15th Triennial World Congress, Barcelona, Spain

IDENTIFICATION OF DYNAMIC ERRORS-IN-VARIABLES MODEL USING PREFILTERED DATA [1]

Kaushik Mahata and **Torsten Söderström**

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

Abstract: Computationally efficient identification of dynamic errors-in-variables model is considered in this paper. The instrumental variable (IV) method is computationally efficient but it suffers from poor small-sample properties of the estimated parameters. The method presented in this work uses the prefiltered data. The input-output data is passed through a pair of user defined prefilters and the output data from the prefilters is subjected to a least-squares like algorithm. Compared to the IV approach, the proposed method shows a significant improvement in the small-sample properties of the MA parameter estimates, without any increase in the computational load. ©IFAC 2002

Keywords: System Identification, Errors-in-variables, Instrumental variable method, Prefiltering.

1. INTRODUCTION

Consider a linear system with noise-free input $u_0(t)$ linked to its true output $y_0(t)$ by the difference equation

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

where $A(q^{-1})$ and $B(q^{-1})$ are co-prime polynomials of known degree n and type

$$\begin{aligned} A(q^{-1}) &= 1 + a_1\, q^{-1} + \cdots + a_n\, q^{-n} \\ B(q^{-1}) &= b_1\, q^{-1} + \cdots + b_n\, q^{-n} \end{aligned} \quad . \tag{2}$$

Note that q^{-1} is the backward shift operator, i.e. $q^{-1}\, x(t) = x(t-1)$. The available input data $u(t)$ and output data $y(t)$ are noise corrupted, i.e.

$$\begin{aligned} u(t) &= u_0(t) + \tilde{u}(t) \\ y(t) &= y_0(t) + \tilde{y}(t) \end{aligned} \quad , \tag{3}$$

where $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually independent zero mean white noise sequences, each independent of $u_0(t)$ and $y_0(t)$, with variances λ_u and λ_y, respectively. The problem under consideration is to identify the vector of system parameters

$$\boldsymbol{\theta} = \begin{bmatrix} a_1 & \ldots & a_n & b_1 & \ldots & b_n \end{bmatrix}^T \tag{4}$$

from the available measurements $\{u(t)\}_{t=1}^{N}$ and $\{y(t)\}_{t=1}^{N}$, assuming λ_u and λ_y are unknown.

The problem introduced here is often referred as a dynamic "errors–in–variables" problem and is found in many disciplines of science, as proved by the several applications collected in (van Huffel, 1997). Several authors have considered this problem with different approaches. For example, we can distinguish between time–domain (Ljung, 1999) and frequency-domain methods (Pintelon

[1] This work was supported by Swedish Research Council for Engineering Sciences under contract 98-654.

and Schoukens, 2001), approaches exploiting deterministic signals (e.g. periodic inputs (Forsell *et al.*, 1999)) or stochastic processes, methods that make use of second-order statistics only or methods based on higher order cumulant statistics (Tugnait, 1992). When only second-order statistics are exploited it is a well-known result that, in general, the identification of errors-in-variables models cannot admit a single solution (Anderson and Deistler, 1984). An overall view of different approaches is presented in (Söderström *et al.*, 2001).

One of the computationally efficient solutions of the errors-in-variables problem is instrumental variable method and related subspace fitting approaches (Stoica *et al.*, 1995), (Söderström and Mahata, 2001). The accuracy of the parameter estimates obtained by IV based methods is often poor, particularly when the system poles are close to the unit circle or the input signal spectrum is peaky. This poor accuracy of IV based methods can mainly be attributed to the fact that they use the estimates of correlation of the data at 'high' lags. In this paper, we propose an approach where the data is prefiltered by two user defined filters. The design of the prefilters enables us to use a least-squares like method, where correlation estimates at lower lags are used. We also demonstrate that significant improvement in the estimate accuracy is achieved for the same computational cost.

