Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

TuA04.4

Variance Analysis for Identification of Cascade Systems

Bo Wahlberg, Håkan Hjalmarsson and Jonas Mårtensson

Abstract— The contribution of this paper is a variance analysis of estimated models of cascade dynamical systems. Models of such systems are important in for example cascade control applications. The model quality is analyzed by means of the asymptotic covariance matrix of the prediction error method parameters estimates. Recent work has shown that identification of two cascaded linear systems, where the transfer functions of the sub-systems are identical, has some fundamental limitations in terms of asymptotic statistical performance. Under this condition, the output from the second sub-system does not influence the quality of the estimated model of the first subsystem. The objective of this paper is to extend this result to the case where the transfer functions of the sub-systems are not completely identical, but do have some common dynamics. We will also study cascaded systems with three sub-systems, and show that a similar variance result also holds for this case. The results are illustrated by some simple FIR examples.

I. INTRODUCTION

System identification is about estimating and validating models of dynamical systems based on measured data. It is very common to have some physical structural information, which should be included in the model structure. The results to be presented are motivated by the discussion on the use of structural system identification in process industry given in [10].

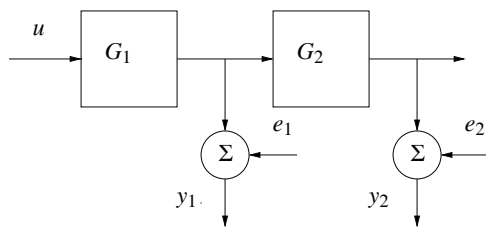Many physical systems have a cascade form illustrated in Fig. 1.

Fig. 1. Cascade system with two subsystems.

The corresponding transfer function model is

$$
\begin{aligned}
y_1(t) &= G_1(q)u(t) + e_1(t) \\
y_2(t) &= G_2(q)G_1(q)u(t) + e_2(t).
\end{aligned}
$$

The input signal is denoted by $u(t)$ and the two output signals are $y_1(t)$ and $y_2(t)$, respectively. The transfer functions are $G_1(q)$ and $G_2(q)$, and are both assumed to be stable. By q we denote the shift operator, $q^{-1}u(t) = u(t-1)$. The signals $e_1(t)$ and $e_2(t)$ are the measurement noise processes. We assume that the dimensions of the input and the two output signals are all one (the scalar case).

Cascade systems are very common in both process control and in control of servo mechanical systems. In mechanical

application $y_1(t)$ is often a rate while $y_2(t)$ is a position. In process control application, the primary output $y_2(t)$ is often a quality variable such as temperature or levels, while the secondary output $y_1(t)$ typically concerns an intermediate variable such as flows or pressures. The quality of the sensors for measuring the two outputs can be quite different. We will model this by the size of the variances of $e_1(t)$ and $e_2(t)$. High variance means a poor measurement quality. The resulting model is often used to design a cascade control system.

We will study the statistical properties of estimated models of cascade system when a Prediction Error Method (PEM), [4], is applied to a measured data set of the form $\{u(t), y_1(t), y_2(t)\}$, $t = 1 \dots N$. This is an asymptotically statistically optimal (efficient) method to solve the structured cascade system identification problem and the corresponding estimates will asymptotically (for large N) achieve the Cramér-Rao lower bound.

The outline of this paper is as follows. First, we will review the case when the true two transfer functions in the cascade structure are identical. These results are based on [8] and have been inspired by a recent geometric approach to variance analysis in system identification developed in [5], [7], [6]. The objective of this paper is to extend these results to more complex cascade structures. Section III deals with systems where the two transfer functions have common dynamics, while Section IV considers identification of cascade systems with three subsystems. Finally, the paper is summarized in Section V.

II. REVIEW

Let $G_1(q, \theta_1)$ and $G_2(q, \theta_2)$ be general transfer function models, with independent parameterizations, of a cascade system as shown in Fig. 1,

$$
\begin{aligned}
y_1(t) &= G_1(q, \theta_1)u(t) + e_1(t) \\
y_2(t) &= G_2(q, \theta_2)G_1(q, \theta_1)u(t) + e_2(t).
\end{aligned}
$$

Furthermore, assume that the measurement noise processes $\{e_1(t)\}$ and $\{e_2(t)\}$ are independent white noises with variances λ_1 and λ_2, respectively. We will assume that these variances are pre-given, but the results can easily be extended to the case when the noise variances also are estimated.

The notation θ_1^o and θ_2^o will be used for the parameters of the true underlying system to be identified, which is assumed to belong to the model set.

