ^0 (and not only for z^0).

Since the difference between $\widehat{x}_N$ and x^0 remains unpredictable ($\widehat{x}_N$ tends to x^* and not to x^0), the natural choice is to consider the degenerate ellipsoid

$$\mathscr{D}\mathscr{E}_\beta\left(\frac{\rho(p)}{N}\right) = \left\{\vartheta = [x'\ z']' \in \Theta : (\widehat{z}_N - z)'(P_\beta^z)^{-1}(\widehat{z}_N - z) \leqslant \frac{\rho(p)}{N}\right\},$$

which is nothing but the ellipsoid $\mathscr{E}_\beta^z(\rho(p)/N)$ extended along the x direction and intersected with Θ.

Then, as a direct consequence of Theorem 2, we have the following theorem saying that $\mathscr{D}\mathscr{E}_\beta(\rho(p)/N)$ is an asymptotic p-confidence region for ϑ^0.

Theorem 3. *We have that*

$$\lim_{N\to\infty}\mathbb{P}\left\{\vartheta^0 \in \mathscr{D}\mathscr{E}_\beta\left(\frac{\rho(p)}{N}\right)\right\} = p,$$

where $\rho(p)$ *is the inverse of the function* $p = \int_0^\rho f_{\chi^2}(x)\,\mathrm{d}x$ *and* $f_{\chi^2}(x)$ *is the probability density of a* χ^2 *random variable with* $n-d$ *degrees of freedom.*

4. Use of the new asymptotic results in practice

Consider an identification problem with a finite number, say N, of data points. We recall in the following the standard way the asymptotic theory is used in practice. Then, we show that this way of proceeding is valid not only in the case of full excitation ($\widehat{\vartheta}_N \approx \vartheta^0$) but also in poor excitation conditions ($\widehat{\vartheta}_N$ far from ϑ^0). In other words, the standard procedure is robust, as it will be seen thanks to the new asymptotic theory developed in the previous section.

The standard procedure can be outlined as follows. After estimating $\widehat{\vartheta}_N$, compute the associated prediction error $\varepsilon(t, \widehat{\vartheta}_N)$ and then estimate a model $\sum_{i=1}^{\infty} \gamma_i e(t-i)$ describing such a prediction error in term of the white noise $e(t)$, and estimate the variance of $e(t)$ (let $\widehat{\lambda}^2$ denote such an estimate). Here, γ_i's are the coefficients estimated from data and, depending on the context in the discussion to follow, they represent either an estimate of the α_i's (see Eq. (6)) or an estimate of the β_i's (see Lemma 1). Then, for a given confidence probability p, we compute the ellipsoid $\widehat{\mathscr{E}}_\gamma(\rho(p)/N)$ along the line traced in Section 2, namely,

$$\widehat{\mathscr{E}}_\gamma\left(\frac{\rho(p)}{N}\right) = \left\{\vartheta : (\widehat{\vartheta}_N - \vartheta)' \widehat{P}_\gamma^{-1} (\widehat{\vartheta}_N - \vartheta) \leqslant \frac{\rho(p)}{N}\right\}, \tag{9}$$

where

$$\widehat{P}_\gamma = \left(\frac{1}{N}\sum_{t=1}^{N} \varphi_r(t)\varphi_r(t)'\right)^{-1} \frac{\widehat{\lambda}^2}{N} \sum_{t=1}^{N} \varphi_r^\gamma(t)\varphi_r^\gamma(t)' \times \left(\frac{1}{N}\sum_{t=1}^{N} \varphi_r(t)\varphi_r(t)'\right)^{-1},$$

$\varphi_r^\gamma(t) = \sum_{i=0}^{\infty} \gamma_i \varphi_r(t-i)$ and $\rho(p)$ is such that $p = \int_0^{\rho(p)} f_{\chi^2}(x)\,\mathrm{d}x$ where $f_{\chi^2}(x)$ is the probability density of a χ^2 random variable with n degrees of freedom. To motivate the validity of this way of proceeding, suppose first that the regressor $\varphi_r(t)$ excites well all the directions in the parameter space (full excitation case). Then, $\widehat{\vartheta}_N \approx \vartheta^0$ so that the γ_i's become an estimate of the α_i's and $\widehat{\mathscr{E}}_\gamma(\rho(p)/N) \approx \mathscr{E}_\alpha(\rho(p)/N)$ so that Theorem 1 applies to conclude that a reliable estimate of a p-confidence region for ϑ^0 has been computed.

