Expressions for the covariance matrix of covariance data

Torsten Söderström
Division of Systems and Control,
Department of Information Technology, Uppsala University,
P O Box 337, SE-75105 Uppsala, Sweden

May 27, 2009

Abstract

In several estimation methods used in system identification, a first step is to estimate the covariance functions of the measured inputs and outputs for a small set of lags. These covariance elements can be set up as a vector. The report treats the problem of deriving and computing the asymptotic covariance matrix of this vector, when the number of underlying input-output data is large. The derived algorithm is derived under fairly general assumptions. It is assumed that the input and output are linked through a linear finite-order system. Further, the input is assumed to be modelled as an ARMA model of a fixed, but arbitrary order. Finally, it is allowed that the both the input and the output are not measured directly, but with some white measurement noise, thus including typical errors-in-variables situations in the analysis.

1 Introduction

There are many system identification methods where the parameter estimates are formed by a nonlinear transformation of a finite set of input-output covariance data. This applies to the least squares estimate, the instrumental variable estimate, and to several methods applied to errors-in-variables, such as bias-eliminating least squares, [12], [13], the Frisch scheme, [1], [2], and covariance matching [9], [4], [11].

Denote the input signal by $u(t)$ and the output signal by $y(t)$. For the methods considered, the parameter vector estimate is generally a nonlinear function of covariance data in the following way,

$$\hat{\boldsymbol{\theta}} = \mathbf{f}(\hat{\mathbf{r}}) \tag{1}$$

where the covariance vector $\mathbf{r}$ has the form

$$\hat{\mathbf{r}} \triangleq \begin{pmatrix} \hat{\mathbf{r}}_y \\ \hat{\mathbf{r}}_u \\ \hat{\mathbf{r}}_{yu} \end{pmatrix}. \tag{2}$$

and

$$\hat{\mathbf{r}}_y = \begin{pmatrix} \hat{r}_y(0) \\ \vdots \\ \hat{r}_y(p_y) \end{pmatrix}, \quad \hat{\mathbf{r}}_u = \begin{pmatrix} \hat{r}_u(0) \\ \vdots \\ \hat{r}_u(p_u) \end{pmatrix}, \quad \hat{\mathbf{r}}_{yu} = \begin{pmatrix} \hat{r}_{yu}(p_1) \\ \vdots \\ \hat{r}_{yu}(p_2) \end{pmatrix} \tag{3}$$

Here $\hat{r}_y(\tau)$ is the *estimated* output covariance function

$$\hat{r}_y(\tau) = \frac{1}{N}\sum_{t=1}^{N} y(t+\tau)y(t) \tag{4}$$

where N is the number of data points.

Example 1.1. Consider a first order linear regression model

$$y(t) + ay(t-1) = bu(t-1) + \varepsilon(t) \tag{5}$$

and let the parameters a and b be estimated using least squares. Then it is the estimation vector is

$$\begin{aligned} \hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} &= \begin{pmatrix} \hat{r}_y(0) & -\hat{r}_{yu}(0) \\ -\hat{r}_{yu}(0) & \hat{r}_u(0) \end{pmatrix}^{-1} \begin{pmatrix} -\hat{r}_y(1) \\ \hat{r}_{yu}(1) \end{pmatrix} \\ &= \frac{1}{\hat{r}_y(0)\hat{r}_u(0) - \hat{r}_{yu}^2(0)} \begin{pmatrix} -\hat{r}_u(0)\hat{r}_y(1) + \hat{r}_{yu}(0)\hat{r}_{yu}(1) \\ -\hat{r}_{yu}(0)\hat{r}_y(1) - \hat{r}_y(0)\hat{r}_{yu}(1) \end{pmatrix} \end{aligned} \tag{6}$$

In this case we have apparently

$$p_y = 1,\ p_u = 0,\ p_1 = 0,\ p_2 = 1 \tag{7}$$

and the function $\mathbf{f}(\cdot)$ is shown in explicit form in (6). □

Using the general form (5) of the estimator, it is straightforward to use linearization to get an expression for the asymptotic covariance matrix of the parameter estimates. Linearization leads to

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \mathbf{f}_r \left(\hat{\mathbf{r}} - \mathbf{r} \right) \tag{8}$$

where $\mathbf{f}_r$ denotes the Jacobian matrix $\frac{\partial \mathbf{f}}{\partial \mathbf{r}}$. Introducing the normalized covariance matrix of the covariance vector as

$$\mathbf{R} = \lim_{N \to \infty} N \mathsf{E} \left(\hat{\mathbf{r}} - \mathbf{r} \right) \left(\hat{\mathbf{r}} - \mathbf{r} \right)^T \tag{9}$$

it follows that the asymptotic normalized covariance matrix of the parameter vector is

$$\mathsf{cov}(\hat{\boldsymbol{\theta}}) = \lim_{N \to \infty} N \mathsf{E} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right)^T = \mathbf{f}_r \mathbf{R} \mathbf{f}_r^T \tag{10}$$

To handle a general case to derive the covariance matrix in (10), it is therefore needed to find the covariance matrix $\mathbf{R}$, (9), of the covariance data. For methods like the covariance matching approach for errors-in-variables identification, this is the feasible way to go, [8]. For several other methods, this very general approach is possible to use, although more direct methods can also be applied, [3], [6], [10].

In order to be able to handle the general case of estimators of the form (1), we will in this report derive expressions for the covariance matrix $\mathbf{R}$, that hold under weak general assumptions.

2 Problem setup and basic assumptions

Consider a linear and single-input single-output (SISO) system given by

$$A(q^{-1}) y_0(t) = B(q^{-1}) u_0(t), \tag{11}$$

where $u_0(t)$ and $y_0(t)$ are the noise-free input and output, respectively. Further, $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the backward shift operator, described as

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_n q^{-n}, \\ B(q^{-1}) &= b_1 q^{-1} + \cdots + b_m q^{-m}. \end{aligned} \tag{12}$$

In order to be able to easily handle the errors-in-variables problem, [7], we assume that the measurements are noise-corrupted. Hence, we assume that the observations are corrupted by additive measurement noises $\tilde{u}(t)$ and $\tilde{y}(t)$. The available signals are in discrete-time and of the form

$$u(t) \;=\; u_0(t) + \tilde{u}(t), \quad y(t) = y_0(t) + \tilde{y}(t). \tag{13}$$

The following assumptions are introduced.

A1. The dynamic system (11) is asymptotically stable, observable and controllable.

A2. The polynomial degrees n and m are a priori known.

A3. The true input $u_0(t)$ is a zero-mean stationary ergodic random signal, that is persistently exciting of sufficiently high order.

A4. The input noise $\tilde{u}(t)$ and the output noise $\tilde{y}(t)$ are both independent of $u_0(t)$ and mutually independent white noise sequences of zero mean, and variances λ_u and λ_y, respectively. [The extension of the method to cope with arbitrarily correlated output noise is possible.]

A5. The data are Gaussian distributed (both noise-free data and measurement noise).

A6. The noise-free input signal, $u_0(t)$ is an ARMA process, say

$$u_0(t) = \frac{C(q^{-1})}{D(q^{-1})} e(t) \tag{14}$$

where $C(q^{-1})$ and $D(q^{-1})$ are polynomials, and $e(t)$ is a white noise sequence of zero mean and unit variance.

