Asymptotic Accuracy Analysis of Bias-Eliminating Least Squares Estimates for Identification of Errors in Variables Systems

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

Wei Xing Zheng
School of QMMS, University of Western Sydney,
Penrith South DC NSW 1797, Australia

October 3, 2006

Abstract

The bias-eliminating least squares (BELS) method is one of the consistent estimators for identifying dynamic errors-in-variables systems. The attraction of the BELS method lies in its good accuracy and its modest computational cost. In this report, we investigate the asymptotic accuracy properties of the BELS estimates. It is shown that the estimated system parameters and the estimated noise variances are asymptotically Gaussian distributed. An explicit expression for the normalized covariance matrix of the estimated parameters is derived and supported by some numerical examples.

1 Introduction

System identification and parameter estimation for stochastic errors-in-variables (EIV) systems, where the input as well as the output measurements are noisy, have been a topic of active research for several decades. This class of system models frequently appears in various problems of practical interest, such as time series econometric models, blind channel equalization in communications, multivariate calibration in analytical chemistry, *etc*. See [11] for more descriptions.

Till now, many solutions to the EIV system identification problem have been proposed with different approaches. For example, the Koopmans-Levin (KL) method [1], the prediction error method [4], frequency domain methods [3], and methods based on higher order cumulate statistics [10], *etc*. See [8] and references therein for a comprehensive survey in this respect.

The focus of this report is placed on the bias-eliminating least squares (BELS) algorithms, which are developed according to the bias compensation principle, see [13], [12]. Since the BELS methods usually give quite good estimation accuracy (more accurate than standard IV methods and often comparable to the use of a prediction error method) but with a modest computational load, they seem to belong to the class of the more interesting and efficient approaches for errors-in-variables identification.

Noticeably, however, a statistical analysis of the accuracy of the BELS methods has been missing in the literature. There is no doubt that such an accuracy analysis can highly facilitate evaluation of and comparison with different identification approaches. Besides, one may also get insight into important issues like how different user choices in the algorithms can influence the accuracy, and when the estimation problem is hard to solve and only a low accuracy can be expected.

In this report, we will make an accuracy analysis of the parameter estimates for the BELS method. We will follow the analysis approach used for the Frisch method in [5]. First, the dynamic errors-in-variables problem is formulated in Section 2 and notations are described in Section 3. After a brief review of the BELS methods in Section 4, we give linearization results of the key equations utilized by the BELS algorithm for large data number N in Section 5. The asymptotic normalized covariance matrix of the BELS estimated parameters is presented as a theorem in Section 6. Finally, an numerical example is studied in Section 7 before concluding in Section 8.

2 Problem Formulation

Consider a linear and finite order system given by

$$A(q^{-1})y_0(t) = B(q^{-1})u_0(t), \tag{2.1}$$

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

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \ldots + a_{n_a}\, q^{-n_a} \\ B(q^{-1}) &= b_1\, q^{-1} + \ldots + b_{n_b}\, q^{-n_b}. \end{aligned} \tag{2.2}$$

Measurements of both the input and the output are contaminated by additive noise, so the available measurements $u(t)$ and $y(t)$ are

$$u(t) = u_0(t) + \tilde{u}(t), \quad y(t) = y_0(t) + \tilde{y}(t). \tag{2.3}$$

We assume that the system is asymptotically stable, observable and controllable; the system orders n_a and n_b are a priori known; the noise-free input $u_0(t)$ is persistently exciting of sufficient order; the noise signals $\tilde{u}(t)$ and $\tilde{y}(t)$ are independent of $u_0(t)$; and $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually independent white noise sequences of zero mean, and variances λ_u and λ_y, respectively.

The problem of identifying this error-in-variables system is concerned with consistently estimating the parameter vector

$$\begin{aligned} \vartheta &= (\theta^T; \quad \lambda^T)^T \\ &= (a_1 \ldots a_{n_a} \quad b_1 \ldots b_{n_b}; \quad \lambda_y \quad \lambda_u)^T \end{aligned} \tag{2.4}$$

from the measured noisy data $\{u(t), y(t)\}_{t=1}^{N}$.

The task of this report is to derive the asymptotic covariance matrix of the estimated parameters for the BELS estimator:

$$P = \lim_{N \to \infty} E\left\{N \text{cov}(\hat{\vartheta} - \vartheta_0)(\hat{\vartheta} - \vartheta_0)^T\right\} \tag{2.5}$$

where $\hat{\vartheta}$ is the estimate of ϑ, and ϑ_0 denotes the true value.

3 Notation

We now introduce several notations. First, the regressor vector is defined by

$$\begin{aligned} \varphi(t) &= (-y(t-1)\ldots - y(t-n_a), u(t-1)\ldots u(t-n_b))^T & (3.1)\\ \varphi(t) &= \varphi_0(t) + \tilde{\varphi}(t) & (3.2) \end{aligned}$$

where $\varphi_0(t)$ and $\tilde{\varphi}(t)$ denotes the noise-free term and the noise contribution to the regressor vector, respectively.

For convenience, we express the system parameter vector in partitioned form as

$$\theta = \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_{n_a} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_{n_b} \end{pmatrix}. \tag{3.3}$$

We will use the conventions

$$\tilde{\theta} = \hat{\theta} - \theta_0, \quad \tilde{\lambda}_y = \hat{\lambda}_y - \lambda_y^0, \quad \tilde{\lambda}_u = \hat{\lambda}_u - \lambda_u^0 \tag{3.4}$$

where $\theta_0, \lambda_y^0, \lambda_u^0$ denote the true value of $\theta, \lambda_y, \lambda_u$, respectively. $\hat{\theta}, \hat{\lambda}_y, \hat{\lambda}_u$ are the corresponding estimates, and $\tilde{\theta}, \tilde{\lambda}_y, \tilde{\lambda}_u$ are the relevant estimation errors.

Furthermore, $\hat{\theta}_{LS,N}$ and $\hat{V}_{LS,N}$ will be used to express the least squares (LS) parameter estimate and the minimum value of the LS loss function for a finite number of data N, while $\hat{\theta}_{LS,\infty}$ and $\hat{V}_{LS,\infty}$ represent the quantities when $N \to \infty$. The differences of these two cases are as follows:

$$\tilde{\hat{\theta}}_{LS} = \hat{\theta}_{LS,N} - \hat{\theta}_{LS,\infty}, \quad \tilde{V}_{LS} = \hat{V}_{LS,N} - \hat{V}_{LS,\infty}. \tag{3.5}$$

The cross-covariance matrices and vectors and their estimates are given by

$$\begin{aligned} R_\varphi &= E\varphi(t)\varphi^T(t), & (3.6)\\ \hat{R}_\varphi &= \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t), & (3.7)\\ r_{\varphi y} &= R_\varphi\hat{\theta}_{LS,\infty} = E\varphi(t)y(t), & (3.8)\\ \hat{r}_{\varphi y} &= \hat{R}_\varphi\hat{\theta}_{LS,N} = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t). & (3.9) \end{aligned}$$

The covariance matrix of the noise terms and its estimate are denoted as

$$R_{\tilde{\varphi}} = E\tilde{\varphi}(t)\tilde{\varphi}^T(t) = \begin{pmatrix} \lambda_y^0 I_{n_a} & \mathbf{0} \\ \mathbf{0} & \lambda_u^0 I_{n_b} \end{pmatrix} \tag{3.10}$$

$$\hat{R}_{\tilde{\varphi}} = \begin{pmatrix} \hat{\lambda}_y I_{n_a} & \mathbf{0} \\ \mathbf{0} & \hat{\lambda}_u I_{n_b} \end{pmatrix} \tag{3.11}$$

$$\tilde{R}_{\tilde{\varphi}} = \hat{R}_{\tilde{\varphi}} - R_{\tilde{\varphi}} \tag{3.12}$$

For R_φ, there exists the relation $R_\varphi = R_{\varphi_0} + R_{\tilde{\varphi}}$, where R_{φ_0} is the covariance matrix of the noise-free term, *i.e.*, $R_{\varphi_0} = E\varphi_0(t)\varphi_0^T(t)$.

4 Review of the BELS method

BELS algorithms are built upon the bias compensation principle. That is, remove the noise-contribution parts from the covariance matrix R_φ to get consistent estimates.

$$\hat{\theta}_{BELS} = (\hat{R}_\varphi - \hat{R}_{\tilde{\varphi}})^{-1}\hat{r}_{\varphi y} \tag{4.1}$$

From equation (3.10), we know that $R_{\tilde{\varphi}}$ contains two unknown parameters, *i.e.* the variances of the input and output noises λ_u and λ_y. So, in addition to the modified normal equation (4.1), (at least) two more relations for λ_u and λ_y are needed. One such relation can be derived from the minimal value of the least squares criterion:

$$V_{LS} \;=\; \min_\theta E(y(t) - \varphi^T(t)\theta)^2 = \lambda_y + \theta_0^T R_{\tilde{\varphi}}\hat{\theta}_{LS} \tag{4.2}$$

(see equation (12) in [7]). To get a second relation for λ_u and λ_y, an extended model structure is considered in BELS. For this purpose introduce extended versions of $\varphi(t)$, θ and θ_0 as

$$\bar{\varphi} \;=\; \begin{pmatrix} \varphi \\ \underline{\varphi} \end{pmatrix}, \quad \bar{\theta} = \begin{pmatrix} \theta \\ \underline{\theta} \end{pmatrix}, \quad \bar{\theta}_0 = \begin{pmatrix} \theta_0 \\ 0 \end{pmatrix}. \tag{4.3}$$

The possible model extension includes, for example, appending an additional A parameter, which yields

$$\underline{\varphi}(t) \;=\; -y(t - n_a - 1), \quad \underline{\theta} = a_{n_a+1} \tag{4.4}$$

or, appending an additional B parameter, which gives

$$\underline{\varphi}(t) \;=\; u(t - n_b - 1), \quad \underline{\theta} = b_{n_b+1} \tag{4.5}$$

In the extended model, by replacing $\varphi(t)$ with $\bar{\varphi}(t)$, the least squares estimate $\hat{\bar{\theta}}_{LS}$, the covariance matrix $R_{\bar{\varphi}}$ and the covariance matrix of the noise $R_{\bar{\tilde{\varphi}}}$ can be calculated in the same way as that $\hat{\theta}_{LS}$, R_φ and $R_{\tilde{\varphi}}$ in the normal model.