Given a data set $\{u(t), y_1(t), y_2(t)\}$, $t = 1 \dots N$, the PEM estimates of the model parameters θ_1 and θ_2 are given by

$$
\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} = \arg\min_{\theta_1, \theta_2} \left(\frac{1}{N} \sum_{t=1}^{N} \frac{[y_1(t) - G_1(q, \theta_1)u(t)]^2}{\lambda_1} \right.
$$
$$
\left. + \frac{1}{N} \sum_{t=1}^{N} \frac{[y_2(t) - G_2(q, \theta_2)G_1(q, \theta_1)u(t)]^2}{\lambda_2} \right).
$$

Let

$$\psi(t) = \begin{pmatrix} \dfrac{G_1'(q,\theta_1^o)u(t)}{\sqrt{\lambda_1}} & \dfrac{G_2(q,\theta_2^o)G_1'(q,\theta_1^o)u(t)}{\sqrt{\lambda_2}} \\[4mm] 0 & \dfrac{G_2'(q,\theta_2^o)G_1(q,\theta_1^o)u(t)}{\sqrt{\lambda_2}} \end{pmatrix},$$

where prime denotes differentiation with respect to the parameter vectors. The asymptotic covariance matrix of the parameter estimate is then given by

$$\mathrm{Cov}\begin{pmatrix} \hat\theta_1 \\ \hat\theta_2 \end{pmatrix} \sim M^{-1},$$

where

$$M = N\mathrm{E}\{\psi(t)\psi^T(t)\}$$

See e.g. [4] for details. Introduce

$$A = \frac{N}{\lambda_1}\mathrm{E}\{[G_1'(q,\theta_1^o)u(t)][G_1'(q,\theta_1^o)u(t)]^T\},$$

$$B = \frac{N}{\lambda_2}\mathrm{E}\{[G_2(q,\theta_2^o)G_1'(q,\theta_1^o)u(t)] \\ \times [G_2(q,\theta_2^o)G_1'(q,\theta_1^o)u(t)]^T\},$$

$$C = \frac{N}{\lambda_2}\mathrm{E}\{[G_2(q,\theta_2^o)G_1'(q,\theta_1^o)u(t)] \\ \times [G_2'(q,\theta_2^o)G_1(q,\theta_1^o)u(t)]^T\},$$

$$D = \frac{N}{\lambda_2}\mathrm{E}\{[G_2'(q,\theta_2^o)G_1(q,\theta_1^o))u(t)] \\ \times [G_2'(q,\theta_2^o)G_1(q,\theta_1^o)u(t)]^T\},$$

which give

$$M = \begin{pmatrix} A+B & C \\ C^T & D \end{pmatrix}.$$

We will use an A,B,C,D notation to specify matrix elements, which will not necessary be the same throughout the paper.

The key result in [8], which is inspired by a recent geometric approach to variance analysis in system identification developed in [5], [6], [7] is as follows: Assume that

> **Condition 1:** $G_2(q,\theta_2^o)G_1'(q,\theta_1^o) = G_2'(q,\theta_2^o)G_1(q,\theta_1^o)$

This is the case if the two true transfer functions are **identical**, $G_2(q,\theta_2^o) = G_1(q,\theta_1^o)$, and that we are using **the same model structure** for both models, $G_2'(q,\theta_2^o) = G_1'(q,\theta_1^o)$.

This assumption implies that M will have the following block structure

$$M = \begin{pmatrix} A+D & D \\ D & D \end{pmatrix}$$

where

$$A = \frac{N}{\lambda_1}\mathrm{E}\{[G_1'(q,\theta_1^o)u(t)][G_1'(q,\theta_1^o)u(t)]^T\},$$

$$D = \frac{N}{\lambda_2}\mathrm{E}\{[G_2'(q,\theta_1^o)G_1(q,\theta_1^o)u(t)] \\ \times [G_2'(q,\theta_1^o)G_1(q,\theta_1^o)u(t)]^T\}. \tag{1}$$

Using the transformation matrix

$$T = \begin{pmatrix} I & -I \\ 0 & I \end{pmatrix}$$

we obtain

$$TMT^T = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix}$$

and

$$M^{-1} = T^T\begin{pmatrix} A^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix}T = \begin{pmatrix} A^{-1} & -A^{-1} \\ -A^{-1} & A^{-1}+D^{-1} \end{pmatrix}.$$

Hence

$$\mathrm{Cov}(\hat\theta_1) \sim A^{-1},$$
$$\mathrm{Cov}(\hat\theta_2) \sim A^{-1}+D^{-1}.$$

If Condition 1 holds we thus can conclude:

- Since the matrix A^{-1} is the asymptotic covariance matrix of $\hat\theta_1$ when only $y_1(t)$ is available, the quality of the estimate of θ_1 is never improved by also measuring $y_2(t)$.

- The covariance of the estimate of θ_2 equals $A^{-1}+D^{-1}$, and is thus always larger than or equal to the covariance of $\hat\theta_1$, since

$$A^{-1}+D^{-1} \geq A^{-1}$$

for positive definite matrices. This is true even if one has noise free measurements of $y_2(t)$. The asymptotic covariance matrix of $\hat\theta_2$ is also always larger than or equal to D^{-1}, $(A^{-1}+D^{-1} \geq D^{-1})$, which equals the asymptotic covariance matrix of $\hat\theta_2$ if $G_1(q)$ is known (or equally $\lambda_1 = 0$).