2. IDENTIFICATION METHODS

2.1 *The instrumental variable estimator*

Any scalar or vector quantity $x(t)$, which is a function of the data $\{u(t)\}_{t=1}^{N}$, $\{y(t)\}_{t=1}^{N}$ is composed of two parts, i.e.

$$x(t) = x_0(t) + \tilde{x}(t),$$

where we denote the contribution of noise free part of the data, i.e. $\{u_0(t)\}_{t=1}^{N}$ and $\{y_0(t)\}_{t=1}^{N}$ by $x_0(t)$, while $\tilde{x}(t)$ is used to denote the contribution of the measurement noise sequences $\{\tilde{u}(t)\}_{t=1}^{N}$ and $\{\tilde{y}(t)\}_{t=1}^{N}$. For example, consider the regressor vector

$$\begin{aligned}\boldsymbol{\psi}(t) &= \big[\, -y(t-1)\ \ldots\ -y(t-n) \\ &\qquad u(t-1)\ \ldots\ u(t-n)\,\big]^T \\ &= \boldsymbol{\psi}_0(t) + \tilde{\boldsymbol{\psi}}(t), \end{aligned} \tag{5}$$

where

$$\begin{aligned}\boldsymbol{\psi}_0(t) &= \big[\, -y_0(t-1)\ \ldots\ -y_0(t-n) \\ &\qquad u_0(t-1)\ \ldots\ u_0(t-n)\,\big]^T, \end{aligned} \tag{6}$$

$$\begin{aligned}\tilde{\boldsymbol{\psi}}(t) &= \big[\, -\tilde{y}(t-1)\ \ldots\ -\tilde{y}(t-n) \\ &\qquad \tilde{u}(t-1)\ \ldots\ \tilde{u}(t-n)\,\big]^T. \end{aligned} \tag{7}$$

Let us introduce the extended regressor vector

$$\boldsymbol{\phi}(t) = \big[\, -y(t)\ \boldsymbol{\psi}^T(t)\,\big]^T. \tag{8}$$

It follows from (1), (2) and (6) that

$$\boldsymbol{\psi}_0^T(t)\boldsymbol{\theta} = y_0(t), \tag{9}$$

and

$$\begin{aligned} v(t) &= y(t) - \boldsymbol{\psi}^T(t)\boldsymbol{\theta} \\ &= A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t) \\ &= C(q^{-1})e(t), \end{aligned} \tag{10}$$

where $e(t)$, the innovation of $v(t)$, is a zero mean white noise with variance λ. The polynomial $C(q^{-1})$ and λ can be found by the spectral factorization

$$\begin{aligned}\lambda C(q)C(q^{-1}) &= \lambda_y A(q)A(q^{-1}) \\ &\quad + \lambda_u B(q)B(q^{-1}) \end{aligned} \tag{11}$$

From (9) we see that, for any data-dependent vector $\mathbf{z}(t)$ of dimension $p \geq 2n$, which is correlated with $\boldsymbol{\psi}_0(t)$, we have

$$E\mathbf{z}(t)\boldsymbol{\psi}_0^T(t)\boldsymbol{\theta} = E\mathbf{z}(t)y_0(t). \tag{12}$$

Now, if it is possible to construct $\mathbf{z}(t)$ in such a way that

$$R_{\mathbf{z}\tilde{\boldsymbol{\phi}}} = E\mathbf{z}(t)\tilde{\boldsymbol{\phi}}^T(t) = 0, \tag{13}$$

we can construct an instrumental variable estimate $\hat{\boldsymbol{\theta}}_{iv}$ of $\boldsymbol{\theta}$ as

$$\hat{\boldsymbol{\theta}}_{iv} = \hat{R}_{\mathbf{z}\boldsymbol{\psi}}^{\dagger}\hat{R}_{\mathbf{z}y}. \tag{14}$$

Note that, we use $\hat{X}$ to denote an estimate of X. We stress that $\hat{\boldsymbol{\theta}}_{iv}$ exists if $R_{\mathbf{z}\boldsymbol{\psi}_0} = E\mathbf{z}(t)\boldsymbol{\psi}_0^T(t)$ is full rank. This is the persistence of excitation like condition on the noise-free input $u_0(t)$. This assumption is not restrictive in the sense that most of the inputs satisfy this condition. We refer the reader to the discussion on generic consistency in (Söderström and Stoica, 1989), (Söderström and Stoica, 1983) and (Stoica *et al.*, 1995) for details . We have following result for $\hat{\boldsymbol{\theta}}_{iv}$.

Lemma 1. If the measurement noise sequences are gaussian, $\hat{\boldsymbol{\theta}}_{iv}$ is asymptotically Gaussian distributed as

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{iv} - \boldsymbol{\theta}) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0, P_{iv}), \tag{15}$$

$$P_{iv} = \lambda R_{\mathbf{z}\boldsymbol{\psi}}^{\dagger}\mathrm{cov}\left[C(q^{-1})z(t)\right] R_{\mathbf{z}\boldsymbol{\psi}}^{\dagger T}, \tag{16}$$

where $X^{\dagger}$ denotes the pseudo-inverse of X.