Next consider least squares estimation in the extended linear regression model

$$y(t) \;=\; \bar{\varphi}^T(t)\bar{\theta} + \varepsilon(t),$$

which leads to

$$R_{\bar{\varphi}}\hat{\bar{\theta}}_{LS} \;=\; r_{\bar{\varphi}y}.$$

Similar to (4.1), it holds that

$$R_{\bar{\varphi}}\hat{\bar{\theta}}_{LS} \;=\; r_{\bar{\varphi}_0 y_0} + r_{\tilde{\bar{\varphi}}\tilde{y}} = R_{\bar{\varphi}_0}\bar{\theta}_0 = (R_{\bar{\varphi}} - R_{\tilde{\bar{\varphi}}})\bar{\theta}_0. \tag{4.6}$$

Set

$$H \;=\; (0,\ldots,1) \in R^{n_a+n_b+1},\; J = \begin{pmatrix} I_{n_a+n_b} \\ \mathbf{0} \end{pmatrix},\; \bar{\theta}_0 = J\theta_0. \tag{4.7}$$

Observe that $H\bar{\theta}_0 = 0$. Equation (4.6) implies

$$H\hat{\bar{\theta}}_{LS} \;=\; HR_{\bar{\varphi}}^{-1}(R_{\bar{\varphi}} - R_{\tilde{\bar{\varphi}}})\bar{\theta}_0 = -HR_{\bar{\varphi}}^{-1}R_{\tilde{\bar{\varphi}}}J\theta_0. \tag{4.8}$$

See [12] and [7] for details.

Summing up so far the BELS algorithm consists of the following equations to determine the system parameter vector θ and noise variance vector λ, where $\lambda = (\lambda_y \quad \lambda_u)^T$.

$$R_{\varphi}\hat{\theta}_{LS} = (R_{\varphi} - R_{\tilde{\varphi}}(\lambda))\,\theta \tag{4.9}$$

$$V_{LS} = \lambda_y + \hat{\theta}_{LS}^T R_{\tilde{\varphi}}(\lambda)\theta \tag{4.10}$$

$$H\hat{\bar{\theta}}_{LS} = -HR_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}(\lambda)J\theta. \tag{4.11}$$

Equations (4.9)-(4.11) turn out to be bilinear in the unknowns θ and λ. There are different ways to solve these equations. In [7] , a variable projection algorithm has been proposed which has better convergence property than the traditional BELS algorithm [13] and [12].

5 Linearization

To analyze the estimation accuracy, we will examine how the estimates $\hat{\vartheta}$ deviates from the true parameter vector ϑ_0 for large data sets (large N). The technique for doing so is to linearize the above three equations (4.9)-(4.11) by assuming that $\hat{\vartheta}$ is close to ϑ_0 (*i.e.* estimation error is small) when N is large. Linearization results of equations (4.9)-(4.11) are summarized in the following three lemmas, while their proofs are given in Appendixes A, B and C, respectively.

Lemma 5.1. Linearizing equation (4.9) leads to

$$R_{\varphi_0}\tilde{\theta} + \begin{pmatrix} -\mathbf{a} \\ \mathbf{0} \end{pmatrix}\tilde{\lambda}_y + \begin{pmatrix} \mathbf{0} \\ -\mathbf{b} \end{pmatrix}\tilde{\lambda}_u = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varepsilon(t) - E\varphi(t)\varepsilon(t), \tag{5.1}$$

where

$$\begin{aligned} \varepsilon(t) &= y(t) - \varphi^T(t)\theta_0 \\ &= A(q^{-1})(y_0(t) + \tilde{y}(t)) - B(q^{-1})(u_0(t) + \tilde{u}(t)) \\ &= A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \end{aligned} \tag{5.2}$$

Lemma 5.2. Linearizing equation (4.10) leads to

$$
\begin{aligned}
& r_{\varphi y}^{T} R_{\varphi}^{-1} R_{\tilde{\varphi}} \tilde{\theta} + \left(1 + r_{\varphi y}^{T} R_{\varphi}^{-1} \begin{pmatrix} \mathbf{a} \\ \mathbf{0} \end{pmatrix}\right) \tilde{\lambda}_y + r_{\varphi y}^{T} R_{\varphi}^{-1} \begin{pmatrix} \mathbf{0} \\ \mathbf{b} \end{pmatrix} \tilde{\lambda}_u \\
& = \frac{1}{N} \sum_{t=1}^{N} \varepsilon_{LS}(t) \varepsilon(t) - E \varepsilon_{LS}(t) \varepsilon(t),
\end{aligned} \tag{5.3}
$$

where

$$
\begin{aligned}
\varepsilon_{LS}(t) &= y(t) - \varphi^T(t) \hat{\theta}_{LS,\infty} \\
&= A_{LS}(q^{-1}) y(t) - B_{LS}(q^{-1}) u(t) \\
&= \left\{ \frac{A_{LS}(q^{-1}) B(q^{-1}) - B_{LS}(q^{-1}) A(q^{-1})}{A(q^{-1})} \right\} u_0(t) \\
&\quad + A_{LS}(q^{-1}) \tilde{y}(t) - B_{LS}(q^{-1}) \tilde{u}(t).
\end{aligned} \tag{5.4}
$$

Lemma 5.3. Linearizing equation (4.11) leads to

$$
\begin{aligned}
& -H R_{\bar{\varphi}}^{-1} \bar{R}_{\tilde{\varphi}} J \tilde{\theta} - H R_{\bar{\varphi}}^{-1} \begin{pmatrix} \mathbf{a} \\ \mathbf{0} \\ 0 \end{pmatrix} \tilde{\lambda}_y - H R_{\bar{\varphi}}^{-1} \begin{pmatrix} \mathbf{0} \\ \mathbf{b} \\ 0 \end{pmatrix} \tilde{\lambda}_u \\
& = H R_{\bar{\varphi}}^{-1} \left(\frac{1}{N} \sum_{t=1}^{N} \bar{\varphi}(t) \varepsilon(t) - E \bar{\varphi}(t) \varepsilon(t) \right).
\end{aligned} \tag{5.5}
$$

Proof See Appendix A-C.

As can be seen, each of the three equations (4.9)-(4.11) is linearized into the generic form

$$
\alpha_\theta \tilde{\theta} + \alpha_{\lambda_y} \tilde{\lambda}_y + \alpha_{\lambda_u} \tilde{\lambda}_u \approx \beta_s \tag{5.6}
$$

where the coefficients $\alpha_\theta, \alpha_{\lambda_y}, \alpha_{\lambda_u}$ are deterministic variables, while β_s is a random term which has zero mean and a variance that decreases when N increases.

Moreover, we note that the term β_s in equations (5.1) and (5.5) depends only on $\varepsilon(t)$, the disturbances coming from the input and output noises. However, β_s in equation (5.3) is related not only to $\varepsilon(t)$ but also to $\varepsilon_{LS}(t)$ which depends on the noise-free input $u_0(t)$.

6 Asymptotic Covariance Matrix

Now we are ready to state the main theoretical result of the report.

Theorem 6.1. Assume that the white noise $\tilde{y}(t)$ has moments $E\tilde{y}(t) = 0,\ E\tilde{y}^2(t) = \lambda_y$ and $E\tilde{y}^4(t) = \mu_y$, and similarly for $\tilde{u}(t)$: $E\tilde{u}(t) = 0,\ E\tilde{u}^2(t) = \lambda_u$ and $E\tilde{u}^4(t) = \mu_u$. Under the given assumptions in Section 2 and the central limit theorem in [2], it follows that the BELS estimated parameter $\hat{\vartheta}$ is asymptotically Gaussian distributed

$$
\sqrt{N}(\hat{\vartheta} - \vartheta_0) \xrightarrow{dist} \mathcal{N}(\mathbf{0}, P), \tag{6.1}
$$

where

$$
\begin{aligned}
P &= \lim_{N\to\infty} NE\left\{N\mathrm{cov}(\hat{\vartheta}-\vartheta_0)(\hat{\vartheta}-\vartheta_0)^T\right\} \\
&= G^{-1}\lim_{N\to\infty} NE\beta\beta^T G^{-T} \\
&= G^{-1}QG^{-T}.
\end{aligned}
\tag{6.2}
$$

The coefficients $\alpha_\theta, \alpha_{\lambda_y}, \alpha_{\lambda_u}$, cf (5.6), appear as block elements of G. The block elements of Q are covariance matrices of the random terms β_s. That is,

$$
G=\begin{pmatrix}
R_{\varphi_0} & \begin{pmatrix}-\mathbf{a}\\ \mathbf{0}\end{pmatrix} & \begin{pmatrix}\mathbf{0}\\ -\mathbf{b}\end{pmatrix} \\
r_{\varphi y}^T R_\varphi^{-1} R_{\tilde{\varphi}} & 1+r_{\varphi y}^T R_\varphi^{-1}\begin{pmatrix}\mathbf{a}\\ \mathbf{0}\end{pmatrix} & r_{\varphi y}^T R_\varphi^{-1}\begin{pmatrix}\mathbf{0}\\ \mathbf{b}\end{pmatrix} \\
-HR_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J & -HR_{\bar{\varphi}}^{-1}\begin{pmatrix}\mathbf{a}\\ \mathbf{0}\\ 0\end{pmatrix} & -HR_{\bar{\varphi}}^{-1}\begin{pmatrix}\mathbf{0}\\ \mathbf{b}\\ 0\end{pmatrix}
\end{pmatrix}
\tag{6.3}
$$

$$
Q = Q^G + Q^{NG} \tag{6.4}
$$

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

$$
Q = \begin{pmatrix} Q_{11}^G & Q_{12}^G & Q_{13}^G \\ Q_{21}^G & Q_{22}^G & Q_{23}^G \\ Q_{31}^G & Q_{32}^G & Q_{33}^G \end{pmatrix}, \quad \begin{pmatrix} Q_{11}^{NG} & Q_{12}^{NG} & Q_{13}^{NG} \\ Q_{21}^{NG} & Q_{22}^{NG} & Q_{23}^{NG} \\ Q_{31}^{NG} & Q_{32}^{NG} & Q_{33}^{NG} \end{pmatrix}. \tag{6.5}
$$