If it is known in advance that $G_2 = G_1$, this should of course be imposed in the model structure,

$$y_1(t) = G_1(q,\theta_1)u(t)+e_1(t)$$
$$y_2(t) = G_1(q,\theta_1)G_1(q,\theta_1)u(t)+e_2(t)$$

The asymptotic covariance matrix M^{-1} of $\hat\theta_1$ can then be calculated using

$$\psi(t) = \begin{pmatrix} \dfrac{G_1'(q,\theta_1^o)u(t)}{\sqrt{\lambda_1}} & \dfrac{2G_1'(q,\theta_1^o)G_1(q,\theta_1^o)u(t)}{\sqrt{\lambda_2}} \end{pmatrix},$$

in $M = N\mathrm{E}\{\psi(t)\psi^T(t)\}$. This implies that $M = A+4D$ with the notation above. Since

$$(A+4D)^{-1} < A^{-1}$$

for positive definite covariance matrices A and D, the variance of $\hat\theta_1$ can here be considerably smaller than for the case with separate parameterizations.

III. GENERALIZATIONS

The objective of this paper is to generalize and extend the variance results reviewed in the previous section. To start, consider the condition

> **Condition 2:** $G_2(q,\theta_2^o)G_1'(q,\theta_1^o) = kG_2'(q,\theta_2^o)G_1(q,\theta_1^o)$

where k is a constant. This is for example the case if

$$G_2(q,\theta_2^o) = \beta G_1(q,\theta_1^o), \quad G_1'(q,\theta_1^o) = \alpha G_2'(q,\theta_2^o), \quad k = \alpha\beta$$

that is $G_2(q, \theta_2^o)$ is **proportional** to $G_1(q, \theta_1^o)$ and the models have the same structure. Then

$$M = \begin{pmatrix} A + k^2 D & kD \\ kD & D \end{pmatrix}$$

and the asymptotic covariance matrix equals

$$M^{-1} = \begin{pmatrix} A^{-1} & -kA^{-1} \\ -kA^{-1} & k^2 A^{-1} + D^{-1} \end{pmatrix}.$$

Here A and D are defined by (1). We thus have the same result as in the previous section, e.g. the quality of the estimate of θ_1 is not improved by measuring $y_2(t)$.

Consider now the more general case when

$$\begin{aligned} y_1(t) &= G_1(q, \theta_1)u(t) + e_1(t) \\ y_2(t) &= G_{22}(q, \theta_{22})G_{21}(q, \theta_{21})G_1(q, \theta_1)u(t) + e_2(t), \end{aligned}$$

i.e.

$$G_2(q, \theta_2) = G_{22}(q, \theta_{22})G_{21}(q, \theta_{21}), \quad \theta_2 = (\theta_{21}^T, \theta_{22}^T)^T$$

The transfer function in the second block in Fig. 1 is factorized into two transfer functions. Now

$$\psi(t) =$$
$$\begin{pmatrix} \dfrac{G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_1}} & \dfrac{G_{21}(q, \theta_{22}^o)G_{21}(q, \theta_{21}^o)G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} \\ 0 & \dfrac{G_{22}(q, \theta_{22}^o)G_{21}'(q, \theta_{21}^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} \\ 0 & \dfrac{G_{22}'(q, \theta_{22}^o)G_{21}(q, \theta_{21}^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} \end{pmatrix}.$$

Assume that

Condition 3: $G_{21}(q, \theta_{21}^o)G_1'(q, \theta_1^o) = G_{21}'(q, \theta_{21}^o)G_1(q, \theta_1^o)$

This is the case if $G_2(q, \theta_2^o)$ has $G_1(q, \theta_1^o)$ as a factor, and hence extends the proportional case discussed above by allowing for **common dynamics.** We then obtain that the asymptotic covariance matrix

$$M = \begin{pmatrix} A + B & B & C \\ B & B & C \\ C^T & C^T & F \end{pmatrix},$$

where

$$A = \frac{N}{\lambda_1}\mathrm{E}\{[G_1'(q, \theta_1^o)u(t)][G_1'(q, \theta_1^o)u(t)]^T\},$$

$$\begin{aligned} B = \frac{N}{\lambda_2}\mathrm{E}\{ &[G_{22}(q, \theta_{22}^o)G_{21}'(q, \theta_{21}^o)G_1(q, \theta_1^o)u(t)] \\ \times &[G_{22}(q, \theta_{22}^o)G_{21}'(q, \theta_{21}^o)G_1(q, \theta_1^o)u(t)]^T\}, \end{aligned}$$

$$\begin{aligned} C = \frac{N}{\lambda_2}\mathrm{E}\{ &[G_{22}(q, \theta_{22}^o)G_{21}(q, \theta_{21}^o)G_1'(q, \theta_1^o)u(t)] \\ \times &[G_{22}'(q, \theta_{22}^o)G_{21}(q, \theta_{21}^o)G_1(q, \theta_1^o)u(t)]^T\}, \end{aligned}$$

$$\begin{aligned} F = \frac{N}{\lambda_2}\mathrm{E}\{ &[G_{22}'(q, \theta_{22}^o)G_{21}(q, \theta_{21}^o)G_1(q, \theta_1^o)u(t)] \\ \times &[(G_{22}'(q, \theta_{22}^o)G_{21}(q, \theta_{21}^o)G_1(q, \theta_1^o)u(t)]^T\}. \end{aligned}$$

