New results on the asymptotic theory of system identification for the assessment of the quality of estimated models [1]

S. Bittanti[2] M.C. Campi[3] S. Garatti[2]

Abstract

In this paper the problem of estimating uncertainty regions for identified models is considered. Usually, one resorts to the asymptotic theory of system identification, by means of which ellipsoidal uncertainty regions can be constructed for the uncertain parameters. We show that these uncertainty regions supplied by the asymptotic theory can be unreliable in certain situations precisely characterized in the paper. Then, we investigate on the conditions of validity of the asymptotic theory, and we prove a new statement of more general applicability. Thanks to this statement, we can identify for which standard classes of models (ARMAX, Box Jenkins, etc.) the asymptotic theory can be safely used to assess the estimation quality. These results are of interest in many applications, including iterative controller design schemes.

1 Introduction

Consider a data-generating dynamical system S and a model $\widehat{S}$ of it estimated from data. It has been fully recognized in the literature that the estimated model $\widehat{S}$ is useless without a statement about its quality. In other words, it is fundamental to characterize the error model, i.e. the distance between S and $\widehat{S}$ (see e.g [3], [7], [8], [11] and [12]).

The most commonly used tool for evaluating the error model is the asymptotic theory of system identification. It returns ellipsoidal confidence regions in the space of parameters such that the true system parameters belong to this ellipsoid with a specified probability (see e.g. [10] and [13]). The main advantage in using the asymptotic theory is that the confidence regions can be easily computed from the available data.

On the other hand, asymptotic theory is rigourously correct only when the number of collected data tends to infinity, while in practical applications only a finite number of data points is available. Therefore, the asymptotic theory is used in practice as a heuristic tool for the model quality evaluation. It is a common experience that it returns sensible results in many cases, but not always, as it has been recently shown in [4], [5] and [15].

More precisely, experience suggests that the asymptotic theory returns reliable confidence regions especially when uncertainty is limited (i.e. the estimated model is quite near to the true system).

On the contrary, we will show (Section 3) that, in presence of wide uncertainty (i.e. the estimated model could be very far from the true system), the asymptotic theory may return confidence regions which are completely unreliable.

Note that model quality assessment plays a fundamental role exactly when uncertainty is very scattered, otherwise the error model is small and the estimated model can be safely used instead of the true system.

The aim of this paper is to pinpoint in a clear-cut way the situations where the asymptotic theory may fail to provide sensible results. The major contributions are the following ones.

Contributions of the paper

i) By way of an example, we explain why the asymptotic theory may fail in presence of a high level of uncertainty

ii) We investigate on the conditions of validity of the asymptotic theory, and we prove a new asymptotic result of more general applicability

iii) Thanks to this new statement, we identify the model classes among the standard ones (ARX, ARMAX, Box-Jenkins, etc.) such that the asymptotic theory can be safely used even in presence of a high level of uncertainty

Poorly exciting inputs are the main causes for the presence of wide uncertainty. As a consequence, our results are of interest whenever the system identification has to be performed in conditions of poor excitation, as, for example, when the system is operated in closed-loop. In particular, for iterative design schemes this is even

[1] Paper supported by the National Research Project "New Techniques of Identification and Adaptive Control for Industrial Systems" and by the European Research Project "NACO - Non linear and Adaptive Control".

[2] Department of Electronics and Information, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy, {bittanti,sgaratti}@elet.polimi.it

[3] Department of Electrical Engineering for Automation - University of Brescia, Via Branze 38, 25123 Brescia, Italy, campi@ing.unibs.it

more true because at the first iteration the closed-loop bandwidth is very restricted (see [1], [6], [9], [14]).

Structure of the paper

In Section 2 the assumptions used in the paper are stated and a brief summary of the asymptotic theory is given. This allows us to keep the paper as self contained as possible. Section 3 delivers the example as explained in the point i) above. After a mid-paper conclusion section (Section 4), Section 5 contains the main result (see point ii)) while its heuristic applicability is discussed in Section 6. Finally, in Section 7 we establish for which classes of models asymptotic theory can be safely used.

For conciseness, the proof of all results are omitted. The interested reader can find them in [2].