The blocks in the symmetric matrix Q^G are as follows

$$
Q_{11}^G = \sum_\tau \left(r_{\varphi\varepsilon}(\tau)r_{\varphi\varepsilon}^T(-\tau) + R_\varphi(\tau)r_\varepsilon(\tau)\right) \tag{6.6}
$$

$$
Q_{12}^G = \sum_\tau \left(r_{\varphi\varepsilon_{LS}}(\tau)r_\varepsilon(\tau) + r_{\varphi\varepsilon}(\tau)r_{\varepsilon\varepsilon_{LS}}(\tau)\right) \tag{6.7}
$$

$$
Q_{13}^G = \sum_\tau \left(r_{\varphi\varepsilon}(\tau)r_{\bar{\varphi}\varepsilon}^T(-\tau) + R_{\varphi\bar{\varphi}}(\tau)r_\varepsilon(\tau)\right) R_{\bar{\varphi}}^{-1}H^T \tag{6.8}
$$

$$
Q_{22}^G = \sum_\tau \left(r_{\varepsilon_{LS}}(\tau)r_\varepsilon(\tau) + r_{\varepsilon_{LS}\varepsilon}(\tau)r_{\varepsilon\varepsilon_{LS}}(\tau)\right) \tag{6.9}
$$

$$
Q_{23}^G = \sum_\tau \left(r_{\varepsilon_{LS}\varepsilon}(\tau)r_{\bar{\varphi}\varepsilon}^T(-\tau) + r_{\varepsilon_{LS}\bar{\varphi}}^T(\tau)r_\varepsilon(\tau)\right) R_{\bar{\varphi}}^{-1}H^T \tag{6.10}
$$

$$
Q_{33}^G = HR_{\bar{\varphi}}^{-1}\left(\sum_\tau \left(r_{\bar{\varphi}\varepsilon}(\tau)r_{\bar{\varphi}\varepsilon}^T(-\tau) + R_{\bar{\varphi}}(\tau)r_\varepsilon(\tau)\right)\right) R_{\bar{\varphi}}^{-1}H^T \tag{6.11}
$$

The blocks of the matrix Q_{NG} are as follows

$$
Q_{11}^{NG} = \begin{pmatrix} (\mu_y - 3\lambda_y^2)\mathbf{a}\mathbf{a}^T & \mathbf{0} \\ \mathbf{0} & (\mu_u - 3\lambda_u^2)\mathbf{b}\mathbf{b}^T \end{pmatrix}, \tag{6.12}
$$

$$
\begin{aligned}
Q_{12}^{NG} &= \begin{pmatrix} -(\mu_y - 3\lambda_y^2)\mathbf{a}(\bar{\mathbf{a}}_{LS}^T\bar{\mathbf{a}}) \\ (\mu_u - 3\lambda_u^2)\mathbf{b}(\mathbf{b}_{LS}^T\mathbf{b}) \end{pmatrix}, &(6.13)\\
Q_{13}^{NG} &= \begin{pmatrix} (\mu_y - 3\lambda_y^2)\mathbf{a}\mathbf{a}^T & \mathbf{0}_{n_a\times n_b} & \mathbf{0}_{n_a\times 1} \\ \mathbf{0}_{n_b\times n_a} & (\mu_u - 3\lambda_u^2)\mathbf{b}\mathbf{b}^T & \mathbf{0}_{n_b\times 1} \end{pmatrix} R_{\bar{\varphi}}^{-1}H^T, &(6.14)\\
Q_{22}^{NG} &= (\mu_y - 3\lambda_y^2)(\bar{\mathbf{a}}_{LS}^T\bar{\mathbf{a}})^2 + (\mu_u - 3\lambda_u^2)(\mathbf{b}_{LS}^T\mathbf{b})^2, &(6.15)\\
Q_{23}^{NG} &= \begin{pmatrix} -(\mu_y - 3\lambda_y^2)\mathbf{a}^T(\bar{\mathbf{a}}_{LS}^T\bar{\mathbf{a}}) & (\mu_u - 3\lambda_u^2)\mathbf{b}^T(\mathbf{b}_{LS}^T\mathbf{b}) & 0 \end{pmatrix} R_{\bar{\varphi}}^{-1}H^T, &(6.16)\\
Q_{33}^{NG} &= HR_{\bar{\varphi}}^{-1} \begin{pmatrix} (\mu_y - 3\lambda_y^2)\mathbf{a}\mathbf{a}^T & \mathbf{0}_{n_a\times n_b} & \mathbf{0}_{n_a\times 1} \\ \mathbf{0}_{n_b\times n_a} & (\mu_u - 3\lambda_u^2)\mathbf{b}\mathbf{b}^T & \mathbf{0}_{n_b\times 1} \\ \mathbf{0}_{1\times n_a} & \mathbf{0}_{1\times n_b} & 0 \end{pmatrix} R_{\bar{\varphi}}^{-1}H^T. &(6.17)
\end{aligned}
$$

Proof. See Appendix D.

The covariance elements in the blocks of Q^G satisfy

$$
r_\varepsilon(k) = \begin{cases} \lambda_y \sum_i a_i a_{i+k} + \lambda_u \sum_i b_i b_{i+k} \\ 0 \qquad |k| > \max(n_a, n_b - 1) \end{cases} \tag{6.18}
$$

$$
r_{\varepsilon\varepsilon_{LS}}(k) = \begin{cases} \lambda_y \sum_i a_i a_{(LS)i+k} + \lambda_u \sum_i b_i b_{(LS)i+k} \\ 0 \qquad |k| > \max(n_a, n_b - 1) \end{cases} \tag{6.19}
$$

$$
\begin{aligned}
r_{\varphi\varepsilon}(k) &= \Big(-\lambda_y(a_{1-k}\dots a_{n_a-k})^T, -\lambda_u(b_{1-k}\dots b_{n_b-k})^T\Big)^T \\
r_{\bar{\varphi}\varepsilon}(k) &= \Big(-\lambda_y(a_{1-k}\dots a_{n_a-k})^T, -\lambda_u(b_{1-k}\dots b_{n_b-k})^T, \\
&\qquad -\lambda_y a_{n_a+1-k} \quad \text{or} \quad \lambda_u b_{n_a+1-k}\Big)^T
\end{aligned} \tag{6.20}
$$

where the conventions

$$
\begin{aligned}
a_i &= \begin{cases} 1 & i = 0 \\ 0 & i > n_a,\ i < 0 \end{cases} &(6.21)\\
b_i &= \quad 0 \quad i > n_b,\ i \le 0. &(6.22)
\end{aligned}
$$

are used. Note that the summations over τ in the elements of Q are over all values making the terms nonzero. However, due to the condition of $r_\varepsilon(\tau)$, $r_{\varepsilon\varepsilon_{LS}}(\tau)$, and $r_{\varphi\varepsilon}(\tau)$ and $r_{\bar{\varphi}\varepsilon}(\tau)$ will be zero when $|\tau| > \max(n_a, n_b - 1)$ or $|\tau| > \max(n_a + 1, n_b)$, respectively. Thus, each sum is finite and will have only a modest number of nonzero terms.

7 Numerical Examples

We now present some examples to illustrate the theoretical formulas derived in the preceding section. For simplicity, only Gaussian data are considered here.

Example 1. Consider a first-order system given by

$$
(1 - 0.8q^{-1})y_0(t) = 1.0q^{-1}u_0(t). \tag{7.1}
$$

The noise-free input $u_0(t)$ is an ARMA(1,1) process

$$
(1 - 0.5q^{-1})u_0(t) = (1 + 0.7q^{-1})e(t), \tag{7.2}
$$

where $e(t)$ is a zero-mean white noise with unit variance. The white measurement noises at the input and output sides both have unit variance, too. For each simulation result, $N = 10000$ data points are used and $M = 1000$ realizations are made.

First, consider that an additional A parameter is appended, and in this case $\underline{\varphi}(t) = -y(t-2)$. The theoretical normalized asymptotic covariance matrix of the estimation error is

$$P = \begin{pmatrix} 0.084 & & & \\ 0.228 & 1.391 & & \\ -0.248 & -1.490 & 6.326 & \\ 0.303 & 1.839 & -2.684 & 8.289 \end{pmatrix}.$$

The corresponding normalized covariance matrix obtained from the Monte-Carlo simulation is as follows

$$P_s = \begin{pmatrix} 0.084 & & & \\ 0.227 & 1.405 & & \\ -0.248 & -1.485 & 6.392 & \\ 0.294 & 1.832 & -2.682 & 8.288 \end{pmatrix}.$$

Second, consider the case of appending an additional B parameter, *i.e.*, $\underline{\varphi}(t) = u(t-2)$. The theoretical normalized asymptotic covariance matrix is

$$P = \begin{pmatrix} 0.191 & & & \\ 0.791 & 4.349 & & \\ -0.864 & -4.734 & 9.688 & \\ 1.747 & 9.395 & -11.226 & 27.263 \end{pmatrix}.$$

The Monte-Carlo simulation result is

$$P_s = \begin{pmatrix} 0.198 & & & \\ 0.803 & 4.284 & & \\ -0.866 & -4.739 & 9.770 & \\ 1.787 & 9.355 & -11.283 & 27.593 \end{pmatrix}.$$

In order to compare P_s with the theoretical covariance matrix P, by Appendix B9 in [9], the following relation

$$\frac{|P_s(i,j) - P(i,j)|}{|P(i,j)|} \leq \frac{3}{\sqrt{M}} \tag{7.3}$$

holds with 95% probability for diagonal elements and to non-diagonal elements whose absolute value is not too small. As can be seen, the theoretical results are in good agreement with the simulation results for both cases. It also shows that adding an extra A parameter can result in much better estimation accuracy than choosing an extra B parameter in the extended models.