Proof. See (Söderström and Mahata, 2001). ■

2.2 *Estimation using prefiltered data*

The often encountered poor accuracy of the instrumental variable estimator is a direct consequence of (13). The elements of $\mathbf{z}(t)$ must be sufficiently delayed with respect to the elements of $\psi(t)$ in order to satisfy (13). For instance, we can use $\mathbf{z}(t) = \psi(t-\tau)$ for $\tau \geq n$. Since it is difficult to accurately estimate the correlation of the data at higher lags, often poor estimates of $R_{\mathbf{z}\psi}$ and $R_{\mathbf{z}y}$ may result, which leads to poor accuracy of the estimates. In this section, an attempt will be made to remedy this problem by passing the data through a pair of user defined filters.

Let us assume that the input-output data are passed through a pair of FIR filters $G_1(q^{-1})$ and $G_2(q^{-1})$, each of order $m > n$. Let us denote for $i = 1,2$

$$u_i(t) = G_i(q^{-1})u(t), \quad (17)$$
$$y_i(t) = G_i(q^{-1})y(t), \quad (18)$$
$$v_i(t) = G_i(q^{-1})v(t), \quad (19)$$
$$\psi_i(t) = G_i(q^{-1})\psi(t), \quad (20)$$
$$\phi_i(t) = G_i(q^{-1})\phi(t). \quad (21)$$

It is straightforward to see from (9) that for $i = 1,2$

$$\psi_{i0}^T(t)\theta = y_{i0}(t). \quad (22)$$

Thus, if it is possible to design $G_1(q^{-1})$ and $G_2(q^{-1})$ such that

$$R_{\tilde{\phi}_1\tilde{\phi}_2} = 0, \quad (23)$$

we can obtain two estimators of θ as

$$\hat{\theta}_1 = \hat{R}^{\dagger}_{\phi_1\psi_2}\hat{R}_{\phi_1 y_2}, \quad (24)$$
$$\hat{\theta}_2 = \hat{R}^{\dagger}_{\phi_2\psi_1}\hat{R}_{\phi_2 y_1}, \quad (25)$$

provided $R_{\phi_{10}\psi_{20}}$ and $R_{\phi_{20}\psi_{10}}$, respectively, are full rank. We have the following result concerning the design of $G_1(q^{-1})$ and $G_2(q^{-1})$ in order to satisfy (23).

Lemma 2. Let us define the polynomial $H(q)$ as

$$H(q) = q^m G_1(q^{-1})G_2(q) = \sum_{i=0}^{2m} h_i q^i. \quad (26)$$

Then (23) is satisfied if

$$h_i = 0, \quad m-n \leq i \leq m+n. \quad (27)$$

Proof. See appendix A. ∎

Thus, we can start with $H(q)$, the polynomial of order $2m$ satisfying (27) and factor it accordingly to obtain $G_1(q^{-1})$ and $G_2(q^{-1})$. We have the following asymptotic result for the accuracy of the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$.

Lemma 3. Define the polynomials

$$C(q^{-1})G_1(q^{-1})G_2(q^{-1}) = G(q^{-1}), \quad (28)$$
$$A^2(q^{-1})H^2(q^{-1}) = \sum_{i=0}^{2m+2n} \beta_i^{(aa)} q^{-i}, \quad (29)$$
$$A(q^{-1})B(q^{-1})H^2(q^{-1}) = \sum_{i=0}^{2m+2n} \beta_i^{(ab)} q^{-i}, \quad (30)$$
$$B^2(q^{-1})H^2(q^{-1}) = \sum_{i=0}^{2m+2n} \beta_i^{(bb)} q^{-i}. \quad (31)$$

If the measurement noise sequences are Gaussian, $\hat{\theta}_i$, $i = 1,2$ are asymptotically Gaussian distributed as

$$\sqrt{N}(\hat{\theta}_i - \theta) \xrightarrow{N\to\infty} \mathcal{N}(0, P_i), \quad (32)$$
$$P_i = R^{\dagger}_{\phi_i\psi_j} Q R^{\dagger T}_{\phi_i\psi_j}, i \neq j \in \{1,2\} \quad (33)$$

where

$$Q = \lambda \mathrm{cov}\left[G(q^{-1})\phi(t)\right] + Q_e \quad (34)$$

and Q_e is given by (B.19)-(B.23) in appendix.

