On data preprocessing for subspace methods

Dietmar Bauer
Institute f. Econometrics, Operations Research and System Theory,
TU Wien, Argentinierstr. 8, A-1040 Wien,
e mail: Dietmar.Bauer@tuwien.ac.at

Abstract

In modern data analysis often the first step is to perform some data preprocessing, e.g. detrending or elimination of periodic components of known period length. This is normally done using least squares regression. Only afterwards black box models are estimated using either pseudo-maximum-likelihood methods, prediction error methods or subspace algorithms. In this paper it is shown, that for subspace methods this is essentially the same as including the corresponding input variables, e.g. a constant or a trend or a periodic component, as additional input variables. Here essentially means, that the estimates only differ through the choice of initial values.

Keywords: subspace algorithms, linear systems, estimation, identification

1 Introduction

It has become standard in modern time series analysis to perform some form of preprocessing on the data prior to identification. However often the effects of this preprocessing are not dealt with in the identification phase. As an example consider the analysis of so called subspace methods: Many algorithms have been proposed by different authors, which are all subsumed under the name 'subspace algorithms', as they show certain similarities. To name just the most popular we cite `CCA` (Larimore, 1983), `MOESP` (Verhaegen, 1994) and `N4SID` (Van Overschee and DeMoor, 1994). The properties of these algorithms have been analysed in a number of papers (for references see e.g. Bauer *et al.*, 1999). However all cited references refer to the case, that no preprocessing prior to identification takes place. Also normally some persistency conditions on the input are imposed, which excludes e.g. the constant as an input variable. Therefore essentially it is assumed, e.g. that the only source for a nonzero mean of the output is due to the filtered input. When this does not hold and no preprocessing is performed, the estimates will be biased and not consistent. If preprocessing is performed the properties of the data change, which is reflected in the asymptotic variances of the estimated system. It seems to be more convenient to perform both the preprocessing and the actual identification in one step, in order to unify the estimation and testing procedure.

The aim of this paper is to show, that the preprocessing and the estimation phase in fact can be unified by including the constant or similar terms as input variables and use generalised inverses i.e. regularisation techniques in the respective regressions. In fact, it will be shown, that this leads to essentially the same results in the sense, that there is no difference in the attained accuracy, i.e. that the asymptotic distribution of the estimated systems are identical. This shows, that some of the preprocessing might be included into the subspace algorithm without any problem, and thus can also be analysed along the same lines.

The organisation of the paper is as follows: In the next section the model set and assumptions are stated. Section 3 then gives a short description of the main steps of the procedure. Section 4 states the main results of the paper. In section 5 a numerical example is presented and section 6 concludes.

2 Model set and assumptions

In this paper we deal with discrete time, finite dimensional, time invariant, state space systems of the form

$$\begin{array}{rcl} x_{t+1} & = & Ax_t + Bz_t + K\varepsilon_t \\ y_t & = & Cx_t + Dz_t + \varepsilon_t \end{array} \tag{1}$$

Here y_t denotes the s dimensional observed output, z_t denotes the m dimensional observed input, ε_t the s dimensional white noise. $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, C \in \mathbb{R}^{s\times n}, D \in \mathbb{R}^{s\times m}, K \in \mathbb{R}^{n\times s}$ are real matrices. The system usually will be described as (A, B, C, D, K). We will assume throughout, that ε_t is i.i.d. with zero mean and variance matrix $\Omega > 0$, having finite fourth moments. Throughout we will assume that the system is stable, i.e. $|\lambda_{max}(A)| < 1$ holds, where λ_{max} denotes an eigenvalue of maximum modulus. It will also be assumed, that the system is strictly minimum-phase, i.e. that $|\lambda_{max}(A - KC)| < 1$.

Corresponding to the input we will assume, that the input can be partitioned into two parts: $u_t \in \mathbb{R}^{m_i}$ and

$p_t \in \mathbb{R}^{m-m_i}$. Here u_t accounts for the identification input, whereas p_t denotes the additional inputs due to the preprocessing. These additional inputs will be restricted to the following choices:

- a constant term, i.e. $p_t = 1, \forall t$
- a periodic component, i.e. $p_{t,1} = sin(\omega t + \phi)$ for known $\omega \in (-\pi, \pi]$ and ϕ. In this case also the lagged variable $p_{t,2} = p_{t-1,1}$ has to be included
- a time trend, i.e. $p_t = t, \forall t$

These choices include the typical preprocessing like detrending and eliminating periodic components of known periodicity. Naturally, all these terms could be used in combination. Especially the inclusion of the trend makes the inclusion of a constant necessary in order to account for unknown initial effects. The key feature of these inputs is that they are persistent of order one. This is the reason for including the lagged term for the periodic components. It will be clear from the text, why this is needed. Note that in a similar fashion also more complicated preprocessing can be dealt with, as long as it is done using regression onto deterministic processes which are persistent of order one.

