Statistical analysis of the Frisch scheme for identifying errors-in-variables systems

Torsten Söderström
Division of Systems and Control, Department of Information Technology,
Uppsala University, P O Box 337, SE-751 05 Uppsala, Sweden.
Email: ts@it.uu.se

January 24, 2006

Abstract

Several estimation methods have been proposed for identifying errors-in-variables systems, where both input and output measurements are corrupted by noise. One of the promising approaches is the so called Frisch scheme. This paper provides an accuracy analysis of the Frisch scheme applied to system identification. The estimates of the system parameters and the noise variances are shown to be asymptotically Gaussian distributed. An explicit expression for the covariance matrix of the asymptotic distribution is given as well. Numerical simulations support the theoretical results. A comparison with the Cramer-Rao lower bound is also given in examples, and it is shown that the Frisch scheme gives a performance close to the Cramer-Rao bound for large signal-to-noise ratios.

1 Introduction

Many different solutions have been presented for system identification of linear dynamic systems from noise–corrupted output measurements see, for example, [7], [11]. Estimation of the parameters for linear dynamic systems when also the input is affected by noise ('errors–in–variables' models) is recognized as a more difficult problem.

The class of scientific disciplines which makes use of such representations is very broad, as proved by the several applications collected in [13], [14], such as time series modelling, array signal processing for direction–of–arrival estimation, blind channel equalization, multivariate calibration in analytical chemistry, image processing, astronomical data reduction, etc. In case of static systems, errors–in–variables representations are closely related to other well–known topics such as *latent variables* models and *factor* models [4].

Some comparisons between different approaches for errors-in-variables modelling are given in [10] and references therein.

The so called Frisch scheme is one of the more interesting approaches for the errors-in-variables identification. It has its roots in [3], where a regression problem was treated. It has been proposed for identifying dynamic systems in [1] and was further elaborated in [2]. So far, theoretical analysis has been limited to consistency. The aim of this paper is to provide such an analysis concerning the accuracy of the estimates obtained using the Frisch scheme.

2 Problem statement and notional setup

2.1 Setup

As a typical model example, consider the system depicted in Figure 1 with noise-corrupted input and output measurements.

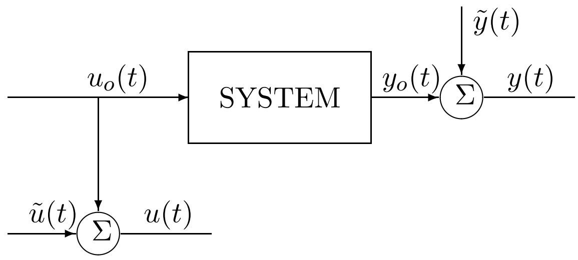

Figure 1: The basic setup for an error-in-variables problem.

The noise–free input is denoted by $u_o(t)$ and the undisturbed output by $y_o(t)$. They are linked through the linear difference equation

$$A(q^{-1})\, y_o(t) = B(q^{-1})\, u_o(t), \tag{1}$$

where $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the backward shift operator q^{-1}, *i.e.* $q^{-1}\, x(t) = x(t-1)$ *etc.* More precisely,

$$\begin{array}{rcl} A(q^{-1}) & = & 1 + a_1 q^{-1} + \cdots + a_{na} q^{-na} \\ B(q^{-1}) & = & b_1 q^{-1} + \cdots + b_{nb} q^{-nb} \end{array} . \tag{2}$$

We assume that the observations are corrupted by additive measurement noises $\tilde{u}(t)$ and $\tilde{y}(t)$. The available signals are of the form

$$\begin{array}{rcl} u(t) & = & u_o(t) + \tilde{u}(t) \\ y(t) & = & y_o(t) + \tilde{y}(t) \end{array} . \tag{3}$$

The general problem is to determine the system characteristics, *i.e.* the transfer function

$$G(q^{-1}) = \frac{B(q^{-1})}{A(q^{-1})}. \tag{4}$$

In other words, the estimation problem is as follows. Given the noisy input-output data $u(1), y(1), \ldots u(N), y(N)$, determine an estimate of the extended parameter vector

$$\vartheta = (a_1 \ldots a_{na}\; b_1 \ldots b_{nb}\; \lambda_y\; \lambda_u)^T , \tag{5}$$

where λ_y and λ_u denote the variances of $\tilde{y}(t)$ and $\tilde{u}(t)$, respectively.

There are several estimation methods that can produce an estimate of ϑ:

- The bias-eliminating least squares method (BELS) is described in [16], [15].
- The Frisch scheme goes back to [3]. Its application to system identification is presented in [1]. An alternative implementation was proposed in [2]. This is the method that will be considered in this paper.
- A prediction error method or a maximum likelihood method can be applied. Its use for the errors-in-variables problem is described in [8], [10].

Once an estimation method is specified, it is of interest to examine its statistical properties. In this paper we will focus on the asymptotic covariance matrix

$$P = \lim_{N\to\infty} \mathsf{E} N \mathrm{cov}(\hat{\vartheta} - \vartheta_o)(\hat{\vartheta} - \vartheta_o)^T \tag{6}$$

where $\hat{\vartheta}$ is the estimate of ϑ, and ϑ_o denotes the true value.

2.2 Assumptions

In order to proceed, some further assumptions must be introduced.

A1. The dynamic system (1) is asymptotically stable, *i.e.* $A(z)$ has all zeros outside the unit circle.
A2. All the system modes are observable and controllable, *i.e.* $A(z)$ and $B(z)$ have no common factors.
A3. The polynomial degrees na and nb are *a priori* known.
A4. The processes $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually uncorrelated, and uncorrelated with the noise–free signals $u_o(t)$ and $y_o(t)$.
A5. The sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are zero–mean white noise sequences with variances λ_u and λ_y, respectively.
A6. The true input $u_o(t)$ is a zero–mean stationary ergodic random signal, that is persistently exciting at least of order $na + nb$.

In case $u_o(t)$ has non-zero mean, one can use the deviations of all signals from the mean values. Hence, this assumption is not restrictive; rather if $\mathsf{E}u_o(t) \neq 0$, this fact may give additional information. For the Frisch scheme, it is essential that both the input and the output noise are uncorrelated in time.

2.3 Notations

The following notations will be convenient. The system parameter vector to be estimated is

$$\theta = (a_1 \dots a_{na}\ b_1 \dots b_{nb})^T. \tag{7}$$

Similarly we introduce the regressor vector

$$\varphi(t) = (-y(t-1) \dots - y(t-na)\ u(t-1) \dots u(t-nb))^T. \tag{8}$$

Further, we will use the conventions:

- θ_o denotes the true parameter vector, and $\hat{\theta}$ its estimate.
- Similarly, we let $A_o(q^{-1}),\ B_o(q^{-1}),\ \lambda_u^o,\ \lambda_y^o, \vartheta_o$ denote the true values of $A(q^{-1}),\ B(q^{-1}),\ \lambda_u,\ \lambda_y, \vartheta$, respectively.
- $\varphi_o(t)$ denotes the noise-free part of the regressor vector:

$$\varphi_o(t) = (-y_o(t-1) \dots - y_o(t-na)\ u_o(t-1) \dots u_o(t-nb))^T. \tag{9}$$

- $\tilde{\varphi}(t)$ denotes the noise-contribution to the regressor vector. This means that

$$\tilde{\varphi}(t) = (-\tilde{y}(t-1) \dots - \tilde{y}(t-na)\ \tilde{u}(t-1) \dots \tilde{u}(t-nb))^T. \tag{10}$$

Sometimes it is very convenient to add a leading element to θ and to φ. For this reason we also introduce the extended regressor vector as

$$\overline{\varphi}(t) = \begin{pmatrix} -y(t) \\ \varphi(t) \end{pmatrix}, \tag{11}$$

and the extended parameter vector

$$\overline{\theta} = \begin{pmatrix} 1 \\ \theta \end{pmatrix}. \tag{12}$$

At other times it is useful to work with partitioned parameter and regression vectors. For this reason we introduce also the partitions of θ and $\varphi(t)$ as

$$\theta = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \qquad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_{na} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_{nb} \end{pmatrix}, \tag{13}$$

and

$$\varphi(t) = \begin{pmatrix} \varphi_y(t) \\ \varphi_u(t) \end{pmatrix}, \qquad \varphi_y(t) = \begin{pmatrix} -y(t-1) \\ \vdots \\ -y(t-na) \end{pmatrix}, \qquad \varphi_u(t) = \begin{pmatrix} u(t-1) \\ \vdots \\ u(t-nb) \end{pmatrix}. \tag{14}$$

Extended versions of the partioned vectors will also be handy:

$$\overline{\theta} = \begin{pmatrix} \overline{\mathbf{a}} \\ \mathbf{b} \end{pmatrix}, \qquad \overline{\mathbf{a}} = \begin{pmatrix} 1 \\ \mathbf{a} \end{pmatrix}, \tag{15}$$

$$\overline{\varphi}(t) = \begin{pmatrix} \overline{\varphi}_y(t) \\ \varphi_u(t) \end{pmatrix}, \qquad \overline{\varphi}_y(t) = \begin{pmatrix} -y(t) \\ \varphi_y(t) \end{pmatrix}. \tag{16}$$

Cross-covariance matrices between two vectors $x(t)$ and $y(t)$ are denoted

$$R_{xy} = \mathsf{E} x(t) y^T(t), \tag{17}$$

and their natural estimates are denoted as

$$\hat{R}_{xy} = \frac{1}{N} \sum_{t=1}^{N} x(t) y^T(t). \tag{18}$$

The covariance matrices are often partioned in a way compatible with the partioning of the vectors. For example,

$$\hat{R}_{\overline{\varphi}} = \begin{pmatrix} \hat{R}_{\overline{\varphi}_y} & \hat{R}_{\overline{\varphi}_y \varphi_u} \\ \hat{R}_{\varphi_u \overline{\varphi}_y} & \hat{R}_{\varphi_u} \end{pmatrix}. \tag{19}$$

3 The Frisch scheme

3.1 Basic relations

The relations presented in this subsection are fundamental when deriving the Frisch estimator.

First we note that

$$\overline{\varphi}_o^T(t) \overline{\theta}_o = -A_o(q^{-1}) y_o(t) + B_o(q^{-1}) u_o(t) = 0. \tag{20}$$

Further it holds that

$$R_{\varphi} = R_{\varphi_o} + R_{\tilde{\varphi}}, \qquad R_{\overline{\varphi}} = R_{\overline{\varphi}_o} + R_{\tilde{\overline{\varphi}}}. \tag{21}$$

It follows from (20) that

$$R_{\overline{\varphi}_o} \overline{\theta}_o = \mathsf{E} \overline{\varphi}_o \overline{\varphi}_o^T \overline{\theta}_o = \mathbf{0}. \tag{22}$$

Hence the matrix $R_{\overline{\varphi}_o}$ is singular (positive semidefinite), with at least one eigenvalue equal to zero. The corresponding eigenvector is $\overline{\theta}_o$. One can show that under the general assumptions **A2** and **A6**, the matrix $R_{\overline{\varphi}_o}$ will in fact have only one eigenvalue in the origin.

The noise covariance matrix has a simple structure, as

$$R_{\tilde{\overline{\varphi}}} = \begin{pmatrix} \lambda_y I_{na+1} & \mathbf{0} \\ \mathbf{0} & \lambda_u I_{nb} \end{pmatrix}. \tag{23}$$

The relation (22) is the basis for the Frisch method. The idea is to have appropriate estimates of the noise variances and then determine the parameter vector θ from

$$\left(\hat{R}_{\overline{\varphi}} - \hat{R}_{\tilde{\overline{\varphi}}} \right) \hat{\overline{\theta}} = \mathbf{0}. \tag{24}$$

3.2 Determining $\hat{\lambda}_y$ and $\hat{\theta}$

Assume for the time being that an estimate $\hat{\lambda}_u$ of the input noise variance is available. Then the output noise variance λ_y is determined so that the matrix appearing in (24) is singular. More specifically, we have the following result.

Lemma 3.1. Let the estimate $\hat{\lambda}_u$ satisfy

$$0 \leq \hat{\lambda}_u \leq \lambda_{\min}\left(\hat{R}_{\varphi_u} - \hat{R}_{\varphi_u\overline{\varphi}_y}\hat{R}^{-1}_{\overline{\varphi}_y}\hat{R}_{\overline{\varphi}_y\varphi_u}\right), \tag{25}$$

where $\lambda_{\min}(R)$ denotes the minimal eigenvalue of the general symmetric matrix R.