The crucial fact is that formula (9) is also motivated in case of poor excitation where $\widehat{\vartheta}_N$ is far from ϑ^0 (the case where estimating the mismatch between $\widehat{\vartheta}_N$ and ϑ^0 is in fact more significant) as we next discuss.

The situation of poor excitation where $\widehat{\vartheta}_N$ is far from ϑ^0 can be interpreted as a perturbed version of the singular setting where $\widehat{\vartheta}_N$ converges to some ϑ^* different from ϑ^0. As we have seen in Section 3, Theorem 3, the degenerate ellipsoid $\mathscr{DE}_\beta$ should be used in this case and, correspondingly, the β_i's are the coefficients to be computed for the construction of $\mathscr{DE}_\beta$. In this respect, note that the γ_i's are estimates of the β_i's in this case since $\widehat{\vartheta}_N \approx \vartheta^*$ (see Lemma 1). Moreover, due to the poor excitation condition, the ellipsoid $\widehat{\mathscr{E}}_\gamma(\rho(p)/N)$ turns out to be elongated in the poor excitation direction, so that $\widehat{\mathscr{E}}_\gamma \approx \mathscr{DE}_\beta$. This motivates the use of the standard asymptotic approach for model quality assessment also in this case.

Perhaps, it should be finally noted that the $\rho(p)$ in (9) refers to a χ^2 with n degrees of freedom while a $n-d$ degrees of freedom χ^2 distribution should be used, as stated by Theorem 3. This is a slight inaccuracy of the standard asymptotic theory which is remarked here, though its effects are quite negligible as it results in a slight over-bounding of the confidence region. See the simulation results in the next section.

5. Simulation results

The simulation example of the present section serves the purpose of illustrating the theory and it is not intended as a real application example. Correspondingly, the simplest possible situation has been selected. While the situation is artificial, the drawn conclusions bear a breath of general applicability.

We considered a first order data-generating system with $\vartheta^0 = [-a^0\ b^0]' = [0.9\ 0.1]'$ and $V(z^{-1}) = 1 + 0.5z^{-1}$. That is

$$y(t) = 0.9y(t-1) + 0.1u(t-1) + e(t) + 0.5e(t-1),$$

where $e(t) = WGN(0, 1)$. As is obvious, during the simulation, ϑ^0 was assumed to be unknown. The only a priori information we assumed was that $\vartheta^0 \in \Theta = \{[a\ b]' : a^2 + b^2 \leqslant 20\}$.

To identify this system, the plant was operated in open-loop with $u(t) = r(t)$, and the IV technique was used with $\varphi(t) = [y(t-1)\quad u(t-1)]'$ and $\zeta(t) = \varphi_r(t) = [y_r(t-1)\quad r(t-1)]'$, where

$$y_r(t) = 0.9y_r(t-1) + 0.1r(t-1).$$

As input signal, we used $u(t) = 1 + \xi(t)$, where $\xi(t) = WGN(0, 10^{-6})$. Note that the variance of $\xi(t)$ is very small as compared to the noise variance so that the input $u(t)$ is poorly exciting ($u(t)$ is nearly exciting of order 1 while two parameters had to be identified).

The identification was repeated 500 times, by using $N = 5000$ data points each time. In each experiment a parameter vector $\widehat{\vartheta}_N^i = [-\widehat{a}_N^i\ \widehat{b}_N^i]'$, $i = 1 \ldots 500$, was identified and a 95%-confidence region was estimated as $\widehat{\mathscr{E}}_\gamma^i(\rho(0.95)/N)$ (see Section 4). The true parameter ϑ^0 turned out to belong to $\widehat{\mathscr{E}}_\gamma^i(\rho(0.95)/N)$ in 489 cases out of 500, that is, with empirical frequency of 97.8%.

Remark 6. Note that the estimated ellipsoid $\widehat{\mathscr{E}}_\gamma^i(\rho(0.95)/N)$ turns out to be an over-bound of a 95%-confidence region (the empirical fraction of $\vartheta^0 \in \widehat{\mathscr{E}}_\gamma^i(\rho(0.95)/N)$ was 97.8%), because as explained in Section 4 the value $\rho(0.95)$ was computed referring to a χ^2 with 2 (instead of 1) degrees of freedom.

As an interesting comparison, we further computed the 95% confidence region with the true parameters α_i's ($\sum_{i=0}^{\infty}\alpha_i z^{-i} = V(z^{-1})$). The results are summarized in Table 1. As it appears, using the true parameters α_i's leads to wrong results (the success rate of $\vartheta^0 \in \widehat{\mathscr{E}}_\alpha^i(\rho(0.95)/N)$ was of only 48.4%). The reason for such a bias is explained in Section 4.

