

Prefilter design for errors in variables model identification

Kaushik Mahata

Abstract—The bias compensated least squares approach for errors-in-variables model identification is examined in a new framework, where it is allowed to prefilter the observed input-output data prior to the estimation process. A statistical analysis of the estimation algorithm is presented. Subsequently, it is shown how these prefilters and the weighting matrix can be tuned in order to optimize the estimation accuracy. According to the numerical simulation results, the covariance matrix of the estimated parameter vector is very close to the Cramér-Rao lower bound for the estimation problem.

I. INTRODUCTION

Consider a linear system with noise-free input $\dot{u}(t)$ linked to its true output $\ddot{y}(t)$ via the difference equation

$$A(q^{-1}) \dot{y}(t) = B(q^{-1}) \dot{u}(t), \quad (1)$$

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

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

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

$$u(t) = \dot{u}(t) + \tilde{u}(t), \quad y(t) = \ddot{y}(t) + \tilde{y}(t), \quad (3)$$

where $\tilde{u}(t)$ and $\tilde{y}(t)$ are mutually independent zero mean white noise sequences, each independent of $\dot{u}(t)$ and $\ddot{y}(t)$, with variances σ_I^2 and σ_O^2 , respectively. The problem under consideration is to identify the vector of system parameters

$$\hat{\theta} = [b_n \dots b_1 a_n \dots a_1]^\top \quad (4)$$

from the available measurements $\{u(t)\}_{t=1}^N$ and $\{y(t)\}_{t=1}^N$, assuming σ_I^2 and σ_O^2 are unknown.

This problem is often referred to as a dynamic errors-in-variables (EIV) problem, and is found in many disciplines of science, as proved by the several applications collected in [1]. An overall view of different approaches is presented in [2].

The accurate solutions for EIV problems involve numerical search techniques, and are computationally quite expensive. The maximum-likelihood solution [3] is known to be optimal in terms of accuracy, but suffers from several numerical difficulties. The prediction error method [4] can handle very general model structure, but suffers from

similar numerical difficulties encountered by the maximum-likelihood approach. Frisch scheme [5], [6], [7] is an attractive alternative in the sense that the search involved here is one dimensional. However, in terms of accuracy, Frisch scheme is often far from the Cramér-Rao Bound [3], particularly when the signal to noise ratio (SNR) is not large. The BELS approach [8] yields estimates as accurate as Frisch scheme. However, it is known to suffer from convergence problems in finite samples. In this paper we propose a modified bias compensated least squares approach, which can be seen as an extension of [9] with prefiltering of the data. We show the covariance of the resulting estimates are very close to the Cramér-Rao bound.

II. ESTIMATION ALGORITHM

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

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

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

$$\begin{aligned} \psi(t) &:= [u(t-n) \dots u(t-1) \\ &\quad -y(t-n) \dots -y(t-1)]' \quad (5) \\ &= \dot{\psi}(t) + \tilde{\psi}(t), \end{aligned} \quad (6)$$

where $\dot{\psi}(t)$ is the noise-free part of $\psi(t)$, and $\tilde{\psi}(t)$ is the noise contribution in $\psi(t)$:

$$\begin{aligned} \dot{\psi}(t) &= [\dot{u}(t-n) \dots \dot{u}(t-1) \\ &\quad -\dot{y}(t-n) \dots -\dot{y}(t-1)]' \\ \tilde{\psi}(t) &= [\tilde{u}(t-n) \dots \tilde{u}(t-1) \\ &\quad -\tilde{y}(t-n) \dots -\tilde{y}(t-1)]'. \end{aligned}$$

Note that we denote the transpose of a matrix X by X' . Now due to (1) and (2) we know that

$$\dot{y}(t) = \dot{\psi}'(t) \hat{\theta}.$$

Now if we pass the input output data through a filter $F(q^{-1})$ it holds that

$$F(q^{-1})\dot{y}(t) = F(q^{-1})\dot{\psi}'(t) \hat{\theta}. \quad (7)$$

In the sequel we use the following notations. We define $\phi_I(t)$, $\phi_O(t)$, $\dot{\vartheta}(t)$, $\hat{\theta}_I$, $\hat{\theta}_O$ and $\ddot{\vartheta}$ as

$$\begin{aligned}\phi_I(t) &= F(q^{-1})[u(t-n) \dots u(t-1)]', \\ \dot{\theta}_I &= [b_n \dots b_1]' \\ \phi_O(t) &= F(q^{-1})[y(t-n) \dots y(t)]', \\ \dot{\theta}_O &= [a_n \dots a_1 1]' \\ \phi(t) &= [\phi'_I(t) \phi'_O(t)]' \\ \ddot{\vartheta} &= [\dot{\theta}'_I \dot{\theta}'_O \dot{\sigma}_I \dot{\sigma}_O]'.\end{aligned}$$

Therefore (1) and (7) imply that

$$\dot{\phi}'_I(t) \dot{\theta}_I = \dot{\phi}'_O(t) \dot{\theta}_O. \quad (8)$$

Now consider a vector valued stationary process $\varphi(t)$ of dimension $m > 2n + 1$, which satisfy the *persistence of excitation* condition that the matrix $\mathcal{E}\{\dot{\varphi}(t) \cdot F(q^{-1})\dot{\psi}'(t)\}$ is nonsingular. We consider the general case where $\varphi(t)$ is a linear function of the observed data. For different methods $\varphi(t)$ may take various forms. Let the joint input-output data vector be denoted by

$$z(t) := [u(t) \ y(t)]',$$

and we assume that $\varphi(t)$ is given by

$$\varphi(t) = G_1(q^{-1})y(t) + G_2(q^{-1}), \quad (9)$$

for some *user chosen* stable filters $G_i(e^{-i\omega}) \in \mathbb{C}^{m \times 1}$, is stable.