Define

$$\hat{\lambda}_y = \lambda_{\min}\left(\hat{R}_{\overline{\varphi}_y} - \hat{R}_{\overline{\varphi}_y\varphi_u}\left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb}\right)^{-1}\hat{R}_{\varphi_u\overline{\varphi}_y}\right). \tag{26}$$

Then the matrix

$$C = \begin{pmatrix} \hat{R}_{\overline{\varphi}_y} & \hat{R}_{\overline{\varphi}_y\varphi_u} \\ \hat{R}_{\varphi_u\overline{\varphi}_y} & \hat{R}_{\varphi_u} \end{pmatrix} - \begin{pmatrix} \hat{\lambda}_y I_{na+1} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix} \tag{27}$$

is positive semidefinite with one eigenvalue in the origin.

Proof. See Appendix A.

An essential part of the Frisch algorithm is based on Lemma 3.1. Generically the matrix C will have a single eigenvalue in the origin. Assume that an estimate $\hat{\lambda}_u$ of the input noise variance is available (how this estimate is to be found will be described in the Section 3.3). The estimate $\hat{\lambda}_y$ is then found from (26). The estimate of the parameter vector θ is next determined by solving the last $na + nb$ equations of

$$C\hat{\overline{\theta}} = \mathbf{0}, \tag{28}$$

that is,

$$\left(\hat{R}_{\varphi} - \begin{pmatrix} \hat{\lambda}_y I_{na} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix}\right)\hat{\theta} = \hat{r}_{\varphi y}. \tag{29}$$

3.3 Determination of $\hat{\lambda}_u$

What remains is to determine $\hat{\lambda}_u$. Different alternatives have been proposed:

- In [1], the function $\hat{\lambda}_y(\hat{\lambda}_u)$ is evaluated both for the nominal model and for an extended model, adding one A or one B parameter (or both). The functions correspond to curves in the $(\hat{\lambda}_u, \hat{\lambda}_y)$ plan. Under quite mild identifiability conditions the curves will present a single common point, which defines the estimates.
- Another alternative is to compute residuals, and compare their statistical properties with what can be predicted from the model. This alternative was proposed in [2] and is the option analysed in this paper. It is described below.

Define the residuals

$$\varepsilon(t, \hat{\theta}) = \hat{A}(q^{-1})y(t) - \hat{B}(q^{-1})u(t) \tag{30}$$

and compute sample covariance elements

$$\hat{r}_\varepsilon(k) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t,\hat{\theta})\varepsilon(t+k,\hat{\theta}). \tag{31}$$

Compute also theoretical covariance elements $\hat{r}_{\varepsilon_o}(k)$ based on the model

$$\varepsilon_o(t) = \hat{A}(q^{-1})\hat{\tilde{y}}(t) - \hat{B}(q^{-1})\hat{\tilde{u}}(t), \tag{32}$$

where

$$\mathsf{E}\hat{\tilde{y}}^2(t) = \hat{\lambda}_y, \qquad \mathsf{E}\hat{\tilde{u}}^2(t) = \hat{\lambda}_u.$$

Next, define a criterion for comparing $\{\hat{r}_\varepsilon(k)\}$ and $\{\hat{r}_{\varepsilon_o}(k)\}$. A fairly general way to do this is to take

$$V_N(\hat{\lambda}_u) = \delta^T W \delta \tag{33}$$

where W is a user chosen, positive definite weighting matrix and the vector δ is

$$\delta = \begin{pmatrix} \hat{r}_\varepsilon(0) - \hat{r}_{\varepsilon_o}(0) \\ \vdots \\ \hat{r}_\varepsilon(m) - \hat{r}_{\varepsilon_o}(m) \end{pmatrix}. \tag{34}$$

The maximum lag m used in (34) is to be chosen by the user.

Such a form was first proposed in [2] although the description followed a slightly different form. Further, the weighting matrix inherent in [2] was

$$W = \begin{pmatrix} m+1 & 0 & \dots & & & \\ 0 & 2m & 0 & \dots & & \\ & 0 & 2(m-1) & 0 & & \\ & & 0 & \ddots & & \\ & & & 0 & 2 \end{pmatrix}. \tag{35}$$

The estimate $\hat{\lambda}_u$ is determined as the minimizing element of the criterion

$$\hat{\lambda}_u = \arg\min_{\lambda_u} V_N(\lambda_u). \tag{36}$$

We hence have

$$\left.\frac{d}{d\lambda_u} V_N(\lambda_u)\right|_{\lambda_u=\hat{\lambda}_u} = 0. \tag{37}$$

Note that in the asymptotic case ($N \to \infty$) all elements of the vector δ become zero for $\hat{\lambda}_u = \lambda_u$, which also lead to $\hat{\lambda}_y = \lambda_y$, $\hat{\theta} = \theta_o$. Hence this is a global minimum point of $V(\hat{\lambda}_u)$ when $N \to \infty$. In the practical use of the algorithm for finite N, the search for the mininum point has not shown to be problematic.

In summary the Frisch scheme algorithm consists of the equations (26), (29) and (37). In its implementation, there is an optimization over one variable, $\hat{\lambda}_u$, in (36). In the evaluation of the loss function $V_N(\hat{\lambda}_u)$, also (26) and (29) are used to get $\hat{\lambda}_y$ and $\hat{\theta}$, respectively.

It turns out that the first element of δ is always zero.

Lemma 3.2. It holds that

$$\hat{r}_\varepsilon(0) = \hat{r}_{\varepsilon_o}(0). \tag{38}$$

Proof. See Appendix A.

In what follows we will therefore exclude the element for time argument 0 in δ, and instead of (34) use

$$\delta = \begin{pmatrix} \hat{r}_\varepsilon(1) - \hat{r}_{\varepsilon_o}(1) \\ \vdots \\ \hat{r}_\varepsilon(m) - \hat{r}_{\varepsilon_o}(m) \end{pmatrix}. \tag{39}$$

4 Linearization

In Section 3 we have given the Frisch algorithm for estimating the parameter vector ϑ. Here we will examine how the estimate $\hat{\vartheta}$ deviates from the true value ϑ_o for large data sets (large N). The technique for doing so is to linearize the equations (26), (29) and (36) for large N. We assume that $\hat{\vartheta}$ is close to ϑ_o and seek relations of the type

$$A_\theta(\hat{\theta} - \theta_o) + A_y(\hat{\lambda}_y - \lambda_y^o) + A_u(\hat{\lambda}_u - \lambda_u^o) \approx A_s \tag{40}$$

where A_s is a random term, of zero mean and with a covariance matrix of order $O(1/N)$. The factors $A_\theta,\ A_y,\ A_u$ are deterministic terms obtained from the linearization. Their explicit forms will appear in the following. We shall linearize the three equations one by one in the subsequent subsections.

4.1 Linearization of (29)

We have the following result.

Lemma 4.1. Linearizing (29) leads to

$$\begin{aligned} R_{\varphi_o}(\hat{\theta} - \theta_o) - \begin{pmatrix} \mathbf{a}_o \\ \mathbf{0} \end{pmatrix} (\hat{\lambda}_y - \lambda_y^o) - \begin{pmatrix} \mathbf{0} \\ \mathbf{b}_o \end{pmatrix} (\hat{\lambda}_u - \lambda_u^o) \\ \approx \frac{1}{N} \sum_{t=1}^{N} \varphi(t)\varepsilon(t, \theta_o) + \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix}, \end{aligned} \tag{41}$$

where

$$\varepsilon(t, \theta_o) = A_o(q^{-1})\tilde{y}(t) - B_o(q^{-1})\tilde{u}(t). \tag{42}$$

Proof. See Appendix B.

Corollary. It follows that at the true parameter values the sensitivity derivatives are

$$\frac{d\hat{\theta}}{d\hat{\lambda}_y} = R_{\varphi_o}^{-1} \begin{pmatrix} \mathbf{a}_o \\ \mathbf{0} \end{pmatrix}, \tag{43}$$

$$\frac{d\hat{\theta}}{d\hat{\lambda}_u} = R_{\varphi_o}^{-1} \begin{pmatrix} \mathbf{0} \\ \mathbf{b}_o \end{pmatrix}. \tag{44}$$

4.2 Linearization of (26)

We have the following result.

Lemma 4.2. Linearizing (26) leads to

$$\bar{\mathbf{a}}_o^T \bar{\mathbf{a}}_o (\hat{\lambda}_y - \lambda_y^o) + \mathbf{b}_o^T \mathbf{b}_o (\hat{\lambda}_u - \lambda_u^o) \approx \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta_o) - \mathsf{E}\varepsilon^2(t, \theta_o) \tag{45}$$

where $\varepsilon(t, \theta_o)$ is as in (42).

Proof. See Appendix B.

Corollary. It follows that for the true parameter values the sensitivity derivative is at the true parameter values

$$\frac{d\hat{\lambda}_y}{d\hat{\lambda}_u} = -\frac{\mathbf{b}_o^T \mathbf{b}_o}{\bar{\mathbf{a}}_o^T \bar{\mathbf{a}}_o}. \tag{46}$$

4.3 Linearization of (37)

We now introduce the conventions

$$a_i^o = \begin{cases} 1 & i = 0 \\ 0 & i > na, \ i < 0 \end{cases}, \tag{47}$$

$$b_i^o = 0 \quad i > nb, \ i \leq 0, \tag{48}$$

which will be used in the sequel.

We then have the following result.

Lemma 4.3. Introduce the vector

$$\gamma = W\beta, \tag{49}$$

where the vector β is given elementwise as

$$\beta_k = -\sum_i b_i^o b_{i+k}^o + \sum_i a_i^o a_{i+k}^o \frac{\mathbf{b}_o^T \mathbf{b}_o}{\bar{\mathbf{a}}_o^T \bar{\mathbf{a}}_o}, \qquad k = 1, \ldots, m. \tag{50}$$

Then linearizing (37) gives

$$\sum_{k=1}^{m} \gamma_k \sum_i a_i^o a_{i+k}^o \left(\hat{\lambda}_y - \lambda_y^o\right) + \sum_{k=1}^{m} \gamma_k \sum_i b_i^o b_{i+k}^o \left(\hat{\lambda}_u - \lambda_u^o)\right)$$

$$\approx \sum_{k=1}^{m} \gamma_k \left[\frac{1}{N} \sum_{t=1}^{N} \varepsilon(t,\theta_o)\varepsilon(t+k,\theta_o) - \lambda_y^o \sum_i a_i^o a_{i+k}^o - \lambda_u^o \sum_i b_i^o b_{i+k}^o \right] \tag{51}$$

Proof. See Appendix B.

5 Asymptotic distribution

The main result is as follows.

Theorem 5.1. Assume that the white noise $\tilde{y}(t)$ has moments $\mathsf{E}\tilde{y}(t) = 0$, $\mathsf{E}\tilde{y}^2(t) = \lambda_y^o$ and $\mathsf{E}\tilde{y}^4(t) = \mu_y^o$, and similarly for $\tilde{u}(t)$: $\mathsf{E}\tilde{u}(t) = 0$, $\mathsf{E}\tilde{u}^2(t) = \lambda_u^o$ and $\mathsf{E}\tilde{u}^4(t) = \mu_u^o$. Under the given assumptions of Section 2.2 the parameter estimates $\hat{\vartheta}$ are asymptotically Gaussian distributed

$$\sqrt{N}(\hat{\vartheta} - \vartheta_o) \xrightarrow{\text{dist}} \mathcal{N}(0, P), \tag{52}$$

where

$$P = R^{-1} Q R^{-T} \tag{53}$$

and

$$R = \begin{pmatrix} R_{\varphi_o} & \begin{pmatrix} -\mathbf{a}_o \\ \mathbf{0} \end{pmatrix} & \begin{pmatrix} \mathbf{0} \\ -\mathbf{b}_o \end{pmatrix} \\ \mathbf{0} & \overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o & \mathbf{b}_o^T \mathbf{b}_o \\ \mathbf{0} & \sum_{k=1}^m \gamma_k \sum_i a_i^o a_{i+k}^o & \sum_{k=1}^m \gamma_k \sum_i b_i^o b_{i+k}^o \end{pmatrix}, \tag{54}$$

$$Q = Q^{\mathrm{G}} + Q^{\mathrm{NG}} \tag{55}$$

In the case of Gaussian measurement noise, only the term Q^{G} applies, and Q^{NG} vanishes. The two terms are given in partioned form as

$$Q^{\mathrm{G}} = \begin{pmatrix} Q_{11}^{\mathrm{G}} & Q_{12}^{\mathrm{G}} & Q_{13}^{\mathrm{G}} \\ Q_{21}^{\mathrm{G}} & Q_{22}^{\mathrm{G}} & Q_{23}^{\mathrm{G}} \\ Q_{31}^{\mathrm{G}} & Q_{32}^{\mathrm{G}} & Q_{33}^{\mathrm{G}} \end{pmatrix}, \quad Q^{\mathrm{NG}} = \begin{pmatrix} Q_{11}^{\mathrm{NG}} & Q_{12}^{\mathrm{NG}} & Q_{13}^{\mathrm{NG}} \\ Q_{21}^{\mathrm{NG}} & Q_{22}^{\mathrm{NG}} & Q_{23}^{\mathrm{NG}} \\ Q_{31}^{\mathrm{NG}} & Q_{32}^{\mathrm{NG}} & Q_{33}^{\mathrm{NG}} \end{pmatrix}, \tag{56}$$

