

# On Identification of Parallel Cascade Serial Systems

Per Hägg\* Bo Wahlberg\* Henrik Sandberg\*

\* Automatic Control Lab and ACCESS, School of Electrical Engineering, KTH, SE-100 44 Stockholm, Sweden  
E-mail: {per.hagg, bo.wahlberg, henrik.sandberg}@ee.kth.se

**Abstract:** We consider identification of systems with a parallel serial (cascade) structure with multiple-input and multiple-output signals. The statistical properties of estimated models are studied with respect to input signals and possible sensor locations. The quality of the estimates are analyzed by means of the asymptotic covariance matrix of the estimated parameters. This is an extension of previous work on identification of cascaded linear systems. The key result concerns systems where the sub-systems have common dynamics. An interesting observation is that for this case the variance for the parameters belonging to the unmeasured subsystem always is larger than for the other sub-systems. This is not true for general parameters. The variance results can be used for optimal input and sensor location design. The results are illustrated by some simple FIR examples and numerical evaluations.

*Keywords:* System identification, Prediction error methods, MIMO, Covariance matrices.

## 1. INTRODUCTION

System identification is about the estimation and validation of mathematical models of dynamical systems from experimental data. It is common to have some physical structural information to be included in the choice of the model structure. We will consider systems with a parallel serial structure with multiple-input and multiple-output signals, as shown in Fig. 1. This class of systems generalizes cascaded systems. This problem is partly motivated by the discussion on the use of structural system identification given in Wahlberg et al. (2007), and collaboration with Honeywell Prague Laboratory on boiler control, see e.g., Baramov et al. (2007).

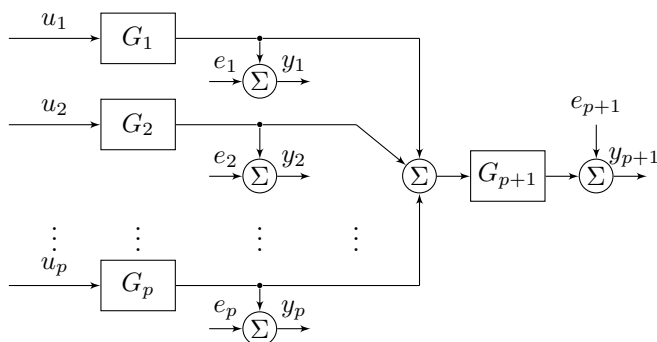


Fig. 1. A parallel serial (PS) structure.

Classical methods in system identification consider single-input single-output (SISO) systems, and can often be extended to multiple-input multiple-output (MIMO) systems, see e.g., Zhu (1998) for a process control perspective. It is important to take a priori information of the structure of the system into account in order to reduce the overall identification task. A priori information also affects the choice of sensors and the design of input signal excitation. The paper Gevers et al. (2006) gives a thorough variance analysis for identification of multi-input systems, together with its implications for input design, while Dayal and MacGregor (1997) consider multi-output process identification. The paper Wahlberg et al. (2009a) analyzes variance properties of cascade systems with three subsystems. The main result is that if the true transfer functions of the first and second subsystem are identical, the output signal information from the second and third subsystems will not affect the asymptotic variance of the estimated model of the first subsystem. Identification of cascaded systems is further studied in Sandberg et al. (2010) and Wahlberg et al. (2009b).

The aim of the current paper is to extend the results in Wahlberg et al. (2009a) to systems with a Parallel Serial (PS) structure as depicted in Fig. 1.

The corresponding transfer functions are

$$\begin{aligned} y_1(t) &= G_1(q)u_1(t) + e_1(t), \\ &\vdots \\ y_p(t) &= G_p(q)u_p(t) + e_p(t), \\ y_{p+1}(t) &= G_{p+1}(q) \sum_{i=1}^p G_i(q)u_i(t) + e_{p+1}(t), \end{aligned}$$

\* This work was partially supported by the Swedish Foundation for Strategic Research, the Swedish Research Council, the Linnaeus Center ACCESS at KTH and the European Commission through the WIDE project.

where  $q$  denotes the forward shift operator, *i.e.*,  $q^{-1}u(t) = u(t-1)$  using normalized sampling time. The system has  $p$  inputs and  $p+1$  outputs. We assume that all transfer functions  $G_k(q)$ ,  $k = 1, \dots, (p+1)$ , are stable.