We will extend the method to find M^{-1} by using

$$T = \begin{pmatrix} I & -I & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \tag{2}$$

to obtain

$$TMT^T = \begin{pmatrix} A & 0 & 0 \\ 0 & B & C \\ 0 & C^T & F \end{pmatrix}.$$

Introduce

$$R = \begin{pmatrix} R_{11} & R_{12} \\ R_{12}^T & R_{22} \end{pmatrix} = \begin{pmatrix} B & C \\ C^T & F \end{pmatrix}^{-1},$$

to give

$$\begin{aligned} M^{-1} &= T^T \begin{pmatrix} A^{-1} & 0 \\ 0 & R \end{pmatrix} T \\ &= \begin{pmatrix} A^{-1} & -A^{-1} & 0 \\ -A^{-1} & A^{-1} + R_{11} & R_{12} \\ 0 & R_{12}^T & R_{22} \end{pmatrix}. \end{aligned} \tag{3}$$

This means that

$$\begin{aligned} \mathrm{Cov}\left(\hat{\theta}_1\right) &\sim A^{-1}, \\ \mathrm{Cov}\left(\hat{\theta}_{21}\right) &\sim A^{-1} + R_{11}, \\ \mathrm{Cov}\left(\hat{\theta}_{22}\right) &\sim R_{22}. \end{aligned}$$

and the **key results** under Condition 3 are:

- Since the matrix A^{-1} is the asymptotic covariance matrix of $\hat{\theta}_1$ when only $y_1(t)$ is available, the quality of the estimate of θ_1 is also in this case not improved by measuring $y_2(t)$.

- The covariance of the estimate of $\theta_2 = (\theta_{21}^T \, \theta_{22}^T)^T$ equals

$$\begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} B & C \\ C^T & F \end{pmatrix}^{-1}.$$

The second term in this expression is nothing else but the asymptotic covariance matrix of $\hat{\theta}_2 = (\hat{\theta}_{21} \, \hat{\theta}_{22})^T$ when $G_1(q)$ is known (or equally $\lambda_1 = 0$). Hence, the quality of the estimate of θ_{21} is for $\lambda_1 > 0$ always worse than the estimate of θ_1. However, the quality of the estimate of θ_{22} is independent of $y_1(t)$ and hence always of the same quality as if $G_1(q)$ is known.

For the constrained model structure, $(G_{21} = G_1)$

$$G_2(q, \theta_2) = G_{22}(q, \theta_{22})G_1(q, \theta_1)$$

we have

$$\psi(t) =$$
$$\begin{pmatrix} \dfrac{G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_1}} & \dfrac{2G_{22}(q, \theta_{22}^o)G_1(q, \theta_1^o)G_1'(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} \\ 0 & \dfrac{G_{22}'(q, \theta_{22}^o)G_1(q, \theta_1^o)G_1(q, \theta_1^o)u(t)}{\sqrt{\lambda_2}} \end{pmatrix}$$

and

$$M = \begin{pmatrix} A + B & C \\ C^T & F \end{pmatrix},$$

where

$$A = \frac{N}{\lambda_1} \mathrm{E}\{[G_1'(q,\theta_1^o)u(t)][G_1'(q,\theta_1^o)u(t)]^T\},$$

$$B = \frac{N}{\lambda_2} \mathrm{E}\{[2G_{22}(q,\theta_{22}^o)G_1'(q,\theta_1^o)G_1(q,\theta_1^o)u(t)]$$
$$\times [2G_{22}(q,\theta_{22}^o)G_1'(q,\theta_1^o)G_1(q,\theta_1^o)u(t)]^T\},$$

$$C = \frac{N}{\lambda_2} \mathrm{E}\{[2G_{22}(q,\theta_{22}^o)G_1(q,\theta_{21}^o)G_1'(q,\theta_1^o)u(t)]$$
$$\times [G_{22}'(q,\theta_{22}^o)G_1(q,\theta_1^o)G_1(q,\theta_1^o)u(t)]^T\},$$

$$F = \frac{N}{\lambda_2} \mathrm{E}\{[G_{22}'(q,\theta_{22}^o)G_1(q,\theta_1^o)G_1(q,\theta_1^o)u(t)]$$
$$\times [(G_{22}'(q,\theta_{22}^o)G_1(q,\theta_1^o)G_1(q,\theta_1^o)u(t)]^T\}.$$

It is now possible to verify that we now obtain a lower variance for $\hat{\theta}_1$ than for the unconstrained case.