2 Asymptotic theory

In this section we provide a compendium of the asymptotic theory of system identification with the objective of clarifying the context of our results. For a more comprehensive description of the subject, we refer the reader to the literature (see e.g. [10] and [13]).

2.1 Mathematical setting

Let

$$\mathcal{M}_\vartheta = \Big\{ \widehat{y}(t,\vartheta) = W_u(z^{-1},\vartheta)u(t) + W_y(z^{-1},\vartheta)y(t),\ \vartheta \in \Theta \subseteq \mathbb{R}^n \Big\}$$

be a parameterized set of predictor models, where $W_u(z^{-1},\vartheta)$ and $W_y(z^{-1},\vartheta)$ satisfy the following assumption.

Assumption 1 *$W_u(z^{-1},\vartheta)$ and $W_y(z^{-1},\vartheta)$ are rational strictly proper transfer functions whose coefficients are functions of a parameter $\vartheta \in \Theta \subset \mathbb{R}^n$, Θ compact. The coefficients are four times differentiable and the fourth derivatives are continuous. Moreover, $W_u(z^{-1},\vartheta)$ and $W_y(z^{-1},\vartheta)$ are asymptotically stable, $\forall\vartheta \in \Theta$.*

Processes u and y are generated according to the following scheme.

Assumption 2 *Processes u and y are given by*

$$\begin{aligned} u(t) &= G_u(z^{-1})r(t) + H_u(z^{-1})e(t) \\ y(t) &= G_y(z^{-1})r(t) + H_y(z^{-1})e(t), \end{aligned}$$

where $G_u(z^{-1})$, $G_y(z^{-1})$, $H_u(z^{-1})$ and $H_y(z^{-1})$ are asymptotically stable rational transfer functions. $e(t)$ is a zero mean independent process with constant variance equal to $\lambda^2 > 0$ and such that $\sup_t \mathbb{E}[|e(t)|^{4+\delta}] < \infty$, for some $\delta > 0$. $r(t)$ is a wide sense stationary, ergodic, stochastic, external input sequence. $e(t)$ and $r(t)$ are independent.

Remark 1 *The results given below can be proved even if $r(t)$ is a bounded deterministic external input sequence. Considering a stationary, ergodic reference in Assumption 2 has been preferred since it simplifies the presentation.*

We also require that the data-generating system belongs to the class of models $\mathcal{M}_\vartheta$, that is:

Assumption 3 *There exists a parameter ϑ^o which is an interior point of Θ such that*

$$y(t) = W_u(z^{-1},\vartheta^o)u(t) + W_y(z^{-1},\vartheta^o)y(t) + e(t).$$

Parameter ϑ is estimated by the minimization of the standard quadratic cost:

$$V_N(\vartheta) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t,\vartheta)^2,$$

where N is the number of data points and $\varepsilon(t,\vartheta) = y(t) - \widehat{y}(t,\vartheta)$ is the prediction error.
Thus,

$$\widehat{\vartheta}_N = \arg\min_{\vartheta\in\Theta} V_N(\vartheta).$$

The asymptotic cost criterion is $\overline{V}(\vartheta) = \mathbb{E}[\varepsilon(t,\vartheta)^2]$, while Θ^* denotes its set of minimizers, that is, $\Theta^* = \Big\{\arg\min_{\vartheta\in\Theta} \overline{V}(\vartheta)\Big\}$. Finally, let $\psi(t,\vartheta)$ be equal to $-\frac{\mathrm{d}}{\mathrm{d}\vartheta}\varepsilon(t,\theta)$.
In the asymptotic theory it is assumed that the minimizer of $\overline{V}(\vartheta)$ is unique:

Assumption 4 *The set Θ^* has cardinality equal to 1.*

Remark 2 *Under Assumption 3, it is easy to demonstrate that the parameter ϑ^o always belongs to the set Θ^*. Therefore, Assumption 4 can be rewritten as $\Theta^* = \{\vartheta^o\}$.*

2.2 Asymptotic theory

Let

$$Q_N = \frac{\frac{1}{N}\sum_{t=1}^{N} \psi(t,\widehat{\vartheta}_N)\psi'(t,\widehat{\vartheta}_N)}{\frac{1}{N}\sum_{t=1}^{N} \varepsilon(t,\widehat{\vartheta}_N)^2}$$

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

$$\mathcal{E}(r) = \Big\{\vartheta : (\widehat{\vartheta}_N - \vartheta)' Q_N (\widehat{\vartheta}_N - \vartheta) \le r\Big\}, \qquad (1)$$

where r is a real positive number called the size of the ellipsoid.
We have the following result.