Let us define the matrices \dot{R}_I and \dot{R}_O as

$$\dot{R}_I = \mathcal{E}\{\varphi(t)\phi'_I(t)\}, \quad \dot{R}_O = \mathcal{E}\{\varphi(t)\phi'_O(t)\}. \quad (10)$$

Now recall that $\tilde{y}(t)$ and $\tilde{u}(t)$ are independent of each other. Hence there exist matrices P_I and P_O such that

$$\mathcal{E}\{\varphi(t)\dot{\theta}'_I(t)\} = \dot{\sigma}_I P_I, \quad \mathcal{E}\{\varphi(t)\dot{\theta}'_O(t)\} = \dot{\sigma}_O P_O, \quad (11)$$

where

$$P_I = \frac{1}{2\pi} \int_{-\pi}^{\pi} G_2(e^{-i\omega})F(e^{i\omega})[e^{in\omega} \dots e^{i\omega}] d\omega,$$

$$P_O = \frac{1}{2\pi} \int_{-\pi}^{\pi} G_1(e^{-i\omega})F(e^{i\omega})[e^{in\omega} \dots e^{i\omega} 1] d\omega.$$

Since we know how $\varphi(t)$ is constructed, we can compute P_I and P_O .

Since the true input-output data $\dot{y}(t)$ and $\dot{u}(t)$ are independent of the measurement noise $\tilde{u}(t)$ and $\tilde{y}(t)$, it follows from (8), (10) and (11) that

$$(\dot{R}_I - \dot{\sigma}_I P_I) \dot{\theta}_I = (\dot{R}_O - \dot{\sigma}_O P_O) \dot{\theta}_O. \quad (12)$$

To estimate $\ddot{\vartheta}$ from the data, we first obtain estimates \hat{R}_I and \hat{R}_O of the covariances \dot{R}_I and \dot{R}_O as

$$\hat{R}_I = \frac{1}{N} \sum_{t=1}^N \varphi(t)\phi'_I(t), \quad \hat{R}_O = \frac{1}{N} \sum_{t=1}^N \varphi(t)\phi'_O(t). \quad (13)$$

Then we can estimate $\ddot{\vartheta}$ by solving a *bilinear least squares* optimization problem

$$\hat{\vartheta} = \arg \min_{\vartheta} \|\hat{R}_I \theta_I - \hat{R}_O \theta_O - \sigma_I P_I \theta_I + \sigma_O P_O \theta_O\|_W^2, \quad (14)$$

subject to the constraint that the last element of θ_O is unity. Note that

$$\vartheta = [\theta'_I \ \theta'_O \ \sigma_I \ \sigma_O]'$$

is a dummy parameter on which serves as the variable of optimization. Also for any positive definite weighting matrix W we define $\|x\|_W^2 = x'Wx$. The approach in (14) can be seen as another way of bias compensation, which is fairly simple in principle. This approach is used in [9] without any prefiltering.

III. ESTIMATION ACCURACY

In this section we give an explicit expression for the limiting covariance matrix (as $N \rightarrow \infty$) of the parameter estimates in (14). Let us define $\hat{\theta}_I$, $\hat{\theta}_O$, $\hat{\sigma}_I$ and $\hat{\sigma}_O$ as the estimates of $\dot{\theta}_I$, $\dot{\theta}_O$, $\dot{\sigma}_I$ and $\dot{\sigma}_O$, respectively, obtained from $\hat{\vartheta}$:

$$\hat{\vartheta} = [\hat{\theta}'_I \ \hat{\theta}'_O \ \hat{\sigma}_I \ \hat{\sigma}_O]'.$$

In the following we use the following notations.

$$\begin{aligned}\hat{\theta}_I &= \dot{\theta}_I + \tilde{\theta}_I, & \hat{\theta}_I &= \dot{\theta}_I + \tilde{\theta}_I, \\ \hat{\sigma}_I &= \dot{\sigma}_I + \tilde{\sigma}_I, & \hat{\sigma}_I &= \dot{\sigma}_I + \tilde{\sigma}_I,\end{aligned}$$

Therefore, for example, $\tilde{\theta}_I$ denote the error in the estimate $\hat{\theta}_I$. Recall that $[\dot{\theta}'_I \ \dot{\theta}'_O] := [\dot{\theta}' \ 1]$. In similar manner we define $\hat{\theta}$ (recall that the last element of $\hat{\theta}_O$ is constrained to be unity):

$$[\hat{\theta}'_I \ \hat{\theta}'_O] := [\hat{\theta}' \ 1].$$

Lemma 1: Define the process $\zeta(t)$ and the matrix Y as

$$\begin{aligned}\zeta(t) &:= [\dot{\psi}'(t) \ -\frac{1}{\sigma_I} B(q^{-1})\tilde{u}(t) \ \frac{1}{\sigma_O} A(q^{-1})\tilde{y}(t)]', \\ Y &:= \mathcal{E}\{\varphi(t) \cdot F(q^{-1})\zeta'(t)\}.\end{aligned} \quad (15)$$