Several important questions have to be answered and user choices have to be made when applying system identification to this special structure.

- Which input signals should be applied and how should the total power be distributed among the  $p$  inputs?
- If only a subset of the  $p+1$  output signals can be measured and with different sensor qualities. Where should the sensors be located for best result?
- Does the dynamics of the sub-systems affect the identification result? For cascaded systems we know that common dynamics give fundamental quality limitations

A standard subspace method could be applied to MIMO input-output data, but it is not straightforward to impose the known structure of the problem, see for example Van Overschee and De Moor (1996) or Ljung (1999). Examples on how to incorporate some known structure in subspace identification can be found in Lyzell et al. (2009) and Hägg et al. (2010). It is also possible to apply a Prediction Error Method (PEM) or the Maximum Likelihood (ML) method to the structured MIMO problem. But because of the transfer function products, such as  $G_{p+1}(q)G_1(q)$ , simple linear model structures such as OE or ARX are not directly applicable. However, structured PEM and ML are statistically efficient methods also for this class of identification problems.

The main contribution of this paper is to derive variance results for identification of parallel serial structured systems and discuss the implications in input and sensor design. We will mainly consider the case of two parallel sub-systems together with a common serial sub-system. The technique used in the paper can be used to analyze more general structures.

The outline of the paper is as follows. In Section 2 we will start by studying low order FIR systems. The general case is studied in Section 3. A numerical example is given in Section 4, while Section 5 concludes the paper.

## 2. VARIANCE ANALYSIS: FIR EXAMPLE

Consider the special case with three sub-systems in Fig. 2.

Here

$$\begin{aligned} y_1(t) &= G_1(q, \theta_1)u_1(t) + e_1(t), \\ y_2(t) &= G_2(q, \theta_2)u_2(t) + e_2(t), \\ y_3(t) &= G_3(q, \theta_3)(G_1(q, \theta_1)u_1(t) + G_2(q, \theta_2)u_2(t)) + e_3(t), \end{aligned} \quad (1)$$

where the subsystems are first order FIR transfer functions

$$\begin{aligned} G_1(q, \theta_1) &= 1 + b_1q^{-1}, & \theta_1 &= b_1, \\ G_2(q, \theta_2) &= 1 + b_2q^{-1}, & \theta_2 &= b_2, \\ G_3(q, \theta_3) &= 1 + b_3q^{-1}, & \theta_3 &= b_3. \end{aligned} \quad (2)$$

The objective is to identify the FIR parameters  $b_1$ ,  $b_2$  and  $b_3$  from measurement of  $\{u_1(t), u_2(t), y_1(t), y_2(t), y_3(t)\}$ ,

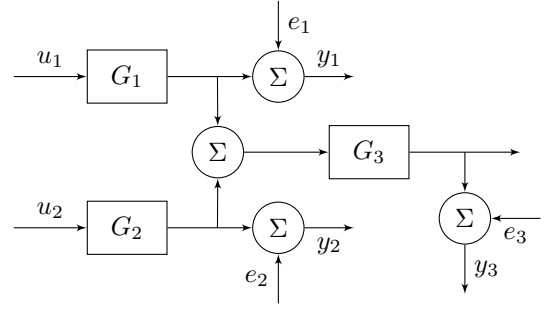


Fig. 2. Three subsystem parallel serial system.

$t = 1, \dots, N$ . We assume that the true system can be described by the model structure (1) with the true FIR parameters  $b_1^0$ ,  $b_2^0$  and  $b_3^0$ , respectively. Furthermore, we assume that  $e_1(t)$ ,  $e_2(t)$  and  $e_3(t)$  are independent white measurement noise processes with given variances  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ , respectively.

To start, the input signals  $u_1(t)$  and  $u_2(t)$  are chosen as independent white noise processes with variances 1. Given the data set  $\{u_1(t), u_2(t), y_1(t), y_2(t), y_3(t)\}$ ,  $t = 1, \dots, N$  the PEM estimate of the model parameters are given by, see Ljung (1999) ,

$$\begin{aligned} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{b}_3 \end{pmatrix} &= \arg \min_{b_1, b_2, b_3} \left[ \frac{1}{N} \sum_{t=1}^N \frac{[y_1(t) - u_1(t) - b_1u_1(t-1)]^2}{\lambda_1} \right. \\ &+ \frac{1}{N} \sum_{t=1}^N \frac{[y_2(t) - u_2(t) - b_2u_2(t-1)]^2}{\lambda_2} \\ &+ \left. \frac{1}{N} \sum_{t=1}^N \frac{1}{\lambda_3} [y_3(t) - u_1(t) - u_2(t) - (b_1 + b_3)u_1(t-1) - \right. \\ &\quad \left. (b_2 + b_3)u_2(t-1) - b_1b_3u_1(t-2) - b_2b_3u_2(t-2)]^2 \right] \end{aligned} \quad (3)$$