Theorem 1 *Let $p \in [0,1)$. Under Assumptions 1, 2, 3 and 4, it follows that*

$$\lim_{N\to\infty} \mathbb{P}\Big\{\vartheta^o \in \mathcal{E}\Big(\frac{\alpha(p)}{N}\Big)\Big\} = p,$$

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

The above theorem suggests how to select r so as to obtain an ellipsoidal confidence region for ϑ^o of preassigned asymptotic probability. The proof of Theorem 1 can be found in chapter 9 of [10].
The following result is obtained immediately from Theorem 1.

Theorem 2 *Under Assumptions 1, 2, 3 and 4, for any sequence α_N which tends to ∞ as $N \to \infty$, we have that*

$$\lim_{N\to\infty} \mathbb{P}\left\{\vartheta^o \in \mathcal{E}\left(\frac{\alpha_N}{N}\right)\right\} = 1.$$

Remark 3 *As a natural choice for α_N, consider $\alpha_N = \alpha(p)(1+\beta_N)$ with $\beta_N \to \infty$ as $N \to \infty$, that is, the ellipsoid size is inflated by $1+\beta_N$ with respect to Theorem 1. If $\frac{\beta_N}{N} \to 0$, when $N \to \infty$, the ellipsoid size still tends to zero, though with a slower rate than the ellipsoid of Theorem 1. Theorem 2 says that, no matter how slow such an inflation takes place, the true parameter ϑ^o will asymptotically belong to the ellipsoid with confidence* 1.

In real applications, the asymptotic theory is often used so as to generate confidence regions for the system parameters, even if, as is obvious, such a theory applies only approximately since the evaluation is based on a finite number of data points. Though it is common experience that the results are still reliable in many cases even for a moderate data sample, it is also true that in other situations even when N is large the asymptotic theory may fail to provide sensible indications.
The goal of the present paper is to give a clear-cut view of the situations in which this actually occurs and to pinpoint the model classes for which the asymptotic theory can be safely used. We start in the next section with an example clarifying where the trouble can come from in the use of the asymptotic theory.

3 An example where the asymptotic theory provides wrong results

Consider the following data-generating system:

$$y(t) = \frac{b^o z^{-1}}{1+a^o z^{-1}} u(t) + (1+h^o z^{-1})e(t),$$

where $a^o = -0.7$, $b^o = 0.3$, $h^o = 0.5$ and $e(t) = WGN(0,1)$ (WGN = White Gaussian Noise). In addition, the plant is operated in closed loop as shown in Figure 1. It is a trivial task to verify that the closed loop system is stable.
We have identified a full order model when the reference signal $r(t) = \mathrm{WGN}(0, 10^{-6})$, independent of $e(t)$,

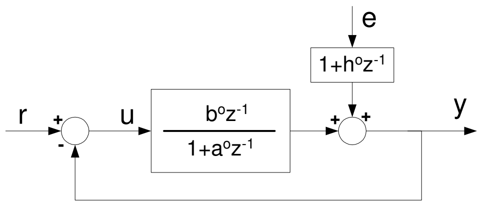

Figure 1: The real plant

and $N = 10000$ (note that the reference variance is very small as compared to the noise variance - poor excitation). A confidence region has also been estimated using the asymptotic theory.
The amplitude Bode diagrams of the identified model and of the real system u to y transfer functions have

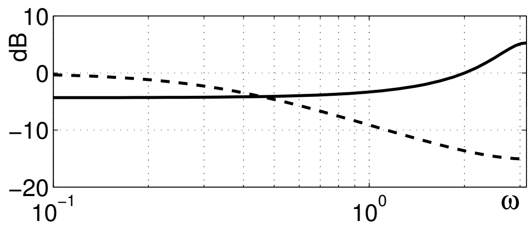

Figure 2: Amplitude Bode plot of the real plant (- -) and of the estimated model (—)

been plotted in Figure 2.
From the plot, it is clear that there is a wide mismatch between the real plant and the identified model. This is not surprising, since the reference signal is really poorly exciting. Correspondingly, we expect the asymptotic theory to return a wide uncertainty region.
Figure 3 displays in the frequency domain the confidence region $\mathcal{E}(\frac{\alpha(p)}{N})$ with $p = 0.99$. Surprisingly,