The blocks of the symmetric matrix Q^{G} are as follows

$$Q_{11}^{\mathrm{G}} = \sum_\tau R_{\varphi_o}(\tau) r_\varepsilon(\tau) + \sum_\tau \left[R_{\tilde{\varphi}}(\tau) r_\varepsilon(\tau) + r_{\tilde{\varphi}\varepsilon}(\tau) r_{\tilde{\varphi}\varepsilon}^T(-\tau) \right], \tag{57}$$

$$Q_{12}^{\mathrm{G}} = 2 \sum_\tau r_{\tilde{\varphi}\varepsilon}(\tau) r_\varepsilon(\tau), \tag{58}$$

$$Q_{13}^{\mathrm{G}} = \sum_{k=1}^{m} \gamma_k \left\{ \sum_{\tau} \left[r_{\tilde{\varphi}\varepsilon}(\tau) r_\varepsilon(\tau-k) + r_{\tilde{\varphi}\varepsilon}(\tau-k) r_\varepsilon(\tau) \right] \right\}, \tag{59}$$

$$Q_{22}^{\mathrm{G}} = 2\sum_{\tau} r_\varepsilon^2(\tau), \tag{60}$$

$$Q_{23}^{\mathrm{G}} = 2\sum_{k=1}^{m} \gamma_k \sum_{\tau} \left[r_\varepsilon(\tau) r_\varepsilon(\tau-k) \right], \tag{61}$$

$$Q_{33}^{\mathrm{G}} = \sum_{k=1}^{m}\sum_{j=1}^{m} \gamma_k \gamma_j \sum_{\tau} \left[r_\varepsilon(\tau) r_\varepsilon(\tau+k-j) + r_\varepsilon(\tau-j) r_\varepsilon(\tau+k) \right]. \tag{62}$$

The covariance elements satisfy

$$r_\varepsilon(k) = \begin{cases} \lambda_y^o \sum_i a_i^o a_{i+k}^o + \lambda_u^o \sum_i b_i^o b_{i+k}^o & |k| \le \max(na, nb-1) \\ 0 & \text{elsewhere} \end{cases}, \tag{63}$$

$$r_{\tilde{\varphi}\varepsilon}(k) = -\begin{pmatrix} \lambda_y^o \begin{pmatrix} a_{1-k}^o \\ \vdots \\ a_{na-k}^o \end{pmatrix} \\ \lambda_u^o \begin{pmatrix} b_{1-k}^o \\ \vdots \\ b_{nb-k}^o \end{pmatrix} \end{pmatrix}. \tag{64}$$

Note that here the conventions (47) and (48) are applied. The summations over τ in (57) – (62) are over all values making the terms nonzero. Due to the condition in (63), each sum will have only a modest number of nonzero terms.

The blocks of the matrix Q^{NG} are as follows

$$Q_{11}^{\mathrm{NG}} = \begin{pmatrix} (\mu_y^o - 3(\lambda_y^o)^2)\mathbf{a}_o\mathbf{a}_o^\top & \mathbf{0} \\ \mathbf{0} & (\mu_u^o - 3(\lambda_u^o)^2)\mathbf{b}_o\mathbf{b}_o^\top \end{pmatrix}, \tag{65}$$

$$Q_{12}^{\mathrm{NG}} = \begin{pmatrix} -\mathbf{b}_o(\overline{\mathbf{a}}_o^\top\overline{\mathbf{a}}_o)(\mu_y^o - 3(\lambda_y^o)^2 \\ \mathbf{b}_o(\mathbf{b}_o^\top\mathbf{b}_o)(\mu_u^o - 3(\lambda_u^o)^2) \end{pmatrix}, \tag{66}$$

$$Q_{13}^{\mathrm{NG}} = \begin{pmatrix} -(\mu_y^o - 3(\lambda_y^o)^2)\mathbf{a}_o \sum_k \gamma_k \sum_j a_j^o a_{j+k}^o \\ (\mu_u^o - 3(\lambda_u^o)^2)\mathbf{b} \sum_k \gamma_k \sum_j a_j^o a_{j+k}^o \end{pmatrix}, \tag{67}$$

$$Q_{22}^{\mathrm{NG}} = (\mu_y^o - 3(\lambda_y^o)^2)(\overline{\mathbf{a}}_o^\top\overline{\mathbf{a}}_o)^2 + (\mu_u^o - 3(\lambda_u^o)^2)(\mathbf{b}_o^\top\mathbf{b}_o)^2, \tag{68}$$

$$\begin{aligned} Q_{23}^{\mathrm{NG}} &= (\mu_y^o - 3(\lambda_y^o)^2)\sum_k \gamma_k \sum_j a_j^o a_{j+k}^o (\overline{\mathbf{a}}_o^\top\overline{\mathbf{a}}_o) \\ &= (\mu_u^o - 3(\lambda_u^o)^2)\sum_k \gamma_k \sum_j b_j^o b_{j+k}^o (\mathbf{b}_o^\top\mathbf{b}_o), \end{aligned} \tag{69}$$

$$
\begin{aligned}
Q_{33}^{\rm NG} &= \sum_k \sum_j \gamma_k \gamma_j \Bigg[(\mu_y^o - 3(\lambda_y^o)^2) \left(\sum_i a_i^o a_{i+k}^o \right) \left(\sum_\ell a_\ell^o a_{\ell+j}^o \right) \\
&\quad + (\mu_u^o - 3(\lambda_u^o)^2) \left(\sum_i b_i^o b_{i+k}^o \right) \left(\sum_\ell b_\ell^o b_{\ell+j}^o \right) \Bigg]. \qquad (70)
\end{aligned}
$$

Proof: See Appendix C.

6 Numerical illustration

We first illustrate the findings in the paper by comparing the theoretical expressions of the covariance matrix with simulations. Next, we compare the theoretical expressions with the Cramer-Rao lower bound.

We will consider two systems.

System S1. This system is of first order and given by

$$
(1 - 0.8q^{-1}) y_o(t) = 2.0 q^{-1} u_o(t). \qquad (71)
$$

Further, the noise-free input is assumed to be an ARMA(1,1) process

$$
(1 - 0.5q^{-1}) u_o(t) = (1 + 0.7q^{-1}) v(t), \qquad \mathsf{E} v(t) v(s) = \lambda_v^o \delta_{t,s}. \qquad (72)
$$

The noise levels are

$$
\lambda_u^o = 1, \qquad \lambda_y^o = 1, \qquad \lambda_v^o = 1, \qquad (73)
$$

resulting in signal-to-noise ratios on the input and output sides, respectively

$$
\mathrm{SNR}_u = 5.82 \text{ dB}, \qquad \mathrm{SNR}_y = 10.55 \text{ dB}. \qquad (74)
$$

□

System S2. This system is of second order, given by

$$
(1 - 1.5q^{-1} + 0.7q^{-2}) y_o(t) = (2.0q^{-1} + 1.0q^{-2}) u_o(t), \qquad (75)
$$

and the noise-free input $u_o(t)$ is still assumed to be described by (72). The noise levels are

$$
\lambda_u^o = 1, \qquad \lambda_y^o = 4 \qquad (76)
$$

resulting in signal-to-noise ratios on the input and output sides, respectively

$$
\mathrm{SNR}_u = 5.82 \text{ dB}, \qquad \mathrm{SNR}_y = 14.40 \text{ dB}. \qquad (77)
$$

□

Example 1. We first compare the theoretical expression (53) for the covariance matrix with sample covariance matrices obtained from a Monte Carlo simulation using Gaussian distributed noise. We used $m = 1$ in the criterion (39), and hence W has no significance in this particular case. We considered System S2 and run 500 realizations each for different numbers of data points: $N = 200, 500, 1000, 3000, 10000$. In this way we can illustrate what is means that the covariance matrix of $\hat{\theta}$ *asymptotically* approaches P, (53). The results are displayed in Figure 2. There we display also the theoretical expressions given by (53). Due to the finite number of realizations, the estimated variances are random variables. As shown in Appendix B.9 of [11], one can expect that the estimated variances mostly lie within one standard deviation from the theoretical value. This interval is marked by dotted lines in Figure 2.

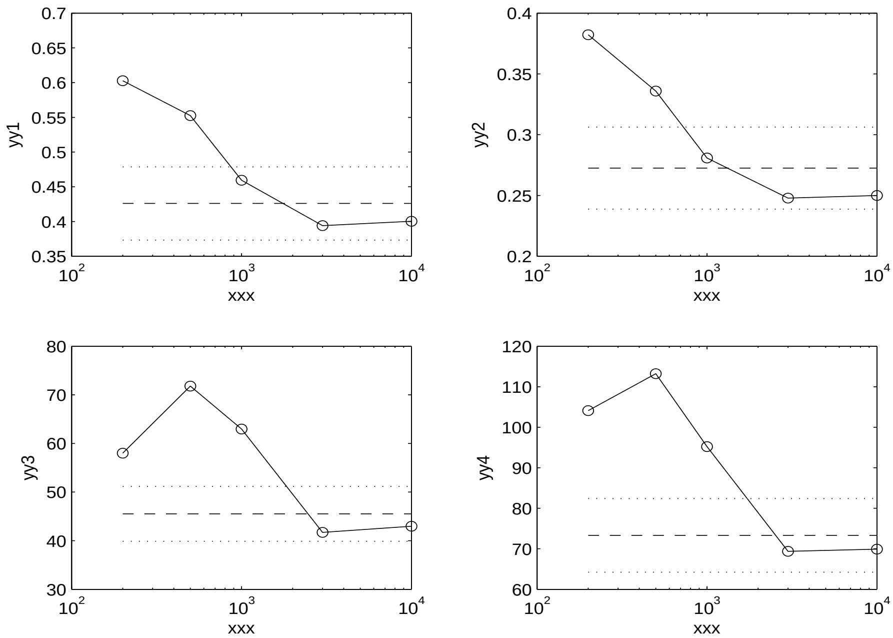

Figure 2: Normalized variances of the parameter estimates, $N\text{var}(\hat{\theta}_k)$ versus N for the parameters of System S2. Solid lines: Results from Monte Carlo simulations. Dashed lines: Theoretical results derived from (53). Dotted lines: Theoretical variance $\pm$ 1 standard deviation.

It can be seen in Figure 2 that the empirical variances fall into the expected interval when N is one or a few thousand. As the theoretical expresssion (53) anyway is asymptotic in N, the agreement between theory and simulation is very satisfactory. $\square$

Example 2. We next examined how the parameter variances vary with the signal-to-noise-ratio. More specifically, we varied $\lambda_v^o = \mathsf{E}v^2(t)$ in (72), and thereby the variance of the unperturbed input, $\mathsf{E}u_o^2(t)$. We show in Figures 3 and 4 how the diagonal elements of P varies with λ_v^o. For further comparisons we provide in these figures also the performance of some other methods:

- Instrumental variable (IV) method, with delayed inputs as instruments, [8]. This is a quite simple estimator which gives crude estimates with low accuracy.

- Bias-eliminating least squares (BELS), [16]. Similarly to the Frisch method, BELS is based on the modified normal equations (29), but the way of estimating the noise variances $\hat{\lambda}_u$ and $\hat{\lambda}_y$ is different. The accuracy analysis of BELS can be found in [5].
- A prediction error method (PEM), [7]. In the errors-in-variables context it was called the joint input-output method in [8].
- The Cramer-Rao bound gives a lower bound on P for any unbiased estimate. This bound is achieved (asymptotically when $N \to \infty$ for the maximum likelihood method. The Cramer-Rao bound can be computed using either a polynomial-based framework, see [6], or using a state-space based formalism, with details given in [9]. (Both approaches give identical results).