Then as $N \rightarrow \infty$, the estimation error (with a slight abuse of notation) $\tilde{\vartheta} := [\tilde{\theta}' \ \tilde{\sigma}_I \ \tilde{\sigma}_O]'$ is given by

$$\tilde{\vartheta} = -(Y'WY)^{-1}Y'W\epsilon, \quad (16)$$

where the *residual* vector ϵ is defined as

$$\epsilon = \hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O - \dot{\sigma}_I P_I \dot{\theta}_I + \dot{\sigma}_O P_O \dot{\theta}_O. \quad (17)$$

Proof: See Appendix I. ■

In Lemma 2 we examine the asymptotic distribution of ϵ in (17).

Lemma 2: Let the measurement noise sequences $\tilde{u}(t)$ and $\tilde{y}(t)$ be Gaussian distributed. Define σ_C and $C(q^{-1})$ via the spectral factorization

$$\sigma_C C(q^{-1}) C(q) = \sigma_I B(q^{-1}) B(q) + \sigma_O A(q^{-1}) A(q). \quad (18)$$

Then as $N \rightarrow \infty$, the residual $\sqrt{N}\epsilon$ is asymptotically Gaussian with a mean zero and covariance matrix

$$\text{cov}\{\sqrt{N}\epsilon\} = \sigma_C \text{cov}\{F(q^{-1})C(q^{-1})\varphi(t)\} + T_3 \quad (19)$$

where

$$T_3 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \chi(\omega) \chi'(\omega) d\omega \quad (20)$$

and

$$\chi(\omega) = \sigma_O G_1(e^{-i\omega}) A(e^{i\omega}) - \sigma_I G_2(e^{-i\omega}) B(e^{i\omega}).$$

Proof: See Appendix II. ■

Remark: We point out that the asymptotic Gaussian nature of ϵ is valid even if $\tilde{u}(t)$ and $\tilde{y}(t)$ are not Gaussian. However, under this condition there will be another term in (19) involving the fourth order moments of $\tilde{y}(t)$ and $\tilde{u}(t)$. This term is in fact not very difficult to evaluate. In this paper this additional term is not considered in order to simplify the analysis in the subsequent sections. ■

Combining the result in Lemma 1 and Lemma 2 we get the following main result.

Theorem 1: Let Y be defined as in (15), and $\tilde{u}(t)$ and $\tilde{y}(t)$ be Gaussian sequences. Then as $N \rightarrow \infty$, the normalized estimation error $\sqrt{N}\hat{\vartheta} = \sqrt{N}[\hat{\theta}' \quad \tilde{\sigma}_I \quad \tilde{\sigma}_O]'$ is asymptotically Gaussian with zero mean and covariance matrix

$$\Sigma = (Y'WY)^{-1}Y'W[\sigma_C \text{cov}\{F(q^{-1})C(q^{-1})\varphi(t)\} + T_3]WY(Y'WY)^{-1}. \quad (21)$$

Proof: The proof is straightforward from Lemma 1 and Lemma 2. ■

IV. TUNING THE USER'S CHOICES

In this section we explore the possibilities of tuning different user's choices so that the covariance matrix Σ of the parameter estimates is as small as possible, see Footnote 1. In particular we comment upon the choice of W and the implementation of $\varphi(t)$ in the state space form (9). The theory of best linear unbiased estimator (BLUE) can be extended [10, p.67], see also [11, p.555], in this case to show that the optimum¹ choice of W for a given choice of $\varphi(t)$ is given by

$$W = [\text{cov}\{F(q^{-1})C(q^{-1})\varphi(t)\} + T_3]^{-1}, \quad (22)$$

and covariance matrix Σ_* of the associated optimum estimate $\sqrt{N}\hat{\vartheta}$ is given by

$$\Sigma_* = \{Y'[\sigma_C \text{cov}\{F(q^{-1})C(q^{-1})\varphi(t)\} + T_3]^{-1}Y\}^{-1}. \quad (23)$$

The remaining part of this section will be devoted in exploring a suitable $\varphi(t)$ and $F(q^{-1})$ such that Σ_* is optimized.

Recall that T_3 is contributed only by the noise in the measurement. It is also readily verified that T_3 is proportional to the square of the measurement noise variances. For moderately large signal to noise ratio the contribution of T_3 in (23) is significantly small compared to $\text{cov}\{F(q^{-1})C(q^{-1})\varphi(t)\}$. Therefore in a practical scenario it is reasonable neglect the contribution of T_3 in the asymptotic covariance of $\hat{\vartheta}$. Thus, we get an approximate expression

$$\Sigma_* \approx \sigma_C \{Y'[\text{cov}\{F(q^{-1})C(q^{-1})\varphi(t)\}]^{-1}Y\}^{-1} \quad (24)$$

¹ The optimality is in the sense that $\Sigma - \Sigma_*$ is a positive definite matrix. Here onwards we are concerned with the same notion of optimality of covariance matrices. For any two symmetric matrices X_1 and X_2 we write $X_1 > X_2$ to denote $X_1 - X_2$ is positive definite.

which is considerably simpler to work with. For example, we can apply some wellknown results from the theory of optimal instrumental variable methods [12], [13].