For a general random process $x(t)$, we define its covariance function $r_x(\tau)$ as:

$$r_x(\tau) \;=\; \mathsf{E}\left\{x(t+\tau)x(t)\right\}, \quad \tau = 0,\ \pm 1,\ \pm 2,\ \ldots. \tag{15}$$

where E is the expectation operator.

Further, introduce the conventions

$$a_i \;=\; \begin{cases} 1 & i = 0, \\ 0 & i < 0, i > n, \end{cases} \tag{16}$$

$$b_i \;=\; 0 \quad i < 1, i > m. \tag{17}$$

3 General considerations

As the vector $\hat{\mathbf{r}}$, (2) split up into three parts, its covariance matrix $\mathbf{R}$ will naturally split into corresponding partitions. For simplicity of notation, we consider first the upper left block of $\mathbf{R}$. We can then write

$$\hat{\mathbf{r}} = \frac{1}{N}\sum_{t=1}^{N}\boldsymbol{\varphi}(t)y(t) \tag{18}$$

where

$$\boldsymbol{\varphi}(t) = \begin{pmatrix} y(t) \\ \vdots \\ y(t-p_y) \end{pmatrix} \tag{19}$$

Furthermore, we let $\mathbf{r}$ denote the true value of $\hat{\mathbf{r}}$:

$$\mathbf{r} = \mathsf{E}\boldsymbol{\varphi}(t)y(t) = \begin{pmatrix} r_y(0) \\ \vdots \\ r_y(p_y) \end{pmatrix} \tag{20}$$

We now consider the asymptotic normalized covariance matrix $\mathbf{R}$ of $\hat{\mathbf{r}}$, (9). Using (18) and (20) we get

$$\begin{aligned} \mathbf{R} &= \lim_{N\to\infty} N\mathsf{E}[\hat{\mathbf{r}}\hat{\mathbf{r}}^T - \mathbf{r}\mathbf{r}^T] \\ &= \lim_{N\to\infty} N\mathsf{E}\left[\frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\boldsymbol{\varphi}(t)y(t)\boldsymbol{\varphi}^T(s)y(s) - \mathbf{r}\mathbf{r}^T\right] \end{aligned} \tag{21}$$

Using Assumption **A5** we can apply the general rule for the product of Gaussian variables

$$\mathsf{E}x_1x_2x_3x_4 = \mathsf{E}x_1x_2Ex_3x_4 + \mathsf{E}x_1x_3Ex_2x_4 + \mathsf{E}x_1x_4Ex_2x_3 \tag{22}$$

Then using the result (22) in (21) leads to

$$\mathbf{R} = \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\left\{\left[\mathsf{E}\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^T(s)\right][\mathsf{E}y(t)y(s)] + [\mathsf{E}\boldsymbol{\varphi}(t)y(s)]\left[\mathsf{E}\boldsymbol{\varphi}^T(s)y(t)\right]\right\} \tag{23}$$

Due to Assumption **A6** the covariance function $r_y(\tau)$ decays exponentially with τ. Therefore we can write

$$\| \lim_{N\to\infty}\frac{1}{N}\sum_{\tau=-N}^{N}|\tau|\mathbf{R}_{\boldsymbol{\varphi}}(\tau)r_y(\tau) \| \leq \lim_{N\to\infty}\frac{1}{N}\sum_{\tau=0}^{N}2\tau C\alpha^{\tau} = 0 \tag{24}$$

for some $|\alpha| < 1$. Using this result, we get

$$\begin{aligned}\mathbf{R} &= \lim_{N\to\infty} \frac{1}{N} \sum_{\tau=-N}^{N} (N - |\tau|) \left[\mathbf{R}_{\boldsymbol{\varphi}}(\tau) r_y(\tau) + \mathbf{r}_{\boldsymbol{\varphi} y}(\tau)\mathbf{r}^T_{\boldsymbol{\varphi} y}(-\tau)\right] \\ &= \sum_{\tau=-\infty}^{\infty} \left[\mathbf{R}_{\boldsymbol{\varphi}}(\tau) r_y(\tau) + \mathbf{r}_{\boldsymbol{\varphi} y}(\tau)\mathbf{r}^T_{\boldsymbol{\varphi} y}(-\tau)\right] \qquad (25)\end{aligned}$$

In order to proceed one will need a technique for computing sums of the form

$$\sum_{\tau=-\infty}^{\infty} r_y(\tau) r_y(\tau + k) \qquad (26)$$

where k is an arbitrary integer. Using Assumption **A6** it follows that the noise-free output is an ARMA process. More details are given in Section 5.

4 A mathematical result

Lemma 4.1. Consider the ARMA processes

$$\begin{aligned} x_1(t) &= \frac{B_1(q^{-1})}{A_1(q^{-1})} e(t) \qquad & x_2(t) &= \frac{B_2(q^{-1})}{A_2(q^{-1})} e(t) \\ x_3(t) &= \frac{C_1(q^{-1})}{D_1(q^{-1})} v(t) \qquad & x_4(t) &= \frac{C_2(q^{-1})}{D_2(q^{-1})} v(t) \end{aligned} \qquad (27)$$

where $e(t)$ and $v(t)$ are zero mean white noise sequences, possibly correlated, and of unit variances. Then it holds that

$$\begin{aligned}\sum_{\tau=-\infty}^{\infty} r_{x_1 x_2}(\tau) r_{x_3 x_4}(\tau) &= \mathsf{E} \left[\frac{B_2(q^{-1})}{A_2(q^{-1})} \frac{C_1(q^{-1})}{D_1(q^{-1})} e_0(t)\right] \\ &\quad \times \left[\frac{B_1(q^{-1})}{A_1(q^{-1})} \frac{C_2(q^{-1})}{D_2(q^{-1})} e_0(t)\right] \qquad (28)\end{aligned}$$

where $e_0(t)$ is white noise with unit variance.

Proof. The trick, which is given in [10], is to imbed the problem into something more general. For this purpose, set

$$\beta_k = \sum_{\tau=-\infty}^{\infty} r_{x_1 x_2}(\tau) r_{x_3 x_4}(\tau + k) \qquad (29)$$

for an arbitrary integer k. (We are interested to determine β_0, but will derive how to find β_k for any value of k.) Use now the following definition of a spectrum,

$$\phi(z) = \sum_{\tau=-\infty}^{\infty} r(\tau) z^{-\tau}. \tag{30}$$

Then it holds

$$r(0) = \frac{1}{2\pi i} \oint \phi(z) \frac{dz}{z} \tag{31}$$

where the integration is counterclockwise around the unit circle, see, for example, [5].