Note that both PEM and ML estimates are considerably more complex computationally than the Frisch method. The errors-in-variables context is one of the cases when PEM and ML give different results, see [9] for details.

The IV method applied to System S2 does not give identifiability. We therefore, in this example, studied a modified system, S2', where the right hand side polynomial of (72) is modified to $(1 + 0.7q^{-1} + 0.2q^{-2})$.

The numerical results are given in Figures 3 and 4.

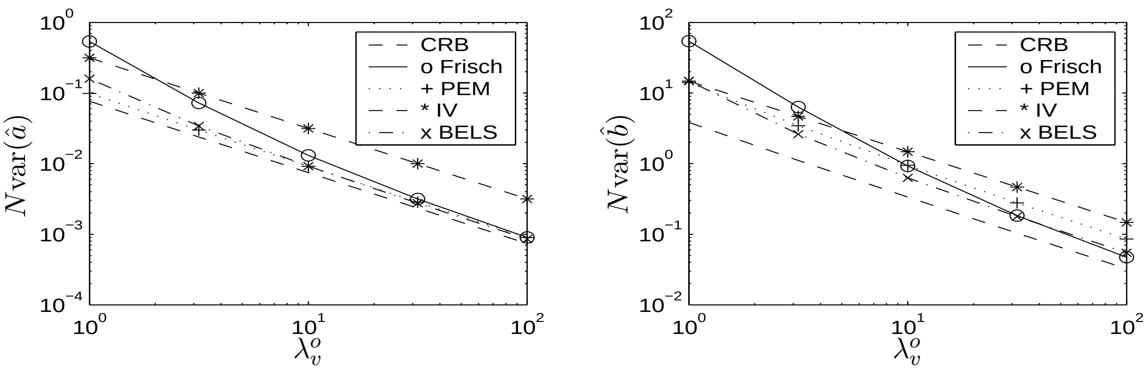

Figure 3: Normalized variances of $\hat{a}$ and $\hat{b}$, System S1: the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed). (The true variances are obtained after division by N.)

From Figures 3 and 4 we see that

- For all methods the accuracy improves (P decreases) when the signal-to-noise ratio increases. This is fairly natural.
- The IV method gives in most cases the by far worst performance.

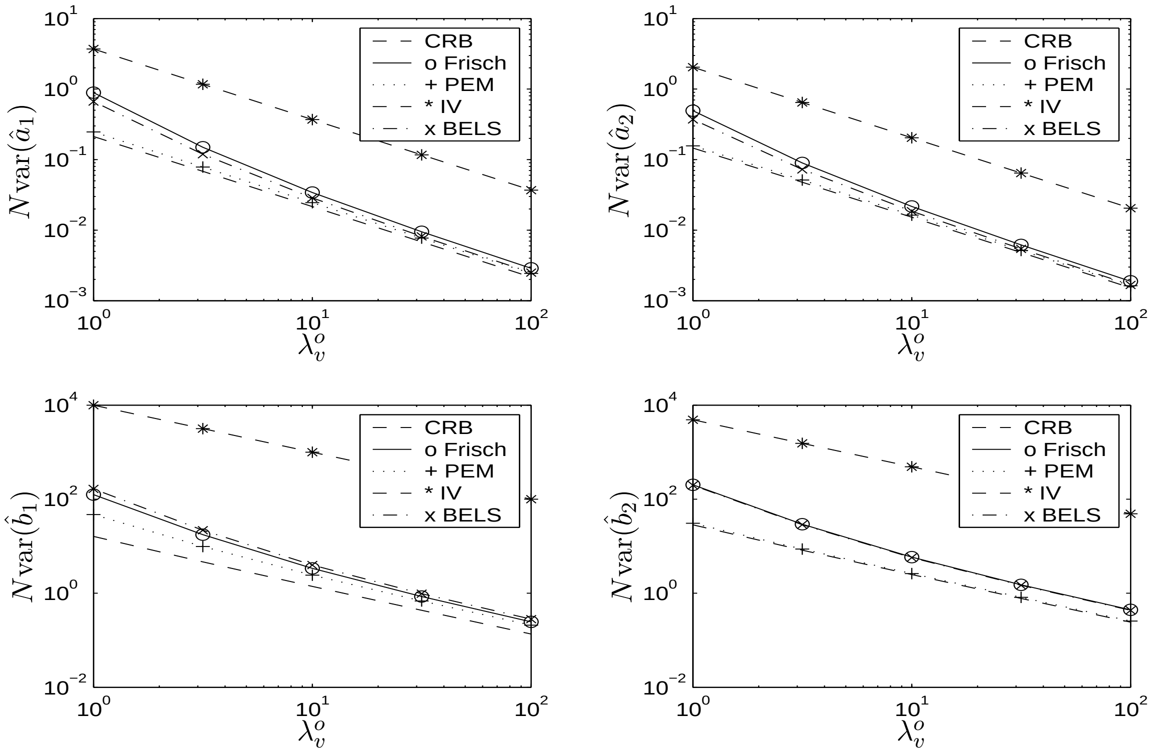

Figure 4: Normalized variances of $\hat{a}_1$, $\hat{a}_2$, $\hat{b}_1$ and $\hat{b}_2$, System S2': the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed). (The true variances are obtained after division by N.)

- The Frisch scheme and the BELS method have rather similar behavior. In comparison to IV, they give variances that are rather close to the Cramer-Rao bound.

□

Example 3. In this example we compare again the theoretical covariance expressions with the Cramer-Rao lower bound. Instead of comparing the accuracy of the parameter estimates, we compute the corresponding accuracy of the estimated transfer function $\hat{G}(e^{i\omega}) = \hat{B}(e^{i\omega})/\hat{A}(e^{i\omega})$ as a function of frequency ω.

For the two given systems S1 and S2 the covariance matrices P, given by (53) and the Cramer-Rao bound $P_{\rm CRB}$ were evaluated numerically. As an illustration, we show in Figures 5 and 6 how the true transfer function as well as its theoretical standard deviations for the Frisch scheme and the Cramer-Rao lower bound vary with frequency. More precisely, we plotted the *normalized* relative error

$$\frac{\sigma_G}{|G|} = \frac{\sqrt{\mathsf{E}|\Delta G|^2}}{|G|} \tag{78}$$

versus angular frequency ω. In (78) the error ΔG is defined as

$$\Delta G = G(e^{i\omega}, \hat{\theta}) - G(e^{i\omega}, \theta_o) \tag{79}$$

and is assumed to be small. When evaluating (78) the error ΔG is linearized around $\theta = \theta_o$ and is therefore a linear transformation of $\hat{\theta} - \theta_o$. The evaluation is based on the theoretical covariance matrix P, (53). Note that the relative error is obtained by dividing the expression (78) with $\sqrt{N}$. It is seen that the statistical error when the Frisch scheme is used is indeed larger than the Cramer-Rao lower bound, but the difference is rather small, in particular for low frequencies. □

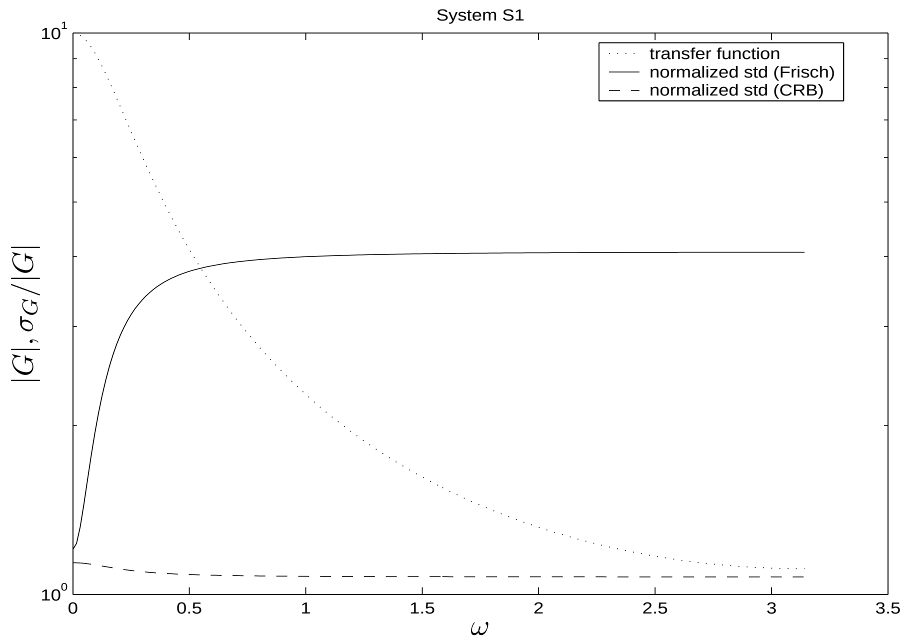

Figure 5: Frequency response of the transfer function (dotted) for System S1, standard deviation of the normalized relative error using the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed).

7 Conclusions

The Frisch approach for identifying error-in-variables systems has been analysed. In particular, the asymptotic covariance matrix of the parameter estimates has been derived. This matrix has also been compared to the Cramer-Rao lower bound, and it has been shown that the differences are small when the signal-to-noise ratio is high.

Acknowledgements

This work was supported by The Swedish Research Council under contract 621-2002-4671.

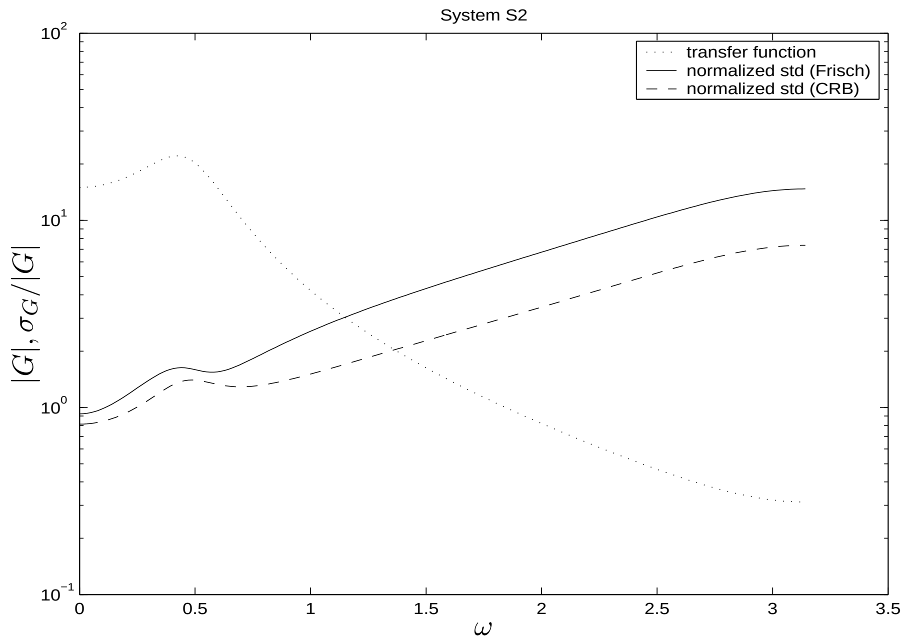

Figure 6: Frequency response of the transfer function (dotted) for System S2, standard deviation of the normalized relative error using the Frisch scheme (solid) and the Cramer-Rao lower bound (dashed).

References

[1] S. Beghelli, R.P. Guidorzi, and U. Soverini. The Frisch scheme in dynamic system identification. *Automatica*, 26:171–176, 1990.

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

[3] R. Frisch. Statistical confluence analysis by means of complete regression systems. Technical Report 5, University of Oslo, Economics Institute, Oslo, Norway, 1934.

[4] W. A. Fuller. *Measurement Error Models.* Wiley, New York, NY, 1987.

[5] M. Hong, T. Söderström, and W. X. Zheng. Accuracy analysis of bias-eliminating least squares estimates for errors-in-variables identification. In *14th IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

[6] E. Karlsson, T. Söderström, and P. Stoica. Computing the Cramér-Rao lower bound for noisy input-output systems. *Signal Processing*, 80(11):2421–2447, November 2000.

[7] L. Ljung. *System Identification - Theory for the User, 2nd edition.* Prentice Hall, Upper Saddle River, NJ, USA, 1999.

[8] T. Söderström. Identification of stochastic linear systems in presence of input noise. *Automatica*, 17:713–725, 1981.

[9] T. Söderström. On computing the Cramer-Rao bound and covariance matrices for PEM estimates in linear state space models. In *14th IFAC Symposium on System Identification*, Newcastle, Australia, March 29-31 2006.