Lemma 3: Let us define the sequences $\xi(t)$ and $\eta(t)$ as

$$\xi(t) = C(q)F(q)\varphi(t), \quad \eta(t) = C^{-1}(q^{-1})\zeta(t). \quad (25)$$

Then the covariance matrix Σ_* satisfy the following:

$$\begin{aligned} \Sigma_* &= \sigma_C [\mathcal{E}\{\eta(t)\xi'(t)\} [\mathcal{E}\{\xi(t)\xi'(t)\}]^{-1} \times \\ &\quad \mathcal{E}\{\xi(t)\eta'(t)\}]^{-1} \\ &> \sigma_C [\mathcal{E}\{\eta(t)\eta'(t)\}]^{-1} := \sigma_C \Sigma_0. \end{aligned} \quad (26)$$

Furthermore, consider two choices $\varphi_1(t)$ and $\varphi_2(t)$ of the correlator $\varphi(t)$ having the nested structure

$$\varphi_2(t) = [\varphi'_1(t) \quad \beta'(t)]'$$

for some $\beta(t)$, a linear function of the data. Let Σ_j ($j = 1, 2$) denote the value of Σ_* obtained by substituting $\varphi_j(t)$ for $\varphi(t)$ in (25) and (26). Then

$$\Sigma_2 \leq \Sigma_1. \quad (27)$$

Proof: The proof of (26) is identical to Theorem 4.2 in [12]. The expression (27) is proved in a way similar to Lemma 2 in [13], see also [10, p.305]. ■

In case of instrumental variable method one way to achieve optimality in a similar scenario² is to set $F(q^{-1}) = C^{-1}(q^{-1})$ and find a realization (9) such that $\varphi(t) = \eta(t)$. Unfortunately, in an errors-in-variables scenario this strategy fails because the data are corrupted by measurement noise. We can never separate noise from the true signal. Now consider the situation when we keep on increasing the dimension of $\varphi(t)$ by appending more signals obtained from the data. Clearly from (27) this will result in a monotonically decreasing sequence of covariances matrices (of the parameter estimates) bounded from below, see (26). This indicates convergence of the sequence of covariance matrices to a limiting value as the dimension of $\varphi(t)$ is increased.

In the following we assume $\dot{u}(t)$ is a stationary stochastic process. Then we have the following lemma which gives the optimal choice of $\varphi(t)$ and the filter $F(q^{-1})$.

Lemma 4: Let $\hat{\zeta}(t)$ be the optimal Wiener filter estimate of $\zeta(t)$ obtained from the observed data $z(t)$. Then

$$\Sigma_* \geq \sigma_C [\mathcal{E}\{C^{-1}(q^{-1})\hat{\zeta}(t) \cdot C^{-1}(q^{-1})\hat{\zeta}(t)\}]^{-1}. \quad (28)$$

Furthermore, Σ_* achieves the lower bound in (28) when $F(q^{-1}) = C^{-1}(q^{-1})$ and

$$\varphi(t) = C^{-1}(q^{-1})\hat{\zeta}(t) \quad (29)$$

Proof: Let us denote

$$\tilde{\zeta}(t) = \hat{\zeta}(t) - \zeta(t) \quad (30)$$

Since the minimum variance estimation error $\tilde{\zeta}(t)$ satisfies

$$\mathcal{E}\{\tilde{\zeta}(t_1) \cdot z'(t_2)\} = 0, \quad \forall t_1, t_2, \quad (31)$$

² Consider for example optimal instrumental variable method for ARMAX models [12].

Since $\xi(t)$ is obtained by filtering $z(t)$ via a linear filter, using (31) we get

$$\Sigma_* = \sigma_C [\mathcal{E}\{\hat{\eta}(t)\xi'(t)\}[\mathcal{E}\{\xi(t)\xi'(t)\}]^{-1} \mathcal{E}\{\xi(t)\hat{\eta}'(t)\}]^{-1}. \quad (32)$$

where

$$\hat{\eta}(t) = C^{-1}(q^{-1})\zeta(t). \quad (33)$$

Now by the property of the covariance matrices, it follows that

$$\mathcal{E}\left\{\begin{bmatrix} \hat{\eta}(t) \\ \xi(t) \end{bmatrix} \begin{bmatrix} \hat{\eta}'(t) & \xi'(t) \end{bmatrix}\right\} \geq 0. \quad (34)$$

Then taking Schur complement in (34) and using (32) we get (28). In addition, the equality holds in (28) when

$$\xi(t) = K\hat{\eta}(t) \quad (35)$$

for any nonsingular and nondynamic matrix K . Setting $K = I$ and choosing $F(q^{-1}) = C^{-1}(q^{-1})$ in (35) we get (29). ■

Wiener filtering of the data $z(t)$ to compute $\hat{\zeta}(t)$ is a noncausal operation. If we are willing to compromise the estimation accuracy slightly, then the Wiener filtering step can be replaced by a causal Kalman filtering procedure.