Now form

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} \beta_k z^{-k} &= \sum_{k=-\infty}^{\infty} \sum_{\tau=-\infty}^{\infty} r_{x_1 x_2}(\tau) r_{x_3 x_4}(\tau + k) z^{-k} \\
&= \sum_{\tau=-\infty}^{\infty} r_{x_1 x_2}(\tau) z^{\tau} \sum_{k=-\infty}^{\infty} r_{x_3 x_4}(\tau + k) z^{-(k+\tau)} \\
&= \sum_{\tau=-\infty}^{\infty} r_{x_1 x_2}(\tau) z^{\tau} \sum_{p=-\infty}^{\infty} r_{x_3 x_4}(p) z^{-p} \\
&= \phi_{x_1 x_2}(z^{-1}) \phi_{x_3 x_4}(z) \\
&= \phi_{x_2 x_1}(z) \phi_{x_3 x_4}(z) \\
&= \frac{B_2(z)}{A_2(z)} \frac{B_1(z^{-1})}{A_1(z^{-1})} \frac{C_1(z)}{D_1(z)} \frac{C_2(z^{-1})}{D_2(z^{-1})} \\
&= \frac{B_2(z)}{A_2(z)} \frac{C_1(z)}{D_1(z)} \times \frac{B_1(z^{-1})}{A_1(z^{-1})} \frac{C_2(z^{-1})}{D_2(z^{-1})},
\end{aligned}$$

from which the result (28) follows after invoking (30). □

5 The covariance matrix of $\hat{\mathbf{r}}_y$

Using Assumption **A6** it follows that the output signal can be written as

$$\begin{aligned}
y(t) &= y_0(t) + \tilde{y}(t) \\
&= \frac{B(q^{-1})}{A(q^{-1})} \frac{C(q^{-1})}{D(q^{-1})} e(t) + \tilde{y}(t)
\end{aligned} \tag{32}$$

We can now write the generic element of the upper left partition of $\mathbf{R}$ as follows, using (25), (32) and Assumption **A4**, ($\mu, \nu = 0, \ldots, p_y$):

$$\begin{aligned}
\mathbf{R}_{\mu\nu} &= \sum_{\tau=-\infty}^{\infty} \left[r_y(\tau+\mu-\nu) r_y(\tau) + r_y(\tau-\mu) r_y(\tau+\nu) \right] \\
&= \sum_{\tau=-\infty}^{\infty} \left[\{ r_{y_0}(\tau+\mu-\nu) + \lambda_y \delta_{\tau+\mu-\nu,0} \} \{ r_{y_0}(\tau) + \lambda_y \delta_{\tau,0} \} \right. \\
&\quad \left. + \{ r_{y_0}(\tau-\mu) + \lambda_y \delta_{\tau-\mu,0} \} \{ r_{y_0}(\tau+\nu) + \lambda_y \delta_{\tau+\nu,0} \} \right] \\
&= \lambda_y^2 \delta_{\mu,\nu} + \lambda_y \left[r_{y_0}(\mu-\nu) + r_{y_0}(\nu-\mu) + r_{y_0}(\mu+\nu) + r_{y_0}(-\nu-\mu) \right] \\
&\quad + \sum_{\tau=-\infty}^{\infty} \left[r_{y_0}(\tau+\mu-\nu) r_{y_0}(\tau) + r_{y_0}(\tau-\mu) r_{y_0}(\tau+\nu) \right] \qquad (33)
\end{aligned}$$

Introduce the notations

$$r_{y_0}(k) = \mathsf{E} \left[\frac{B(q^{-1})}{A(q^{-1})} \frac{C(q^{-1})}{D(q^{-1})} e(t+k) \right] \left[\frac{B(q^{-1})}{A(q^{-1})} \frac{C(q^{-1})}{D(q^{-1})} e(t) \right] \qquad (34)$$

$$\beta_k = \mathsf{E} \left[\frac{B^2(q^{-1})}{A^2(q^{-1})} \frac{C^2(q^{-1})}{D^2(q^{-1})} e(t+k) \right] \left[\frac{B^2(q^{-1})}{A^2(q^{-1})} \frac{C^2(q^{-1})}{D^2(q^{-1})} e(t) \right] \qquad (35)$$

Then it follows from Lemma 4.1 that the sought matrix element can be computed as

$$\mathbf{R}_{\mu\nu} = \lambda_y^2 \delta_{\mu,\nu} + 2\lambda_y \left[r_{y_0}(\mu-\nu) + r_{y_0}(\mu+\nu) \right] + \beta_{\mu-\nu} + \beta_{\mu+\nu} \qquad (36)$$

In order to compute $\mathbf{R}_{\mu\nu}$ for $0 \le \mu, \nu \le p_y$, one will need to compute $r_{y_0}(k)$ and $\beta(k)$ for $k = 0, \ldots, 2p_y$. This can be done in any standard way to compute the covariance function of an ARMA process, see [5] for examples.

6 The remaining matrix elements of $\mathbf{R}$

The full vector $\hat{\mathbf{r}}$ is partitioned according to (2). The normalized covariance matrix $\mathbf{R}$ has a corresponding partitioning as

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \mathbf{R}_{13} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \mathbf{R}_{23} \\ \mathbf{R}_{31} & \mathbf{R}_{32} & \mathbf{R}_{33} \end{pmatrix} \qquad (37)$$

The elements in $\mathbf{R}_{11}$ are given in Section 5. As $\mathbf{R}$ is symmetric, it remains to find the elements of the block matrices $\mathbf{R}_{12},\ \mathbf{R}_{13},\ \mathbf{R}_{22}, \mathbf{R}_{23}$ and $\mathbf{R}_{33}$.

The block $\mathbf{R}_{12}$ can be written as

$$\begin{aligned}\mathbf{R}_{12} &= \lim_{N\to\infty} N\mathsf{E}[\hat{\mathbf{r}}_y - \mathbf{r}_y][\hat{\mathbf{r}}_u - \mathbf{r}_u]^T \\ &= \lim_{N\to\infty} N\left[\frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\boldsymbol{\varphi}_y(t)y(t)\boldsymbol{\varphi}_u^T(s)u(s) - \mathbf{r}_y\mathbf{r}_u^T\right] \end{aligned} \tag{38}$$

where $\boldsymbol{\varphi}_y(t)$ is as given by (19) and

$$\boldsymbol{\varphi}_u(t) = \begin{pmatrix} u(t) & \dots & u(t-p_u) \end{pmatrix}^T \tag{39}$$

Using Assumptions **A5** and **A6**, and proceeding as in Section 3 one obtains

$$= \sum_{\tau=-\infty}^{\infty}\left[\mathbf{R}_{\boldsymbol{\varphi}_y\boldsymbol{\varphi}_u}(\tau)r_{yu}(\tau) + \mathbf{r}_{\boldsymbol{\varphi}_y u}(\tau)\mathbf{r}_{\boldsymbol{\varphi}_u y}^T(-\tau)\right] \tag{40}$$

compare (25). In contrast to the developments in Section 5, there will in this case be no contribution for the measurement noise variances, as $\tilde{u}(t)$ and $\tilde{y}(s)$ are assumed to be uncorrelated for all t and s. A generic element ($0 \le \mu \le p_y,\ 0 \le \nu \le p_u$) of the matrix $\mathbf{R}_{12}$ can be written as

$$(\mathbf{R}_{12})_{\mu\nu} = \sum_{\tau=-\infty}^{\infty}\left[r_{yu}(\tau-\mu+\nu)r_{yu}(\tau) + r_{yu}(\tau-\mu)r_{yu}(\tau+\nu)\right] \tag{41}$$