[10] T. Söderström, U. Soverini, and K. Mahata. Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing*, 82(8):1139–1154, August 2002.

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

[12] G W Stewart. *Introduction to Matrix Computations*. Academic Press, New York, NY, USA, 1974.

[13] S. Van Huffel, editor. *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modelling*. SIAM, Philadelphia, USA, 1997.

[14] S. Van Huffel and Ph. Lemmerling, editors. *Total Least Squares and Errors-in-Variables Modelling. Analysis, Algorithms and Applications*. Kluwer, Dordrecht, The Netherlands, 2002.

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

[16] W. X. Zheng and C. B. Feng. Unbiased parameter estimation of linear systems in presence of input and output noise. *International Journal of Adaptive Control and Signal Processing*, 3:231–251, 1989.

A Proofs for Section 3

A.1 Proof of Lemma 3.1

Using Lemma A.3 in [11], the matrix C in (27) is positive semidefinite with one eigenvalue in the origin precisely when the matrix

$$\overline{C} = \hat{R}_{\overline{\varphi}_y} - \hat{\lambda}_y I_{na+1} - \hat{R}_{\overline{\varphi}_y \varphi_u} \left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb} \right)^{-1} \hat{R}_{\varphi_u \overline{\varphi}_y}$$

is positive semidefinite with one eigenvalue in the origin. However, by construction

$$\lambda_{\min}(\overline{C}) = -\hat{\lambda}_y + \lambda_{\min} \left(\hat{R}_{\overline{\varphi}_y} - \hat{R}_{\overline{\varphi}_y \varphi_u} \left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb} \right)^{-1} \hat{R}_{\varphi_u \overline{\varphi}_y} \right) = 0.$$

Hence, the smallest eigenvalue of $\overline{C}$ is in the origin. We need also to check that the variance estimate $\hat{\lambda}_y$ is nonnegative. However,

$$\begin{aligned}
&\hat{\lambda}_y \geq 0 \Leftrightarrow \\
&\hat{R}_{\overline{\varphi}_y} - \hat{R}_{\overline{\varphi}_y \varphi_u} \left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb} \right)^{-1} \hat{R}_{\varphi_u \overline{\varphi}_y} \text{ nonnegative definite } \Leftrightarrow \\
&\begin{pmatrix} \hat{R}_{\overline{\varphi}_y} & \hat{R}_{\overline{\varphi}_y \varphi_u} \\ \hat{R}_{\varphi_u \overline{\varphi}_y} & \left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb} \right) \end{pmatrix} \text{ nonnegative definite } \Leftrightarrow \\
&\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb} - \hat{R}_{\varphi_u \overline{\varphi}_y} \hat{R}_{\overline{\varphi}_y}^{-1} \hat{R}_{\overline{\varphi}_y \varphi_u} \text{ nonnegative definite,}
\end{aligned}$$

which is true due to the given bound (25) on $\hat{\lambda}_u$. This completes the proof.

A.2 Proof of Lemma 3.2

Straightforward calculations give

$$\hat{r}_{\varepsilon}(0) \quad = \quad \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \hat{\theta}) = \frac{1}{N} \sum \left[\overline{\varphi}^T \hat{\overline{\theta}} \right]^2 = \hat{\overline{\theta}}^T \hat{R}_{\overline{\varphi}} \hat{\overline{\theta}}$$

and

$$\hat{r}_{\varepsilon_o}(0) \quad = \quad \hat{\lambda}_y^2 \hat{\overline{\mathbf{a}}}^T \hat{\overline{\mathbf{a}}} + \hat{\lambda}_u^2 \hat{\mathbf{b}}^T \hat{\mathbf{b}} = \hat{\overline{\theta}}^T \begin{pmatrix} \hat{\lambda}_y I_{na+1} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix} \hat{\overline{\theta}}.$$

It now follows from (28) that

$$\hat{r}_{\varepsilon}(0) - \hat{r}_{\varepsilon_o}(0) = \hat{\overline{\theta}}^T \left(\hat{R}_{\overline{\varphi}} - \begin{pmatrix} \hat{\lambda}_y I_{na+1} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix} \right) \hat{\overline{\theta}} = 0.$$

This concludes the proof.

B Proofs for Section 4

B.1 Proof of Lemma 4.1

Starting with (29), straightforward calculations give

$$\begin{aligned}
\hat{\theta} - \theta_o &= \left(\hat{R}_\varphi - \begin{pmatrix} \hat{\lambda}_y I_{na} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix}\right)^{-1} \left(\hat{r}_{\varphi y} - \left(\hat{R}_\varphi - \begin{pmatrix} \hat{\lambda}_y I_{na} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{nb} \end{pmatrix}\right)\theta_o\right) \\
&\approx R_{\varphi_o}^{-1} \left(\hat{r}_{\varphi y} - \hat{R}_\varphi \theta_o + \begin{pmatrix} \hat{\lambda}_y \mathbf{a}_o \\ \hat{\lambda}_u \mathbf{b}_o \end{pmatrix}\right) \\
&= R_{\varphi_o}^{-1} \left(\frac{1}{N}\sum_{t=1}^{N} \varphi(t)[A_o(q^{-1})y(t) - B_o(q^{-1})u(t)] + \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix} + \begin{pmatrix} (\hat{\lambda}_y - \lambda_y^o)\mathbf{a}_o \\ (\hat{\lambda}_u - \lambda_u^o)\mathbf{b}_o \end{pmatrix}\right) \\
&= R_{\varphi_o}^{-1} \left(\frac{1}{N}\sum_{t=1}^{N} \varphi(t)[A_o(q^{-1})\tilde{y}(t) - B_o(q^{-1})\tilde{u}(t)] + \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix} + \begin{pmatrix} (\hat{\lambda}_y - \lambda_y^o)\mathbf{a}_o \\ (\hat{\lambda}_u - \lambda_u^o)\mathbf{b}_o \end{pmatrix}\right)
\end{aligned}$$

which is easily rewritten as (41).

B.2 Proof of Lemma 4.2

To perform the linearization we will utilize the following result. Let C and $\tilde{C}$ be some symmetric matrices, and assume that C is positive semidefinite. Let C have one distinct smallest eigenvalue $\lambda_{\min}(C)$, and the corresponding eigenvector be x. We seek an expression of how the smallest eigenvalue is changed when the small term $\varepsilon\tilde{C}$ is added to C, and will look for modifications that are linear in the small number ε. Then, see [12],

$$\lambda_{\min}(C + \varepsilon\tilde{C}) = \lambda_{\min}(C) + \varepsilon\frac{x^T\tilde{C}x}{x^Tx} + O(\varepsilon^2). \tag{80}$$

From (27) and (28) we have for the case of exact covariance matrices and exact noise variances, cf (22),

$$\begin{pmatrix} R_{\overline{\varphi}_y} - \lambda_y^o I_{na+1} & -R_{\overline{\varphi}_y\varphi_u} \\ -R_{\varphi_u\overline{\varphi}_y} & R_{\varphi_u} - \lambda_u^o I_{nb} \end{pmatrix} \begin{pmatrix} \overline{\mathbf{a}}_o \\ \mathbf{b}_o \end{pmatrix} = 0. \tag{81}$$

From the lower part of (81) we get

$$\mathbf{b}_o = (R_{\varphi_u} - \lambda_u^o I_{nb})^{-1} R_{\varphi_u\overline{\varphi}_y}\overline{\mathbf{a}}_o. \tag{82}$$

Inserting this in the upper part of (81) then leads to

$$\left(R_{\overline{\varphi}_y} - \lambda_y^o I_{na+1} - R_{\overline{\varphi}_y\varphi_u}(R_{\varphi_u} - \lambda_u^o I_{nb})^{-1} R_{\varphi_u\overline{\varphi}_y}\right)\overline{\mathbf{a}}_o = 0. \tag{83}$$

Now applying the result (80) to the definition (26) of $\hat{\lambda}_y$ gives with

$$\begin{aligned}
C &= R_{\overline{\varphi}_y} - R_{\overline{\varphi}_y\varphi_u}(R_{\varphi_u} - \lambda_u^o I_{nb})^{-1} R_{\varphi_u\overline{\varphi}_y} \\
\tilde{C} &= \hat{R}_{\overline{\varphi}_y} - \hat{R}_{\overline{\varphi}_y\varphi_u}\left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb}\right)^{-1} \hat{R}_{\varphi_u\overline{\varphi}_y} - R_{\overline{\varphi}_y} + R_{\overline{\varphi}_y\varphi_u}(R_{\varphi_u} - \lambda_u^o I_{nb})^{-1} R_{\varphi_u\overline{\varphi}_y}
\end{aligned}$$

neglecting higher order terms

$$\begin{aligned}\hat{\lambda}_y - \lambda_y^o &\approx \frac{1}{\overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o} \overline{\mathbf{a}}_o^T \Big[\hat{R}_{\overline{\varphi}_y} - \hat{R}_{\overline{\varphi}_y \varphi_u} \left(\hat{R}_{\varphi_u} - \hat{\lambda}_u I_{nb} \right)^{-1} \hat{R}_{\varphi_u \overline{\varphi}_y} \qquad (84)\\ &\qquad -R_{\overline{\varphi}_y} + R_{\overline{\varphi}_y \varphi_u} \left(R_{\varphi_u} - \lambda_u^o I_{nb} \right)^{-1} R_{\varphi_u \overline{\varphi}_y} \Big] \overline{\mathbf{a}}_o \\ &= \frac{1}{\overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o} \overline{\mathbf{a}}_o^T \left[\hat{R}_{\overline{\varphi}_y} - R_{\overline{\varphi}_y} \right] \mathbf{a}_o - \frac{1}{\overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o} \overline{\mathbf{a}}_o^T X \overline{\mathbf{a}}_o, \qquad (85)\end{aligned}$$

where

$$\begin{aligned}X &= \left(\hat{R}_{\overline{\varphi}_y \varphi_u} - R_{\overline{\varphi}_y \varphi_u} \right) \left(R_{\varphi_u} - \lambda_u^o I_{nb} \right)^{-1} R_{\varphi_u \overline{\varphi}_y} \\ &\quad + R_{\overline{\varphi}_y \varphi_u} \left(R_{\varphi_u} - \lambda_u^o I_{nb} \right)^{-1} \left(\hat{R}_{\varphi_u \overline{\varphi}_y} - R_{\varphi_u \overline{\varphi}_y} \right) \\ &\quad - R_{\overline{\varphi}_y \varphi_u} \left(R_{\varphi_u} - \lambda_u^o I_{nb} \right)^{-1} \left(\hat{R}_{\varphi_u} - R_{\varphi_u} - (\hat{\lambda}_u - \lambda_u^o) I_{nb} \right) \left(R_{\varphi_u} - \lambda_u^o I_{nb} \right)^{-1} R_{\varphi_u \overline{\varphi}_y}.\end{aligned}$$

Applying (82) will now give

$$\begin{aligned}\overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o (\hat{\lambda}_y - \lambda_y^o) &\approx \overline{\mathbf{a}}_o^T \left[\hat{R}_{\overline{\varphi}_y} - R_{\overline{\varphi}_y} \right] \mathbf{a}_o - \overline{\mathbf{a}}_o^T \left(\hat{R}_{\overline{\varphi}_y \varphi_u} - R_{\overline{\varphi}_y \varphi_u} \right) \mathbf{b}_o \\ &- \mathbf{b}_o^T \left(\hat{R}_{\varphi_u \overline{\varphi}_y} - R_{\varphi_u \overline{\varphi}_y} \right) \overline{\mathbf{a}}_o + \mathbf{b}_o^T \left[(\hat{R}_{\varphi_u} - R_{\varphi_u}) - (\hat{\lambda}_u - \lambda_u^o) I_{nb} \right] \mathbf{b}_o. \qquad (86)\end{aligned}$$

Next noting that

$$\overline{\varphi}_y^T(t) \overline{\mathbf{a}}_o = A_o(q^{-1}) y(t), \qquad \varphi_u^T(t) \mathbf{b}_o = B_o(q^{-1}) u(t) \qquad (87)$$

we find

$$\begin{aligned}\overline{\mathbf{a}}_o^T \overline{\mathbf{a}}_o (\hat{\lambda}_y - \lambda_y^o) &\approx -\mathbf{b}_o^T \mathbf{b}_o (\hat{\lambda}_u - \lambda_u^o) + \Bigg[\frac{1}{N} \sum_{t=1}^{N} [A_o(q^{-1}) y(t) - B_o(q^{-1}) u(t)]^2 \\ &\qquad - \mathsf{E}[A_o(q^{-1}) y(t) - B_o(q^{-1}) u(t)]^2 \Bigg] \\ &= -\mathbf{b}_o^T \mathbf{b}_o (\hat{\lambda}_u - \lambda_u^o) + \left[\frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t, \theta_o) - \mathsf{E}\varepsilon^2(t, \theta_o) \right]. \qquad (88)\end{aligned}$$

B.3 Proof of Lemma 4.3

When linearizing (37), it is useful to use the representation (33) of the loss function V_N. We can easily rewrite (37) as

$$\delta^T W \beta = 0 \qquad (89)$$

where $\beta = \delta'$ is the derivate of δ with respect to $\hat{\lambda}_u$.