If $\dot{u}(t)$ is not a stationary stochastic process we can not apply Lemma 4. In that case we can use

$$\varphi(t) = C^{-1}(q^{-1})[z'(t) \cdots z'(t-d)]' \quad (36)$$

for suitable choice of d . Lemma 3 ensures that the covariance of the resulting optimally weighted [see (22)] estimate of ϑ decreases monotonically with increase in d . This is true even when the filtering by $C^{-1}(q^{-1})$ is not performed in (36). However, when we perform filtering by $C^{-1}(q^{-1})$, it is possible to achieve a good performance with a significantly small d .

V. A CASE STUDY

In this section we illustrate the analytical results derived in the paper. Here $\dot{u}(t)$ to be a ARMA process obtained by filtering a zero mean unity variance white noise process through a filter

$$\frac{1 + 0.2q^{-1}}{1 - 0.8q^{-1}}.$$

We consider the situation when it is a priori known that $\dot{u}(t)$ is a ARMA process. We also show results when the ARMA nature of $\dot{u}(t)$ is not assumed in the estimation process. The unknown system is chosen such that

$$\begin{aligned} A(q^{-1}) &= 1 - 1.5q^{-1} + 0.7q^{-2}, \\ B(q^{-1}) &= 1.0q^{-1} + 0.5q^{-2}, \end{aligned}$$

Here we compare the performance of the proposed algorithms with that of the Frisch scheme and the Cramér-Rao Bound [3]. We point out that, the performance of Frisch scheme is similar to the prediction error method [2]. The Cramér-Rao bound indicates the best possible performance

achievable by an unbiased estimator, which is known to be achieved only by the maximum likelihood method for EIV models [3].

We choose $F(q^{-1}) = C^{-1}(q^{-1})$ for different choices of $\varphi(t)$. In one class of estimators we consider $\varphi(t)$ of the form given in (36) for $d \in \{5, 7, 9\}$. The estimator obtained by setting $d = 5$, etc in (36) is denoted by Es-5, etc. These estimators are compared with the suboptimal estimator computed by setting

$$\varphi(t) = C^{-1}(q^{-1})\bar{\zeta}(t),$$

where $\bar{\zeta}(t)$ is the Kalman filter estimate of $\zeta(t)$. The associated estimator is denoted by Es-K. The Frisch scheme estimator is also computed, which is denoted by Es-Fr. Recall that it was argued that T_3 is contributed by the measurement noise only, and is significantly small compared to $\text{cov}\{F(q^{-1})C(q^{-1})\varphi(t)\}$. Therefore, it is interesting to investigate the effect of large measurement noise variance on the proposed estimators. Hence we set data signal to noise ratio to 5 dB. The results from 100 independent Monte-Carlo simulations are shown in Table I. It is interesting to note that the overall performance of the proposed estimators are very close to the Cramér-Rao bound even with considerably large measurement noise variances. On the other hand the Frisch scheme estimate of $B(q^{-1})$ breaks down at this noise level. With $\varphi(t)$ as in (36), the estimation accuracy increases monotonically with increasing d . As predicted by the theory, the estimate Es-K outperforms all. However the difference between Es-K and Es-9 is marginal.

Estimator		b_1	b_2	a_1	a_2
	True value	1	0.5	-1.5	0.7
	CRB	0.0351	0.0685	0.0009	0.0004
Es-5	Mean	0.9921	0.5210	-1.4969	0.6983
	Variance (empirical)	0.0449	0.1064	0.0014	0.0006
	Variance (analytical)	0.0385	0.0808	0.0010	0.0005
Es-7	Mean	0.9969	0.5124	-1.4974	0.6984
	Variance (empirical)	0.0421	0.0987	0.0013	0.0006
	Variance (analytical)	0.0337	0.0750	0.0010	0.0005
Es-9	Mean	0.9964	0.5124	-1.4975	0.6984
	Variance (empirical)	0.0415	0.0948	0.0012	0.0005
	Variance (analytical)	0.0367	0.0736	0.0009	0.0005
Es-K	Mean	1.0001	0.5056	-1.4990	0.6995
	Variance (empirical)	0.0406	0.0933	0.0011	0.0005
	Variance (analytical)	0.0370	0.0731	0.0009	0.0005
Es-Fr	Mean	1.3629	0.3536	-1.4391	0.6998
	Variance (empirical)	3.2673	1.3821	0.1056	0.0468

TABLE I
SIMULATION RESULTS FOR A SNR= 5 dB.

VI. CONCLUSIONS

In this paper we consider the bias compensated least squares identification of dynamic errors-in-variables models from a new perspective. Here we allow prefiltering of the data. The regressor vector $\phi(t)$ is composed of the filtered and delayed version of the observed data, while the correlator vector $\varphi(t)$ is obtained via a vector valued filtering of the observed data. We propose a modification of conventional bias compensation algorithm to cope with this filtering of the data. An asymptotic accuracy analysis of the estimates are presented, where we show that the estimates are asymptotically Gaussian, with a covariance matrix given in Theorem 1. Subsequently, the problem of choosing the user defined filters to optimize the estimation accuracy is treated in detail. The results in Section IV give some interesting insight on this aspect. The algorithms are tested in numerical simulations where the proposed algorithms gives estimates having accuracy close to the Cramér-Rao bound for both low and high signal to noise ratio.