Invoking Lemma 4.1 and defining

$$\beta_k^{(2)} = \mathsf{E}\left[\frac{B(q^{-1})}{A(q^{-1})}\frac{C^2(q^{-1})}{D^2(q^{-1})}e(t+k)\right]\left[\frac{B(q^{-1})}{A(q^{-1})}\frac{C^2(q^{-1})}{D^2(q^{-1})}e(t)\right] \tag{42}$$

we can write, ($0 \le \mu \le p_y,\ 0 \le \nu \le p_u$)

$$(\mathbf{R}_{12})_{\mu\nu} = \beta_{-\mu+\nu}^{(2)} + \beta_{\mu+\nu}^{(2)} \tag{43}$$

In a similar fashion one obtains for the block $\mathbf{R}_{13}$

$$\begin{aligned}\mathbf{R}_{13} &= \lim_{N\to\infty} N\mathsf{E}[\hat{\mathbf{r}}_y - \mathbf{r}_y][\hat{\mathbf{r}}_{yu} - \mathbf{r}_{yu}]^T \\ &= \lim_{N\to\infty} N\left[\frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\boldsymbol{\varphi}_y(t)y(t)\boldsymbol{\varphi}_{yu}^T(s)y(s) - \mathbf{r}_y\mathbf{r}_{yu}^T\right] \end{aligned} \tag{44}$$

where this time $\boldsymbol{\varphi}_y(t)$ is given by (19) and it holds

$$\boldsymbol{\varphi}_{yu}(t) = \begin{pmatrix} u(t-p_1) & \dots & u(t-p_2) \end{pmatrix}^T \tag{45}$$

This leads, as above, to

$$\mathbf{R}_{13} = \sum_{\tau=-\infty}^{\infty} \left[\mathbf{R}_{\boldsymbol{\varphi}_y \boldsymbol{\varphi}_{yu}}(\tau) r_y(\tau) + \mathbf{r}_{\boldsymbol{\varphi}_y y}(\tau) \mathbf{r}^T_{\boldsymbol{\varphi}_{yu} y}(-\tau) \right] \tag{46}$$

compare to (25). The generic element of $\mathbf{R}_{13}$ ($0 \le \mu \le p_y,\ p_1 \le \nu \le p_2$) can be written as

$$\begin{aligned} (\mathbf{R}_{13})_{\mu\nu} &= \lambda_y r_{yu}(-\mu+\nu) + \lambda_y r_{yu}(\mu+\nu) \\ &\quad + \sum_{\tau=-\infty}^{\infty} \left[r_{yu}(\tau-\mu+\nu) r_{y_0}(\tau) \,+\, r_{yu}(\tau+\nu) r_{y_0}(\tau-\mu) \right] \\ &= \lambda_y \left[r_{yu}(-\mu+\nu) + r_{yu}(\mu+\nu) \right] + \beta^{(3)}_{\nu-\mu} + \beta^{(3)}_{\mu+\nu} \end{aligned} \tag{47}$$

where

$$\beta^{(3)}_k = \mathsf{E} \left[\frac{B^2(q^{-1})}{A^2(q^{-1})} \frac{C^2(q^{-1})}{D^2(q^{-1})} e(t+k) \right] \left[\frac{B(q^{-1})}{A(q^{-1})} \frac{C^2(q^{-1})}{D^2(q^{-1})} e(t) \right] \tag{48}$$

For the block $\mathbf{R}_{22}$ it holds

$$\begin{aligned} \mathbf{R}_{22} &= \lim_{N\to\infty} N \mathsf{E} [\hat{\mathbf{r}}_u - \mathbf{r}_u][\hat{\mathbf{r}}_u - \mathbf{r}_u]^T \\ &= \lim_{N\to\infty} N \left[\frac{1}{N^2} \sum_{t=1}^{N} \sum_{s=1}^{N} \boldsymbol{\varphi}_u(t) u(t) \boldsymbol{\varphi}_u^T(s) u(s) - \mathbf{r}_u \mathbf{r}_u^T \right] \end{aligned} \tag{49}$$

where $\boldsymbol{\varphi}_u(t)$ is as given by (39). Therefore

$$\mathbf{R}_{22} = \sum_{\tau=-\infty}^{\infty} \left[\mathbf{R}_{\boldsymbol{\varphi}_u}(\tau) r_u(\tau) + \mathbf{r}_{\boldsymbol{\varphi}_u u}(\tau) \mathbf{r}^T_{\boldsymbol{\varphi}_u u}(-\tau) \right] \tag{50}$$

The generic element ($0 \le \mu \le p_u,\ 0 \le \nu \le p_u$) of the matrix $\mathbf{R}_{22}$ becomes, cf. (36)

$$\begin{aligned} (\mathbf{R}_{22})_{\mu\nu} &= \sum_{\tau=-\infty}^{\infty} \left[(\lambda_u \delta_{\mu,\tau+\nu} + r_{u_0}(\tau+\mu-\nu)) \ (\lambda_u \delta_{\tau,0} + r_{u_0}(\tau)) \right. \\ &\quad \left. + \ (\lambda_u \delta_{\tau-\mu,0} + r_{u_0}(\tau-\mu)) \ (\lambda_u \delta_{-\tau-\nu,0} + r_{u_0}(-\tau-\nu)) \right] \\ &= \lambda_u^2 \delta_{\mu,\nu} + 2\lambda_u \left[r_{u_0}(\mu-\nu) + r_{u_0}(\mu+\nu) \right] \\ &\quad + \sum_{\tau=-\infty}^{\infty} \left[r_{u_0}(\tau+\mu-\nu) r_{u_0}(\tau) \ + r_{u_0}(\tau-\mu) r_{u_0}(-\tau-\nu) \right] \\ &= \lambda_u^2 \delta_{\mu,\nu} + 2\lambda_u \left[r_{u_0}(\mu-\nu) + r_{u_0}(\mu+\nu) \right] + \beta^{(4)}_{\mu-\nu} + \beta^{(4)}_{\mu+\nu} \end{aligned} \tag{51}$$

where

$$\beta_k^{(4)} = \mathsf{E}\left[\frac{C^2(q^{-1})}{D^2(q^{-1})}e(t+k)\right]\left[\frac{C^2(q^{-1})}{D^2(q^{-1})}e(t)\right] \tag{52}$$

The block $\mathbf{R}_{23}$ can be written as

$$\begin{aligned}\mathbf{R}_{23} &= \lim_{N\to\infty} N\mathsf{E}[\hat{\mathbf{r}}_u - \mathbf{r}_u][\hat{\mathbf{r}}_{yu} - \mathbf{r}_{yu}]^T \\ &= \lim_{N\to\infty} N\left[\frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\boldsymbol{\varphi}_u(t)u(t)\boldsymbol{\varphi}_{yu}^T(s)y(s) - \mathbf{r}_u\mathbf{r}_{yu}^T\right] \end{aligned}\tag{53}$$

where $\boldsymbol{\varphi}_u(t)$ is as given by (39) and $\boldsymbol{\varphi}_{yu}(t)$ is defined in (45). Using Assumptions **A5** and **A6**, and proceeding as in Section 4 one obtains

$$\mathbf{R}_{23} = \sum_{\tau=-\infty}^{\infty}\left[\mathbf{R}_{\boldsymbol{\varphi}_u\boldsymbol{\varphi}_{yu}}(\tau)r_{uy}(\tau) + \mathbf{r}_{\boldsymbol{\varphi}_u y}(\tau)\mathbf{r}_{\boldsymbol{\varphi}_{yu}u}^T(-\tau)\right] \tag{54}$$