Table 1
Reliability of the estimated confidence region

	ϑ^0 in $\mathscr{E}$	ϑ^0 out of $\mathscr{E}$	% of success
$\widehat{\mathscr{E}}_\alpha\left(\frac{\rho(0.95)}{N}\right)$	242	258	48.4
$\widehat{\mathscr{E}}_\gamma\left(\frac{\rho(0.95)}{N}\right)$	489	11	97.8

6. Concluding remarks

In this paper, a new asymptotic result, valid also in a singular case, has been developed for an IV identification setting. Grounded on this new result, we have shown that the asymptotic theory can be safely used for model quality assessment, even in the case of poor excitation and moderate data samples.

Acknowledgements

Paper supported by the Italian National Research Project "New Techniques of Identification and Adaptive Control for Industrial Systems" and partially by CNR-IEIIT.

Appendix A

A.1. Complements to Remark 3

Note that, P_α is invertible provided that Q_α is nonsingular. Here, we show that $Q_\alpha > 0$.

Let v be a generic vector of $\mathbb{R}^n$ and consider $v'Q_\alpha v = \lambda^2 \cdot v'\mathbb{E}[\varphi_r^\alpha(t)\varphi_r^\alpha(t)']v = \lambda^2 \cdot v'\mathbb{E}[(\varphi_r^\alpha(t) - \mathbb{E}[\varphi_r^\alpha(t)])(\varphi_r^\alpha(t) - \mathbb{E}[\varphi_r^\alpha(t)])']v + \lambda^2 \cdot v'\mathbb{E}[\varphi_r^\alpha(t)]\mathbb{E}[\varphi_r^\alpha(t)]'v$.

Since $\varphi_r^\alpha(t) = V(z^{-1})\varphi_r(t)$, we obtain, through the Parseval identity,

$$v'Q_\alpha v = \frac{\lambda^2}{2\pi}\int_{-\pi}^{\pi} v'\Phi_r(\mathrm{e}^{\mathrm{j}\omega})v \cdot |V(\mathrm{e}^{\mathrm{j}\omega})|^2\,\mathrm{d}\omega + \lambda^2 V(1)^2 \cdot v'\mathbb{E}[\varphi_r(t)]\mathbb{E}[\varphi_r(t)]'v,$$

where $\Phi_r(\mathrm{e}^{\mathrm{j}\omega})$ is the spectrum of the n-dimensional process $\varphi_r(t)$.

This implies that

$$v'Q_\alpha v \geqslant \min_{\omega\in[-\pi,\pi]} |V(\mathrm{e}^{\mathrm{j}\omega})|^2 \cdot \frac{\lambda^2}{2\pi}\int_{-\pi}^{\pi} v'\Phi_r(\mathrm{e}^{\mathrm{j}\omega})v\;\mathrm{d}\omega + \lambda^2 V(1)^2 \cdot v'\mathbb{E}[\varphi_r(t)]\mathbb{E}[\varphi_r(t)]'v.$$

Applying now the assumption that $V(z^{-1})$ has no zeroes on the unit circle we have $\min_{\omega\in[-\pi,\pi]}|V(\mathrm{e}^{\mathrm{j}\omega})|^2 = k > 0$. Since, in addition, $\lambda^2 > 0$ and $E[\varphi_r(t)\varphi_r(t)'] > 0$ by assumption, we conclude that

$$v'Q_\alpha v \geqslant k\lambda^2\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} v'\Phi_r(\mathrm{e}^{\mathrm{j}\omega})v\;\mathrm{d}\omega + v'\mathbb{E}[\varphi_r(t)]\mathbb{E}[\varphi_r(t)]'v\right) = k\lambda^2 \cdot v'\mathbb{E}[\varphi_r(t)\varphi_r(t)']v > 0 \quad \forall v \neq 0,$$

i.e. Q_α is positive definite.

A.2. Proof of Lemma 1

Let T be the $n \times n$ rotation matrix such that $T\vartheta = [x'\ z']'$. Referring Eq. (5) to the x, z coordinates (i.e. $T\mathbb{E}[\varphi_r(t)\varphi_r(t)']T'T(\vartheta - \vartheta^0) = 0$), we obtain

$$\mathbb{E}\begin{bmatrix} \varphi_r^x(t)\varphi_r^x(t)' & \varphi_r^x(t)\varphi_r^z(t)' \\ \varphi_r^z(t)\varphi_r^x(t)' & \varphi_r^z(t)\varphi_r^z(t)' \end{bmatrix}\begin{bmatrix} x - x^0 \\ z - z^0 \end{bmatrix} = 0.$$

Since $[x'\quad z']'$ is a solution of this equation if and only if $z = z^0$, while each value of x is feasible, it follows that $\mathbb{E}[\varphi_r^z(t)\varphi_r^z(t)']$ must be nonsingular, while $\mathbb{E}[\varphi_r^x(t)\varphi_r^x(t)']$ must be equal to zero so that $\varphi_r^x(t) = 0$, almost surely.