APPENDIX I PROOF LEMMA 1

In this appendix we denote $\hat{R}_I = R_I + \tilde{R}_I$ and $\hat{R}_O = R_O + \tilde{R}_O$. Therefore, for example, \tilde{R}_I denote the error in \hat{R}_I . Define $\theta = [\theta'_I \ \theta'_{O(1:n)}]'$. Now partition R_O and P_O as

$$R_O := [S_o \ s_o], \quad P_O := [Q_o \ q_o],$$

where s_o and q_o are column vectors. Then it is straightforward to see using (14) that

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta} \|\hat{R}_I \theta_I - \hat{R}_O \theta_O - \hat{\sigma}_I P_I \theta_I + \hat{\sigma}_O P_O \theta_O\|_W^2 \\ &= \arg \min_{\theta} \|[\hat{R}_I - \hat{\sigma}_I P_I \quad -\hat{S}_o + \hat{\sigma}_O Q_o] \theta - s_o + \hat{\sigma}_O q_o\|_W^2, \end{aligned}$$

where we use $\theta_{O(n+1)} = 1$. Let us denote

$$\begin{aligned} \dot{X} &:= [R_I - \hat{\sigma}_I P_I \quad -\dot{S}_o + \hat{\sigma}_O Q_o] \\ &= \mathcal{E}\{\varphi(t) \cdot F(q^{-1})\psi'(t)\}. \end{aligned} \quad (37)$$

Then using standard result for least squares estimation error [10, p.285] we get

$$\begin{aligned} \dot{X}' W \dot{X} \tilde{\theta} &= -\dot{X}' W \left\{ \hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O - \hat{\sigma}_I P_I \dot{\theta}_I + \hat{\sigma}_O P_O \dot{\theta}_O \right\} \\ &= -\dot{X}' W \left\{ (R_I + \tilde{R}_I) \dot{\theta}_I - (R_O + \tilde{R}_O) \dot{\theta}_O \right. \\ &\quad \left. - (\hat{\sigma}_I + \tilde{\sigma}_I) P_I \dot{\theta}_I + (\hat{\sigma}_O + \tilde{\sigma}_O) P_O \dot{\theta}_O \right\} \\ &= -\dot{X}' W \left\{ \tilde{R}_I \dot{\theta}_I - \tilde{R}_O \dot{\theta}_O - \tilde{\sigma}_I P_I \dot{\theta}_I + \tilde{\sigma}_O P_O \dot{\theta}_O \right\}, \end{aligned}$$

where we use (12) in the last equality. Therefore, we get after rearranging

$$\dot{X}' W [\dot{X} \quad -P_I \dot{\theta}_I \quad P_O \dot{\theta}_O] [\tilde{\theta}' \ \tilde{\sigma}_I \ \tilde{\sigma}_O]' = -\dot{X}' W \epsilon, \quad (38)$$

where

$$\begin{aligned} \epsilon &:= \tilde{R}_I \dot{\theta}_I - \tilde{R}_O \dot{\theta}_O \\ &= \tilde{R}_I \dot{\theta}_I - \tilde{R}_O \dot{\theta}_O + R_I \dot{\theta}_I - R_O \dot{\theta}_O \\ &\quad - \hat{\sigma}_I P_I \dot{\theta}_I + \hat{\sigma}_O P_O \dot{\theta}_O \\ &= \hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O - \hat{\sigma}_I P_I \dot{\theta}_I + \hat{\sigma}_O P_O \dot{\theta}_O. \end{aligned} \quad (39)$$

In the second equality in (39) we use (12). Again by (14) we have

$$\begin{aligned} [\hat{\sigma}_I \ \hat{\sigma}_O]' &= \arg \min_{\sigma_I, \sigma_O} \|\hat{R}_I \hat{\theta}_I - \hat{R}_O \hat{\theta}_O\|_W^2 \\ &\quad - \sigma_I P_I \hat{\theta}_I + \sigma_O P_O \hat{\theta}_O\|_W^2, \\ &= \arg \min_{\sigma_I, \sigma_O} \|\hat{R}_I \hat{\theta}_I - \hat{R}_O \hat{\theta}_O - \\ &\quad [P_I \hat{\theta}_I \quad -P_O \hat{\theta}_O] [\sigma_I \ \sigma_O]'\|_W^2. \end{aligned}$$

Consequently, using the same standard result for least squares estimation error [10, p.285] we have

$$\begin{aligned} &[P_I \dot{\theta}_I \quad -P_O \dot{\theta}_O]' W [P_I \dot{\theta}_I \quad -P_O \dot{\theta}_O] [\tilde{\sigma}_I \ \tilde{\sigma}_O]' \\ &= [P_I \dot{\theta}_I' \quad -P_O \dot{\theta}_O'] W \{\hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O - \\ &\quad \hat{\sigma}_I P_I \dot{\theta}_I + \hat{\sigma}_O P_O \dot{\theta}_O\} \\ &= -[-P_I \dot{\theta}_I \quad P_O \dot{\theta}_O]' W \times \\ &\quad \{R_I \dot{\theta}_I + \tilde{R}_I \dot{\theta}_I - R_O \dot{\theta}_O - \tilde{R}_O \dot{\theta}_O - \\ &\quad \hat{\sigma}_I P_I \dot{\theta}_I + \hat{\sigma}_O P_O \dot{\theta}_O\} \\ &= -[-P_I \dot{\theta}_I \quad P_O \dot{\theta}_O] W \{\epsilon + \dot{X} \tilde{\theta}\}, \end{aligned} \quad (40)$$