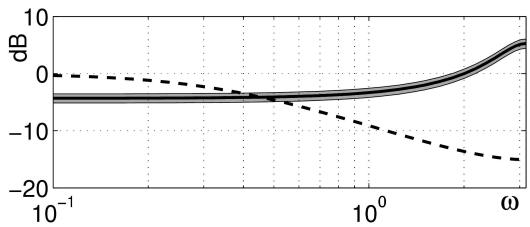

Figure 3: Uncertainty region of the estimated model

the confidence region concentrates around the identified model, showing that the estimated uncertainty is completely unreliable in this case.

Explanation
Let us briefly explain the mechanism that made the asymptotic theory unreliable in the present situation.
The explanation becomes easier if we assume that the reference signal is exactly equal to zero. Further below we return to the case when $r(t)$ has a small variance.

A simple computation shows that:

$$\overline{V}(\vartheta) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Big|\frac{1+h^o z^{-1}}{1+hz^{-1}}\cdot\frac{1+(a+b)z^{-1}}{1+(a^o+b^o)z^{-1}}\cdot \frac{1+a^o z^{-1}}{1+az^{-1}}\Big|^2_{z=e^{j\omega}}\,\mathrm{d}\omega,$$

where $\vartheta = [a\ b\ h]'$.
The minimal value of $\overline{V}(\vartheta)$ is 1 and it is easy to see that the minimum is achieved if and only if every monomial at the numerator is cancelled by another monomial at the denominator. This happens only in the following two cases:

$$\begin{array}{rcl|rcl} a_1^*+b_1^* &=& a^o+b^o & a_2^*+b_2^* &=& a^o+b^o \\ a_1^* &=& a^o & a_2^* &=& h^o \\ h_1^* &=& h^o & h_2^* &=& a^o. \end{array}$$

Thus, there are just two distinct minima of the asymptotic cost criterion, one of which corresponding to the true system.
Turn now to the case where $r(t)$ is a $WGN(0,10^{-6})$, that is, to the actual situation. Here, the minimizer of the asymptotic cost criterion $\overline{V}(\vartheta)$ is unique, as the asymptotic theory prescribes, and coincides with ϑ^o. The other minimum becomes a local minimum. Yet, the difference between the two minima will be very small. When identification is performed, a limited number of data points are available and, by minimizing $V_N(\vartheta)$, it may happen that the estimate gets trapped in the minimum which does not correspond to the real plant parameter.

In order to explain why the asymptotic theory fails to provide a reliable confidence region, it is, at this point, necessary to recall an aspect of the asymptotic theory which is relevant to the present discussion (see [10] and [13] for details).
Theorems 1 and 2 are both based on the following fundamental expansion:

$$\begin{aligned} 0 &= \sqrt{N}\frac{\mathrm{d}}{\mathrm{d}\vartheta}V_N(\widehat{\vartheta}_N) \\ &= \sqrt{N}\tfrac{\mathrm{d}}{\mathrm{d}\vartheta}V_N(\vartheta^o) + \tfrac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}V_N(\xi_N)\sqrt{N}(\widehat{\vartheta}_N-\vartheta^o). \quad (2)\end{aligned}$$