A generic element ($0 \le \mu \le p_u,\ p_1 \le \nu \le p_2$) of the matrix $\mathbf{R}_{23}$ can be written as

$$(\mathbf{R}_{23})_{\mu\nu} = \sum_{\tau=-\infty}^{\infty}\left[r_u(\tau-\mu+\nu)r_{yu}(-\tau) + r_{yu}(-\tau+\mu)r_u(\tau+\nu)\right] \tag{55}$$

Invoking Lemma 4.1 and defining

$$\beta_k^{(5)} = \mathsf{E}\left[\frac{B(q^{-1})}{A(q^{-1})}\frac{C^2(q^{-1})}{D^2(q^{-1})}e(t+k)\right]\left[\frac{C^2(q^{-1})}{D^2(q^{-1})}e(t)\right] \tag{56}$$

we can write

$$\begin{aligned}(\mathbf{R}_{23})_{\mu\nu} &= \lambda_u\left[r_{yu}(\nu-\mu) + r_{yu}(\nu+\mu)\right] \\ &\quad + \sum_{\tau=-\infty}^{\infty}\left[r_{u_0}(\tau-\mu+\nu)r_{yu}(-\tau) + r_{yu}(-\tau+\mu)r_{u_0}(\tau+\nu)\right] \\ &= \lambda_u\left[r_{yu}(\nu-\mu) + r_{yu}(\nu+\mu)\right] + \beta_{-\mu+\nu}^{(5)} + \beta_{\mu+\nu}^{(5)}\end{aligned} \tag{57}$$

Finally, the block $\mathbf{R}_{33}$ can be written as

$$\begin{aligned}\mathbf{R}_{33} &= \lim_{N\to\infty} N\mathsf{E}[\hat{\mathbf{r}}_{yu} - \mathbf{r}_{yu}][\hat{\mathbf{r}}_{yu} - \mathbf{r}_{yu}]^T \\ &= \lim_{N\to\infty} N\left[\frac{1}{N^2}\sum_{t=1}^{N}\sum_{s=1}^{N}\boldsymbol{\varphi}_{yu}(t)y(t)\boldsymbol{\varphi}_{yu}^T(s)y(s) - \mathbf{r}_{yu}\mathbf{r}_{yu}^T\right]\end{aligned} \tag{58}$$

where $\boldsymbol{\varphi}_{yu}(t)$ is as given by (45). Proceeding as before one obtains

$$\mathbf{R}_{33} = \sum_{\tau=-\infty}^{\infty} \Big[\mathbf{R}_{\boldsymbol{\varphi}_{yu}}(\tau) r_y(\tau) + \mathbf{r}_{\boldsymbol{\varphi}_{yu} y}(\tau) \mathbf{r}^T_{\boldsymbol{\varphi}_{yu} y}(-\tau) \Big] \tag{59}$$

compare (25). A generic element ($p_1 \leq \mu \leq p_2,\ p_1 \leq \nu \leq p_2$) of the matrix $\mathbf{R}_{33}$ can be written as

$$\begin{aligned} (\mathbf{R}_{33})_{\mu\nu} &= \sum_{\tau=-\infty}^{\infty} [r_u(\tau-\mu+\nu) r_y(\tau) \;+\; r_{uy}(\tau-\mu) r_{yu}(\tau+\nu)] \\ &= \lambda_y \lambda_u \delta_{\mu,\nu} + \lambda_u r_{y_0}(\mu-\nu) + \lambda_y r_{u_0}(\nu-\mu) \\ &\quad + \sum_{\tau=-\infty}^{\infty} [r_{u_0}(\tau-\mu+\nu) r_{y_0}(\tau) \;+\; r_{uy}(\tau-\mu) r_{yu}(\tau+\nu)] \end{aligned} \tag{60}$$

Invoking Lemma 4.1 and defining

$$\beta_k^{(6)} = \mathsf{E} \left[\frac{B(q^{-1})}{A(q^{-1})} \frac{C^2(q^{-1})}{D^2(q^{-1})} e(t+k) \right] \left[\frac{B(q^{-1})}{A(q^{-1})} \frac{C^2(q^{-1})}{D^2(q^{-1})} e(t) \right] \tag{61}$$

$$\gamma_k^{(6)} = \mathsf{E} \left[\frac{B^2(q^{-1})}{A^2(q^{-1})} \frac{C^2(q^{-1})}{D^2(q^{-1})} e(t+k) \right] \left[\frac{C^2(q^{-1})}{D^2(q^{-1})} e(t) \right] \tag{62}$$

we can write

$$(\mathbf{R}_{33})_{\mu\nu} = \lambda_y \lambda_u \delta_{\mu,\nu} + \lambda_u r_{y_0}(\mu-\nu) + \lambda_y r_{u_0}(\nu-\mu) + \beta^{(6)}_{-\mu+\nu} + \gamma^{(6)}_{\mu+\nu} \tag{63}$$

7 Extensions

The results of Sections 5 and 6 can be somewhat extended.

- First consider the relaxation of Assumption A4 on the measurement noise, $\tilde{y}(t)$ and $\tilde{u}(t)$ to be white. If these noise sequences instead are autocorrelated, and modelled as ARMA processes, but independent of the noise-free input, one can proceed much along the lines already used. The difference would be that sums such as (26), will now contain more terms. Writing

$$y(t) = y_0(t) + \tilde{y}(t) \tag{64}$$

the proper expression for such sums change from

$$\sum_{\tau=-\infty}^{\infty} r_y(\tau) r_y(\tau+k) = \sum_{\tau=-\infty}^{\infty} \left[\left(r_{y_0}(\tau) + \lambda_y^2 \delta_{\tau,0} \right) \left(r_{y_0}(\tau+k) + \lambda_y^2 \delta_{\tau+k,0} \right) \right] \tag{65}$$

to the more general result

$$\sum_{\tau=-\infty}^{\infty} r_y(\tau) r_y(\tau+k) = \sum_{\tau=-\infty}^{\infty} \left[\left(r_{y_0}(\tau) + r_{\tilde{y}}(\tau) \right) \left(r_{y_0}(\tau+k) + r_{\tilde{y}}(\tau+k) \right) \right] \tag{66}$$

The different resulting sums can then be evaluated using Lemma 4.1, as before.

- Another extension is to relax Assumption A5 of Gaussian distributions. This assumption was the key to derive the important relation (22. In what follows we will instead use the fact that the signals depend linearly on noise sources.

We consider a generic element of the matrix block $\mathbf{R}_{11}$ in the nonGaussian case. Then it holds

$$(\mathbf{R}_{11})_{\mu\nu} = \lim_{N\to\infty} N \left[\frac{1}{N^2} \sum_{t=1}^{N} \sum_{s=1}^{N} y(t) y(t-\mu) y(s) y(s-\nu) - r_y(\mu) r_y(\nu) \right] \tag{67}$$

Write the output signal as a weighted sum of a white noise source as

$$y(t) = \sum_{i=0}^{\infty} h_i e(t-i) \tag{68}$$

where $e(t)$ is a white noise, of zero mean, and not necessarily Gaussian distibuted. Set further

$$\lambda^2 = \mathsf{E} e^2(t), \quad \mu = \mathsf{E} e^4(t) \tag{69}$$

For future use, introduce the convention that $h_i = 0$ for $i < 0$. This implies that we can let the sum in (68) start at $i = -\infty$. In the Gaussian case it holds $\mu = 3\lambda^4$.