Proof. See appendix B. ∎

3. NUMERICAL ILLUSTRATION

To illustrate numerically the identification methods introduced in the previous section, we shall consider the second order system with

$$A(q^{-1}) = 1 - 1.5q^{-1} + 0.7q^{-2}, \quad (35)$$
$$B(q^{-1}) = 1.0q^{-1} + 0.5q^{-2}. \quad (36)$$

The true input $u_0(t)$ is an ARMA process given by

$$u_0(t) = \frac{1 + 2q^{-1} + q^{-2}}{1 - 1.8q^{-1} + 0.9q^{-2}}\epsilon(t), \quad (37)$$

where $\epsilon(t)$ is a zero mean white noise sequence with variance 0.25. The variance of both the measurement noise sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ is 4. As an illustrative example, we consider the polynomial $H(q^{-1})$ as

$$H(q^{-1}) = 1 - q^{-6}. \quad (38)$$

Note that, the value of m is set to 3 and (27) is satisfied with this choice of $H(q^{-1})$. Among several possibilities of factorising $H(q^{-1})$ into a

Estimate		a_1	a_2	b_1	b_2
$\hat{\theta}_1$	Mean	-1.50	0.70	1.00	0.49
	S.D(n)	0.037	0.029	0.165	0.208
	S.D(t)	0.031	0.024	0.120	0.156
$\hat{\theta}_2$	Mean	-1.50	0.70	1.00	0.49
	S.D(n)	0.036	0.028	0.127	0.162
	S.D(t)	0.031	0.022	0.120	0.156
$\hat{\theta}_{iv}$ $(p = 6)$	Mean	-1.50	0.70	0.88	0.64
	S.D(n)	0.022	0.029	0.792	0.944
	S.D(t)	0.015	0.022	0.556	0.661
$\hat{\theta}_{iv}$ $(p = 10)$	Mean	-1.50	0.70	0.99	0.51
	S.D(n)	0.014	0.013	0.239	0.274
	S.D(t)	0.013	0.013	0.215	0.246

Table 1. Simulation results form 500 Monte-Carlo simulations.

Estimate	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_{iv}$ $(p = 6)$	$\hat{\theta}_{iv}$ $(p = 10)$
	30579	30579	31741	49173

Table 2. Computational load involved in different algorithms, in terms of Matlab flops.

pair of polynomials of order 3, see (26), we have arbitrarily taken $G_1(q^{-1})$ and $G_2(q^{-1})$ as

$$G_1(q^{-1}) = 1 + 2q^{-1} + 2q^{-2} + q^{-3}, \quad (39)$$

$$G_2(q^{-1}) = 1 - 2q^{-1} + 2q^{-2} - q^{-3}. \quad (40)$$

As candidate algorithms, along with the proposed algorithm using prefiltered data, we consider the IV algorithms with the instrument vector given by

$$\mathbf{z}(t) = \left[\, u(t-3) \;\ldots\; u(t-p-2) \,\right]^T \quad (41)$$

for different values of p, the dimension of the instrument vector. We have summarised the simulation results in Table 1. The results are based on 500 independent Monte-Carlo simulations. The number of data points N in each simulation was 300. For each method we have given the sample mean and sample standard deviation denoted by SD(n). The standard deviations of the parameters obtained from the Monte-Carlo simulations are compared with the corresponding asymptotic standard deviation denoted by SD(t) calculated using Lemma 1 and Lemma 3.

In Table 2, we have compared the number of Matlab floating point operations involved in each of the algorithms.

4. DISCUSSION

From Table 1, we see that the IV estimates of the A parameters are reasonably accurate, but the corresponding accuracy of the B parameters are very poor for $p = 6$. Increasing the instrument dimension from 6 to 10, an improvement in the B parameter estimation can be noticed. On the other hand, using the approach using the prefiltered data, the improvement in the B parameter estimation is significant. One demerit of the proposed algorithm is the slight loss of accuracy in A parameter estimation. If we compare the computation involved in the algorithms in Table 2, we see that increasing the dimension of the instrument from 6 to 10, the computational load is increased significantly, while the proposed algorithm is of less computational complexity.