Example 2. Consider a second-order system given by

$$(1 - 1.5q^{-1} + 0.7q^{-2})y_0(t) \;=\; (1.0q^{-1} + 0.5q^{-2})u_0(t), \tag{7.4}$$

where $u_0(t)$ and the variances of the measurement noises are kept as the same as in Example 1. Assume $\underline{\varphi}(t) = -y(t-3)$. Then the theoretical normalized asymptotic covariance matrix is

$$P \;=\; \begin{pmatrix} 0.34 & & & & & \\ -0.26 & 0.22 & & & & \\ -1.01 & 0.63 & 14.36 & & & \\ 1.71 & -1.16 & -14.52 & 17.50 & & \\ -0.04 & 0.06 & -3.74 & 2.67 & 6.03 & \\ -0.07 & -0.01 & 7.38 & -5.75 & -3.62 & 11.36 \end{pmatrix},$$

and the Monte-Carlo simulation result is

$$P_s \;=\; \begin{pmatrix} 0.33 & & & & & \\ -0.26 & 0.22 & & & & \\ -1.02 & 0.65 & 13.32 & & & \\ 1.72 & -1.19 & -13.91 & 17.20 & & \\ -0.09 & 0.10 & -3.33 & 2.24 & 5.85 & \\ -0.05 & -0.02 & 6.42 & -5.06 & -3.33 & 10.51 \end{pmatrix}.$$

Again, the agreement between theory and simulation is fairly good.

For BELS, the user choice on the construction of the extended models is very important. The analysis here shows that adding an extra A parameter can result much better estimation accuracy than choosing an extra B parameter in the extended models.

8 Conclusions

In this report, we have analyzed the accuracy of the BELS estimates for identifying the errors-in-variables systems. The normalized asymptotic covariance matrix of the estimated system parameters and the estimated noise variances has been derived. The numerical examples demonstrate the correctness of the theoretical results. It suggests that, in the extended model of the BELS estimator, choose to add the regressor which contains more information on the system and/or the noises in order to get better estimates. The analysis can be extend to the BELS algorithm under more general noise conditions.

Acknowledgment: This research was partially supported by The Swedish Research Council, contract 621-2005-4207.

A Linearization of equation (4.9)

Consider equation (4.9) for a finite N:

$$\hat{R}_{\varphi}\hat{\theta} \;=\; \hat{R}_{\varphi}\hat{\theta}_{LS,N} + \hat{R}_{\tilde{\varphi}}\hat{\theta} = \hat{r}_{\varphi y} + \hat{R}_{\tilde{\varphi}}\hat{\theta} \tag{A.1}$$

and for $N \to \infty$:

$$R_\varphi \theta_0 \quad = \quad R_\varphi \hat{\theta}_{LS,\infty} + R_{\tilde{\varphi}} \theta_0 = r_{\varphi y} + R_{\tilde{\varphi}} \theta_0. \tag{A.2}$$

In equation (A.1), replace $\hat{R}_\varphi$, $\hat{r}_{\varphi y}$ and $\hat{\theta}$ with $R_\varphi + \tilde{R}_\varphi$, $r_{\varphi y} + \tilde{r}_{\varphi y}$, and $\theta_0 + \tilde{\theta}$, respectively. It gives

$$(R_\varphi + \tilde{R}_\varphi)(\theta_0 + \tilde{\theta}) \quad = \quad (r_{\varphi y} + \tilde{r}_{\varphi y}) + (R_{\tilde{\varphi}} + \tilde{R}_{\tilde{\varphi}})(\theta_0 + \tilde{\theta}).$$

Assume that $\hat{\vartheta}$ is close to ϑ_0. Then $\tilde{R}_\varphi$, $\tilde{r}_{\varphi y}$, $\tilde{\theta}$, and $\tilde{R}_{\tilde{\varphi}}$ are all small. We can neglect the second order terms and use (A.2) to get

$$\begin{aligned}
& R_\varphi \tilde{\theta} + \tilde{R}_\varphi \theta_0 = \tilde{r}_{\varphi y} + R_{\tilde{\varphi}} \tilde{\theta} + \tilde{R}_{\tilde{\varphi}} \theta_0 \\
& \Rightarrow (R_\varphi - R_{\tilde{\varphi}})\tilde{\theta} - \tilde{R}_{\tilde{\varphi}} \theta_0 = \tilde{r}_{\varphi y} - \tilde{R}_\varphi \theta_0 \\
& \Rightarrow \begin{pmatrix} R_{\varphi_0} & \begin{matrix} -\mathbf{a} & \mathbf{0} \\ \mathbf{0} & -\mathbf{b} \end{matrix} \end{pmatrix} \begin{pmatrix} \tilde{\theta} \\ \tilde{\lambda}_y \\ \tilde{\lambda}_u \end{pmatrix} = (\hat{r}_{\varphi y} - r_{\varphi y}) - (\hat{R}_\varphi - R_\varphi)\theta_0
\end{aligned}$$

Finally we have

$$\begin{aligned}
\beta_1 & \triangleq (\hat{r}_{\varphi y} - r_{\varphi y}) - (\hat{R}_\varphi - R_\varphi)\theta_0 \\
& = \frac{1}{N} \sum \varphi(t) \left(y(t) - \varphi^T(t)\theta_0 \right) \\
& \quad - E\varphi(t) \left(y(t) - \varphi^T(t)\theta_0 \right) \\
& = \frac{1}{N} \sum \varphi(t)\varepsilon(t) - E\varphi(t)\varepsilon(t).
\end{aligned} \tag{A.3}$$

B Linearization of equation (4.10)

Consider equation (4.10) for a finite N:

$$\hat{V}_{LS,N} = \hat{\lambda}_y + \hat{\theta}^T_{LS,N} \hat{R}_{\tilde{\varphi}} \hat{\theta} \tag{B.1}$$

and for $N \to \infty$:

$$V_{LS,\infty} = \lambda_y + \hat{\theta}^T_{LS,\infty} R_{\tilde{\varphi}} \theta_0 \tag{B.2}$$

where $\hat{\theta}_{LS,N} = \hat{R}_\varphi^{-1} \hat{r}_{\varphi y}$, $\hat{\theta}_{LS,\infty} = R_\varphi^{-1} r_{\varphi y}$. Here we assume that $\hat{R}_\varphi^{-1}$ exists (R_φ^{-1} exists due to the assumptions made in Section 2).
$V_{LS,\infty}$ can be further expressed as

$$\begin{aligned}
V_{LS,\infty} & = E(y(t) - \varphi^T(t)\hat{\theta}_{LS,\infty})^2 \\
& = E\big(y(t) - \varphi^T(t) R_\varphi^{-1} r_{\varphi y})(y(t) - \varphi^T(t) R_\varphi^{-1} r_{\varphi y}\big) \\
& = r_y - r_{\varphi y}^T R_\varphi^{-1} r_{\varphi y}.
\end{aligned} \tag{B.3}$$

Similarly, for $\hat{V}_{LS,N}$, we can get

$$\hat{V}_{LS,N} \quad = \quad \hat{r}_y - \hat{r}_{\varphi y}^T \hat{R}_\varphi^{-1} \hat{r}_{\varphi y}. \tag{B.4}$$

Equation (B.1) can be expressed as

$$\begin{aligned} &(V_{LS,\infty} + \tilde{V}_{LS}) \\ = \quad &(\lambda_y + \tilde{\lambda}_y) + (\hat{\theta}^T_{LS,\infty} + \tilde{\hat{\theta}}^T_{LS})(R_{\tilde{\varphi}} + \tilde{R}_{\tilde{\varphi}})(\theta_0 + \tilde{\theta}). \end{aligned}$$

By neglecting the second and higher order terms, using (B.2) results in

$$\tilde{V}_{LS} = \tilde{\lambda}_y + \hat{\theta}^T_{LS,\infty} R_{\tilde{\varphi}} \tilde{\theta} + \hat{\theta}^T_{LS,\infty} \tilde{R}_{\tilde{\varphi}} \theta_0 + \tilde{\hat{\theta}}^T_{LS} R_{\tilde{\varphi}} \theta_0.$$

Then we use (B.3) and (B.4) to get

$$\begin{aligned} \Rightarrow \quad & r^T_{\varphi y} R^{-1}_{\varphi} R_{\tilde{\varphi}} \tilde{\theta} + \left(1 + r^T_{\varphi y} R^{-1}_{\varphi} \begin{pmatrix} \mathbf{a} \\ \mathbf{0} \end{pmatrix}\right) \tilde{\lambda}_y \\ & + r^T_{\varphi y} R^{-1}_{\varphi} \begin{pmatrix} \mathbf{0} \\ \mathbf{b} \end{pmatrix} \tilde{\lambda}_u \\ & = (\hat{r}_y - r_y) - (\hat{r}^T_{\varphi y} \hat{R}^{-1}_{\varphi} \hat{r}_{\varphi y} - r^T_{\varphi y} R^{-1}_{\varphi} r_{\varphi y}) \\ & - (\hat{r}^T_{\varphi y} \hat{R}^{-1}_{\varphi} - r^T_{\varphi y} R^{-1}_{\varphi}) R_{\tilde{\varphi}} \theta_0. \end{aligned}$$

Let

$$\begin{aligned} \beta_2 \quad \triangleq \quad & (\hat{r}_y - r_y) - (\hat{r}^T_{\varphi y} \hat{R}^{-1}_{\varphi} \hat{r}_{\varphi y} - r^T_{\varphi y} R^{-1}_{\varphi} r_{\varphi y}) \\ & - (\hat{r}^T_{\varphi y} \hat{R}^{-1}_{\varphi} - r^T_{\varphi y} R^{-1}_{\varphi}) R_{\tilde{\varphi}} \theta_0. \end{aligned}$$