The asymptotic (in  $N$ ) covariance matrix of the parameter estimates, which in this case corresponds to the Cramér-Rao lower bound, is given by

$$P_\theta = \frac{1}{N} [E\{\Psi(t)\Psi^T(t)\}]^{-1},$$

where

$$\Psi(t) = \begin{bmatrix} \frac{u_1(t-1)}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{u_2(t-1)}{\sqrt{\lambda_2}} \\ 0 & 0 \\ \frac{u_1(t-1) + b_3^0u_1(t-2)}{\sqrt{\lambda_3}} \\ \frac{u_2(t-1) + b_3^0u_2(t-2)}{\sqrt{\lambda_3}} \\ \frac{u_1(t-1) + u_2(t-1) + b_1^0u_1(t-2) + b_2^0u_2(t-2)}{\sqrt{\lambda_3}} \end{bmatrix},$$

is the normalized predictor gradient. The expressions for the parameter variances, that is the diagonal elements of  $P_\theta$  are lengthy and difficult to access. We will instead look at some special cases. The notation  $\sim$  is used to stress the asymptotic relations.

## 2.1 Preliminaries

If only the outputs from the parallel subsystems  $y_1(t)$  and  $y_2(t)$  are used the asymptotic variances of the estimates of  $b_1$  and  $b_2$  equal

$$\begin{aligned}\text{Var } \hat{b}_1 &\sim \frac{\lambda_1}{N}, \\ \text{Var } \hat{b}_2 &\sim \frac{\lambda_2}{N}.\end{aligned}\quad (4)$$

This corresponds to setting  $\lambda_3 = \infty$  in the general expression for  $\Psi(t)$  and  $P_\theta$ .

In case the quality of the two sensors  $y_1(t)$  and  $y_2(t)$  are much worse than for the last output  $y_3(t)$ , *i.e.*, letting  $\lambda_1 \rightarrow \infty$  and  $\lambda_2 \rightarrow \infty$  we have

$$\begin{aligned}\text{Var } \hat{b}_1 &\sim \frac{\lambda_3}{N} \frac{(2 + (b_1^0)^2 + (b_2^0)^2) - \frac{(1 + b_2^0 b_3^0)^2}{1 + (b_3^0)^2}}{(b_3^0 - b_1^0)^2 + (b_3^0 - b_2^0)^2}, \\ \text{Var } \hat{b}_3 &\sim \frac{\lambda_3}{N} \frac{1 + (b_3^0)^2}{(b_3^0 - b_1^0)^2 + (b_3^0 - b_2^0)^2},\end{aligned}$$

which will be large if  $b_1^0 \approx b_2^0 \approx b_3^0$ . The result is valid for  $\hat{b}_2$  as well due to symmetry, and is related to the work by Wahlberg et al. (2009a), where this result is proved for cascade structured systems.

## 2.2 Subsystems with common dynamics

From Wahlberg et al. (2009a) it is known that for cascade structured systems with the two identical transfer functions there is no improvement in the estimation of the first subsystem obtained from the second measurement. For this parallel serial system the variance of the parameter estimates when the three subsystems are identical, *i.e.*,  $b_2^0 = b_3^0 = b_1^0$  becomes

$$\text{Var } \hat{b}_1 \sim \frac{1}{N} \left( \lambda_1 - \frac{\lambda_1^2 (1 + (b_1^0)^2)}{2\lambda_3 + (\lambda_1 + \lambda_2)(1 + (b_1^0)^2)} \right) \leq \frac{\lambda_1}{N}.\quad (5)$$

Hence, for PS systems the estimate of  $b_1$  is improved from the second measurement, *c.f.*, (4).