Now using (67) and (68)

$$\begin{aligned} (\mathbf{R}_{11})_{\mu\nu} &= \lim_{N\to\infty} \left[\frac{1}{N} \mathsf{E} \sum_t \sum_s \sum_i \sum_j \sum_k \sum_\ell h_i e(t-i) h_j e(t-\mu-j) \right. \\ &\qquad \left. \times h_k e(s-k) h_\ell e(s-\nu-\ell) - N r_y(\mu) r_y(\nu) \right] \\ &= \lim_{N\to\infty} \left[\frac{1}{N} \sum_t \sum_s \sum_i \sum_j \sum_k \sum_\ell h_i h_j h_k h_\ell \right. \\ &\qquad \times \mathsf{E} \left[e(t-i) e(t-\mu-j) e(s-k) s(s-\nu-\ell) \right] \\ &\qquad \left. - N r_y(\mu) r_y(\nu) \right] \end{aligned} \tag{70}$$

Next consider the expectation of the product of the noise values at four time instants. As $e(t)$ has zero mean, the contribution can only be nonzero if the time arguments are either pairwise equal or all equal. Utilizing this principle, we have

$$
\begin{aligned}
& \mathsf{E}\left[e(t-i)e(t-\mu-j)e(s-k)s(s-\nu-\ell)\right] \\
= \; & \lambda^4 \delta_{t-i,t-\mu-j}\delta_{s-k,s-\nu-\ell} + \lambda^4 \delta_{t-i,s-k}\delta_{t-\mu-j,s-\nu-\ell} + \lambda^4 \delta_{t-i,s-\nu-\ell}\delta_{s-k,t-\mu-j} \\
& +(\mu - 3\lambda^4)\delta_{t-i,t-\mu-j}\delta_{s-k,s-\nu-\ell}\delta_{t-i,s-k}
\end{aligned} \tag{71}
$$

Therefore we have

$$
\begin{aligned}
& \frac{1}{N}\mathsf{E}\sum_t \sum_s \sum_i \sum_j \sum_k \sum_\ell h_i e(t-i) h_j e(t-\mu-j) \\
& \qquad \times h_k e(s-k) h_\ell e(s-\nu-\ell) \\
= \; & \frac{1}{N}\left(\sum_t \sum_j h_{\mu-j} h_j \lambda^2\right)\left(\sum_s \sum_\ell h_{\nu+\ell} h_\ell \lambda^2\right) \\
& + \frac{1}{N}\sum_t \sum_s \sum_k \sum_\ell \left[h_{t-s+k} h_{t-s-\mu+\nu+\ell} h_k h_\ell \lambda^4 \right. \\
& \left. + h_{t-s+\nu+\ell} h_{t-s+\mu+k} h_k h_\ell \lambda^4\right] \\
+ \; & \frac{1}{N}\sum_t \sum_s \sum_k \sum_\ell h_{t-s+\nu+\ell} h_{t-s+\nu+\ell-\mu} h_{\nu+\ell} h_\ell (\mu - 3\lambda^4)
\end{aligned} \tag{72}
$$

The first term in (72) matches precisely the last term in (70). The second and third terms in (72) gives precisely the total contribution in the Gaussian case, as detailed in Section 5, see (36). In the non-Gaussian case we have thus one additional term that can be written as, cf (24),

$$
\begin{aligned}
(\mathbf{R}_{11}^{\mathrm{NG}})_{\mu\nu} &= \lim_{N\to\infty} (\mu - 3\lambda^4) \frac{1}{N} \sum_{\tau=-N}^{N} (N - |\tau|) \sum_\ell h_{\tau+\nu+\ell} h_{\tau+\nu+\ell-\mu} h_{\nu+\ell} h_\ell \\
&= (\mu - 3\lambda^4) \sum_{\tau=-\infty}^{\infty} \sum_\ell h_{\tau+\nu+\ell} h_{\tau+\nu+\ell-\mu} h_{\nu+\ell} h_\ell \\
&= (\mu - 3\lambda^4) \sum_\ell h_{\nu+\ell} h_\ell \sum_{\tau=-\infty}^{\infty} h_{\tau+\nu+\ell} h_{\tau+\nu+\ell-\mu} \\
&= (\mu - 3\lambda^4) \sum_\ell h_{\nu+\ell} h_\ell \sum_k h_k h_{k-\mu} \\
&= (\mu - 3\lambda^4) \frac{r_y(\nu)}{\lambda^2} \frac{r_y(\mu)}{\lambda^2} \\
&= \frac{\mu - 3\lambda^4}{\lambda^4} r_y(\mu) r_y(\nu)
\end{aligned} \tag{73}
$$

The result (73) can be generalized for the whole matrix $\mathbf{R}$. Assuming that in the input model (14),

$$\mathsf{E}e(t) = 0, \;\; \mathsf{E}e^2(t) = \sigma^2, \;\; \mathsf{E}e^4(t) = \mu \tag{74}$$

the expressions for $\mathbf{R}$ will, when $e(t)$ is nonGaussian distributed, have an additional term

$$\mathbf{R}^{\text{NG}} = \frac{\mu - 3\sigma^4}{\sigma^4}\mathbf{r}\mathbf{r}^T \tag{75}$$

8 Numerical illustration

We now illustrate the derived results in some simple examples.

Example 8.1. In a first example, we illustrate $\mathbf{R}$ numerically. We considered a system with the parameters

$$\begin{aligned} &A(q^{-1}) = 1 - 0.8q^{-1}, \; B(q^{-1}) = 2.0q^{-1}, \; C(q^{-1}) = 1, \; D(q^{-1}) = 1, \\ &\lambda_y = 1, \; \lambda_u = 2, \; \lambda_e = 1, \\ &p_y = 3, \; p_u = 2, \; p_1 = -2, \; p_2 = 4 \end{aligned} \tag{76}$$

Compared to the case treated in Sections 5 - 6, the elements $r_y(0)$ and $r_u(0)$ were omitted.

We run $M = 1000$ Monte-Carlo simulations, each of length $N = 1000$. The matrix $\mathbf{R}$ was computed according to the expressions derived in Sections 5 - 6. Denoting the estimated $\mathbf{r}$ vector in realization j by $\hat{\mathbf{r}}_j$, we also computed an estimated covariance matrix as

$$\hat{\mathbf{R}} = \frac{1}{M}\sum_{j=1}^{M} \left(\hat{\mathbf{r}}_j - \mathbf{r}\right)\left(\hat{\mathbf{r}}_j - \mathbf{r}\right)^T \tag{77}$$

The diagonal elements of the matrices $\mathbf{R}$ and $\hat{\mathbf{R}}$ are compared below. There are in total $p_y + p_u - p_1 + p_2 + 1 = 12$ diagonal elements.

k	$10^{-3}\mathbf{R}_{k,k}$	$10^{-3}\hat{\mathbf{R}}_{k,k}$
1	1.1178	1.1359
2	1.0274	1.0459
3	0.9331	0.9331
4	0.0090	0.0089
5	0.0090	0.0088
6	0.0363	0.0334
7	0.0363	0.0341
8	0.0363	0.0345
9	0.0403	0.0391
10	0.0440	0.0431
11	0.0445	0.0425
12	0.0437	0.0428

The two matrices coincide well, with some small deviations due to both N and M being finite. □

We proceed with a few more examples. In these $\hat{\boldsymbol{\theta}}$ will be the least squares estimate, and it is certainly easy to derive the covariance matrix of $\hat{\boldsymbol{\theta}}$ directly, [10]. We rather choose to illustrate that the formalism given by (10) leads to the same result.