Because

$$\begin{aligned} & \hat{r}^T_{\varphi y} \hat{R}^{-1}_{\varphi} \hat{r}_{\varphi y} - r^T_{\varphi y} R^{-1}_{\varphi} r_{\varphi y} \approx (\hat{r}_{\varphi y} - r_{\varphi y})^T R^{-1}_{\varphi} r_{\varphi y} \\ & - r^T_{\varphi y} R^{-1}_{\varphi} (\hat{R}_{\varphi} - R_{\varphi}) R^{-1}_{\varphi} r_{\varphi y} + r^T_{\varphi y} R^{-1}_{\varphi} (\hat{r}_{\varphi y} - r_{\varphi y}), \\ & \hat{r}^T_{\varphi y} \hat{R}^{-1}_{\varphi} - r^T_{\varphi y} R^{-1}_{\varphi} \approx \\ & (\hat{r}_{\varphi y} - r_{\varphi y})^T R^{-1}_{\varphi} - r^T_{\varphi y} R^{-1}_{\varphi} (\hat{R}_{\varphi} - R_{\varphi}) R^{-1}_{\varphi}, \\ & r_{\varphi y} = (R_{\varphi} - R_{\tilde{\varphi}}) \theta_0, \end{aligned}$$

β_2 can be represented as

$$\begin{aligned} \beta_2 \quad = \quad & (\hat{r}_y - r_y) - (\hat{r}_{\varphi y} - r_{\varphi y})^T R^{-1}_{\varphi} (2 r_{\varphi y} + R_{\tilde{\varphi}} \theta_0) \\ & + r^T_{\varphi y} R^{-1}_{\varphi} (\hat{R}_{\varphi} - R_{\varphi}) R^{-1}_{\varphi} (r_{\varphi y} + R_{\tilde{\varphi}} \theta_0) \\ = \quad & (\hat{r}_y - r_y) - (\hat{r}_{\varphi y} - r_{\varphi y})^T (2\theta_0 - R^{-1}_{\varphi} R_{\tilde{\varphi}} \theta_0) \\ & + r^T_{\varphi y} R^{-1}_{\varphi} (\hat{R}_{\varphi} - R_{\varphi}) \theta_0. \end{aligned}$$

Further, since

$$\begin{aligned} & \hat{\theta}_{LS,\infty} = R^{-1}_{\varphi} r_{\varphi y} = R^{-1}_{\varphi} (R_{\varphi} - R_{\tilde{\varphi}}) \theta_0 = \theta_0 - R^{-1}_{\varphi} R_{\tilde{\varphi}} \theta_0 \\ & \varphi(t)^T \theta_0 = y(t) - y(t) + \varphi(t)^T \theta_0 = y(t) - \varepsilon(t) \\ & \varphi(t)^T \hat{\theta}_{LS,\infty} = y(t) - y(t) + \varphi(t)^T \theta_{LS,\infty} \\ & \quad = y(t) - \varepsilon_{LS}(t), \end{aligned}$$

we have

$$\begin{aligned}\beta_2 &= (\hat{r}_y - r_y) - (\hat{r}_{\varphi y} - r_{\varphi y})^T(\theta_0 + \hat{\theta}_{LS,\infty}) + \hat{\theta}_{LS,\infty}^T(\hat{R}_\varphi - R_\varphi)\theta_0 \\ &= \left(\frac{1}{N}\sum y^2(t) - Ey^2(t)\right) - \left(\frac{1}{N}\sum y(t)\varphi(t)^T - Ey(t)\varphi(t)^T\right)(\theta_0 + \hat{\theta}_{LS,\infty}) \\ &\quad + \hat{\theta}_{LS,\infty}^T\left(\frac{1}{N}\sum \varphi(t)\varphi(t)^T - E\varphi(t)\varphi(t)^T\right)\theta_0 \\ &= \frac{1}{N}\sum \varepsilon_{LS}(t)\varepsilon(t) - E\varepsilon_{LS}(t)\varepsilon(t). \qquad \text{(B.5)}\end{aligned}$$

C Linearization of equation (4.11)

Consider equation (4.11) for a finite N:

$$H\hat{\bar{\theta}}_{LS,N} = -H\hat{R}_{\bar{\varphi}}^{-1}\hat{\bar{R}}_{\tilde{\varphi}}J\hat{\theta} \qquad \text{(C.1)}$$

and for $N \to \infty$:

$$H\hat{\bar{\theta}}_{LS,\infty} = -HR_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\theta_0. \qquad \text{(C.2)}$$

Because $(R_{\bar{\varphi}} + \tilde{R}_{\bar{\varphi}})^{-1} \approx R_{\bar{\varphi}}^{-1} - R_{\bar{\varphi}}^{-1}\tilde{R}_{\bar{\varphi}}R_{\bar{\varphi}}^{-1}$, it follows from (C.1) that

$$H\hat{\bar{\theta}}_{LS,N} = -H(R_{\bar{\varphi}}^{-1} - R_{\bar{\varphi}}^{-1}\tilde{R}_{\bar{\varphi}}R_{\bar{\varphi}}^{-1})(\bar{R}_{\tilde{\varphi}} + \tilde{\bar{R}}_{\tilde{\varphi}})J(\theta_0 + \tilde{\theta}).$$

Neglecting the second and higher order terms gives

$$\begin{aligned}H\hat{\bar{\theta}}_{LS,N} &= -HR_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\theta_0 + HR_{\bar{\varphi}}^{-1}\tilde{R}_{\bar{\varphi}}R_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\theta_0 \\ &- HR_{\bar{\varphi}}^{-1}\tilde{\bar{R}}_{\tilde{\varphi}}J\theta_0 - HR_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\tilde{\theta}\end{aligned}$$

$$\begin{aligned}&\Rightarrow H(\hat{\bar{\theta}}_{LS,N} - \hat{\bar{\theta}}_{LS,\infty}) - HR_{\bar{\varphi}}^{-1}\tilde{R}_{\bar{\varphi}}R_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\theta_0 \\ &= -HR_{\bar{\varphi}}^{-1}\tilde{\bar{R}}_{\tilde{\varphi}}J\theta_0 - HR_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\tilde{\theta} \\ &\Rightarrow -HR_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\tilde{\theta} - HR_{\bar{\varphi}}^{-1}\begin{pmatrix}\mathbf{a}\\ \mathbf{0}\\ 0\end{pmatrix}\tilde{\lambda}_y - HR_{\bar{\varphi}}^{-1}\begin{pmatrix}\mathbf{0}\\ \mathbf{b}\\ 0\end{pmatrix}\tilde{\lambda}_u \\ &= H(\hat{R}_{\bar{\varphi}}^{-1}\hat{r}_{\bar{\varphi}y} - R_{\bar{\varphi}}^{-1}r_{\bar{\varphi}y}) - HR_{\bar{\varphi}}^{-1}(\hat{R}_{\bar{\varphi}} - R_{\bar{\varphi}})R_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\theta_0.\end{aligned}$$

Let

$$\begin{aligned}\beta_3 &\triangleq H(\hat{R}_{\bar{\varphi}}^{-1}\hat{r}_{\bar{\varphi}y} - R_{\bar{\varphi}}^{-1}r_{\bar{\varphi}y}) - HR_{\bar{\varphi}}^{-1}(\hat{R}_{\bar{\varphi}} - R_{\bar{\varphi}})R_{\bar{\varphi}}^{-1}\bar{R}_{\tilde{\varphi}}J\theta_0 \\ &= HR_{\bar{\varphi}}^{-1}\Big((\hat{r}_{\bar{\varphi}y} - r_{\bar{\varphi}y}) - (\hat{R}_{\bar{\varphi}} - R_{\bar{\varphi}})R_{\bar{\varphi}}^{-1}(r_{\bar{\varphi}y} + \bar{R}_{\tilde{\varphi}}\bar{\theta}_0)\Big).\end{aligned}$$

Then using the relation $r_{\bar{\varphi}y} = (R_{\bar{\varphi}} - \bar{R}_{\tilde{\varphi}})\bar{\theta}_0$ (which can easily be proved), we get

$$\begin{aligned}\beta_3 &= HR_{\bar{\varphi}}^{-1}\left((\hat{r}_{\bar{\varphi}y} - r_{\bar{\varphi}y}) - (\hat{R}_{\bar{\varphi}} - R_{\bar{\varphi}})\bar{\theta}_0\right) \\ &= HR_{\bar{\varphi}}^{-1}\left(\frac{1}{N}\sum \bar{\varphi}(t)(y(t) - \bar{\varphi}^T(t)\bar{\theta}_0) - E\bar{\varphi}(t)(y(t) - \bar{\varphi}^T(t)\bar{\theta}_0)\right) \\ &= HR_{\bar{\varphi}}^{-1}\left(\frac{1}{N}\sum \bar{\varphi}(t)(y(t) - \varphi^T(t)\theta_0) - E\bar{\varphi}(t)(y(t) - \varphi^T(t)\theta_0)\right) \\ &= HR_{\bar{\varphi}}^{-1}\left(\frac{1}{N}\sum \bar{\varphi}(t)\varepsilon(t) - E\bar{\varphi}(t)\varepsilon(t)\right). \qquad \text{(C.3)}\end{aligned}$$

D Derivation of elements $Q_{i,j}$

First, we have

$$
\begin{aligned}
Q_{11} &= \lim_{N\to\infty} EN\beta_1\beta_1^T \\
&= \lim_{N\to\infty} EN\left(\frac{1}{N}\sum_t \varphi(t)\varepsilon(t) - E\varphi(t)\varepsilon(t)\right)\left(\frac{1}{N}\sum_s \varepsilon(s)\varphi^T(s) - E\varepsilon(s)\varphi^T(s)\right) \\
&= \lim_{N\to\infty}\Bigg[EN\left(\frac{1}{N}\sum_t \varphi(t)\varepsilon(t)\right)\left(\frac{1}{N}\sum_s \varepsilon(s)\varphi^T(s)\right) - EN\left(\frac{1}{N}\sum_t \varphi(t)\varepsilon(t)\right)E\varepsilon(s)\varphi^T(s) \\
&\quad -E\varphi(t)\varepsilon(t)EN\left(\frac{1}{N}\sum_s \varepsilon(s)\varphi^T(s)\right) + EN\left(E\varphi(t)\varepsilon(t)E\varepsilon(s)\varphi^T(s)\right)\Bigg] \\
&= \lim_{N\to\infty}\left[\frac{1}{N}\sum_t\sum_s\left(E\varphi(t)\varepsilon(t)\varepsilon(s)\varphi^T(s) - E\varphi(t)\varepsilon(t)E\varepsilon(s)\varphi^T(s)\right)\right] \qquad \text{(D.1)}
\end{aligned}
$$