However, if  $b_2^0 = b_3^0 = b_1^0$  and  $y_2(t)$  is not measured ( $\lambda_2 = \infty$ ) then

$$\begin{aligned}\text{Var } \hat{b}_1 &\sim \frac{\lambda_1}{N}, \\ \text{Var } \hat{b}_2 &\sim \frac{1}{N} \left( \lambda_1 + \frac{2\lambda_3}{1 + (b_1^0)^2} \right), \\ \text{Var } \hat{b}_3 &\sim \frac{1}{N} \left( \lambda_1 + \frac{\lambda_3}{1 + (b_1^0)^2} \right).\end{aligned}\quad (6)$$

The expressions for  $\hat{b}_1$  and  $\hat{b}_3$  are exactly the expressions obtained by Wahlberg et al. (2009a) for cascade structured systems. An interesting observation is that the variance for the parameter  $\hat{b}_2$  belonging to the unmeasured subsystem always is larger ( $\lambda_3 > 0$ ) than the variance for the two other parameter estimates.

This observation is not true in general. It is quite easy to find cases such that  $\text{Var } \hat{b}_1 > \text{Var } \hat{b}_2$  when  $y_2(t)$  is not

measured. For example

$$\begin{aligned}\text{Var } \hat{b}_1 - \text{Var } \hat{b}_2 &> 0 \\ &\Leftrightarrow \\ \lambda_1 \lambda_3 b_3^0 (b_1^0 - b_2^0) (2 + b_1^0 b_3^0 + b_2^0 b_3^0) &> \\ \lambda_3^2 (2 + (b_1^0)^2 + (b_2^0)^2) &\end{aligned}$$

for this case. If  $\lambda_1 = \lambda_3 = 1$  and  $b_1^0 = 3$ ,  $b_2^0 = -0.5$ ,  $b_3^0 = 2$  then  $\text{Var } \hat{b}_1 - \text{Var } \hat{b}_2 \approx \frac{0.40}{N}$ . Hence, if one only can measure one of the two output signals from the first sub-systems, it is not always best to follow the intuition and measure the output from the system corresponding to the parameter one wants to identify.

If instead only two sub-systems are equal,  $b_3^0 = b_1^0$ ,  $b_2^0 \neq b_1^0$  and  $y_2(t)$  is not measured ( $\lambda_2 = \infty$ ) then

$$\text{Var } \hat{b}_1 \sim \frac{1}{N} \left[ \lambda_1 - \frac{\lambda_1^2}{\lambda_1 + \lambda_3 \frac{(b_1^0 - b_2^0)^2 + (b_1^0)^2 + 1}{(b_1^0)^2 + 1} (b_1^0 - b_2^0)} \right],$$

which always is smaller than the expression (6) obtained when all three subsystems are equal.

If it is known in advance that  $G_3 = G_2 = G_1$ , one should incorporate this information to further constrain the model. In this case this means that only one parameter,  $b_1$ , should be estimated. The asymptotic variance, when  $y_2$  is not measured, is

$$\text{Var } \hat{b}_1 \sim \frac{1}{N} \frac{\lambda_1}{1 + \frac{8\lambda_1(1 + (b_1^0)^2)}{\lambda_3}}.$$

which could be considerably lower than (4). Hence, additional information about the structure should always be incorporated.

## 2.3 Effects of input signals

Now consider the case when the same input signal is used as inputs to both sub-systems, *i.e.*,  $u_1(t) = u_2(t)$  and that  $y_2(t)$  is not measured ( $\lambda_2 = \infty$ ). The asymptotic variance for  $\hat{b}_1$  then equals

$$\text{Var } \hat{b}_1 \sim \frac{\lambda_1}{N}.$$

Hence, nothing is gained by measuring  $y_3(t)$  if the objective is to only estimate  $b_1$ .

However, if the aim of the identification is to identify  $b_3$ , this input signal could be good for some special cases. Consider the case when the second input is not measured, *i.e.*,  $\lambda_2 = \infty$  and  $\lambda_1 = \lambda_3 = b_1^0 = 1$ ,  $b_2^0 = 2$  and  $b_3^0 = 3$ . When the input signals are independent white noise sequences, we have

$$\text{Var } \hat{b}_3 \sim \frac{1}{N} \frac{110}{71} \approx \frac{1.55}{N}$$

and for the case  $u_1(t) = u_2(t)$

$$\text{Var } \hat{b}_3 \sim \frac{1}{N} \frac{10}{9} \approx \frac{1.11}{N}.$$

Here the variance is smaller when using the same input signal to both sub-systems. If we instead have  $\lambda_1 = \lambda_3 =$