This equation is nothing but the Taylor expansion of $\frac{\mathrm{d}}{\mathrm{d}\vartheta}V_N$ (where all terms are inflated by the coefficient $\sqrt{N}$ and ξ_N is a point between ϑ^o and $\widehat{\vartheta}_N$). The evaluation of the confidence region for $\widehat{\vartheta}_N-\vartheta^o$ is carried out by observing that: first, $\sqrt{N}\frac{\mathrm{d}}{\mathrm{d}\vartheta}V_N(\widehat{\vartheta}_N)$ is asymptotically a zero mean Gaussian random variable; second, $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}V_N(\xi_N)$ converges to $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}\overline{V}(\vartheta^o)$, since $\widehat{\vartheta}_N \to \vartheta^o$ so that ξ_N is squeezed towards ϑ^o. The quantity $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}\overline{V}(\vartheta^o)$ is further approximated by $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}V_N(\widehat{\vartheta}_N)$ leading to the asymptotic Theorems 1 and 2.
The last introduced approximation is acceptable provided that $\widehat{\vartheta}_N$ is sufficiently near to ϑ^o. However, in the previous example, $\widehat{\vartheta}_N$ is far from ϑ^o, since it has got trapped in a false local minimum.
In such a case, $\sqrt{N}(\widehat{\vartheta}_N-\vartheta^o)$ is very large, due to the inflating term $\sqrt{N}$. Yet, equation (2) holds true since $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}V_N$ is computed at the point ξ_N, between ϑ^o and $\widehat{\vartheta}_N$, where, $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}V_N(\xi_N)$ is almost singular.
Unfortunately, as explained before, ξ_N is not accessible to observations and $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}V_N(\xi_N)$ has to be substituted by $\frac{\mathrm{d}^2}{\mathrm{d}\vartheta^2}V_N(\widehat{\vartheta}_N)$ which turns out to be well positive definite. This leads to the mistaken conclusion that $\widehat{\vartheta}_N-\vartheta^o$ is small and to the unreliable uncertainty region shown in Figure 3.

4 Mid paper conclusions

The results of the previous sections can be summarized as follows.

i) The asymptotic theory requires that the asymptotic cost criterion has a unique minimizer $\vartheta^*=\vartheta^o$ and that $\widehat{\vartheta}_N \approx \vartheta^*$

ii) When there is wide uncertainty, even if ϑ^* is unique (and equal to ϑ^o), for any reasonable sample size N, one cannot assume that $\widehat{\vartheta}_N \approx \vartheta^*$

iii) As a result of ii), a blind application of the asymptotic theory can lead to misleading results (see the example in Section 3)

In the next sections the following results are presented:

v) We prove a new asymptotic result for uncertainty estimation which requires weaker assumptions

vi) We show that the new asymptotic result can be safely used in presence of a high level of uncertainty ($\widehat{\vartheta}_N$ far from ϑ^o) if a suitable additional condition on the model class is satisfied

vii) We establish which standard model classes (ARMAX, Box-Jenkins, etc.) satisfy the additional condition in point vi).

5 A new asymptotic result

In this section we provide a new asymptotic result which is a variant of the standard asymptotic theory. Precisely, we no longer insist that $\widehat{\vartheta}_N \to \vartheta^o$ (as is a consequence of Assumption 4) and, instead, assume that $\widehat{\vartheta}_N \to \vartheta^*$ where ϑ^* is some limiting estimate, possibly different from ϑ^o.
If the model class satisfies a certain condition (Condition 1 below) we show that Theorem 2 still holds (see Theorem 3). Thus, in our framework the fact that $\widehat{\vartheta}_N$ is close to ϑ^o is not crucial as it is in the standard asymptotic theory, and Theorem 3 can be safely applied with

a finite number of data points even in case of a high level of uncertainty. This aspect will be discussed in Section 6.

Assumption 5 *$\widehat{\vartheta}_N \to \vartheta^*$ (not necessary equal to ϑ^o) almost surely, where ϑ^* is a minimizer of $\overline{V}(\vartheta)$ and is an interior point of Θ.*

In the following Theorem 3 we show that the asymptotic Theorem 2 can be preserved in the present setting provided that the following condition is satisfied.

Condition 1 *$\Theta^* = \mathcal{S} \cap \Theta$, where $\mathcal{S}$ is an affine subspace of the parameter space $\mathbb{R}^n$.*

Theorem 3 *Suppose that Assumptions 1, 2, 3 and 5, along with Condition 1, hold true.*
For any sequence α_N which tends to ∞ as $N \to \infty$, we have that (see (1) for the definition of $\mathcal{E}(\cdot)$)

$$\lim_{N\to\infty} \mathbb{P}\left\{\vartheta^o \in \mathcal{E}\left(\frac{\alpha_N}{N}\right)\right\} = 1.$$