Consider now $\varepsilon(t, \vartheta^*) = y(t) - \varphi(t)'\vartheta^*$. It can be rewritten as

$$\varphi(t)'(\vartheta^0 - \vartheta^*) + v(t) = \varphi^x(t)'(x^0 - x^*) + \varphi^z(t)'(z^0 - z^*) + v(t),$$

where $[\varphi^x(t)'\quad \varphi^z(t)']' = T\varphi(t)$. Noting that $z^0 = z^*$ and that $\varphi^x(t)' = \varphi_r^x(t)' + \varphi_e^x(t)' = \varphi_e^x(t)'$ almost surely, we obtain

$$\varepsilon(t, \vartheta^*) = \varphi_e^x(t)'(x^0 - x^*) + v(t).$$

Thus, $\varepsilon(t, \vartheta^*)$ is the stationary output of a dynamical linear system fed by $e(t)$, and $\sum_{i=0}^{\infty}\beta_i e(t-i)$ is the Markov representation of such a process.

A.3. Proof of Theorem 2

Referring the equation in (3) to the x, z coordinates, we have that

$$\frac{1}{N}\sum_{t=1}^{N}\varphi_r^x(t)\varphi^x(t)'(\widehat{x}_N - x^0) + \frac{1}{N}\sum_{t=1}^{N}\varphi_r^x(t)\varphi^z(t)'(\widehat{z}_N - z^0) = \frac{1}{N}\sum_{t=1}^{N}\varphi_r^x(t)v(t),$$

$$\frac{1}{N}\sum_{t=1}^{N}\varphi_r^z(t)\varphi^x(t)'(\widehat{x}_N - x^0) + \frac{1}{N}\sum_{t=1}^{N}\varphi_r^z(t)\varphi^z(t)'(\widehat{z}_N - z^0) = \frac{1}{N}\sum_{t=1}^{N}\varphi_r^z(t)v(t),$$

with $\varphi^x(t)$ and $\varphi^z(t)$ defined as in the proof of Lemma 1.

The first equation is $0=0$ almost surely, since $\varphi_r^x(t)=0$, almost surely. Instead, inflating the second equation by $\sqrt{N}$ yields

$$\begin{aligned}\frac{1}{N}\sum_{t=1}^{N}\varphi_r^z(t)\varphi^z(t)'\sqrt{N}(\widehat{z}_N-z^0)&=\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\varphi_r^z(t)(v(t)+\varphi_e^x(t)'(x^0-\widehat{x}_N))\\&=\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\varphi_r^z(t)(\varphi_e^x(t)'(x^0-x^*)+v(t))\\&\quad+\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\varphi_r^z(t)\varphi_e^x(t)'(x^*-\widehat{x}_N)\end{aligned}\tag{A.1}$$

almost surely, where we have used the fact that $\varphi^x(t)=\varphi_r^x(t)+\varphi_e^x(t)=\varphi_e^x(t)$, almost surely. Note that the term $\varphi_e^x(t)'(x^0-x^*)+v(t)$ is equal to $\sum_{i=0}^{\infty}\beta_i e(t-i)$ (Lemma 1).

If we now suppose that the second term $(1/\sqrt{N})\sum_{t=1}^{N}\varphi_r^z(t)\varphi_e^x(t)'(x^*-\widehat{x}_N)$ can be neglected (a fact whose proof is postponed below) then, following the same rationale as in [10]—chapter 9—it can be proved that the term $\sqrt{N}(\widehat{z}_N-z^0)$ in (A.1) is asymptotically distributed as a $(n-d)$-dimensional Gaussian random variable with zero mean and variance equal to

$$\mathbb{E}[\varphi_r^z(t)\varphi_r^z(t)']^{-1}\lambda^2\mathbb{E}\left[\sum_{i=0}^{\infty}\beta_i\varphi_r^z(t-i)\sum_{j=0}^{\infty}\beta_j\varphi_r^z(t-j)'\right]\times\mathbb{E}[\varphi_r^z(t)\varphi_r^z(t)']^{-1}=P_\beta^z.$$

From this, the theorem thesis easily follows noting that $N(\widehat{z}_N-z)'(P_\beta^z)^{-1}(\widehat{z}_N-z)$ is asymptotically distributed as a χ^2 random variable with $(n-d)$-degrees of freedom.