$b_3^0 = 1, b_2^0 = 2$  and  $b_1^0 = 3$  then we have for the independent input signal case

$$\text{Var } \hat{b}_3 \sim \frac{1}{N} \frac{6}{31} \approx \frac{0.19}{N}$$

and when the same input signal is used as the inputs

$$\text{Var } \hat{b}_3 \sim \frac{1}{N} \frac{2}{9} \approx \frac{0.22}{N}.$$

These simple examples above shows the important of choosing input signals depending on what one wants to identify.

### 3. VARIANCE ANALYSIS: GENERAL NUMBER OF SUBSYSTEMS

In this section we will try to generalize some of the results from the FIR-examples in the previous section. Consider again the PS model structure (1) with arbitrary number of parallel sub-systems

Given a data set  $\{u_1(t), \dots, u_p(t), y_1(t), \dots, y_{p+1}(t)\}$ ,  $t = 1, \dots, N$  the PEM estimates of the model parameters,

$$\begin{aligned} \theta_1 &= [\theta_{1,1}, \theta_{1,2}, \dots, \theta_{1,n_1}]^T, \\ &\vdots \\ \theta_{p+1} &= [\theta_{p+1,1}, \theta_{p+1,2}, \dots, \theta_{p+1,n_{p+1}}]^T, \end{aligned}$$

where  $n_i$  denotes the number of parameters in  $\theta_i$ , are given by

$$\begin{aligned} \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_{p+1} \end{pmatrix} &= \arg \min_{\theta_1, \dots, \theta_{p+1}} \left[ \frac{1}{N} \sum_{t=1}^N \frac{[y_1(t) - G_1(q, \theta_1)u_1(t)]^2}{\lambda_1} + \dots + \right. \\ &\frac{1}{N} \sum_{t=1}^N \frac{[y_p(t) - G_p(q, \theta_p)u_p(t)]^2}{\lambda_p} + \\ &\left. \frac{1}{N} \sum_{t=1}^N \frac{[y_{p+1}(t) - G_{p+1}(q, \theta_{p+1}) (\sum_{i=1}^p G_i(q, \theta_i)u_i(t))]^2}{\lambda_{p+1}} \right]. \end{aligned}$$

Define the  $(p+1) \times (p+1)$  block matrix

$$\Psi(t) = \begin{bmatrix} \frac{G'_1(q, \theta_1)u_1(t)}{\sqrt{\lambda_1}} & 0 & 0 & \dots \\ 0 & \frac{G'_2(q, \theta_2)u_2(t)}{\sqrt{\lambda_2}} & 0 & \dots \\ \vdots & & \ddots & \\ 0 & 0 & \dots & \\ 0 & 0 & \dots & \\ 0 & \frac{G_{p+1}(q, \theta_3)G'_1(q, \theta_1)u_1(t)}{\sqrt{\lambda_{p+1}}} & & \\ 0 & \frac{G_{p+1}(q, \theta_3)G'_2(q, \theta_2)u_2(t)}{\sqrt{\lambda_{p+1}}} & & \\ \vdots & \vdots & & \\ \frac{G'_p(q, \theta_1)u_p(t)}{\sqrt{\lambda_p}} & \frac{G_{p+1}(q, \theta_3)G'_p(q, \theta_2)u_p(t)}{\sqrt{\lambda_{p+1}}} & & \\ 0 & \frac{G'_{p+1}(q, \theta_{p+1})}{\sqrt{\lambda_{p+1}}} \sum_{i=1}^p G_i(q, \theta_i)u_i(t) & & \end{bmatrix}$$

where prime denotes differentiation with respect to the respective parameter vector. The asymptotic covariance matrix of the parameter estimates is then given by

$$\text{Cov} \begin{pmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_{p+1} \end{pmatrix} \sim P_\theta$$

where  $P_\theta = \frac{1}{N} [E\{\Psi(t)\Psi(t)^T\}]^{-1}$ . See, *e.g.*, Ljung (1999) for details.