We will first treat the case of Gaussian distributed measurement noise. Using the following property for jointly Gaussian distributed variables:

$$
Ex_1x_2x_3x_4 = (Ex_1x_2)(Ex_3x_4) + (Ex_1x_3)(Ex_2x_4) + (Ex_1x_4)(Ex_2x_3), \qquad \text{(D.2)}
$$

we get

$$
Q_{11}^G = \lim_{N\to\infty}\frac{1}{N}\sum_t\sum_s\left(E\varphi(t)\varphi^T(s)E\varepsilon(t)\varepsilon(s) + E\varphi(t)\varepsilon(s)E\varepsilon(t)\varphi^T(s)\right).
$$

By changing variables, $\tau = t - s$, the first double sum in Q_{11} can be expressed as

$$
\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_t\sum_s\left(E\varphi(t)\varphi^T(s)E\varepsilon(t)\varepsilon(s)\right) \\
&= \lim_{N\to\infty}\left(\frac{1}{N}\sum_{\tau=1-N}^{N-1}(N-|\tau|)\,R_\varphi(\tau)r_\varepsilon(\tau)\right) \\
&= \sum_{\tau=-\infty}^{\infty} R_\varphi(\tau)r_\varepsilon(\tau) - \lim_{N\to\infty}\frac{1}{N}\sum_{\tau=1-N}^{N-1}|\tau|R_\varphi(\tau)r_\varepsilon(\tau)
\end{aligned}
$$

Since the covariance functions $R_\varphi(\tau)$ and $r_\varepsilon(\tau)$ will decay exponentially when $|\tau| \to \infty$, the second term will converge to zero (see [6] Appendix C for more details). Hence,

$$
\lim_{N\to\infty}\frac{1}{N}\sum_t\sum_s\left(E\varphi(t)\varphi^T(s)E\varepsilon(t)\varepsilon(s)\right) = \sum_{\tau=-\infty}^{\infty} R_\varphi(\tau)r_\varepsilon(\tau)
$$

Applying the same techniques to the second double sum in Q_{11}^G gives finally

$$
Q_{11}^G = \sum_\tau\left(r_{\varphi\varepsilon}(\tau)r_{\varphi\varepsilon}^T(-\tau) + R_\varphi(\tau)r_\varepsilon(\tau)\right) \qquad \text{(D.3)}
$$

A similar process is used to derive $Q_{12}^G, Q_{13}^G, Q_{22}^G, Q_{23}^G$, and Q_{33}^G.

$$
\begin{aligned}
Q_{12}^G &= \lim_{N\to\infty} EN\beta_1\beta_2^T \\
&= \lim_{N\to\infty} EN\left(\frac{1}{N}\sum_t \varphi(t)\varepsilon(t) - E\varphi(t)\varepsilon(t)\right)\left(\frac{1}{N}\sum_s \varepsilon_{LS}(s)\varepsilon(s) - E\varepsilon_{LS}(s)\varepsilon(s)\right)^T \\
&= \lim_{N\to\infty} \left[\frac{1}{N}\sum_t\sum_s E\varphi(t)\varepsilon(t)\varepsilon_{LS}(s)\varepsilon(s) - NE\varphi(t)\varepsilon(t)E\varepsilon_{LS}(s)\varepsilon(s)\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_t\sum_s \Big(E\varphi(t)\varepsilon_{LS}(s)E\varepsilon(t)\varepsilon(s) + E\varphi(t)\varepsilon(s)E\varepsilon(t)\varepsilon_{LS}(s)\Big) \\
&= \sum_\tau \Big(r_{\varphi\varepsilon_{LS}}(\tau)r_\varepsilon(\tau) + r_{\varphi\varepsilon}(\tau)r_{\varepsilon\varepsilon_{LS}}(\tau)\Big) \qquad \text{(D.4)}
\end{aligned}
$$

$$
\begin{aligned}
Q_{13}^G &= \lim_{N\to\infty} EN\beta_1\beta_3^T \\
&= \lim_{N\to\infty} EN\left(\frac{1}{N}\sum_t \varphi(t)\varepsilon(t) - E\varphi(t)\varepsilon(t)\right)\left(\frac{1}{N}\sum_s \varepsilon(s)\bar{\varphi}^T(s) - E\varepsilon(s)\bar{\varphi}^T(s)\right)R_{\bar{\varphi}}^{-1}H^T \\
&= \lim_{N\to\infty} \left[\frac{1}{N}\sum_t\sum_s E\varphi(t)\varepsilon(t)\varepsilon(s)\bar{\varphi}^T(s) - NE\varphi(t)\varepsilon(t)E\varepsilon(s)\bar{\varphi}^T(s)\right] R_{\bar{\varphi}}^{-1}H^T \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_t\sum_s \Big(E\varphi(t)\varepsilon(s)E\varepsilon(t)\bar{\varphi}^T(s) + E\varphi(t)\bar{\varphi}^T(s)E\varepsilon(t)\varepsilon(s)\Big) R_{\bar{\varphi}}^{-1}H^T \\
&= \sum_\tau \Big(r_{\varphi\varepsilon}(\tau)r_{\bar{\varphi}\varepsilon}^T(-\tau) + R_{\varphi\bar{\varphi}}(\tau)r_\varepsilon(\tau)\Big) R_{\bar{\varphi}}^{-1}H^T \qquad \text{(D.5)}
\end{aligned}
$$

$$
\begin{aligned}
Q_{22}^G &= \lim_{N\to\infty} EN\beta_2\beta_2^T \\
&= \lim_{N\to\infty} EN\left(\frac{1}{N}\sum_t \varepsilon_{LS}(t)\varepsilon(t) - E\varepsilon_{LS}(t)\varepsilon(t)\right)\left(\frac{1}{N}\sum_s \varepsilon_{LS}(s)\varepsilon(s) - E\varepsilon_{LS}(s)\varepsilon(s)\right)^T \\
&= \lim_{N\to\infty} \left[\frac{1}{N}\sum_t\sum_s E\varepsilon_{LS}(t)\varepsilon(t)\varepsilon_{LS}(s)\varepsilon(s) - NE\varepsilon_{LS}(t)\varepsilon(t)E\varepsilon_{LS}(s)\varepsilon(s)\right] \\
&= \lim_{N\to\infty} \frac{1}{N}\sum_t\sum_s \Big(E\varepsilon_{LS}(t)\varepsilon_{LS}(s)E\varepsilon(t)\varepsilon(s) + E\varepsilon_{LS}(t)\varepsilon(s)E\varepsilon(t)\varepsilon_{LS}(s)\Big) \\
&= \sum_\tau \Big(r_{\varepsilon_{LS}}(\tau)r_\varepsilon(\tau) + r_{\varepsilon_{LS}\varepsilon}(\tau)r_{\varepsilon\varepsilon_{LS}}(\tau)\Big) \qquad \text{(D.6)}
\end{aligned}
$$

$$
\begin{aligned}
Q_{23}^G &= \lim_{N\to\infty} EN\beta_2\beta_3^T \\
&= \lim_{N\to\infty} EN\left(\frac{1}{N}\sum_t \varepsilon_{LS}(t)\varepsilon(t) - E\varepsilon_{LS}(t)\varepsilon(t)\right)
\end{aligned}
$$

$$\times \left(\frac{1}{N} \sum_s \varepsilon(s)\bar{\varphi}^T(s) - E\varepsilon(s)\bar{\varphi}^T(s) \right) R_{\bar{\varphi}}^{-1} H^T$$

$$= \lim_{N\to\infty} \left(\frac{1}{N} \sum_t \sum_s E\varepsilon_{LS}(t)\varepsilon(t)\varepsilon(s)\bar{\varphi}^T(s) - NE\varepsilon_{LS}(t)\varepsilon(t)E\varepsilon(s)\bar{\varphi}^T(s) \right) R_{\bar{\varphi}}^{-1} H^T$$

$$= \lim_{N\to\infty} \frac{1}{N} \sum_t \sum_s \left[E\varepsilon_{LS}(t)\varepsilon(s)E\varepsilon(t)\bar{\varphi}^T(s) + E\varepsilon_{LS}(t)\bar{\varphi}^T(s)E\varepsilon(t)\varepsilon(s) \right] R_{\bar{\varphi}}^{-1} H^T$$

$$= \sum_\tau \left(r_{\varepsilon_{LS}\varepsilon}(\tau) r_{\bar{\varphi}\varepsilon}^T(-\tau) + r_{\varepsilon_{LS}\bar{\varphi}}^T(\tau) r_\varepsilon(\tau) \right) R_{\bar{\varphi}}^{-1} H^T \tag{D.7}$$

$$\begin{aligned}
Q_{33}^G &= \lim_{N\to\infty} EN\beta_3\beta_3^T \\
&= \lim_{N\to\infty} HR_{\bar{\varphi}}^{-1} \left[EN \left(\frac{1}{N} \sum_t \bar{\varphi}(t)\varepsilon(t) - E\bar{\varphi}(t)\varepsilon(t) \right) \right. \\
&\quad \times \left. \left(\frac{1}{N} \sum_s \varepsilon(s)\bar{\varphi}^T(s) - E\varepsilon(s)\bar{\varphi}^T(s) \right) \right] R_{\bar{\varphi}}^{-1} H^T \\
&= \lim_{N\to\infty} HR_{\bar{\varphi}}^{-1} \left[\frac{1}{N} \sum_t \sum_s E\bar{\varphi}(t)\varepsilon(t)\varepsilon(s)\bar{\varphi}^T(s) - NE\bar{\varphi}(t)\varepsilon(t)E\varepsilon(s)\bar{\varphi}^T(s) \right] R_{\bar{\varphi}}^{-1} H^T \\
&= \lim_{N\to\infty} HR_{\bar{\varphi}}^{-1} \left[\frac{1}{N} \sum_t \sum_s \left(E\bar{\varphi}(t)\varepsilon(s)E\varepsilon(t)\bar{\varphi}^T(s) + E\bar{\varphi}(t)\bar{\varphi}^T(s)E\varepsilon(t)\varepsilon(s) \right) \right] R_{\bar{\varphi}}^{-1} H^T \\
&= HR_{\bar{\varphi}}^{-1} \left(\sum_\tau \left(r_{\bar{\varphi}\varepsilon}(\tau) r_{\bar{\varphi}\varepsilon}^T(-\tau) + R_{\bar{\varphi}}(\tau) r_\varepsilon(\tau) \right) \right) R_{\bar{\varphi}}^{-1} H^T.
\end{aligned} \tag{D.8}$$