where in the second equality we have used a first order approximation; in the third equality we use (12), (39) and $\tilde{\theta}_{o(n+1)} = 0$. From (11) we know

$$\sigma_I P_I \dot{\theta}_I = \mathcal{E}\{\varphi(t) \cdot F(q^{-1})B(q^{-1})\tilde{u}(t)\},$$

and

$$\sigma_O P_O \dot{\theta}_O = \mathcal{E}\{\varphi(t) \cdot F(q^{-1})A(q^{-1})\tilde{y}(t)\}.$$

Hence combining (37), (38) and (40) we get (16) where

$$Y = [\dot{X} \quad -P_I \dot{\theta}_I \quad P_O \dot{\theta}_O] = \mathcal{E}\{\varphi(t) \cdot F(q^{-1})\zeta'(t)\}.$$

APPENDIX II PROOF LEMMA 2

For large N , we know that $\mathcal{E}\{\hat{R}_I\} = R_I$ and $\mathcal{E}\{\hat{R}_O\} = R_O$. Hence

$$\begin{aligned} \mathcal{E}\{\hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O\} &= \sigma_I P_I \dot{\theta}_I - \sigma_O P_O \dot{\theta}_O \\ &\Rightarrow \mathcal{E}\{\epsilon\} = 0. \end{aligned} \quad (41)$$

Hence $\hat{\theta}$ is an asymptotically unbiased estimate of $\dot{\theta}$. By (41) and (39) we get

$$\begin{aligned} \mathcal{E}\{\epsilon \epsilon'\} &= \mathcal{E}\{(\hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O)(\hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O)'\} \\ &\quad - (\sigma_I P_I \dot{\theta}_I - \sigma_O P_O \dot{\theta}_O)(\sigma_I P_I \dot{\theta}_I - \sigma_O P_O \dot{\theta}_O)'. \end{aligned} \quad (42)$$

Denote

$$\begin{aligned} v(t) &= A(q^{-1})y(t) - B(q^{-1})u(t) \\ &= A(q^{-1})\tilde{y}(t) - B(q^{-1})\tilde{u}(t). \end{aligned} \quad (43)$$

Note that $v(t)$ is independent of $\dot{z}(t)$ and is an order n moving average process. In addition the right hand side of (18) gives the spectrum of $v(t)$. Hence, by (18) there exist a white innovations sequence $\varepsilon(t)$ with variance σ_C so that

$$v(t) = C(q^{-1})\varepsilon(t). \quad (44)$$

It follows using (13) and (43) that

$$\begin{aligned} \hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O &= \frac{1}{N} \sum_{t=1}^N \varphi(t) \{ \phi'_I(t) \theta_I - \phi'_O(t) \theta_O \} \\ &= -\frac{1}{N} \sum_{t=1}^N \{ \dot{\varphi}(t) + \tilde{\varphi}(t) \} v(t). \end{aligned} \quad (45)$$

Now in our aim of computing the second order moment of the term in the left hand side of (45) we write $\varphi(t) = \dot{\varphi}(t) + \tilde{\varphi}(t)$. Then retaining the nonzero terms we have

$$\mathcal{E}[(\hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O)(\hat{R}_I \dot{\theta}_I - \hat{R}_O \dot{\theta}_O)'] = T_1 + T_2, \quad (46)$$

where [see (44)]

$$\begin{aligned} T_1 &= \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \mathcal{E} \{ \dot{\varphi}(t) \dot{\varphi}'(s) \} \\ &\quad \mathcal{E} \{ [C(q^{-1})\varepsilon(t)][C(q^{-1})\varepsilon(s)] \} \\ &= \frac{\sigma_c}{N} \text{cov} \{ C(q^{-1})\dot{\varphi}(t) \}, \\ T_2 &= \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \mathcal{E} \{ \tilde{\varphi}(t) \tilde{\varphi}'(s) v(t) v(s) \}. \end{aligned} \quad (47)$$

In (47) we have skipped a few typical standard steps involved in computing such expectations [10, p.285]. Using fourth order moment formula for Gaussian distributions [14] it follows that

$$\begin{aligned} T_2 &= \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N [\mathcal{E} \{ \tilde{\varphi}(t) \cdot v(t) \} \mathcal{E} \{ v(s) \cdot \tilde{\varphi}'(s) \} \\ &\quad + \mathcal{E} \{ \tilde{\varphi}(t) v(s) \} \mathcal{E} \{ v(t) \tilde{\varphi}'(s) \} \\ &\quad + \mathcal{E} \{ \tilde{\varphi}(t) \tilde{\varphi}'(s) \} \mathcal{E} \{ v(t) v(s) \}] \\ &= (\sigma_I P_I \dot{\theta}_I - \sigma_O P_O \dot{\theta}_O)(\sigma_I P_I \dot{\theta}_I - \sigma_O P_O \dot{\theta}_O)' + \frac{T_3}{N} \\ &\quad + \frac{1}{N^2} \sum_{t=1}^N \sum_{s=1}^N \mathcal{E} \{ \tilde{\varphi}(t) \tilde{\varphi}'(s) \} \times \\ &\quad \mathcal{E} \{ C(q^{-1})\varepsilon(t) \cdot C(q^{-1})\varepsilon(s) \} \\ &= (\sigma_I P_I \dot{\theta}_I - \sigma_O P_O \dot{\theta}_O)(\sigma_I P_I \dot{\theta}_I - \sigma_O P_O \dot{\theta}_O)' \\ &\quad + \frac{T_3}{N} + \frac{\sigma_c}{N} \text{cov} \{ C(q^{-1})\tilde{\varphi}(t) \}, \end{aligned} \quad (48)$$