To proceed we need to analyse the vectors δ and $\beta = \delta'$ in some details. Note that δ, but not β will be small. This will be taken into account when searching for the linearization.

For an arbitrary component of δ, (39), we have $(1 \leq k \leq m)$

$$\frac{d\delta_k}{d\hat{\lambda}_u} = \frac{\partial \delta_k}{\partial \hat{\lambda}_u} + \frac{\partial \delta_k}{\partial \hat{\lambda}_y}\frac{\partial \hat{\lambda}_y}{\partial \hat{\lambda}_u} + \frac{\partial \delta_k}{\partial \hat{\theta}}\left(\frac{\partial \hat{\theta}}{\partial \hat{\lambda}_u} + \frac{\partial \hat{\theta}}{\partial \hat{\lambda}_y}\frac{\partial \hat{\lambda}_y}{\partial \hat{\lambda}_u}\right). \tag{90}$$

We now apply the conventions (47) and (48). Invoking the derivatives (43), (44), (46) leads to (at the true parameters)

$$\frac{d\delta_k}{d\hat{\lambda}_u} = -\sum_i b_i^o b_{i+k}^o - \sum_i a_i^o a_{i+k}^o \frac{-\mathbf{b}_o^T \mathbf{b}_o}{\bar{\mathbf{a}}_o^T \bar{\mathbf{a}}_o} + \frac{\partial \delta_k}{\partial \theta} R_{\varphi_o}^{-1}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{b}_o \end{pmatrix} + \frac{-\mathbf{b}_o^T \mathbf{b}_o}{\bar{\mathbf{a}}_o^T \bar{\mathbf{a}}_o}\begin{pmatrix} \mathbf{a}_o \\ \mathbf{0} \end{pmatrix}\right). \tag{91}$$

Using the expression (42) for $\varepsilon(t, \theta_o)$, we find by direct differentiation

$$\begin{aligned}
\frac{\partial \delta_k}{\partial \mathbf{a}}_{|\hat{\vartheta}=\vartheta_o} &= -\frac{1}{N}\sum_{t=1}^{N} \varphi_y^T(t)\varepsilon(t+k, \theta_o) - \frac{1}{N}\sum_{t=1}^{N} \varepsilon(t, \theta_o)\varphi_y^T(t+k) \\
&\quad -\lambda_y^o \begin{pmatrix} a_{1+k}^o & a_{2+k}^o & \dots & a_{na+k}^o \end{pmatrix} - \lambda_y^o \begin{pmatrix} a_{1-k}^o & a_{2-k}^o & \dots & a_{na-k}^o \end{pmatrix} \\
&\to -\mathsf{E}\varphi_{\tilde{y}}^T(t) A_o(q^{-1})\tilde{y}(t+k) - \mathsf{E}A_o(q^{-1})\tilde{y}(t)\varphi_{\tilde{y}}^T(t+k) \\
&\quad -\lambda_y^o \begin{pmatrix} a_{1+k}^o & a_{2+k}^o & \dots & a_{na+k}^o \end{pmatrix} - \lambda_y^o \begin{pmatrix} a_{1-k}^o & a_{2-k}^o & \dots & a_{na-k}^o \end{pmatrix} \\
&= \mathbf{0}.
\end{aligned} \tag{92}$$

In the same fashion it can be proved that

$$\frac{\partial \delta_k}{\partial \mathbf{b}} = \mathbf{0}, \tag{93}$$

and hence

$$\frac{\partial \delta_k}{\partial \theta} = \mathbf{0}. \tag{94}$$

Thus (91) simplifies to (for the true parameter vector, and in the limiting case when $N \to \infty$)

$$\beta_k = \frac{d\delta_k}{d\hat{\lambda}_u} = -\sum_i b_i^o b_{i+k}^o + \sum_i a_i^o a_{i+k}^o \frac{\mathbf{b}_o^T \mathbf{b}_o}{\bar{\mathbf{a}}_o^T \bar{\mathbf{a}}_o}, \qquad k = 1, \dots, m \tag{95}$$

which is (50).

Next we evaluate an arbitrary element of the vector δ, again in the asymptotic case. It follows from (94) that there will be no terms proportional to $\hat{\theta} - \theta$.

$$\begin{aligned}
\delta_k &= \frac{1}{N}\sum_{t=1}^{N} [\hat{A}(q^{-1})y(t) - \hat{B}(q^{-1})u(t)][\hat{A}(q^{-1})y(t+k) - \hat{B}(q^{-1})u(t+k)] \\
&\quad -\hat{\lambda}_y \sum_i \hat{a}_i \hat{a}_{i+k} - \hat{\lambda}_u \sum_i \hat{b}_i \hat{b}_{i+k}
\end{aligned} \tag{96}$$

When simplifying this expression, we allow that terms that are one magnitude smaller may be added or subtracted. For example, we can write

$$\begin{aligned}(\hat{\lambda}_y - \lambda_y^o)\sum_i \hat{a}_i\hat{a}_{i+k} &= (\hat{\lambda}_y - \lambda_y^o)\sum_i a_i^o a_{i+k}^o + (\hat{\lambda}_y - \lambda_y^o)\left(\sum_i \hat{a}_i\hat{a}_{i+k} - \sum_i a_i^o a_{i+k}^o\right) \\ &\approx (\hat{\lambda}_y - \lambda_y^o)\sum_i a_i^o a_{i+k}^o \end{aligned} \tag{97}$$

Here we applied the conventions (47) and (48) also to $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. The reason why the second term can be dropped is that both $\hat{\lambda}_y - \lambda_y^o$ and $\left(\sum_i \hat{a}_i\hat{a}_{i+k} - \sum_i a_i^o a_{i+k}^o\right)$ converge to zero, so this term is negligible compared to the first term. Proceeding in this way, we have

$$\begin{aligned}\delta_k &= \frac{1}{N}\sum_{t=1}^{N}[\hat{A}(q^{-1})y(t) - \hat{B}(q^{-1})u(t)][\hat{A}(q^{-1})y(t+k) - \hat{B}(q^{-1})u(t+k)] \\ &\quad -(\hat{\lambda}_y - \lambda_y^o)\sum_i \hat{a}_i\hat{a}_{i+k} - (\hat{\lambda}_u - \lambda_u^o)\sum_i \hat{b}_i\hat{b}_{i+k} - \lambda_y^o\sum_i \hat{a}_i\hat{a}_{i+k} - \lambda_u^o\sum_i \hat{b}_i\hat{b}_{i+k} \\ &\approx -(\hat{\lambda}_y - \lambda_y^o)\sum_i a_i^o a_{i+k}^o - (\hat{\lambda}_u - \lambda_u^o)\sum_i b_i^o b_{i+k}^o \\ &\quad +\frac{1}{N}\sum_{t=1}^{N}[A_o(q^{-1})\tilde{y}(t) - B_o(q^{-1})\tilde{u}(t)][A_o(q^{-1})\tilde{y}(t+k) - B_o(q^{-1})\tilde{u}(t+k)] \\ &\quad -\lambda_y^o\sum_i a_i^o a_{i+k}^o - \lambda_u^o\sum_i b_i^o b_{i+k}^o \\ &= -(\hat{\lambda}_y - \lambda_y^o)\sum_i a_i^o a_{i+k}^o - (\hat{\lambda}_u - \lambda_u^o)\sum_i b_i^o b_{i+k}^o \\ &\quad +\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta_o)\varepsilon(t+k,\theta_o) - \lambda_y^o\sum_i a_i^o a_{i+k}^o - \lambda_u^o\sum_i b_i^o b_{i+k}^o. \end{aligned} \tag{98}$$

We next note that (89) can be written as

$$\gamma^T\delta = \sum_{k=1}^{m}\gamma_k\delta_k = 0. \tag{99}$$

Inserting the derived expression (98) for δ into (99) finally gives the result (51).

C Proof of Theorem 5.1

It follows directly from Lemmas 4.1, 4.2 and 4.3 that

$$R(\hat{\vartheta} - \vartheta_o) = \begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix} \tag{100}$$

where the stochastic components in the right hand side are

$$T_1 = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varepsilon(t,\theta_o) + \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix}, \tag{101}$$

$$T_2 = \frac{1}{N}\sum_{t=1}^{N}\varepsilon^2(t,\theta_o) - \mathsf{E}\varepsilon^2(t,\theta_o), \tag{102}$$

$$T_3 = \sum_{k=1}^{m}\gamma_k\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t,\theta_o)\varepsilon(t+k,\theta_o) - \lambda_y^o\sum_i a_i^o a_{i+k}^o - \lambda_u^o\sum_i b_i^o b_{i+k}^o\right]. \tag{103}$$

The asymptotic Gaussian distribution of the estimates then follows with standard arguments, see [7], [11]. What remains is find the asymptotic covariance matrix of the T_i variables, or more precisely to evaluate

$$Q = \begin{pmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \\ Q_{31} & Q_{32} & Q_{33} \end{pmatrix} = \lim_{N\to\infty} N\mathsf{E}\begin{pmatrix} T_1 \\ T_2 \\ T_3 \end{pmatrix}\begin{pmatrix} T_1 & T_2 & T_3 \end{pmatrix}. \tag{104}$$

We first make some general observations.

1. The T_i variables all have zero mean,

$$\mathsf{E}T_i = 0, \qquad i = 1,\ 2,\ 3.$$

2. To make the notations more compact, we will in the proof use the abbreviation

$$\varepsilon_t = \varepsilon(t,\theta_o) \tag{105}$$

3. The residual ε_t is an MA process. Its covariance function satisfies

$$r_\varepsilon(k) = \mathsf{E}\varepsilon_{t+k}\varepsilon_t = \begin{cases} \lambda_y^o\sum_i a_i^o a_{i+k}^o + \lambda_u^o\sum_i b_i^o b_{i+k}^o & |k| \le \max(na, nb-1) \\ 0 & \text{elsewhere} \end{cases} \tag{106}$$

which verifies (63).

4. The cross-covariance vector between the noise part of the regressor, $\tilde{\varphi}(t)$ and the residual $\varepsilon(t,\theta_o) = A_o(q^{-1})\tilde{y}(t) - B_o(q^{-1})\tilde{u}(t)$ can be evaluated as follows

$$r_{\tilde{\varphi}\varepsilon}(k) = \mathsf{E}\begin{pmatrix} -\tilde{y}(t+k-1) \\ \vdots \\ -\tilde{y}(t+k-na) \\ \tilde{u}(t+k-1) \\ \vdots \\ \tilde{u}(t+k-nb) \end{pmatrix}\big(A_o(q^{-1})\tilde{y}(t) - B_o(q^{-1})\tilde{u}(t)\big)$$

$$
\begin{aligned}
&= \mathsf{E} \begin{pmatrix} -\tilde{y}(t+k-1) \\ \vdots \\ -\tilde{y}(t+k-na) \\ \tilde{u}(t+k-1) \\ \vdots \\ \tilde{u}(t+k-nb) \end{pmatrix} \left(\sum_{j=0}^{na} a_j^o \tilde{y}(t-j) - \sum_{j=1}^{nb} b_j^o \tilde{u}(t-j) \right) \\
&= - \begin{pmatrix} \lambda_y^o \begin{pmatrix} a_{1-k}^o \\ \vdots \\ a_{na-k}^o \end{pmatrix} \\ \lambda_u^o \begin{pmatrix} b_{1-k}^o \\ \vdots \\ b_{nb-k}^o \end{pmatrix} \end{pmatrix}
\end{aligned} \tag{107}
$$

which verifies (64).

Note that the conventions (47) and (48) are used throughout in (106) and (107). It follows that $r_{\tilde{\varphi}\varepsilon}(k) = 0$ if $|k| \leq \max(na, nb-1)$ does not hold.