Example 8.2. Consider a first order autoregressive model

$$y(t) + ay(t-1) = e_0(t), \quad \mathsf{E}e_0^2(t) = \lambda^2 \tag{78}$$

The least square estimate of a is

$$\hat{a} = -\frac{\hat{r}_y(1)}{\hat{r}_y(0)} \tag{79}$$

It is well known that the asympotic normalized variance of the estimate is

$$\lim_{N\to\infty} N\mathsf{E}\left(\hat{a}-a\right)^2 = \lambda^2 \frac{1}{r_y(0)} = 1 - a^2 \tag{80}$$

Using the formalism of this report, we have in this example

$$\mathbf{r} = \begin{pmatrix} r_y(0) \\ r_y(1) \end{pmatrix}, \quad \mathbf{f}(\mathbf{r}) = -\frac{r_y(1)}{r_y(0)} \tag{81}$$

Hence,

$$\mathbf{f}_{\mathbf{r}} = \begin{pmatrix} \frac{r_y(1)}{r_y^2(0)} & -\frac{1}{r_y(0)} \end{pmatrix} = -\frac{1}{r_y(0)} \begin{pmatrix} a & 1 \end{pmatrix} \tag{82}$$

Following (36) we have

$$\mathbf{R} = \begin{pmatrix} 2\beta_0 & 2\beta_1 \\ 2\beta_1 & \beta_0 + \beta_2 \end{pmatrix} \tag{83}$$

The β coefficients, cf (35), are given by

$$\beta_k = \mathsf{E}\left[\frac{\lambda^2}{(1+aq^{-1})^2} e(t+k)\right]\left[\frac{\lambda^2}{(1+aq^{-1})^2} e(t)\right] \tag{84}$$

and can be found to be

$$\beta_0 = \frac{1+a^2}{(1-a^2)^3}\lambda^4 \tag{85}$$

$$\beta_1 = \frac{-2a}{(1-a^2)^3}\lambda^4 \tag{86}$$

$$\beta_2 = \frac{a^2(3-a^2)}{(1-a^2)^3}\lambda^4 \tag{87}$$

Applying (10) leads to

$$\begin{aligned} \lim_{N\to\infty} N\mathsf{E}\,(\hat{a}-a)^2 &= \frac{(1-a^2)^2}{\lambda^4} \begin{pmatrix} a & 1 \end{pmatrix} \frac{\lambda^4}{(1-a^2)^3} \\ &\quad \times \begin{pmatrix} 2+2a^2 & -4a \\ -4a & 1+4a^2-a^4 \end{pmatrix} \begin{pmatrix} a \\ 1 \end{pmatrix} \\ &= \frac{1}{1-a^2}\left[2a^2(1+a^2) + (1+4a^2-a^4) - 8a^2\right] \\ &= \frac{1}{1-a^2}\left[1-2a^2+a^4\right] = 1-a^2 \end{aligned} \tag{88}$$

which coincides with the previous finding, (80). □

Example 8.3. Consider the system

$$y(t) = b_1 u(t-1) + b_2 u(t-2) + e_0(t) \tag{89}$$

where $e_0(t)$ is white noise of variance λ^2. The least squares estimate of the parameters is

$$\begin{aligned} \hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} &= \begin{pmatrix} \hat{r}_u(0) & \hat{r}_u(1) \\ \hat{r}_u(1) & \hat{r}_u(0) \end{pmatrix}^{-1} \begin{pmatrix} \hat{r}_{yu}(1) \\ \hat{r}_{yu}(2) \end{pmatrix} \\ &= \frac{1}{\hat{r}_u^2(0) - \hat{r}_u^2(1)} \begin{pmatrix} \hat{r}_u(0)\hat{r}_{yu}(1) - \hat{r}_u(1)\hat{r}_{yu}(2) \\ -\hat{r}_u(1)\hat{r}_{yu}(1) + \hat{r}_u(0)\hat{r}_{yu}(2) \end{pmatrix} \end{aligned} \tag{90}$$

Using standard theory, [10], the asymptotic normalized covariance matrix of $\hat{\boldsymbol{\theta}}$ is found to be

$$\mathbf{C} = \frac{\lambda^2}{r_u^2(0) - r_u^2(1)} \begin{pmatrix} r_u(0) & -r_u(1) \\ -r_u(1) & r_u(0) \end{pmatrix} \tag{91}$$

In this example we have obviously

$$\mathbf{r} = \begin{pmatrix} r_u(0) \\ r_u(1) \\ r_{yu}(1) \\ r_{yuy}(2) \end{pmatrix} \tag{92}$$

The Jacobian $\mathbf{f_r}$ is found by straightforward differentiation:

$$\left[r_u^2(0) - r_u^2(1)\right]^2 \mathbf{f_r} = \left(r_u^2(0) - r_u^2(1)\right) \begin{pmatrix} r_{yu}(1) & -r_{yu}(2) & r_u(0) & -r_u(1) \\ r_{yu}(2) & -r_{yu}(1) & -r_u(1) & r_u(0) \end{pmatrix}$$

$$- \begin{pmatrix} r_u(0)r_{yu}(1) - r_u(1)r_{yu}(2) \\ -r_u(1)r_{yu}(1) + r_u(0)r_{yu}(2) \end{pmatrix} \begin{pmatrix} 2r_u(0) & -2r_u(1) & 0 & 0 \end{pmatrix} \tag{93}$$

As it holds

$$r_{yu}(1) = b_1 r_u(0) + b_2 r_u(1), \quad r_{yu}(2) = b_1 r_u(1) + b_2 r_u(0) \tag{94}$$

one gets after some simplifications and calculations

$$\begin{aligned} &\left(r_u^2(0) - r_u^2(1)\right) \mathbf{f_r} \\ &= \begin{pmatrix} -b_1 r_u(0) + b_2 r_u(1) & b_1 r_u(1) - b_2 r_u(0) & r_u(0) & -r_u(1) \\ b_1 r_u(1) - b_2 r_u(0) & -b_1 r_u(0) + b_2 r_u(1) & -r_u(1) & r_u(0) \end{pmatrix} \\ &= r_u(0) \begin{pmatrix} -b_1 & -b_2 & 1 & 0 \\ -b_2 & -b_1 & 0 & 1 \end{pmatrix} + r_u(1) \begin{pmatrix} b_2 & b_1 & 0 & -1 \\ b_1 & b_2 & -1 & 0 \end{pmatrix} \end{aligned} \tag{95}$$