It is also interesting to note that, although the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ have been designed in such a way that they have the same asymptotic properties, their small sample properties differ. As can be seen from Table 1, $\hat{\theta}_2$ performs better than $\hat{\theta}_1$. This can be attributed to the fact that $\hat{\theta}_2$ uses $\psi_1(t)$ as the regressor vector and $\phi_2(t)$ as the instrument vector. Since in the example considered here, the system to be identified is low pass and $G_1(q^{-1})$ is also a low pass filter, the system properties are excited better using $\psi_1(t)$ as the regressor vector.

5. CONCLUSIONS

In this work, we have presented a computationally economic algorithm to identify the dynamic errors-in-variables model. The algorithm shows significant improvement in the accuracy of the system parameter estimates over the instrumental variable method. The method can also be used, when the measurement noise sequences satisfy more generalised constraints. The same algorithm can handle the case when $\tilde{u}(t)$ and $\tilde{y}(t)$ are correlated with each other. In case the sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ are moving average processes of finite order, we can increase the parameter m to account for that. The polynomial $H(q^{-1})$ is to be designed by the user. In the example in the last section $H(q^{-1})$ was selected in an ad-hoc manner. But it is possible to make it data dependent, i.e. one can think of an iterative procedure of adjusting the coefficients of the polynomial $H(q^{-1})$, based on the last estimates of the system parameters, to arrive at statistically more accurate estimates. For example, the expression (3.3) can be used to optimise the covariance matrix of the estimates in some sense to derive such a scheme.

6. REFERENCES

Anderson, B. D. O. and M. Deistler (1984). Identifiability of dynamic errors-in-variables models. *J. Time Series Analysis* **5**, 1–13.

Forsell, U., F. Gustafsson and T. McKelvey (1999). Time-domain identification of dynamic errors-in-variables systems using periodic excitation signals. In: *Proc. IFAC 14th World Congress*. Beijing, P.R.China.

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

Pintelon, R. and J. Schoukens (2001). *System Identification: A Frequency Domain Approach.* IEEE Press and John Wiley.

Söderström, T. and K. Mahata (2001). On subspace fitting and instrumental variable approaches for identification of error-in-variables models. Submitted to International Journal of Control.

Söderström, T. and P. Stoica (1983). *Instrumental Variable Methods for System Identification.* Springer-Verlag, Berlin.

Söderström, T. and P. Stoica (1989). *System Identification.* Prentice Hall International. Hemel Hempstead, UK.

Söderström, T., U. Soverini and K. Mahata (2001). Perspectives on errors-in-variables estimation. In: *3rd International Workshop on TLS and Errors-in-Variables Modelling,.* Leuven, Belgium.

Stoica, P, M Cedervall and A Eriksson (1995). Combined instrumental variable and subspace fitting approach to parameter estimation of noisy input-output systems. *IEEE Transactions on Signal Processing* **43**, 2386–2397.

Tugnait, J. K. (1992). Stochastic system identification with noisy input using cumulant statistics. *IEEE Transactions on Automatic Control* **AC-37**, 476–485.

van Huffel, S. (1997). *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modelling.* SIAM. Philadelphia, USA.

Appendix A. PROOF OF LEMMA 2

We need the following proposition.

Proposition: Consider the stationary vector processes $\mathbf{v}(t)$, $\mathbf{w}_1(t)$ and $\mathbf{w}_2(t)$ such that

$$\mathbf{w}_i(t) = \mathcal{G}_i(q^{-1})\mathbf{v}(t), \quad i = 1, 2; \qquad (A.1)$$

$$\mathbf{r}_v(\tau) = E\mathbf{v}(t)\mathbf{v}^T(t-\tau), \qquad (A.2)$$

$$\mathbf{r}_{12}(\tau) = E\mathbf{w}_1(t)\mathbf{w}_2^T(t-\tau), \qquad (A.3)$$

where $\mathcal{G}_1(q^{-1})$ and $\mathcal{G}_2(q^{-1})$ are FIR filters of orders l and k, respectively. Let us assume

$$q^{-l}\mathcal{G}_1(q)\mathcal{G}_2(q^{-1}) = \mathcal{G}(q^{-1}) = \sum_{i=0}^{l+k} g_i q^{-i}. \qquad (A.4)$$

Then

$$\mathbf{r}_{12}(\tau) = \sum_{i=0}^{l+k} g_i \mathbf{r}_v(l-\tau-i). \qquad (A.5)$$

Proof: Straightforward and omitted.