5. In the evaluation of the covariance elements, we will make use of the following property for jointly Gaussian distributed random variables:

$$
\mathsf{E} x_1 x_2 x_3 x_4 = (\mathsf{E} x_1 x_2)(\mathsf{E} x_3 x_4) + (\mathsf{E} x_1 x_3)(\mathsf{E} x_2 x_4) + (\mathsf{E} x_1 x_4)(\mathsf{E} x_2 x_3). \tag{108}
$$

With these preparations, we are ready to evaluate the covariance blocks Q_{jk}. In the calculations below, the summations over t and s goes generally from 1 to N. The summations over τ goes over values making the terms nonzero. This implies that

$$
|\tau| \leq \max(na, nb-1) \tag{109}
$$

holds. We will first treat the case of Gaussian distributed measurement noise.

Straightforward calculations give

$$
\begin{aligned}
Q_{11}^{\mathrm{G}} &= \lim_{N\to\infty} N \mathsf{E} \left[\frac{1}{N} \sum_t (\varphi_o(t) + \tilde{\varphi}(t)) \varepsilon_t + \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix} \right] \\
&\quad \times \left[\frac{1}{N} \sum_s (\varphi_o(s) + \tilde{\varphi}(s))^T \varepsilon_s + \begin{pmatrix} \lambda_y^o \mathbf{a}_o^T & \lambda_u^o \mathbf{b}_o^T \end{pmatrix} \right]
\end{aligned} \tag{110}
$$

As $\frac{1}{N} \sum_t (\varphi_o(t) + \tilde{\varphi}(t)) \varepsilon_t$ has mean value $-\begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix}$, it follows that

$$
Q_{11}^{\mathrm{G}} = \lim_{N\to\infty} \left\{ \mathsf{E} \frac{1}{N} \left[\sum_t (\varphi_o(t) + \tilde{\varphi}(t)) \varepsilon_t \right] \left[\sum_s (\varphi_o(s) + \tilde{\varphi}(s)) \varepsilon_s \right]^T - N \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix} \begin{pmatrix} \lambda_y^o \mathbf{a}_o^T & \lambda_u^o \mathbf{b}_o^T \end{pmatrix} \right\}. \tag{111}
$$

As $\varphi_o(t)$ is independent of both $\tilde{\varphi}(t')$ and $\varepsilon_{t'}$ for all t and t', it can be concluded that

$$Q_{11}^{\mathrm{G}} = \lim_{N\to\infty} \left[\mathsf{E}\frac{1}{N}\sum_t\sum_s \varphi_o(t)\varphi_o^T(s)\varepsilon_t\varepsilon_s + \mathsf{E}\frac{1}{N}\sum_t\sum_s \tilde{\varphi}(t)\tilde{\varphi}^T(s)\varepsilon_t\varepsilon_s - N\begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix} \begin{pmatrix} \lambda_y^o \mathbf{a}_o^T & \lambda_u^o \mathbf{b}_o^T \end{pmatrix} \right]. \tag{112}$$

We further get using (108)

$$\begin{aligned} \mathsf{E}\left[\frac{1}{N}\sum_t\sum_s \tilde{\varphi}(t)\tilde{\varphi}^T(s)\varepsilon_t\varepsilon_s\right] &= \frac{1}{N}\sum_t\sum_s R_{\tilde{\varphi}}(t-s)r_\varepsilon(t-s) + \frac{1}{N}\sum_t\sum_s \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix} \begin{pmatrix} \lambda_y^o \mathbf{a}_o^T & \lambda_u^o \mathbf{b}_o^T \end{pmatrix} \\ &\quad + \frac{1}{N}\sum_t\sum_s r_{\tilde{\varphi}\varepsilon}(t-s) r_{\tilde{\varphi}\varepsilon}^T(s-t). \end{aligned} \tag{113}$$

The first double sum in (112) can be evaluated by changing variables as

$$\begin{aligned} \lim_{N\to\infty}\left\{\frac{1}{N}\sum_t\sum_s R_{\varphi_o}(t-s)r_\varepsilon(t-s)\right\} &= \lim_{N\to\infty}\left\{\frac{1}{N}\sum_{\tau=-N+1}^{N-1}(N-|\tau|)R_{\varphi_o}(\tau)r_\varepsilon(\tau)\right\} \\ &= \lim_{N\to\infty}\left(\sum_{\tau=-N+1}^{N-1} R_{\varphi_o}(\tau)r_\varepsilon(\tau)\right) - \lim_{N\to\infty}\frac{1}{N}\sum_{\tau=-N+1}^{N-1}|\tau| R_{\varphi_o}(\tau)r_\varepsilon(\tau). \end{aligned} \tag{114}$$

Recall that the covariance functions in (114) decay exponentially to zero, as $|\tau| \to \infty$. The first term in (114) apparently converges to

$$\sum_{\tau=-\infty}^{\infty} R_{\varphi_o}(\tau)r_\varepsilon(\tau).$$

The second term in (114) is bounded with some $C > 0$ and some $\alpha,\ 0 < \alpha < 1$:

$$\left|\frac{1}{N}\sum_{\tau=-N+1}^{N-1}|\tau| R_{\varphi_o}(\tau)r_\varepsilon(\tau)\right| \le \left|\frac{1}{N}\sum_{\tau=-N+1}^{N-1}|\tau| C\alpha^{|\tau|}\right| \le \frac{2C}{N}\sum_{\tau=0}^{N}\tau\alpha^\tau \le \frac{2C}{N}\sum_{\tau=0}^{\infty}\tau\alpha^\tau \le \frac{2C\alpha}{N(1-\alpha)^2}. \tag{115}$$

Hence the magnitude of the second term will be arbitrarily small when $N \to \infty$, and it will converge to zero. Applying the above techniques the different double sums in Q_{11}^{G} gives finally

$$Q_{11}^{\mathrm{G}} = \sum_\tau R_{\varphi_o}(\tau)r_\varepsilon(\tau) + \sum_\tau \left[R_{\tilde{\varphi}}(\tau)r_\varepsilon(\tau) + r_{\tilde{\varphi}\varepsilon}(\tau)r_{\tilde{\varphi}\varepsilon}^T(-\tau)\right]. \tag{116}$$

Using the same techniques the remaining blocks of the Q matrix can also be evaluated. In brief, the derivations are as follows.

$$\begin{aligned} Q_{12}^{\mathrm{G}} &= \lim_{N\to\infty} N\mathsf{E}\left[\frac{1}{N}\sum_t(\varphi_o(t)+\tilde{\varphi}(t))\varepsilon_t - \begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix}\right]\left[\frac{1}{N}\sum_s \varepsilon_s^2 - \mathsf{E}\varepsilon_s^2\right] \\ &= \lim_{N\to\infty}\left[\mathsf{E}\frac{1}{N}\sum_t\sum_s \tilde{\varphi}(t)\varepsilon_t\varepsilon_s^2 - N\begin{pmatrix} \lambda_y^o \mathbf{a}_o \\ \lambda_u^o \mathbf{b}_o \end{pmatrix}\mathsf{E}\varepsilon_s^2\right] \\ &= 2\sum_\tau r_{\tilde{\varphi}\varepsilon}(\tau)r_\varepsilon(\tau), \end{aligned} \tag{117}$$

$$
\begin{aligned}
Q_{13}^{\mathrm{G}} &= \lim_{N\to\infty} N\mathsf{E}\left[\frac{1}{N}\sum_t(\varphi_o(t)+\tilde{\varphi}(t))\varepsilon_t - \begin{pmatrix}\lambda_y^o\mathbf{a}_o\\ \lambda_u^o\mathbf{b}_o\end{pmatrix}\right]\left[\sum_k \gamma_k\left\{\frac{1}{N}\sum_s \varepsilon_s\varepsilon_{s+k} - \mathsf{E}\varepsilon_s\varepsilon_{s+k}\right\}\right] \\
&= \sum_k \gamma_k \lim_{N\to\infty}\left\{\frac{1}{N}\sum_t\sum_s \mathsf{E}\tilde{\varphi}(t)\varepsilon_t\varepsilon_s\varepsilon_{s+k} - N\begin{pmatrix}\lambda_y^o\mathbf{a}_o\\ \lambda_u^o\mathbf{b}_o\end{pmatrix}\mathsf{E}\varepsilon_s\varepsilon_{s+k}\right\} \\
&= \sum_k \gamma_k\left\{\sum_\tau\left[r_{\tilde{\varphi}\varepsilon}(\tau)r_\varepsilon(\tau-k)+r_{\tilde{\varphi}\varepsilon}(\tau-k)r_\varepsilon(\tau)\right]\right\}, \qquad (118)\\
Q_{22}^{\mathrm{G}} &= \lim_{N\to\infty} N\mathsf{E}\left[\frac{1}{N}\sum_t \varepsilon_t^2 - \mathsf{E}\varepsilon_t^2\right]\left[\frac{1}{N}\sum_s \varepsilon_s^2 - \mathsf{E}\varepsilon_s^2\right] \\
&= \lim_{N\to\infty}\left[\mathsf{E}\frac{1}{N}\sum_t\sum_s \varepsilon_t^2\varepsilon_s^2 - N(\mathsf{E}\varepsilon_t^2)^2\right] \\
&= \lim_{N\to\infty}\frac{2}{N}\sum_t\sum_s[\mathsf{E}\varepsilon_t\varepsilon_s]^2 \\
&= 2\sum_\tau r_\varepsilon^2(\tau), \qquad (119)\\
Q_{23}^{\mathrm{G}} &= \lim_{N\to\infty} \mathsf{E}\left[\frac{1}{N}\sum_t \varepsilon_t^2 - \mathsf{E}\varepsilon_t^2\right]\left[\sum_k \gamma_k\left\{\frac{1}{N}\sum_s \varepsilon_s\varepsilon_{s+k} - \mathsf{E}\varepsilon_s\varepsilon_{s+k}\right\}\right] \\
&= \sum_k \gamma_k\left\{\lim_{N\to\infty}\left[\frac{1}{N}\sum_t\sum_s \mathsf{E}\varepsilon_t^2\varepsilon_s\varepsilon_{s+k} - N(\mathsf{E}\varepsilon_t^2)(\mathsf{E}\varepsilon_s\varepsilon_{s+k})\right]\right\} \\
&= 2\sum_k\gamma_k\sum_\tau\left[r_\varepsilon(\tau)r_\varepsilon(\tau-k)\right], \qquad (120)\\
Q_{33}^{\mathrm{G}} &= \lim_{N\to\infty} N\mathsf{E}\left[\sum_k \gamma_k\left\{\frac{1}{N}\sum_t \varepsilon_t\varepsilon_{t+k} - \mathsf{E}\varepsilon_t\varepsilon_{t+k}\right\}\right]\left[\sum_j \gamma_j\left\{\frac{1}{N}\sum_s \varepsilon_s\varepsilon_{s+j} - \mathsf{E}\varepsilon_s\varepsilon_{s+j}\right\}\right] \\
&= \sum_k\sum_j \gamma_k\gamma_j\left\{\lim_{N\to\infty}\left[\frac{1}{N}\sum_t\sum_s \mathsf{E}\varepsilon_t\varepsilon_{t+k}\varepsilon_s\varepsilon_{s+j} - N(\mathsf{E}\varepsilon_t\varepsilon_{t+k})(\mathsf{E}\varepsilon_s\varepsilon_{s+j})\right]\right\} \\
&= \sum_k\sum_j \gamma_k\gamma_j\left\{\lim_{N\to\infty}\frac{1}{N}\sum_t\sum_s\left[(\mathsf{E}\varepsilon_t\varepsilon_s)(\mathsf{E}\varepsilon_{t+k}\varepsilon_{s+j})+(\mathsf{E}\varepsilon_t\varepsilon_{s+j})(\mathsf{E}\varepsilon_s\varepsilon_{t+k})\right]\right\} \\
&= \sum_k\sum_j \gamma_k\gamma_j\sum_\tau\left[r_\varepsilon(\tau)r_\varepsilon(\tau+k-j)+r_\varepsilon(\tau-j)r_\varepsilon(\tau+k)\right]. \qquad (121)
\end{aligned}
$$

These calculations thus verify the expressions (57) - (62), and complete the proof so far.