Next we consider the case of non-Gaussian noise. In this case relation (D.2) cannot be used. Instead, we follow the techniques used in [6], *i.e.* utilize 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 the fourth moment μ. It holds that

$$x_k(t) = H_k(q^{-1})e(t), \quad H_k(q^{-1}) = \sum_{j=0}^\infty h_{kj} q^{-j}, \quad k = 1, 2, 3, 4. \tag{D.9}$$

Then it holds that

$$\begin{aligned}
& Ex_1(t)x_2(t)x_3(t)x_4(t) \\
= & \sum_{i=0}^\infty \sum_{j=0}^\infty \sum_{k=0}^\infty \sum_{l=0}^\infty h_{1i}h_{2j}h_{1i}h_{3k}h_{4l} Ee(t-i)e(t-j)e(t-k)e(t-l).
\end{aligned} \tag{D.10}$$

As the white noise $e(t)$ has zero mean and is uncorrelated at different time points, the expectation in (D.10) is nonzero only when the time arguments are pairwise equal or all equal. Therefore

$$\begin{aligned}
& Ee(t-i)e(t-j)e(t-k)e(t-l) \\
= & \lambda^2[\delta_{i,j}\delta_{k,l} + \delta_{i,k}\delta_{j,l} + \delta_{i,l}\delta_{j,k}] + (\mu - 3\lambda^2)\delta_{i,j}\delta_{j,k}\delta_{k,l}.
\end{aligned}$$

Now we have

$$\begin{aligned}
& Ex_1(t)x_2(t)x_3(t)x_4(t) \\
= \quad & [Ex_1(t)x_2(t)][Ex_3(t)x_4(t)] + [Ex_1(t)x_3(t)][Ex_2(t)x_4(t)] \\
& + [Ex_1(t)x_4(t)][Ex_2(t)x_3(t)] + (\mu - 3\lambda^2)\sum_{i=0}^{\infty} h_{1i}h_{2i}h_{3i}h_{4i}. \qquad \text{(D.11)}
\end{aligned}$$

We see that using the first part of (D.11) leads precisely to the Gaussian formula with Q^G. For the remaining terms (that vanishes in the Gaussian case), we write the additional matrix in block form as

$$Q^{NG} = \begin{pmatrix} Q_{11}^{NG} & Q_{12}^{NG} & Q_{13}^{NG} \\ Q_{21}^{NG} & Q_{22}^{NG} & Q_{23}^{NG} \\ Q_{31}^{NG} & Q_{32}^{NG} & Q_{33}^{NG} \end{pmatrix}. \qquad \text{(D.12)}$$

In what follows we neglect the influence of the noise-free input $u_0(t)$, which is captured in the term Q^G. Also only the "non Gaussian" contributions are treated. Cf (D.1). We have, for $i, j = 1, \ldots, n_a$,

$$\begin{aligned}
(Q_{11})_{i,j} &= \sum_{\tau=-\infty}^{\infty} E\tilde{y}(t-i)A(q^{-1})\tilde{y}(t)A(q^{-1})\tilde{y}(t+\tau)\tilde{y}(t+\tau-j) \\
&= \sum_{\tau} E\tilde{y}(t-i)\tilde{y}(t+\tau-j)\sum_k a_k\tilde{y}(t-k)\sum_l a_l\tilde{y}(t+\tau-l)
\end{aligned}$$

By using the relation (D.11)

$$\begin{aligned}
(Q_{11}^{NG})_{i,j} &= (\mu_y - 3\lambda_y^2)\sum_{\tau=-\infty}^{\infty} a_i\delta_{i,j-\tau}a_j \\
&= (\mu_y - 3\lambda_y^2)a_ia_j.
\end{aligned}$$

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

$$(Q_{11}^{NG}) = \begin{pmatrix} (\mu_y - 3\lambda_y^2)\mathbf{a}\mathbf{a}^T & \mathbf{0} \\ \mathbf{0} & (\mu_u - 3\lambda_u^2)\mathbf{b}\mathbf{b}^T \end{pmatrix}. \qquad \text{(D.13)}$$

Continuing with the block Q_{12} leads to, for $i = 1, \ldots, n_a$,

$$\begin{aligned}
(Q_{12})_i &= -\sum_{\tau=-\infty}^{\infty} E\tilde{y}(t+\tau-i)A(q^{-1})\tilde{y}(t+\tau)A_{LS}(q^{-1})\tilde{y}(t)A(q^{-1})\tilde{y}(t) \\
&= -\sum_{\tau} E\tilde{y}(t+\tau-i)\sum_k a_k\tilde{y}(t+\tau-k)\sum_l a_{LS,l}\tilde{y}(t-l)\sum_m a_m\tilde{y}(t-m)
\end{aligned}$$

From relation (D.11), we get

$$(Q_{12}^{NG})_i = -(\mu_y - 3\lambda_y^2)\sum_{\tau=-\infty}^{\infty} a_ia_{LS,i-\tau}a_{i-\tau}. \qquad \text{(D.14)}$$

Similarly, for $i = n_a + j, \quad j = 1, \ldots, n_b$ one gets

$$(Q_{12}^{NG})_i = (\mu_u - 3\lambda_u^2) \sum_{\tau=-\infty}^{\infty} b_j b_{LS,j-\tau} b_{j-\tau}. \tag{D.15}$$

Summarizing the findings in (D.14) and (D.15) gives

$$Q_{12}^{NG} = \begin{pmatrix} -(\mu_y - 3\lambda_y^2)\mathbf{a}(\bar{\mathbf{a}}_{LS}^T \bar{\mathbf{a}}) \\ (\mu_u - 3\lambda_u^2)\mathbf{b}(\mathbf{b}_{LS}^T \mathbf{b}) \end{pmatrix}. \tag{D.16}$$

where

$$\bar{\mathbf{a}} = \begin{pmatrix} 1 \\ \mathbf{a} \end{pmatrix}, \quad \bar{\mathbf{a}}_{LS} = \begin{pmatrix} 1 \\ \mathbf{a}_{LS} \end{pmatrix}. \tag{D.17}$$

Proceeding to Q_{22}, we get the effect due to the output and input noises as

$$\begin{aligned} Q_{22} &= \sum_{\tau=-\infty}^{\infty} E A_{LS}(q^{-1})\tilde{y}(t+\tau) A(q^{-1})\tilde{y}(t+\tau) A_{LS}(q^{-1})\tilde{y}(t) A(q^{-1})\tilde{y}(t) \\ &\quad + \sum_{\tau=-\infty}^{\infty} E B_{LS}(q^{-1})\tilde{u}(t+\tau) B(q^{-1})\tilde{u}(t+\tau) B_{LS}(q^{-1})\tilde{u}(t) B(q^{-1})\tilde{u}(t) \\ &= \sum_{\tau=-\infty}^{\infty} E \sum_i a_{LS,i}\tilde{y}(t+\tau-i) \sum_k a_k \tilde{y}(t+\tau-k) \sum_l a_{LS,l}\tilde{y}(t-l) \sum_m a_m \tilde{y}(t-m) \\ &\quad + \sum_{\tau=-\infty}^{\infty} E \sum_i b_{LS,i}\tilde{u}(t+\tau-i) \sum_k b_k \tilde{u}(t+\tau-k) \sum_l b_{LS,l}\tilde{u}(t-l) \sum_m b_m \tilde{u}(t-m) \end{aligned}$$

Relation (D.11) gives

$$\begin{aligned} Q_{22}^{NG} &= (\mu_y - 3\lambda_y^2)(\sum_i a_{LS,i} a_i)^2 + (\mu_u - 3\lambda_u^2)(\sum_i b_{LS,i} b_i)^2 \\ &= (\mu_y - 3\lambda_y^2)(\bar{\mathbf{a}}_{LS}^T \bar{\mathbf{a}})^2 + (\mu_u - 3\lambda_u^2)(\mathbf{b}_{LS}^T \mathbf{b})^2. \end{aligned} \tag{D.18}$$

Next try to find the block Q_{13}^{NG}. Firstly, set

$$W = \sum_{\tau=-\infty}^{\infty} E\varphi(t+\tau)\varepsilon(t+\tau)\bar{\varphi}^T(t)\varepsilon(t).$$

Utilizing the relation $\bar{\varphi}^T(t) = (\varphi(t) \quad \underline{\varphi}^T(t))$, it results in $W = (W_1 \quad W_2)$ with

$$W_1 = \sum_{\tau=-\infty}^{\infty} E\varphi(t+\tau)\varepsilon(t+\tau)\varphi^T(t)\varepsilon(t), \tag{D.19}$$

$$W_2 = \sum_{\tau=-\infty}^{\infty} E\varphi(t+\tau)\varepsilon(t+\tau)\underline{\varphi}^T(t)\varepsilon(t). \tag{D.20}$$

Compare to the expression of Q_{11} in (D.1), we have

$$W_1^{NG} = Q_{11}^{NG} = \begin{pmatrix} (\mu_y - 3\lambda_y^2)\mathbf{a}\mathbf{a}^T & \mathbf{0} \\ \mathbf{0} & (\mu_u - 3\lambda_u^2)\mathbf{b}\mathbf{b}^T \end{pmatrix}.$$

For W_2, if the extended vector is $\underline{\varphi}(t) = -y(t-n_a-1)$, for $i = 1, \ldots, n_a$,

$$\begin{aligned}(W_2)_i &= \sum_{\tau=-\infty}^{\infty} E\tilde{y}(t+\tau-i)A(q^{-1})\tilde{y}(t+\tau)\tilde{y}(t-n_a-1)A(q^{-1})\tilde{y}(t) \\ &= \sum_{\tau} E\tilde{y}(t+\tau-i)\sum_k a_k\tilde{y}(t+\tau-k)\tilde{y}(t-na-1)\sum_l a_l\tilde{y}(t-l).\end{aligned}$$

Then

$$\begin{aligned}(W_2^{NG})_i &= (\mu_y - 3\lambda_y^2)\sum_{\tau=-\infty}^{\infty} a_i\delta_{i-\tau,n_a+1}a_{n_a+1} \\ &= (\mu_y - 3\lambda_y^2)a_i a_{n_a+1} \\ &= 0.\end{aligned}$$

The last step comes from the fact that the true value of a_{n_a+1} equals zero. As $\tilde{u}$ and $\tilde{y}$ are white noises with zero means, and both are uncorrelated with each other, it results $(W_2^{NG})_i = 0$, when $i = na+j, \quad j = 1, \ldots, n_b$.