We will now illustrate this result with a simple analytic example.

Example: Consider the model structure

$$y_1(t) = G_1(q,\theta_1)u(t) + e_1(t)$$
$$y_2(t) = G_{22}(q,\theta_{22})G_{21}(q,\theta_{21})G_1(q,\theta_1)u(t) + e_2(t),$$

with first order FIR transfer functions

$$G_1(q,\theta_1) = 1 + b_1 q^{-1}, \quad \theta_1 = b_1,$$
$$G_{21}(q,\theta_{21}) = 1 + b_{21} q^{-1}, \quad \theta_{21} = b_{21},$$
$$G_{22}(q,\theta_{22}) = 1 + b_{22} q^{-1}, \quad \theta_{22} = b_{22},$$

Assume $\{e_1(t)\}$ and $\{e_2(t)\}$ to be white noise with variances λ_1 and λ_2, respectively. Let the input signal $u(t)$ be white noise with variance 1.

We will study the case when $G_{21}(q,\theta_{21}^o) = G_1(q,\theta_1^o)$, i.e. $b_{21}^o = b_1^o$, for which

$$\psi(t) = \begin{pmatrix} \dfrac{u(t-1)}{\sqrt{\lambda_1}} & \dfrac{(1+b_{22}^o q^{-1})(1+b_1^o q^{-1})u(t-1)}{\sqrt{\lambda_2}} \\ 0 & \dfrac{(1+b_{22}^o q^{-1})(1+b_1^o q^{-1})u(t-1)}{\sqrt{\lambda_2}} \\ 0 & \dfrac{(1+b_1^o q^{-1})(1+b_1^o q^{-1})u(t-1)}{\sqrt{\lambda_2}} \end{pmatrix}$$

and

$$M = \begin{pmatrix} a+b & b & c \\ b & b & c \\ c & c & f \end{pmatrix},$$

where

$$a = \frac{N}{\lambda_1}, \quad b = \frac{N}{\lambda_2}(1 + (b_{22}^o + b_1^o)^2 + (b_{22}^o b_1^o)^2)$$

$$c = \frac{N}{\lambda_2}(1 + 2b_1^o(b_{22}^o + b_1^o) + b_{22}^o(b_1^o)^3)$$

$$f = \frac{N}{\lambda_2}(1 + 4(b_1^o)^2 + (b_1^o)^4).$$

Direct calculations using determinants and co-factors now give

$$M^{-1} = \begin{pmatrix} \dfrac{1}{a} & 0 & 0 \\ 0 & \dfrac{1}{a} + \dfrac{f}{bf-c^2} & \dfrac{-c}{bf-c^2} \\ 0 & \dfrac{-c}{bf-c^2} & \dfrac{b}{bf-c^2} \end{pmatrix}.$$

This means that for example

$$\mathrm{Var}\,(\hat{b}_1) \sim \frac{1}{a} = \frac{\lambda_1}{N},$$

which hence is the same variance as estimating b_1 from $y_1(t)$ only.

IV. GENERALIZATION TO THREE CASCADED SUBSYSTEMS

What can be said when estimating more than two cascaded subsystems, for example three

$$y_1(t) = G_1(q,\theta_1)u(t) + e_1(t)$$
$$y_2(t) = G_2(q,\theta_2)G_1(q,\theta_1)u(t) + e_2(t)$$
$$y_3(t) = G_3(q,\theta_3)G_2(q,\theta_2)G_1(q,\theta_1)u(t) + e_3(t), \quad (4)$$

where $e_3(t)$ is white noise with variance λ_3? This structure is illustrated in Fig. 2.
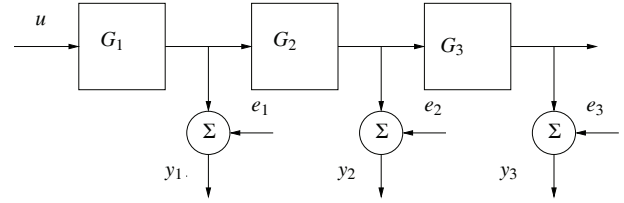

Fig. 2.   Cascade system with three subsystems.

Define the 3×3 block matrix

$$\psi(t) = \begin{pmatrix} \dfrac{G_1'(q,\theta_1^o)u(t)}{\sqrt{\lambda_1}} & \dfrac{G_2(q,\theta_2^o)G_1'(q,\theta^o)u(t)}{\sqrt{\lambda_2}} \\ 0 & \dfrac{G_2'(q,\theta_2^o)G_1(q,\theta_1^o)u(t)}{\sqrt{\lambda_2}} \\ 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \dfrac{G_3(q,\theta_3^o)G_2(q,\theta_2^o)G_1'(q,\theta^o)u(t)}{\sqrt{\lambda_3}} \\ \dfrac{G_3(q,\theta_3^o)G_2'(q,\theta_2^o)G_1(q,\theta_1^o)u(t)}{\sqrt{\lambda_3}} \\ \dfrac{G_3'(q,\theta_3^o)G_2(q,\theta_2^o)G_1(q,\theta_1^o)u(t)}{\sqrt{\lambda_3}} \end{pmatrix}.$$

The asymptotic covariance matrix of the parameter estimates is then given by

$$\mathrm{Cov}\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \hat{\theta}_3 \end{pmatrix} \sim M^{-1},$$

where, as before, $M = N\mathrm{E}\{\psi(t)\psi^T(t)\}$.