Now consider the case when all subsystems are identical and have the same structure. Then

$$G_i(q, \theta_i^0)G'_j(q, \theta_j^0) = G'_i(q, \theta_i^0)G_j(q, \theta_j^0)$$

for  $i, j = 1, \dots, (p+1)$ . If all  $u_i(t)$  and  $u_j(t)$  are independent for  $i \neq j$ , we have

$$P_\theta^{-1} = NE\{\Psi(t)\Psi(t)^T\} = \begin{pmatrix} A_1 + B_1 & 0 & \dots & 0 & B_1 \\ 0 & A_2 + B_2 & & 0 & B_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & A_p + B_p & B_p \\ B_1 & B_2 & \dots & B_p & \sum_{i=1}^p B_i \end{pmatrix},$$

where the covariance matrices  $A_i$  and  $B_i$ , are defined as

$$\begin{aligned} A_i &= \frac{N}{\lambda_i} E\{[G'_i(q, \theta_i)u_i(t)][G'_i(q, \theta_i)u_i(t)]^T\}, \\ B_i &= \frac{N}{\lambda_{p+1}} E\{[G_{p+1}(q, \theta_{p+1})G'_i(q, \theta_i)u_i(t)] \\ &\quad \times [G_{p+1}(q, \theta_{p+1})G'_i(q, \theta_i)u_i(t)]^T\}. \end{aligned}$$

Next, we will consider the case when only  $y_1(t)$  and  $y_{p+1}(t)$  are measured, *i.e.*,  $\lambda_i = \infty$  for  $i = 2, \dots, p$ . Then

$$\lim_{\lambda_i \rightarrow \infty} A_i = 0 \quad \forall i = 2, \dots, p$$

and

$$P_\theta^{-1} = \begin{pmatrix} A_1 + B_1 & 0 & \dots & 0 & B_1 \\ 0 & B_2 & & 0 & B_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & B_p & B_p \\ B_1 & B_2 & \dots & B_p & \sum_{i=1}^p B_i \end{pmatrix}.$$

We need to invert this matrix in order to find the parameter covariance matrix. The matrix can be block diagonalized by the following transformation matrix

$$T = \begin{pmatrix} I & I & \dots & I & -I \\ 0 & I & \dots & 0 & 0 \\ \vdots & & \ddots & \vdots & \\ \vdots & & & I & 0 \\ 0 & 0 & \dots & 0 & I \end{pmatrix},$$

where  $I$  denotes an identity matrix of suitable size. We then have the following block diagonal matrix

$$\bar{P}_\theta^{-1} = TP_\theta^{-1}T^T = \begin{pmatrix} A_1 & 0 & \dots & 0 & 0 \\ 0 & B_2 & & 0 & B_2 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & B_p & B_p \\ 0 & B_2 & \dots & B_p & \sum_{i=1}^p B_i \end{pmatrix}.$$

It is possible to verify that the inverse of the lower right block is

$$\begin{pmatrix} B_2 & 0 & B_2 \\ \vdots & \vdots & \vdots \\ 0 & B_p & B_p \\ B_2 \dots B_p & \sum_{i=1}^p B_i \end{pmatrix}^{-1} = \begin{pmatrix} B_1^{-1} + B_2^{-1} & B_1^{-1} & \dots & B_1^{-1} & -B_1^{-1} \\ B_1^{-1} & B_1^{-1} + B_3^{-1} & \ddots & \vdots & -B_1^{-1} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ B_1^{-1} & B_1^{-1} & \dots & B_1^{-1} + B_p^{-1} & -B_1^{-1} \\ -B_1^{-1} & -B_1^{-1} & \dots & -B_1^{-1} & B_1^{-1} \end{pmatrix}.$$

Now  $P_\theta = T^T \bar{P}_\theta T$ , which gives

$$P_\theta = \begin{pmatrix} A_1^{-1} & A_1^{-1} & \dots & A_1^{-1} & -A_1^{-1} \\ A_1^{-1} & C + B_2^{-1} & C & \dots & C & -C \\ \vdots & C & C + B_3^{-1} & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & C & -C \\ A_1^{-1} & C & \dots & C & C + B_p^{-1} & -C \\ -A_1^{-1} & -C & \dots & -C & -C & C \end{pmatrix}.$$

where  $C = A_1^{-1} + B_1^{-1}$ . We have now proved the following proposition.