As a special case of (A.1), if

$$\mathbf{v}(t) = \left[\, \tilde{y}(t) \;\; \tilde{u}(t) \,\right]^T, \qquad (A.6)$$

$$\mathbf{w}_i(t) = \left[\, \tilde{y}_i(t) \;\; \tilde{u}_i(t) \,\right]^T, i = 1, 2; \qquad (A.7)$$

then we get from (A.5) and (26)

$$\mathbf{r}_v(\tau) = \begin{cases} \mathbf{\Lambda}, & \text{if } \tau = 0 \\ 0, & \text{if } \tau \neq 0 \end{cases}, \qquad (A.8)$$

$$\mathbf{r}_w(\tau) = \begin{cases} \mathbf{\Lambda} h_{m-\tau}, & \text{if } |\tau| \leq m \\ 0, & \text{if } |\tau| > m \end{cases}. \qquad (A.9)$$

Note that, in this specific case, $\mathbf{\Lambda}$ is a diagonal matrix, but it need not be diagonal for (A.9) to hold. Since $R_{\tilde{\phi}_1\tilde{\phi}_2}$ in (23) is composed of the elements of $\{\mathbf{r}_w(\tau)\}_{\tau=-n}^{n}$, it is sufficient to have (27) satisfied in order (23) to hold.

Appendix B. PROOF OF LEMMA 3

In this appendix, we shall assume $R_{x_1x_2}$ is estimated as

$$\hat{R}_{x_1x_2} = \frac{1}{N}\sum_{t=1}^{N} x_1(t)x_2^T(t). \qquad (B.1)$$

Form (24), we get after some steps

$$\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta} = \hat{R}^{\dagger}_{\phi_1\psi_2}\frac{1}{N}\sum_{t=1}^{N}\phi_1(t)\left[\tilde{y}_2(t) - \tilde{\psi}_2^T(t)\boldsymbol{\theta}\right]$$

$$= \hat{R}^{\dagger}_{\phi_1\psi_2}\frac{1}{N}\sum_{t=1}^{N}\phi_1(t)v_2(t). \qquad (B.2)$$

Hence it follows from ergodicity results, see (Söderström and Stoica, 1989), that $\hat{\boldsymbol{\theta}}_1$ is asymptotically Gaussian distributed as,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}) \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0, P_1), \qquad (B.3)$$

where

$$P_1 = R^{\dagger}_{\phi_1\psi_2} Q R^{\dagger T}_{\phi_1\psi_2} \qquad (B.4)$$

and

$$Q = \lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\phi_1(t)v_2(t)v_2(s)\phi_1^T(s)\right\}. \qquad (B.5)$$

It remains to compute Q. In this aim, we decompose $\phi_1(t)$ in (B.5) into its noisy and noise-free parts and retain only the non-zero terms to get

$$Q = Q_a + Q_b, \qquad (B.6)$$

where

$$Q_a = \lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\phi_{10}(t)v_2(t)v_2(s)\phi_{10}^T(s)\right\}$$

$$= \lambda \operatorname{cov}\left[G(q^{-1})\phi_0(t)\right]; \qquad (B.7)$$

$$Q_b = \lim_{N\to\infty} E\left\{\frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}\tilde{\phi}_1(t)v_2(t)v_2(s)\tilde{\phi}_1^T(s)\right\}. \qquad (B.8)$$

Note that, in (B.7) we have used (28). For the detailed derivation of (B.7), we refer the reader to (Söderström and Stoica, 1983). To simplify Q_b, let us apply the property of jointly Gaussian distributed random vectors to get

$$Q_b = Q_c + Q_d + Q_e, \quad \text{(B.9)}$$

where

$$Q_c = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} E\left\{\tilde{\phi}_1(t) v_2(t)\right\} E\left\{v_2(s)\tilde{\phi}_1^T(s)\right\} = 0, \quad \text{(B.10)}$$

$$Q_d = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} E\left\{\tilde{\phi}_1(t)\tilde{\phi}_1^T(s)\right\} E\left\{v_2(t)v_2(s)\right\} = \lambda \text{cov}\left[G(q^{-1})\tilde{\phi}(t)\right], \quad \text{(B.11)}$$

$$Q_e = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} E\left\{\tilde{\phi}_1(t) v_2(s)\right\} E\left\{v_2(t)\tilde{\phi}_1^T(s)\right\} \quad \text{(B.12)}$$