Now consider

Condition 4: $G_2(q,\theta_2^o)G_1'(q,\theta_1^o) = G_2'(q,\theta_2^o)G_1(q,\theta_1^o)$

Hence the true first and second block transfer functions are **identical**, while $G_3(q,\theta_3^o)$ can be arbitrary.

This is very similar to the two cascaded systems case and the intuition is that the estimate of θ_1 should not be improved

by measuring $y_3(t)$, which just a a filtered version of the noise free $y_2(t)$. To verify that this is the case, let

$$A = \frac{N}{\lambda_1} \mathrm{E}\left\{ [G_1'(q,\theta_1^o)u(t)][G_1'(q,\theta_1^o)u(t)]^T \right\},$$

$$B = \frac{N}{\lambda_2} \mathrm{E}\left\{ [G_2'(q,\theta_2^o)G_1(q,\theta_1^o)u(t)] \right.$$
$$\left. \times [G_2'(q,\theta_2^o)G_1(q,\theta_1^o)u(t)]^T \right\},$$

$$C = \frac{N}{\lambda_3} \mathrm{E}\left\{ [G_3(q,\theta_3^o)G_2(q,\theta_2^o)G_1'(q,\theta_1^o)u(t)] \right.$$
$$\left. \times [G_3(q,\theta_3^o)G_2(q,\theta_2^o)G_1'(q,\theta_1^o)u(t)]^T \right\},$$

$$D = \frac{N}{\lambda_3} \mathrm{E}\left\{ [G_3(q,\theta_3^o)G_2(q,\theta_2^o)G_1'(q,\theta_1^o)u(t)] \right.$$
$$\left. \times [G_3'(q,\theta_3^o)G_2(q,\theta_2^o)G_1(q,\theta_1^o)u(t)]^T \right\},$$

$$F = \frac{N}{\lambda_3} \mathrm{E}\left\{ [G_3'(q,\theta_3^o)G_2(q,\theta_2^o)G_1(q,\theta_1^o)u(t)] \right.$$
$$\left. \times [G_3'(q,\theta_3^o)G_2(q,\theta_2^o)G_1(q,\theta_1^o)u(t)]^T \right\},$$

which implies

$$M = \begin{pmatrix} A+B+C & B+C & D \\ B+C & B+C & D \\ D^T & D^T & F \end{pmatrix}.$$

This matrix has almost the same structure as for the common dynamics case, and it is easy to verify using the transformation T, defined by (2), that

$$M^{-1} = T^T \begin{pmatrix} A^{-1} & 0 \\ 0 & S^{-1} \end{pmatrix} T$$
$$= \begin{pmatrix} A^{-1} & -A^{-1} & 0 \\ -A^{-1} & A^{-1}+S_{11} & S_{12} \\ 0 & S_{12}^T & S_{22} \end{pmatrix},$$

where

$$S = \begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix} = \begin{pmatrix} B+C & D \\ D^T & F \end{pmatrix}^{-1}.$$

This means that under Condition 4,

$$\mathrm{Cov}\,(\hat{\theta}_1) \sim A^{-1},$$
$$\mathrm{Cov}\,(\hat{\theta}_2) \sim A^{-1}+S_{11},$$
$$\mathrm{Cov}\,(\hat{\theta}_3) \sim S_{22},$$

which is very similar to the common dynamics case and the same conclusions hold. In particular, the statistical properties of the estimate of θ_1 only depends on the input signal $u(t)$ and the first output signal $y_1(t)$.

We can further simplify the variance result by in addition to Condition 4 also assume

Condition 5: $G_3(q,\theta_3^o)G_2'(q,\theta_2^o) = G_3'(q,\theta_3^o)G_2(q,\theta_2^o)$

Hence under Conditions 4 and 5 all three subsystems are **identical**, and $C = D = F$ give

$$S = \begin{pmatrix} B^{-1} & -B^{-1} \\ -B^{-1} & B^{-1}+C^{-1} \end{pmatrix}.$$