as $N \rightarrow \infty$. Here the first term in the right hand side of the second equality in (48) follows from (43) and (12). For the last term in the right hand side of (48) the calculations are similar as those were involved in evaluating T_1 in (47). It remains to evaluate

$$\begin{aligned} T_3 &:= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^N \sum_{s=1}^N \mathcal{E} \{ \tilde{\varphi}(t) v(s) \} \mathcal{E} \{ v(t) \tilde{\varphi}'(s) \} \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\tau=-N}^N (N - |\tau|) \lambda(\tau) \lambda'(-\tau) \end{aligned}$$

where $\tau = t - s$, and

$$\begin{aligned} \lambda(\tau) &:= \mathcal{E} \{ \tilde{\varphi}(t + \tau) v(t) \} \\ &= \mathcal{E} \{ G_1(q^{-1})\tilde{y}(t + \tau) \cdot A(q^{-1})\tilde{y}(t) \} \\ &\quad - G_2(q^{-1})\tilde{u}(t + \tau) \cdot B(q^{-1})\tilde{u}(t) \} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \chi(\omega) e^{i\omega\tau} d\omega, \end{aligned}$$

where

$$\chi(\omega) = \sigma_O G_1(e^{-i\omega}) A(e^{i\omega}) - \sigma_I G_2(e^{-i\omega}) B(e^{i\omega}).$$

Since $||\lambda(\tau)||$ and $||\lambda(-\tau)||$ decay exponentially with τ we have

$$\begin{aligned} T_3 &= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \sum_{\tau=-\infty}^{\infty} \chi(\omega_1) \chi'(\omega_2) e^{i\tau(\omega_1 - \omega_2)} d\omega_1 d\omega_2 \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \chi(\omega) \chi'(\omega) d\omega \end{aligned}$$

which is (20).

REFERENCES

- [1] S. Van Huffel and P. Lemmerling, Eds., *Total Least Squares and Errors-in-Variables Modelling, Analysis, Algorithms and Applications*. The Netherlands: Kluwer Academic Publishers, 2002.
- [2] T. Söderström, U. Soverini, and K. Mahata, "Perspectives on errors-in-variables estimation for dynamic systems," *Signal Processing*, vol. 82, pp. 1139–1154, 2002.
- [3] E. Karlsson, T. Söderström, and P. Stoica, "The Cramér-Rao lower bound for noisy input output systems," *Signal Processing*, vol. 80, no. 11, pp. 2421–2447, November 2000.
- [4] T. Söderström, "Identification of stochastic linear systems in presence of input noise," *Automatica*, vol. 17, pp. 713–725, 1981.
- [5] S. Beghelli, R. Guidorzi, and U. Soverini, "The Frisch scheme in dynamic system identification," *Automatica*, vol. 26, pp. 171–176, 1990.
- [6] P. Castaldi and U. Soverini, "Identification of dynamic errors-in-variables models," *Automatica*, vol. 32, no. 4, pp. 631–636, April 1996.
- [7] R. Diversi, R. Guidorzi, and U. Soverini, "A new criterion in EIV identification and filtering applications," in *13 th IFAC Symposium on System Identification*, Rotterdam, The Netherlands, August 27–29 2003.
- [8] W. X. Zheng, "Transfer function estimation from noisy input and output data," *International Journal of Adaptive Control and Signal Processing*, vol. 12, pp. 365–380, 1998.
- [9] M. Ekman, M. Hong, and T. Söderström, "A separable nonlinear least-squares approach for identification of linear systems with errors in variables," in *14th IFAC Symposium on System Identification*, Newcastle, Australia, March, 29–31 2006.
- [10] T. Söderström and P. Stoica, *System Identification*. Hemel Hempstead, UK: Prentice Hall International, 1989.
- [11] L. Ljung, *System Identification - Theory for the User, 2nd edition*. Upper Saddle River, NJ, USA: Prentice Hall, 1999.
- [12] P. Stoica and T. Söderström, "Optimal instrumental variable estimation and approximate implementation," *IEEE Transactions on Automatic Control*, vol. AC-28, pp. 757–772, 1983.
- [13] P. Stoica, T. Söderström, and B. Friedlander, "Optimal instrumental variable estimates of the AR parameters of an ARMA process," *IEEE Transactions on Automatic Control*, vol. AC-30, pp. 1066–1074, 1985.
- [14] P. Janssen and P. Stoica, "On the expectation of the product of four matrix-valued Gaussian random variables," *IEEE Transactions on Automatic Control*, vol. AC-33, pp. 867–870, 1988.