To proceed, we need to determine the matrix $\mathbf{R}$. It holds in this example

$$\mathbf{R} = \begin{pmatrix} 2\beta_0^{(4)} & 2\beta_1^{(4)} & 2\beta_1^{(5)} & 2\beta_2^{(5)} \\ 2\beta_1^{(4)} & \beta_0^{(4)} + \beta_2^{(4)} & \beta_0^{(5)} + \beta_2^{(5)} & \beta_1^{(5)} + \beta_3^{(5)} \\ 2\beta_1^{(5)} & \beta_0^{(5)} + \beta_2^{(5)} & \lambda^2 r_u(0) + \beta_0^{(6)} + \gamma_2^{(6)} & \beta_1^{(6)} + \gamma_3^{(6)} \\ 2\beta_2^{(5)} & \beta_1^{(5)} + \beta_3^{(5)} & \beta_1^{(6)} + \gamma_3^{(6)} & \lambda^2 r_u(0) + \beta_0^{(6)} + \gamma_4^{(6)} \end{pmatrix} \tag{96}$$

The next step is to evaluate the different components of the matrix elements. Using the definitions, we find (below we drop in all the relations a common factor $r_u(0)$ for convenience)

$$
\begin{aligned}
\beta_0^{(4)} &= r_u(0), \;\; \beta_1^{(4)} = r_u(1), \;\; \beta_2^{(4)} = r_u(2) \\
\beta_k^{(5)} &= \mathsf{E}\left[(b_1 q^{-1} + b_2 q^{-2}) u(t+k)\right] [u(t)] = b_1 r_u(k-1) + b_2 r_u(k-2) \\
\beta_0^{(5)} &= b_1 r_u(1) + b_2 r_u(2) \\
\beta_1^{(5)} &= b_1 r_u(0) + b_2 r_u(1) \\
\beta_2^{(5)} &= b_1 r_u(1) + b_2 r_u(0) \\
\beta_3^{(5)} &= b_1 r_u(2) + b_2 r_u(1) \\
\beta_k^{(6)} &= \mathsf{E}\left[(b_1 q^{-1} + b_2 q^{-2}) u(t+k)\right] \mathsf{E}\left[(b_1 q^{-1} + b_2 q^{-2}) u(t)\right] \\
&= b_1^2 r_u(k) + b_2^2 r_u(k) + b_1 b_2 r_u(k+1) + b_1 b_2 r_u(k-1) \\
\beta_0^{(6)} &= (b_1^2 + b_2^2) r_u(0) + 2 b_1 b_2 r_u(1) \\
\beta_1^{(6)} &= (b_1^2 + b_2^2) r_u(1) + b_1 b_2 r_u(0) + b_1 b_2 r_u(2) \\
\gamma_k^{(6)} &= \mathsf{E}\left[(b_1 q^{-1} + b_2 q^{-2})^2 u(t+k)\right] [u(t)] \\
&= b_1^2 r_u(k-2) + 2 b_1 b_2 r_u(k-3) + b_2^2 r_u(k-4) \\
\gamma_2^{(6)} &= b_1^2 r_u(0) + 2 b_1 b_2 r_u(1) + b_2^2 r_u(2) \\
\gamma_3^{(6)} &= b_1^2 r_u(1) + 2 b_1 b_2 r_u(0) + b_2^2 r_u(1) \\
\gamma_4^{(6)} &= b_1^2 r_u(2) + 2 b_1 b_2 r_u(1) + b_2^2 r_u(0)
\end{aligned}
$$

We next find that the matrix $\mathbf{R}$ can be written as

$$
\begin{aligned}
\mathbf{R} &= \lambda^2 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & r_u(0) & r_u(1) \\ 0 & 0 & r_u(1) & r_u(0) \end{pmatrix} \\
&\quad + r_u(2) r_u(0) \begin{pmatrix} 0 \\ 1 \\ b_2 \\ b_1 \end{pmatrix} \begin{pmatrix} 0 & 1 & b_2 & b_1 \end{pmatrix} \\
&\quad + 2 r_u^2(0) \begin{pmatrix} 1 \\ 0 \\ b_1 \\ b_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & b_1 & b_2 \end{pmatrix} \\
&\quad + 2 r_u(1) r_u(0) \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ b_2 & b_1 \\ b_1 & b_2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & b_2 & b_1 \\ 1 & 0 & b_1 & b_2 \end{pmatrix} \qquad (97)
\end{aligned}
$$

When multiplying $\mathbf{R}$ from the left by $\mathbf{f_r}$, it turns out that all terms except the first in (97) cancel. Hence,

$$\begin{aligned}
\mathbf{f_r R f_r}^T &= \lambda^2 \mathbf{f_r} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & r_u(0) & r_u(1) \\ 0 & 0 & r_u(1) & r_u(0) \end{pmatrix} \mathbf{f_r}^T \\
&= \frac{\lambda^2}{r_u^2(0) - r_u^2(1)} \begin{pmatrix} r_u(0) & -r_u(1) \\ -r_u(1) & r_u(0) \end{pmatrix} \\
&= \lambda^2 \begin{pmatrix} r_u(0) & r_u(1) \\ r_u(1) & r_u(0) \end{pmatrix}^{-1} = \mathbf{C} \qquad (98)
\end{aligned}$$

which is the same result as obtained before, (91). □

References

[1] S. Beghelli, P. Castaldi, R. Guidorzi, and U. Soverini. A robust criterion for model selection in identification from noisy data. In *Proc. 9th International Conference on Systems Engineering*, pages 480–484, Las Vegas, Nevada, USA, 1993.

[2] R. Diversi, R. Guidorzi, and U. Soverini. A new criterion in EIV identification and filtering applications. In *Proc. 13th IFAC Symposium on System Identification*, pages 1993–1998, Rotterdam, The Netherlands, August 27-29 2003.

[3] M. Hong, T. Söderström, and W. X. Zheng. Accuray analysis of bias-eliminating least squares estimates for errors-in-variables systems. *Automatica*, 43(9):1590–1596, September 2007.

[4] M. Mossberg and T. Söderström. Continuous-time errors-in-variables system identification through covariance matching without input signal modeling. In *2009 American Control Conference*, St. Louis, Missouri, USA, June 10-12 2009. To appear.

[5] T Söderström. *Discrete-time Stochastic Systems - Estimation and Control, 2nd edition*. Springer-Verlag, London, UK, 2002.

[6] T. Söderström. Accuracy analysis of the Frisch estimates for identifying errors-in-variables systems. *IEEE Transactions on Automatic Control*, 52(6):985–997, June 2007.

[7] T. Söderström. Errors-in-variables methods in system identification. *Automatica*, 43(6):939–958, June 2007. Survey paper.

[8] T. Söderström et al. Accurary of parameter estimates using a covariance matching approach for errors-in-variables system identification. *To be decided*, 2009. To be submitted.

[9] T. Söderström, M. Mossberg, and M. Hong. Identifying errors-in-variables systems by using a covariance matching approach. In *SYSID 2009, IFAC 15th Symposium on System Identification*, Saint-Malo, France, July 06-08 2009. Submitted.

[10] T. Söderström and P. Stoica. *System Identification*. Prentice Hall International, Hemel Hempstead, UK, 1989.

[11] T. Söderström, M. Mossberg, and M. Hong. A covariance matching approach for identifying errors-in-variables systems. *Automatica*, 2008. Submitted.

[12] W. X. Zheng. Transfer function estimation form noisy input and output data. *International Journal of Adaptive Control and Signal Processing*, 12:365–380, 1998.

[13] W. X. Zheng. A bias correction method for identification of linear dynamic errors-in-variables models. *IEEE Transactions on Automatic Control*, 47(7):1142–1147, July 2002.