Next we must also consider the case of non-Gaussian noise. In this case equation (108) cannot be used. Instead utilize here that all x_k terms are linear filters operating on a white noise source $e(t)$ (being either $\tilde{y}(t)$ or $\tilde{u}(t)$). Let the noise $e(t)$ have zero mean, variance λ and fourth moment μ. It holds that

$$x_k(t) = H_k(q^{-1})e(t), \qquad H_k(q^{-1}) = \sum_{j=0}^{\infty} h_{kj} q^{-j}, \qquad k = 1, 2, 3, 4. \tag{122}$$

Then it holds that

$$\mathsf{E} x_1(t)x_2(t)x_3(t)x_4(t) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} h_{1i}h_{2j}h_{3k}h_{4\ell}\mathsf{E} e(t-i)e(t-j)e(t-k)e(t-\ell). \tag{123}$$

As the white noise $e(t)$ has zero mean and is uncorrelated at different time points, the expectation in (123) is nonzero only when the time arguments are pairwise equal or all equal. Therefore

$$\begin{aligned} \mathsf{E} e(t-i)e(t-j)e(t-k)e(t-\ell) &= \lambda^2[\delta_{i,j}\delta_{k,\ell} + \delta_{i,k}\delta_{j,\ell} + \delta_{i,\ell}\delta_{j,k}] \\ &\quad +(\mu - 3\lambda^2)\delta_{i,j}\delta_{j,k}\delta_{k,\ell}. \end{aligned} \tag{124}$$

We now have

$$\begin{aligned} \mathsf{E} x_1(t)x_2(t)x_3(t)x_4(t) &= [\mathsf{E} x_1(t)x_2(t)][\mathsf{E} x_3(t)x_4(t)] + [\mathsf{E} x_1(t)x_3(t)][\mathsf{E} x_2(t)x_4(t)] \\ &\quad +[\mathsf{E} x_1(t)x_4(t)][\mathsf{E} x_2(t)x_3(t)] + (\mu - 3\lambda^2)\sum_{i=0}^{\infty} h_{1i}h_{2i}h_{3i}h_{4i}. \end{aligned} \tag{125}$$

Using the first part of (125) leads precisely to the Gaussian formula with Q^{G}.

We now focus on the the remaining terms (that vanishes in the Gaussian case). We write the additional matrix in block form as

$$Q^{\mathrm{NG}} = \begin{pmatrix} Q_{11}^{\mathrm{NG}} & Q_{12}^{\mathrm{NG}} & Q_{13}^{\mathrm{NG}} \\ Q_{21}^{\mathrm{NG}} & Q_{22}^{\mathrm{NG}} & Q_{23}^{\mathrm{NG}} \\ Q_{31}^{\mathrm{NG}} & Q_{32}^{\mathrm{NG}} & Q_{33}^{\mathrm{NG}} \end{pmatrix}. \tag{126}$$

In what follows we neglect the influence of the noisefree input $u_o(t)$, which anyway is captured in the term Q^{G}. Here we have, cf. (110)

$$\begin{aligned} (Q_{11})_{i,j=1:na} &= \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N} \mathsf{E}\varphi_{\tilde{y}}(t)A_o(q^{-1})\tilde{y}(t)\varphi_{\tilde{y}}^{\top}(s)A_o(q^{-1})\tilde{y}(s) \\ &= \sum_{\tau=-\infty}^{\infty} \mathsf{E}\varphi_{\tilde{y}}(t)A_o(q^{-1})\tilde{y}(t)\varphi_{\tilde{y}}^{\top}(t+\tau)A_o(q^{-1})\tilde{y}(t+\tau). \end{aligned} \tag{127}$$

Examining this componentwize gives for $i, j = 1, \ldots, na$,

$$(Q_{11}^{y})_{i,j} = \sum_{\tau=-\infty}^{\infty} \mathsf{E}\tilde{y}(t-i)A_o(q^{-1})\tilde{y}(t)\tilde{y}(t+\tau-j)A_o(q^{-1})\tilde{y}(t+\tau), \tag{128}$$

$$(Q_{11}^{\mathrm{NG}})_{i,j} = (\mu_y^o - 3(\lambda_y^o)^2)\sum_{\tau=-\infty}^{\infty} a_i^o\delta_{i,j-\tau}a_j^o = (\mu_y^o - 3(\lambda_y^o)^2)a_i^o a_j^o. \tag{129}$$

The elements $(Q_{11}^{\mathrm{NG}})_{i,j=na+1:na+nb}$ can be evaluated similarly. The result can be summarized as

$$Q_{11}^{\mathrm{NG}} = \begin{pmatrix} (\mu_y^o - 3(\lambda_y^o)^2)\mathbf{a}_o\mathbf{a}_o^\top & \mathbf{0} \\ \mathbf{0} & (\mu_u^o - 3(\lambda_u^o)^2)\mathbf{b}_o\mathbf{b}_o^\top \end{pmatrix} \tag{130}$$

Continuing with the other blocks leads to

$$(Q_{12})_{i=1:na} = \sum_{\tau=-\infty}^{\infty} \mathsf{E}\varphi_{\tilde{y}}(t+\tau)\varepsilon(t+\tau)\varepsilon^2(t), \tag{131}$$

$$\begin{aligned} (Q_{12})_i &= -\sum_{\tau=-\infty}^{\infty} \mathsf{E}\tilde{y}(t+\tau-i)a_i^o\tilde{y}(t+\tau-i)\varepsilon^2(t) \\ &= -\sum_{\tau=-\infty}^{\infty}\sum_j\sum_k \mathsf{E}\tilde{y}(t+\tau-i)a_i^o\tilde{y}(t+\tau-i)a_j^o\tilde{y}(t-j)a_k^o\tilde{y}(t-k) \\ &= -(\mu_y^o - 3(\lambda_y^o)^2)\sum_{\tau=-\infty}^{\infty} a_i^o(a_{i-\tau}^o)^2, \qquad i=1,\ldots,na. \end{aligned} \tag{132}$$

Similarly, one gets

$$(Q_{12})_{na+i} = (\mu_u^o - 3(\lambda_u^o)^2)\sum_{\tau=-\infty}^{\infty} b_i^o(b_{i-\tau}^o)^2, \qquad i=1,\ldots,nb. \tag{133}$$

Summarizing the findings in (132) and (133) gives

$$Q_{12}^{\mathrm{NG}} = \begin{pmatrix} -\mathbf{b}_o(\overline{\mathbf{a}}_o^\top\overline{\mathbf{a}}_o)(\mu_y^o - 3(\lambda_y^o)^2) \\ \mathbf{b}_o(\mathbf{b}_o^\top\mathbf{b}_o)(\mu_u^o - 3(\lambda_u^o)^2) \end{pmatrix}. \tag{134}$$

Next refer to (118) to find Q_{13}^{NG}. We first focus only on the part that is due to the output noise $\tilde{y}(t)$.

$$\begin{aligned} Q_{13} &= \sum_k \gamma_k \sum_{\tau=-\infty}^{\infty} \mathsf{E}\varphi_{\tilde{y}}(t+\tau)\varepsilon(t+\tau)\varepsilon(t)\varepsilon(t+k), \\ (Q_{13})_i &= -\sum_k \gamma_k \sum_{\tau=-\infty}^{\infty} \mathsf{E}\tilde{y}^2(t+\tau-i)a_i^o \sum_\mu a_\mu^o\tilde{y}(t-\mu)\sum_\nu a_\nu^o\tilde{y}(t+k-\nu), \\ (Q_{13}^{\mathrm{NG}})_i &= -(\mu_y^o - 3(\lambda_y^o)^2)\sum_k \gamma_k \sum_{\tau=-\infty}^{\infty} a_i^o a_{i-\tau}^o a_{i-\tau+k}^o, \qquad i=1,\ldots,na. \end{aligned} \tag{135}$$

We get similarly for the effect due to the input noise that

$$(Q_{13}^{\mathrm{NG}})_{na+i} = (\mu_u^o - 3(\lambda_u^o)^2)\sum_k \gamma_k \sum_{\tau=-\infty}^{\infty} b_i^o b_{i-\tau}^o b_{i-\tau+k}^o, \qquad i=1,\ldots,nb. \tag{136}$$

Summarizing (135) and (136) leads to

$$Q_{13}^{\rm NG} = \begin{pmatrix} -(\mu_y^o - 3(\lambda_y^o)^2)\mathbf{a}_o \sum_k \gamma_k \sum_j a_j^o a_{j+k}^o \\ (\mu_u^o - 3(\lambda_u^o)^2)\mathbf{b}_o \sum_k \gamma_k \sum_j a_j^o a_{j+k}^o \end{pmatrix}. \quad (137)$$

Proceeding to $Q_{22}^{\rm NG}$, we get, cf (119)

$$\begin{aligned} Q_{22}^{\rm NG} &= (\mu_y^o - 3(\lambda_y^o)^2)\left(\sum_i (a_i^o)^2\right)^2 + (\mu_u^o - 3(\lambda_u^o)^2)\left(\sum_i (b_i^o)^2\right)^2 \\ &= (\mu_y^o - 3(\lambda_y^o)^2)(\overline{\mathbf{a}}_o^\top \overline{\mathbf{a}}_o)^2 + (\mu_u^o - 3(\lambda_u^o)^2)(\mathbf{b}_o^\top \mathbf{b}_o)^2. \quad (138) \end{aligned}$$

From (120) one gets, treating first only the part depending on $\tilde{y}(t)$:

$$\begin{aligned} Q_{23} &= \mathsf{E}\sum_k \gamma_k \sum_{\tau=-\infty}^{\infty} \varepsilon^2(t+\tau)\varepsilon(t)\varepsilon(t+k) \\ &= \sum_k \gamma_k \sum_{\tau=-\infty}^{\infty} \sum_i \sum_j \sum_\mu \sum_\nu a_i^o a_j^o a_\mu^o a_\nu^o \mathsf{E}\tilde{y}(t+\tau-i)\tilde{y}(t+\tau-j)\tilde{y}(t-\mu)\tilde{y}(t+k-\nu). \quad (139) \end{aligned}$$

Hence

$$Q_{23}^{\rm NG} = (\mu_y^o - 3(\lambda_y^o)^2)\sum_k \gamma_k \sum_{\tau=-\infty}^{\infty} \sum_i (a_i^o)^2 a_{i-\tau}^o a_{k+i-\tau}^o. \quad (140)$$

Adding the contribution due to $\tilde{u}(t)$, and simplifying the expression leads to

$$\begin{aligned} Q_{23}^{\rm NG} &= (\mu_y^o - 3(\lambda_y^o)^2)\sum_k \gamma_k \sum_j a_j^o a_{j+k}^o (\overline{\mathbf{a}}_o^\top \overline{\mathbf{a}}_o) \\ &= (\mu_u^o - 3(\lambda_u^o)^2)\sum_k \gamma_k \sum_j b_j^o b_{j+k}^o (\mathbf{b}_o^\top \mathbf{b}_o). \quad (141) \end{aligned}$$

Finally, referring to (121) treating first only the part of $Q_{33}^{\rm NG}$ that depends on $\tilde{y}(t)$:

$$\begin{aligned} Q_{33} &= \sum_k \gamma_k \sum_j \gamma_j \sum_{\tau=-\infty}^{\infty} \mathsf{E}\varepsilon(t+\tau)\varepsilon(t+k+\tau)\varepsilon(t)\varepsilon(t+j) \\ &= \sum_k \gamma_k \sum_j \gamma_j \sum_{\tau=-\infty}^{\infty} \sum_i \sum_\ell \sum_\mu \sum_\nu a_i^o a_\ell^o a_\mu^o a_\nu^o \mathsf{E}\tilde{y}(t+\tau-i)\varepsilon(t+\tau+k-\ell)\varepsilon(t-\mu)\varepsilon(t+j-\nu), \quad (142) \end{aligned}$$

$$Q_{33}^{\rm NG} = (\mu_y^o - 3(\lambda_y^o)^2)\sum_k \gamma_k \sum_j \gamma_j \sum_{\tau=-\infty}^{\infty} \sum_i a_i^o a_{i+k}^o a_{i-\tau}^o a_{j+i-\tau}^o. \quad (143)$$

Adding also the contribution due to the input noise gives finally

$$
\begin{aligned}
Q_{33}^{\mathrm{NG}} \quad = \quad & \sum_k \sum_j \gamma_k \gamma_j \Bigg[(\mu_y^o - 3(\lambda_y^o)^2) \left(\sum_i a_i^o a_{i+k}^o \right) \left(\sum_\ell a_\ell^o a_{\ell+j}^o \right) \\
& + (\mu_u^o - 3(\lambda_u^o)^2) \left(\sum_i b_i^o b_{i+k}^o \right) \left(\sum_\ell b_\ell^o b_{\ell+j}^o \right) \Bigg]. \qquad (144)
\end{aligned}
$$