Note that, (B.10) follows from (23) and in (B.11) we have used (28). The derivation of (B.11) is similar to that of (B.7) and can be found in (Söderström and Mahata, 2001). Combining (B.6)-(B.12) we get

$$Q = Q_e + Q_f, \quad \text{(B.13)}$$

where

$$Q_f = \lambda \text{cov}\left[G(q^{-1})\phi(t)\right]. \quad \text{(B.14)}$$

To examine Q_e closely, let us first define the polynomials

$$F(q^{-1}) = A(q^{-1})G_2(q^{-1}), \quad \text{(B.15)}$$

$$L(q^{-1}) = B(q^{-1})G_2(q^{-1}), \quad \text{(B.16)}$$

$$\sum_{i=0}^{2m+n} \Gamma_i q^{-i} = A(q^{-1})H(q^{-1}), \quad \text{(B.17)}$$

$$\sum_{i=0}^{2m+n} \Delta_i q^{-i} = B(q^{-1})H(q^{-1}). \quad \text{(B.18)}$$

Note also that, $\tilde{\phi}_1(t)$ is composed of the delayed and filtered measurement noise sequences $\tilde{y}_1(t)$ and $\tilde{u}_1(t)$. In the aim of evaluating $E\left\{\tilde{\phi}_1(t)v_2(s)\right\}$, let us examine

$$E\tilde{y}_1(t)v_2(t-\tau) = E\left\{G_1(q^{-1})\tilde{y}(t)\right\}\left\{F(q^{-1})\tilde{y}(t-\tau)\right\} = \lambda_y \Gamma_{m-\tau}. \quad \text{(B.19)}$$

Note that, in the last equality we have used the proposition in the appendix A, (B.17) and (26). Similarly by (A.5), (B.18) and (26) we have

$$E\tilde{u}_1(t)v_2(t-\tau) = \lambda_u \Delta_{m-\tau}. \quad \text{(B.20)}$$

Note that Q_e is a $(2n+1)\times(2n+1)$ matrix, which can be splitted into block matrices

$$Q_e = \begin{bmatrix} Q^{(11)} & Q^{(12)} \\ Q^{(21)} & Q^{(22)} \end{bmatrix}, \quad \text{(B.21)}$$

where $Q^{(11)}$ is a $(n+1)\times(n+1)$ matrix and $Q^{(22)}$ is a $n \times n$ matrix. Consider $Q^{(11)}_{\mu\nu}$, which is the element at μth row and νth column of $Q^{(11)}$. From (B.19) for $1 \le \mu \le n+1$ and $1 \le \nu \le n+1$ we get

$$\begin{aligned} Q^{(11)}_{\mu\nu} &= \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} E\left\{\tilde{y}(t-\mu+1)v_2(s)\right\} E\left\{\tilde{y}(s-\nu+1)v_2(t)\right\} \\ &= \lim_{N\to\infty} \frac{\lambda_y^2}{N} \sum_{t=1}^{N} \sum_{s=1}^{N} \Gamma_{m-t+s+\mu-1}\Gamma_{m-s+t+\nu-1} \\ &= \lambda_y^2 \sum_{i=0}^{2m+n} \Gamma_i \Gamma_{2m+\mu+\nu-2-i} \\ &= \lambda_y^2 \beta^{(aa)}_{2m+\mu+\nu-2}, \end{aligned} \quad \text{(B.22)}$$

where we used the convention $\Gamma_i = 0$ for $i < 0$ and $i > 2m+n$. Note that in the last equality, we have used (29). Similarly, for the other blocks of Q_e we can show using (B.19), (B.20), (30) and (31) that

$$Q^{(12)}_{\mu\nu} = \lambda_y \lambda_u \beta^{(ab)}_{2m+\mu+\nu-1}, \quad 1 \le \mu \le n+1,\ 1 \le \nu \le n; \quad \text{(B.23)}$$

$$Q^{(21)}_{\mu\nu} = Q^{(12)}_{\nu\mu}, \quad 1 \le \nu \le n+1,\ 1 \le \mu \le n; \quad \text{(B.24)}$$

$$Q^{(22)}_{\mu\nu} = \lambda_u^2 \beta^{(bb)}_{2m+\mu+\nu}, \quad 1 \le \mu \le n,\ 1 \le \nu \le n. \quad \text{(B.25)}$$

Thus, (B.14), (B.21)-(B.25) together define the matrix Q. This completes the derivation of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_1$. The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_2$ can be obtained from the expression of P_1 by interchanging the subscripts 1 and 2.