*Proposition 1.* Assume that the true subsystems are all identical and that we use the identical sub-model structures. Furthermore, assume that we only measure  $y_1(t)$  and  $y_{p+1}(t)$ . Let

$$A_i = \frac{N}{\lambda_i} E\{[G'_i(q, \theta_i)u_i(t)][G'_i(q, \theta_i)u_i(t)]^T\},$$

$$B_i = \frac{N}{\lambda_{p+1}} E\{[G_{p+1}(q, \theta_{p+1})G'_i(q, \theta_i)u_i(t)] \times [G_{p+1}(q, \theta_{p+1})G'_i(q, \theta_i)u_i(t)]^T\}.$$

The the asymptotic covariance matrix of the parameter estimates equal

$$\begin{aligned} \text{Cov } \hat{\theta}_1 &\sim A_1^{-1}, \\ \text{Cov } \hat{\theta}_i &\sim A_1^{-1} + B_1^{-1} + B_i^{-1}, \quad i = 2, \dots, p \\ \text{Cov } \hat{\theta}_{p+1} &\sim A_1^{-1} + B_1^{-1}. \end{aligned}$$

**The key results when all the subsystems are identical and only  $y_1(t)$  and  $y_{p+1}(t)$  are measured are:**

- Since  $A_1^{-1}$  is the asymptotic covariance matrix of  $\hat{\theta}_1$  when only  $y_1(t)$  is measured, the quality of the estimate of  $\theta_1$  is not improved by also measuring  $y_{p+1}(t)$ .
- The asymptotic variance for  $\hat{\theta}_{p+1}$  is always worse than for  $\hat{\theta}_1$  since  $B_1^{-1}$  is a positive definite matrix.

- The asymptotic variance for  $\hat{\theta}_i$ ,  $i = 2, \dots, p$  is always worse than both the one for  $\hat{\theta}_1$  and the one for  $\hat{\theta}_{p+1}$  since  $B_i^{-1}$  is a positive definite matrix.
- The estimates of  $\hat{\theta}_1$ ,  $\hat{\theta}_i$ ,  $i = 2, \dots, p$  and  $\hat{\theta}_{p+1}$  all contains terms of  $A_1^{-1}$ . Hence if the quality of the measurement of  $y_1(t)$  is bad, *i.e.*,  $\lambda_1$  is large, then we always get bad estimates of  $\hat{\theta}_i$ ,  $i = 2, \dots, p$  and  $\hat{\theta}_{p+1}$  irrespective of how good our measurement of  $y_{p+1}(t)$  is and irrespective of how much input power that is used in  $u_i(t)$  for  $i = 2, \dots, p$ .
- If it is known beforehand that all the systems are identical, this information should be incorporated in the model structure. If this information is used in the PEM parameter estimate, the asymptotic covariance matrix  $P_\theta$  can be calculated using

$$\Psi(t) = \begin{pmatrix} \frac{G'_1(q, \theta_1)u_1(t)}{\lambda_1}, \dots, \\ \frac{G'_1(q, \theta_1)u_p(t)}{\lambda_p}, \frac{2G'_1(q, \theta_1)G_1(q, \theta_1) \sum_{i=1}^p u_i(t)}{\lambda_{p+1}} \end{pmatrix}$$

in  $P_\theta = \frac{1}{N} E\{\Psi(t)\Psi(t)^T\}^{-1}$ . This gives

$$\text{Cov } \hat{\theta}_1 \sim \left( A_1 + 4 \sum_{i=1}^p B_i \right)^{-1}$$

which always is smaller than  $A_1^{-1}$  which is the asymptotic covariance matrix if the structural information is not used.

- The asymptotic covariances  $A_i$  and  $B_i$  are directly proportional to the power in the input signal  $u_i(t)$ . If we assume that we have some constraint on the total power that we can use to excite the system, and our aim is to estimate  $\theta_1$  or  $\theta_{p+1}$ , all input power should be used in  $u_1(t)$ .

If we on the other hand want to estimate  $\theta_i$  for  $i = 2, \dots, p$  the input signal should be divided between  $u_1(t)$  and  $u_i(t)$ . How the power should be distributed is decided by the relations between  $A_1$ ,  $B_1$  and  $B_i$ . If we use independent white noise signals one could for example formulate the following optimization problem to find the signal powers  $\sigma_1^2$  and  $\sigma_i^2$  for  $u_1(t)$  and  $u_i(t)$  respectively

$$\begin{aligned} \min_{\sigma_1^2, \sigma_i^2} \quad & \text{tr} \left( \frac{A_1^{-1}}{\sigma_1^2} + \frac{B_1^{-1}}{\sigma_1^2} + \frac{B_i^{-1}}{\sigma_i^2} \right) \\ \text{s. t.} \quad & \sigma_1^2 + \sigma_i^2 = S_{tot} \end{aligned}$$

where  $S_{tot}$  is the maximum input power and  $A_1$ ,  $B_1$  and  $B_i$  are calculated with white noise inputs with unit variance.