Example: Let

$$\begin{aligned} G_1(q,\theta_1) &= 1+b_1 q^{-1}, & \theta_1 &= b_1, \\ G_2(q,\theta_3) &= 1+b_2 q^{-1}, & \theta_2 &= b_2, \\ G_3(q,\theta_{22}) &= 1+b_3 q^{-1}, & \theta_3 &= b_3, \end{aligned}$$

and $\{e_1(t)\}$, $\{e_2(t)\}$ and $\{e_3(t)\}$ are white noise with variances λ_1, λ_2 and λ_3, respectively. Let the input signal $u(t)$ be white noise with variance 1. Now assume that Condition 4 holds, i.e. $b_2^o = b_1^o \neq 0$, and furthermore set $b_3^o = 0$ to simplify the calculations. This gives

$$\psi(t) = \begin{pmatrix} \frac{u(t-1)}{\sqrt{\lambda_1}} & \frac{(1+b_1^o q^{-1})u(t-1)}{\sqrt{\lambda_2}}, & \frac{(1+b_1^o q^{-1})u(t-1)}{\sqrt{\lambda_3}} \\ 0 & \frac{(1+b_1^o q^{-1})u(t-1)}{\sqrt{\lambda_2}} & \frac{(1+b_1^o q^{-1})u(t-1)}{\sqrt{\lambda_3}} \\ 0 & 0 & \frac{(1+b_1^o q^{-1})^2 u(t-1)}{\sqrt{\lambda_3}} \end{pmatrix}.$$

Let

$$a = \frac{N}{\lambda_1}, \quad b = \frac{N}{\lambda_2}(1+(b_1^o)^2), \quad c = \frac{N}{\lambda_3}(1+(b_1^o)^2),$$

$$d = \frac{N}{\lambda_3}(1+2(b_1^o)^2), \quad f = \frac{N}{\lambda_3}(1+4(b_1^o)^2+(b_1^o)^4),$$

to obtain

$$M = \begin{pmatrix} a+b+c & b+c & d \\ b+c & b+c & d \\ d & d & f \end{pmatrix},$$

Some calculations now give that under Condition 4

$$\mathrm{Var}\,(\hat{b}_1) \sim \frac{1}{a} = \frac{\lambda_1}{N},$$
$$\mathrm{Var}\,(\hat{b}_2) \sim \frac{1}{a} + \frac{f}{(b+c)f-d^2},$$
$$\mathrm{Var}\,(\hat{b}_3) \sim \frac{b+c}{(b+c)f-d^2}.$$

Here the quality of the estimate of b_1 is only a function of λ_1. ∎

A remaining question is if it is possible to obtain similar results under

Condition 6: $G_3(q,\theta_3^o)G_1'(q,\theta_1^o) = G_3'(q,\theta_3^o)G_1(q,\theta_1^o)$

while $G_2(q)$ could be an arbitrary transfer function? The answer is *no*, which can be verified by the following simple example.

Example: Consider again

$$\begin{aligned} G_1(q,\theta_1) &= 1+b_1 q^{-1}, & \theta_1 &= b_1, \\ G_2(q,\theta_3) &= 1+b_2 q^{-1}, & \theta_2 &= b_2, \\ G_3(q,\theta_{22}) &= 1+b_3 q^{-1}, & \theta_3 &= b_3, \end{aligned}$$

and $\{e_1(t)\}$, $\{e_2(t)\}$ and $\{e_3(t)\}$ are white noise with variances λ_1, λ_2 and λ_3, respectively. Let the input signal $u(t)$ be white noise with variance 1.

We will study the case when $G_3(q,\theta_3^o) = G_1(q,\theta_1^o)$, i.e. $b_3^o = b_1^o \neq 0$, and $b_2^o = 0$. The assumption $b_2^o = 0$ is just to simplify the computations. Here

$$M = \begin{pmatrix} a+b+c & b+d & c \\ b+d & f & d \\ c & d & c \end{pmatrix},$$

where

$$a = \frac{N}{\lambda_1}, \quad b = \frac{N}{\lambda_2}, \quad c = \frac{N}{\lambda_3}(1 + (b_1^o)^2),$$

$$d = \frac{N}{\lambda_2}(1 + 2(b_1^o)^2),$$

$$f = \frac{N}{\lambda_2}(1 + (b_1^o)^2) + \frac{N}{\lambda_3}(1 + 4(b_1^o)^2) + (b_1^o)^4).$$

Then

$$\text{Var}\left(\hat{b}_1\right) \sim \frac{fc - d^2}{(a+b)(fc - d^2) - cb^2},$$

$$\text{Var}\left(\hat{b}_2\right) \sim \frac{(a+b)c}{(a+b)(fc - d^2) - cb^2},$$

$$\text{Var}\left(\hat{b}_3\right) \sim \frac{(a+b+c)f - (b+d)^2}{(a+b)(fc - d^2) - cb^2}.$$

For e.g. $b_1^o = 1$, we obtain

$$\text{Var}\left(\hat{b}_1\right) \sim \frac{1}{N} \frac{4/\lambda_2 + 3/\lambda_3}{4/(\lambda_1\lambda_2) + 3/(\lambda_1\lambda_3) + 2/\lambda_2^2 + 3/(\lambda_2\lambda_3)},$$

which clearly is a function of the quality of all three output signals. ∎

The example shows that the ordering of the systems is important for the results to hold. It is also possible to show that only Condition 4 is not sufficient for a separation result, i.e. the variances of parameter estimates will in general depend on all three outputs.