If the extended vector is $\underline{\varphi}(t) = u(t-n_b-1)$, similar reason as above gives $(W_2^{NG})_i = 0$ for $i = 1, \ldots, n_a$. And for $i = na+j, \quad j = 1, \ldots, n_b$, we have

$$(W_2)_i = \sum_{\tau=-\infty}^{\infty} E\tilde{u}(t+\tau-i)B(q^{-1})\tilde{u}(t+\tau)\tilde{u}(t-n_b-1)B(q^{-1})\tilde{u}(t),$$

$$(W_2^{NG})_i = (\mu_u - 3\lambda_u^2)b_i b_{n_b+1} = 0.$$

To sum up, we have $(W_2^{NG}) = \mathbf{0}_{(na+nb)\times 1}$ for all the cases.

Next, consider the structure of Q_{13}. Using (D.19) and (D.20) gives

$$\begin{aligned}Q_{13}^{NG} &= \left(\sum_{\tau=-\infty}^{\infty} E\varphi(t+\tau)\varepsilon(t+\tau)\bar{\varphi}^T(t)\varepsilon(t)\right)^{NG} R_{\bar{\varphi}}^{-1}H^T \\ &= \begin{pmatrix} W_1^{NG} & W_2^{NG}\end{pmatrix} R_{\bar{\varphi}}^{-1}H^T \\ &= \begin{pmatrix} Q_{11}^{NG} & \mathbf{0}\end{pmatrix} R_{\bar{\varphi}}^{-1}H^T \\ &= \begin{pmatrix} (\mu_y - 3\lambda_y^2)\mathbf{a}\mathbf{a}^T & \mathbf{0}_{n_a\times n_b} & \mathbf{0}_{n_a\times 1} \\ \mathbf{0}_{n_b\times n_a} & (\mu_u - 3\lambda_u^2)\mathbf{b}\mathbf{b}^T & \mathbf{0}_{n_b\times 1}\end{pmatrix} R_{\bar{\varphi}}^{-1}H^T. \qquad \text{(D.21)}\end{aligned}$$

For the block Q_{23}^{NG} we note

$$\begin{aligned}Q_{23}^{NG} &= \left(\sum_{\tau=-\infty}^{\infty} E\varepsilon_{LS}(t+\tau)\varepsilon(t+\tau)\bar{\varphi}^T(t)\varepsilon(t)\right)^{NG} R_{\bar{\varphi}}^{-1}H^T \\ &= \left(\sum_{\tau=-\infty}^{\infty} E\varepsilon_{LS}(t+\tau)\varepsilon(t+\tau)\varphi^T(t)\varepsilon(t),\right. \\ &\qquad \left.\sum_{\tau=-\infty}^{\infty} E\varepsilon_{LS}(t+\tau)\varepsilon(t+\tau)\underline{\varphi}^T(t)\varepsilon(t)\right)^{NG} R_{\bar{\varphi}}^{-1}H^T.\end{aligned}$$

By setting

$$W_3 = \sum_{\tau=-\infty}^{\infty} E\varepsilon_{LS}(t+\tau)\varepsilon(t+\tau)\underline{\varphi}^T(t)\varepsilon(t) \quad \text{(D.22)}$$

and considering the expression for Q_{12}^{NG}, we get

$$Q_{23}^{NG} = \left((Q_{12}^{NG})^T \quad W_3^{NG}\right) R_{\bar{\varphi}}^{-1} H^T.$$

When the extended vector is $\underline{\varphi}(t) = -y(t - n_a - 1)$,

$$W_3 = \sum_{\tau=-\infty}^{\infty} E \sum_i a_{LS,i}\tilde{y}(t+\tau-i) \sum_k a_k \tilde{y}(t+\tau-k) \sum_l a_l \tilde{y}(t-l)\tilde{y}(t-na-1),$$

$$W_3^{NG} = (\mu_y - 3\lambda_y^2) \sum_{\tau=-\infty}^{\infty} a_{LS,n_a+1+\tau} a_{n_a+1+\tau} a_{n_a+1} = 0.$$

Similarly, when the extended vector is $\underline{\varphi}(t) = u(t - n_b - 1)$,

$$W_3^{NG} = (\mu_u - 3\lambda_u^2) \sum_{\tau=-\infty}^{\infty} b_{LS,n_b+1+\tau} b_{n_b+1+\tau} b_{n_b+1} = 0.$$

Hence,

$$\begin{aligned} Q_{23}^{NG} &= \left((Q_{12}^{NG})^T \quad W_3^{NG}\right) R_{\bar{\varphi}}^{-1} H^T \\ &= \left(-(\mu_y - 3\lambda_y^2)\mathbf{a}^T(\bar{\mathbf{a}}_{LS}^T \bar{\mathbf{a}}) \quad (\mu_u - 3\lambda_u^2)\mathbf{b}^T(\mathbf{b}_{LS}^T \mathbf{b}) \quad 0 \right) R_{\bar{\varphi}}^{-1} H^T. \end{aligned} \quad \text{(D.23)}$$

Finally, the similar process is done for the block Q_{33}^{NG}. We set

$$W_4 = \sum_{\tau=-\infty}^{\infty} E\underline{\varphi}(t+\tau)\varepsilon(t+\tau)\underline{\varphi}^T(t)\varepsilon(t). \quad \text{(D.24)}$$

For $\underline{\varphi}(t) = -y(t - n_a - 1)$,

$$W_4^{NG} = (\mu_y - 3\lambda_y^2) \sum_{\tau=-\infty}^{\infty} a_{n_a+1}^2 = 0,$$

and for $\underline{\varphi}(t) = u(t - n_b - 1)$,

$$W_4^{NG} = (\mu_u - 3\lambda_u^2) \sum_{\tau=-\infty}^{\infty} b_{n_b+1}^2 = 0.$$

Hence the result is $W_4^{NG} = 0$.

Considering the structure of Q_{33}, then we have

$$\begin{aligned}
Q_{33}^{NG} &= HR_{\bar{\varphi}}^{-1}\left(\sum_{\tau=-\infty}^{\infty} E\bar{\varphi}(t+\tau)\varepsilon(t+\tau)\bar{\varphi}^T(t)\varepsilon(t)\right)^{NG} R_{\bar{\varphi}}^{-1}H^T \\
&= HR_{\bar{\varphi}}^{-1}\begin{pmatrix} \sum_\tau E\varphi(t+\tau)\varepsilon(t+\tau)\varphi^T(t)\varepsilon(t) & \sum_\tau E\varphi(t+\tau)\varepsilon(t+\tau)\underline{\varphi}^T(t)\varepsilon(t) \\ \sum_\tau E\underline{\varphi}(t+\tau)\varepsilon(t+\tau)\varphi^T(t)\varepsilon(t) & \sum_\tau E\underline{\varphi}(t+\tau)\varepsilon(t+\tau)\underline{\varphi}^T(t)\varepsilon(t) \end{pmatrix}^{NG} R_{\bar{\varphi}}^{-1}H^T \\
&= HR_{\bar{\varphi}}^{-1}\begin{pmatrix} Q_{11}^{NG} & W_2^{NG} \\ (W_2^{NG})^T & W_4^{NG} \end{pmatrix} R_{\bar{\varphi}}^{-1}H^T \\
&= HR_{\bar{\varphi}}^{-1}\begin{pmatrix} (\mu_y - 3\lambda_y^2)\mathbf{a}\mathbf{a}^T & \mathbf{0}_{n_a\times n_b} & \mathbf{0}_{n_a\times 1} \\ \mathbf{0}_{n_b\times n_a} & (\mu_u - 3\lambda_u^2)\mathbf{b}\mathbf{b}^T & \mathbf{0}_{n_b\times 1} \\ \mathbf{0}_{1\times n_a} & \mathbf{0}_{1\times n_b} & 0 \end{pmatrix} R_{\bar{\varphi}}^{-1}H^T.
\end{aligned} \tag{D.25}$$

References

[1] K. V. Fernando and H. Nicholson. Identification of linear systems with input and output noise: the Koopmans–Levin method. *IEE Proceedings, Part D*, 132(1):30–36, January 1985.

[2] L. Ljung. Some limit results for functionals of stochastic processes. Technical Report Rep. LiTH-ISY-I-0167, Dept. Elec. Eng. linköping Univ. Linköping, Sweden, 1997. Technical Report.

[3] R. Pintelon and J. Schoukens. *System Identification. A Frequency Domain Approach*. IEEE Press, New York, NY, USA, 2001.

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

[5] T. Söderström. Accuracy analysis of the Frisch estimates for identifying errors-in-variables systems. In *Proc. 44th IEEE CDC/ECC 2005*, Seville, Spain, December 12-15 2005.

[6] T. Söderström. Statistical analysis of the Frisch scheme for identifying errors-in-variables systems. In *Technical Report No:2006-002*, Department of Information Technology, Uppsala University, Uppsala, Sweden, January 2006.

[7] T. Söderström, M. Hong, and W.X. Zheng. Convergence properties of bias-eliminating algorithms for errors-in-variables identification. *International Journal of Adaptive Control and Signal Processing*, 19:703–722, November 2005.

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

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

[10] J. K. Tugnait and Y. Ye. Stochastic system identification with noisy input-output measurement using polyspectra. *IEEE Transactions on Automatic Control*, AC-40:670–683, 1995.

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

[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 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.