#### 4. NUMERICAL EXAMPLE

In this section some of the theoretical results are verified by MATLAB simulations. We use the model structure (1) with first order FIR transfer function (2) and assume that the true system has the same structure. The following parameter values are used

$$b_1^0 = b_2^0 = b_3^0 = -0.6, \\ \lambda_1 = \lambda_2 = \lambda_3 = 1,$$

*i.e.*, we have the case (5), where all systems are equal and all signals are measured.

This system is simulated for 500 samples and the parameters are estimated by solving the optimization problem (3). The optimization problem is solved numerically in MATLAB with the command *fminunc*. Since we only are interested in the variance and not the convergence of the algorithm we initialize the solver with the true parameters.

1000 Monte Carlo simulations are performed and the parameter mean and variance over the Monte Carlo simulations are calculated. The result is shown in Table 1 together with the theoretical values calculated from (5).

Table 1. Theoretical and simulated mean and variance for the three parameters.

	Theoretical		Simulated	
	mean	<i>N</i> Var	mean	<i>N</i> Var
$\hat{b}_1$	-0.60	0.71	-0.60	0.71
$\hat{b}_2$	-0.60	0.71	-0.60	0.72
$\hat{b}_3$	-0.60	0.87	-0.60	0.88

The simulated values are very close to the theoretical values.

#### 5. CONCLUSIONS

The objective of this contribution has been to present some fundamental asymptotic variance analysis results for identification of parallel serial systems. The main result concerns the case when the true sub systems are identical, but is also relevant for systems where the true sub-transfer functions are close to each other. Future work involves optimal input design for parallel serial systems when only the output  $y_{p+1}(t)$  is measured. It is not obvious how to excite all systems at the same time while guaranteeing identifiability of the parallel subsystems.

#### REFERENCES

- Baramov, L., Beneš, M., and Havlena, V. (2007). Model for estimating stress in pressurized boiler components based on interconnections. In *Decision and Control, 2007 46th IEEE Conference on*, 1603–1608.
- Dayal, B.S. and MacGregor, J.F. (1997). Multi-output process identification. *Journal of Process Control*, 7(4), 269–282.
- Gevers, M., Mišković, L., Bonvin, D., and Karimi, A. (2006). Identification of multi-input systems: variance analysis and input design issues. *Automatica*, 42(4), 559–572.
- Hägg, P., Wahlberg, B., and Sandberg, H. (2010). On subspace identification of cascade structured systems. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2843–2848.
- Ljung, L. (1999). *System Identification - Theory For the User*. Prentice Hall, Upper Saddle River, New Jersey, 2nd edition.
- Lyzell, C., Enqvist, M., and Ljung, L. (2009). Handling certain structure information in subspace identification. In *Proceedings of the 15th IFAC Symposium on System Identification*.
- Sandberg, H., Hägg, P., and Wahlberg, B. (2010). System identification and model reduction of cascade systems. *IEEE Transactions on Automatic Control*. Submitted.
- Van Overschee, P. and De Moor, B. (1996). *Subspace Identification for Linear Systems : Theory - Implementation - Applications*. Kluwer Academic Publishers, Boston/London/Dordrecht.
- Wahlberg, B., Hjalmarsson, H., and Mårtensson, J. (2009a). Variance results for identification of cascade systems. *Automatica*, 45(6), 1443–1448.
- Wahlberg, B., Jansson, M., Matsko, T., and Molander, M. (2007). *Experiences from Subspace System Identification - Comments from Process Industry Users and Researchers*, 315–327. Springer.
- Wahlberg, B., Stoica, P., and Babu, P. (2009b). On estimation of cascade systems with common dynamics. In *15th IFAC Symposium on System Identification, SYSID 2009*.
- Zhu, Y. (1998). Multivariable process identification for mpc: the asymptotic method and its applications. *Journal of Process Control*, 8(2), 101–115.