6 Use of Theorem 3 with a finite number of data points

Let us first recapitulate the reasons why applying the asymptotic theory in Section 3 leads to a misleading result.
In that example, if $r(t) = 0$, then $\overline{V}(\vartheta)$ has two global minimizers. If $r(t) \neq 0$ but small, then $\widehat{\vartheta}_N$ is still close to one of these two minimizers, and can possibly be in the vicinity of the minimizer which does not correspond to the true system. If so, the asymptotic theory leads to computing a deceivingly small uncertainty region as explained in Section 3. Moreover, this is recognized to be the general cause of a possible malfunctioning of the asymptotic theory with finite data samples.
Suppose now that a model class is used such that, independently of the level of the excitation in the signals, $\overline{V}(\vartheta)$ is minimized in an affine subspace. Then, when we move to the real identification setting where $N < \infty$, the estimated parameter will be still close to one point in this subspace and Theorem 3 will still hold approximately.
Thus, it appears that the model classes to which the asymptotic theory can be safely applied with finite sample data points are those for which the set of minimizers of $\overline{V}(\vartheta)$ is an affine subspace. Studying this classes is the subject of the next section.

7 Assessment of the model classes for which $\overline{V}$ is minimized in an affine subspace

We treat separately two different situations, namely open-loop identification and closed-loop identification as these two settings give different results.

7.1 Open-loop identification

By "open-loop identification" we mean that the input signal $u(t)$ and the noise signal $e(t)$ are independent. Technically speaking, this reflects to say that $H_u(z^{-1}) = 0$ in Assumption 2.

Theorem 4 *Let $\mathcal{M}_\vartheta$ be the Box-Jenkins (BJ) class of predictor models, i.e.*

$$\begin{aligned}\mathcal{M}_\vartheta \;\; = \Big\{ \;\; & \widehat{y}(t,\vartheta) = (1 - H(z^{-1},\vartheta)^{-1})y(t) + \\ & H(z^{-1},\vartheta)^{-1}G(z^{-1},\vartheta)u(t),\ \vartheta \in \Theta \;\; \Big\},\end{aligned}$$

where G and H are rational transfer functions and ϑ is the vector of the coefficients of the polynomials at the numerator and at the denominator of G and H.
Suppose that the identification is performed in open-loop and that Assumptions 1, 2 and 3 are satisfied. Then, Condition 1 holds true.

Theorem 4 can be applied to Output Error (OE) models as well, since OE is a particular case of BJ. We remind that the OE predictor model class is

$$\mathcal{M}_\vartheta = \Big\{\widehat{y}(t,\vartheta) = G(z^{-1},\vartheta)u(t),\ \vartheta \in \Theta\Big\},$$

where G is a rational transfer function and ϑ is the vector of the coefficients of the polynomials at the numerator and at the denominator of G.
A similar result applies to ARX and ARMAX models too. That is,

$$\begin{aligned}\mathcal{M}_\vartheta \;\; = \Big\{ \;\; & \widehat{y}(t,\vartheta) = \Big(1 - \frac{A(z^{-1},\vartheta)}{C(z^{-1},\vartheta)}\Big)y(t) + \\ & \frac{B(z^{-1},\vartheta)}{C(z^{-1},\vartheta)}u(t),\ \vartheta \in \Theta \;\; \Big\},\end{aligned}$$

where, A, B and C are polynomials in z^{-1} and ϑ is the vector of the coefficients of these polynomials ($C(z^{-1},\vartheta) = 1$ in the ARX case). One should note that in ARX and ARMAX structures, $G(z^{-1},\vartheta)$ and $H(z^{-1},\vartheta)$ are not freely parameterized as assumed in Theorem 4 so that this theorem cannot directly applied in this case. However, the proof of Theorem 4 can be extended with minor amendments to cover the ARX and ARMAX cases as well.
It is perhaps worth mentioning that not all model structures satisfy Condition 1 in open-loop (note that the example in Section 3 is not a counter example here, since the system is operated in closed-loop). An example is given by the model class

$$A(z^{-1},\vartheta)y(t) = G(z^{-1},\vartheta)u(t) + H(z^{-1},\vartheta)e(t),$$

where $e(t) = WGN$, A is a polynomial in z^{-1}, G and H are rational transfer functions and ϑ is the vector of the coefficients of A and of the polynomials at the numerator and at the denominator of G and H. See [2] for details.