V. CONCLUSION

The objective of this contribution has been to generalize some variance analysis results for identification of cascade systems. Models of cascade system are special in that they contain products of transfer functions. We have studied the case when some sub-transfer functions are equal and showed that this leads to some unexpected variance results in the sense that the quality of the estimate of the first subsystem only depends on the first output. From an industrial perspective this means that it is extra important to monitor the quality of estimates of the transfer functions of a cascade system using validation data.

The results presented in Sections III and IV are related. Consider the three subsystems case (4) and assume that Condition 4

$$G_2(q, \theta_2^o)G_1'(q, \theta_1^o) = G_2'(q, \theta_2^o)G_1(q, \theta_1^o)$$

holds. As shown in Section IV, this gives

$$M = \begin{pmatrix} A+B+C & B+C & D \\ B+C & B+C & D \\ D^T & D^T & F \end{pmatrix}$$

for which

$$M^{-1} = \begin{pmatrix} A^{-1} & -A^{-1} & 0 \\ -A^{-1} & A^{-1}+S_{11} & S_{12} \\ 0 & S_{12}^T & S_{22} \end{pmatrix},$$

where

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{12}^T & S_{22} \end{pmatrix} = \begin{pmatrix} B+C & D \\ D^T & F \end{pmatrix}^{-1}.$$

By letting the variance λ_2 of $e_2(t)$ tend to infinity we remove all information from the second output $y_2(t)$, and the problem reduces to the two subsystems case with two output signals. For $\lambda_2 = \infty$ we have $B = 0$ and thus

$$M^{-1} = \begin{pmatrix} A^{-1} & -A^{-1} & 0 \\ -A^{-1} & A^{-1}+R_{11} & R_{12} \\ 0 & R_{12}^T & R_{22} \end{pmatrix},$$

where

$$\begin{pmatrix} R_{11} & R_{12} \\ R_{12}^T & R_{22} \end{pmatrix} = \begin{pmatrix} C & D \\ D^T & F \end{pmatrix}^{-1}.$$

We have recovered the result (3) for two subsystems with some common dynamics. This approach is studied in detail in [9].

It would be interesting to study the identification problem for more complex interconnected block structures, and in particular quality/variance issues. Many tools for modeling of physical systems are based on such representations. This information should then be incorporated in the model structure used in system identification. Identification techniques for such system are by no means not new. For example, [1] considered systems composed of cascade, feed-forward, feedback and multiplicative connections of linear dynamic and zero memory nonlinear elements, and showed that such systems can be identified in terms of the individual component subsystems from measurements of the system input and output only. This includes Wiener and Hammerstein models as special cases. More recent work on identification of general structured models can be found in [2], [3].

REFERENCES

[1] S. A. Billings and S. Y. Fakhouri. Identification of systems containing linear dynamic and static nonlinear elements. *Automatica*, 18(1):15–26, 1982.

[2] K. Hsu, T. Vincent, and K. Poolla. A kernel based approach to structured nonlinear system identification part i: Algorithms. In *Proceedings of the 14th IFAC Symposium on System Identification, SYSID 2007*, Newcastle, Australia, 2006.

[3] K. Hsu, T. Vincent, and K. Poolla. A kernel based approach to structured nonlinear system identification part ii: Convergence and consistency. In *Proceedings of the 14th IFAC Symposium on System Identification, SYSID 2007*, Newcastle, Australia, 2006.

[4] L. Ljung. *System Identification - Theory For the User, 2nd ed.* Prentice Hall, Upper Saddle River, New Jersey, 1999.

[5] J. Mårtensson. *Geometric analysis of stochastic model errors in system identification*. PhD thesis, KTH, October 2007. TRITA-EE 2007:061.

[6] J. Mårtensson and H. Hjalmarsson. A geometric approach to variance analysis in system identification: Linear time-invariant systems. In *46th IEEE Conference on Decision and Control*, New Orleans, USA, December 2007.

[7] J. Mårtensson and H. Hjalmarsson. A geometric approach to variance analysis in system identification: Theory and nonlinear systems. In *46th IEEE Conference on Decision and Control*, New Orleans, USA, December 2007.

[8] B. Wahlberg, H. Hjalmarsson, and J. Mårtensson. On identification of cascade systems. In *17th IFAC World Congress*, October 2007.

[9] B. Wahlberg, H. Hjalmarsson, and J. Mårtensson. Variance results for identification of cascade systems. *Automatica*, 2008. Provisionally accepted.

[10] B. Wahlberg, M. Jansson, T. Matsko, and M. Molander. *Experiences from Subspace System Identification - Comments from Process Industry Users and Researchers*, pages 315–327. Modeling, Estimation and Control, Lecture Notes in Information and Control Sciences, Springer, September 2007.