Turning now back to the term $(1/\sqrt{N})\sum_{t=1}^{N}\varphi_r^z(t)\varphi_e^x(t)'(x^*-\widehat{x}_N)$, the fact that it can be neglected corresponds to say that it tends to zero in distribution. To show this we prove the stronger convergence to zero in probability.

To ease the notation suppose x and z scalar. Then, for $v>0$ and $k>0$ we have

$$\begin{aligned}&\lim_{N\to\infty}\mathbb{P}\left\{\left|\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\varphi_r^z(t)\varphi_e^x(t)(x^*-\widehat{x}_N)\right|>v\right\}\\&\leqslant\lim_{N\to\infty}\mathbb{P}\left\{\left|\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\varphi_r^z(t)\varphi_e^x(t)\right|>k\cup|x^*-\widehat{x}_N|>\frac{v}{k}\right\}\\&\leqslant\lim_{N\to\infty}\mathbb{P}\left\{\left|\frac{1}{\sqrt{N}}\sum_{t=1}^{N}\varphi_r^z(t)\varphi_e^x(t)\right|>k\right\}\\&\quad+\lim_{N\to\infty}\mathbb{P}\left\{|x^*-\widehat{x}_N|>\frac{v}{k}\right\}.\end{aligned}\tag{A.2}$$

The second term is zero $\forall k>0$ since $x^*-\widehat{x}_N\to 0$ almost surely. In the first, $(1/\sqrt{N})\sum_{t=1}^{N}\varphi_r^z(t)\varphi_e^x(t)$ is asymptotically Gaussian, say $\mathscr{G}(0,\sigma^2)$, so that the first term is the tail probability of $\mathscr{G}(0,\sigma^2)$ and it tends to 0 as $k\to\infty$. Thus, the right-hand side of (A.2) is vanishing with $k\to\infty$. Since the left hand side does not depend on k, it remains proven that it is zero.

References

[1] S. Bittanti (Ed.), Proceedings of the IFAC workshop on Adaptation and Learning in Control and Signal Processing (ALCOSP), Cernobbio, Como, Italy, 2001.

[2] S. Bittanti, M.C. Campi, S. Garatti, An iterative controller design scheme based on average robust control, in: Proceedings of the 15th World IFAC Congress, Barcelona, Spain, 2002.

[3] S. Bittanti, M.C. Campi, S. Garatti, New results on the asymptotic theory of system identification for the assessment of the quality of estimated models, in: Proceedings of the 41st IEEE Conference on Decision and Control, Las Vegas, NV, USA, 2002.

[4] S. Bittanti, M.C. Campi, S. Garatti, Some critical implementation issues in iterative robust control design, in: Proceedings of the IFAC Workshop on Adaptation and Learning in Control and Signal Processing (ALCOSP), Yokohama, Japan, 2004.

[5] S. Bittanti, G. Picci (Eds.), Identification, Adaptation, Learning—The Science of Learning Models from Data, Springer, Berlin, 1996.

[6] S. Garatti, M.C. Campi, S. Bittanti, Assessing the quality of identified models through the asymptotic theory—when is the result reliable?, Automatica 40 (2004) 1319–1332.

[7] M. Gevers, A decade of progress in iterative control design: from theory to practice, in: Symposium on Advanced Control of Chemical Processes, Pisa, Italy, 2000, pp. 677–694.

[8] W.S. Lee, B.D.O. Anderson, R.L. Kosut, I.M.Y. Mareels, On robust performance improvement through the windsurfer approach to adaptive robust control, in: Proceedings of the 32nd IEEE Conference on Decision and Control, San Antonio, TX, USA, 1993, pp. 2821–2827.

[9] L. Ljung, Model validation and model error modeling, in: The Åström Symposium on Control, Lund, Sweden, 1999.

[10] L. Ljung, System Identification: Theory for the User, Prentice-Hall, Upper Saddle River, NJ, 1999.

[11] B. Ninness, G. Goodwin, Estimation of model quality, Automatica 31 (1995) 1771–1795.

[12] T. Söderström, P. Stoica, Instrumental Variable Methods for System Identification, Springer, New York, 1983.

[13] T. Söderström, P. Stoica, System Identification, Prentice-Hall, Englewood Cliffs, NJ, 1989.

[14] P.M.J. Van den Hof, R.J.P. Schrama, Identification and control—closed-loop issues, Automatica 31 (1995) 1751–1770.