Example
We have considered again the example (based on BJ model class) presented in Section 3, but now the system has been operated in open-loop. In Figure 4 the fre-

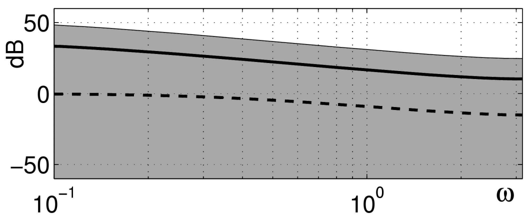

Figure 4: Uncertainty region of the estimated model (Dashed line = real plant frequency response, solid line = estimated frequency response, grey area = confidence region ($p = 0.99$))

quency response of the identified model together with the estimated 99%-confidence region is plotted.
As can be seen, this confidence region covers the gap between the identified model and the true system. Thus, the estimated uncertainty is reliable, in agreement with Theorem 4.

7.2 Closed-loop identification

Suppose now that the system is operated in closed-loop with a controller R as in Figure 5.

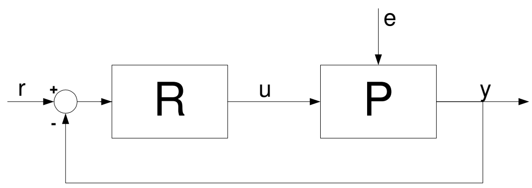

Figure 5: Closed loop system

Theorem 5 *Suppose that the identification is performed in closed-loop and that Assumptions 1, 2 and 3 are satisfied.*
Then, Condition 1 holds true for the ARMAX *and* OE *classes of models.*

It has to be noted that, when identification is performed in closed-loop, the Box-Jenkins structure does not meet Condition 1 in general. In fact, the example presented in Section 3 was based on a Box-Jenkins model.

References

[1] S. Bittanti, M.C. Campi, S. Garatti, *An iterative controller design scheme based on average robust control*, Proc. 15th World IFAC Congress, Barcelona, 2002

[2] S. Bittanti, M.C. Campi, S. Garatti, *Assessing the quality of identified models trough the asymptotic theory - When is the result reliable?*, Internal Report n. 2002.29, Politecnico di Milano - DEI, 2002

[3] S. Bittanti, G. Picci (eds), *Identification, Adaptation, Learning - The science of learning models from data*, Springer Verlag, 1996.

[4] M.C. Campi, E. Weyer, *Finite sample properties of system identification methods* to appear in IEEE Trans. on Automatic Control, short version in Proceedings of the 38th Conference on Decision & Control, Phoenix, USA, 1999, pp. 510-515

[5] M.C. Campi, E. Weyer and Su Ki Ooi, *Nonasymptotic quality assessment of identified models*, Proc. 15th World IFAC Congress, Barcelona, 2002.

[6] M. Gevers, *A decade of progress in iterative control design: from theory to practice*, Proc. Int. Symp. on Advanced Control of Chemical Processes, Pisa, Italy, Vol. II, pp. 677-694 (keynote lecture), June 2000.

[7] B. Ninness, G. Goodwin, *Estimation of model quality*, Automatica, 1995, vol. 31, 12, pp. 1771-1795

[8] H. Hjalmarsson, L. Ljung, *Estimating model variance in the case of undermodeling*, IEEE Trans. Automatic Control, 1992, vol. 37, 7, pp. 1004-1008

[9] 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*, Proc. 32nd IEEE Conf. Decision and Control, San Antonio, TX, 1993.

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

[11] L. Ljung, *Estimating linear time-invariant models of nonlinear time-varying systems*, European Journal of Control, 2001, 7, pp. 203-219.

[12] L. Ljung, *Model validation and model error modeling*, report from the Åström Symposium on Control, Lund, Sweden, 1999.

[13] T. Soderstrom, 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, Vol. 31, 12, pp. 1751-1770, December, 1995

[15] E. Weyer, M.C. Campi, *Non-asymptotic confidence ellipsoids for the least squares estimate* to appear in Automatica, short version in Proceedings of the 39th Conference on Decision & Control, Sydney, Australia, 2000, pp. 2688-2693