Note, that the model as stated is not identifiable for these inputs, which can be seen from a discussion of the constant: Assume that the input has nonzero mean μ_u and the output has mean μ_y. Then we obtain

$$\mu_y = D\mu_z + \sum_{j=0}^{\infty} CA^j B\mu_z$$

Thus it is clear, that the contribution of the constant is equal to $D_p + \sum_{j=0}^{\infty} CA^j B_p$, where B_p and D_p denote the columns of B and D respectively corresponding to the constant. This leads to a linear equation in B_p and D_p showing the nonidentifiability, which will complicate the analysis of the asymptotic distribution.

According to the model the output process $y_t = \bar{y}_t + \theta_y p_t$, where in $\bar{y}_t$ the deterministic components are not contained. Similarily $u_t = \bar{u}_t + \theta_u p_t$. Thus the preprocessing aims at estimating $\bar{y}_t$ and $\bar{u}_t$.

In the following we will always distinguish between these two kinds of inputs mainly for the reason of stating assumptions for the identification inputs. The reason is the different nature of the inputs: All preprocessing inputs used above have the characteristic of being perfectly predictable from one observation, i.e. they are persistently exciting of degree one. For the identification inputs however higher degrees of persistency will be required.

3 Description of the algorithms

Subspace procedures have been described in a number of papers. Therefore we restrict ourselves to a short description only. For details see e.g. the survey in (Bauer, 1998, Chapter 3). Let $Y_{t,f}^+ = [y_t', \cdots, y_{t+f-1}']'$ denote the stacked vector of outputs, where f is a user defined integer. Similarily define $U_{t,f}^+$ and $E_{t,f}^+$ using the inputs and the noise respectively in the place of the outputs. Furthermore let $Z_{t,p}^- = [y_{t-1}', u_{t-1}', \cdots, y_{t-p}', u_{t-p}']'$, where p is a user defined integer. Then it is easy to show the following equation:

$$Y_{t,f}^+ = \mathcal{O}_f \mathcal{K}_p Z_{t,p}^- + \mathcal{O}_f \bar{A}^p x_{t-p} + \mathcal{U}_f U_{t,f}^+ + \mathcal{E}_f E_{t,f}^+. \quad (2)$$

Here $\bar{A} = (A - KC)$, $\bar{B} = (B - KD)$, $\mathcal{O}_f = [C', A'C', \ldots, (A^{f-1})']'$ and

$$\mathcal{K}_p = [[K, \bar{B}], \bar{A}[K, \bar{B}], \ldots, \bar{A}^{p-1}[K, \bar{B}]$$

Further $\mathcal{E}_f$ denotes the block Toeplitz matrix, whose i-th block row is equal to

$$[CA^{i-2}K, \ldots, CK, I_s, 0^{s\times(f-i)s}]$$

where I_s is the s-dimensional identity and $0^{a\times b}$ is a $a \times b$ dimensional nullmatrix. Finally $\mathcal{U}_f$ denotes the block Toeplitz matrix, whose i-th column is equal to

$$[CA^{i-2}B, \ldots, CB, D, 0^{s\times(f-i)m}]$$

This equation is the starting point for all subspace algorithms. An outline can be given as follows:

1. Regress $Y_{t,f}^+$ onto $Z_{t,p}^-$ and $U_{t,f}^+$ in order to obtain estimates $\hat{\beta}_z$ and $\hat{\beta}_u$.
2. $\hat{\beta}_z$ will typically be of full rank, whereas the limit β_z is of rank n. Thus a SVD of $\hat{W}_f^+ \hat{\beta}_z \hat{W}_p^- = \hat{U}\hat{\Sigma}\hat{V}' = \hat{U}_n\hat{\Sigma}_n\hat{V}_n' + \hat{R}$ is performed leading to an estimate $\hat{\mathcal{O}}_f = (\hat{W}_f^+)^{-1}\hat{U}_n, \hat{\mathcal{K}}_p = \hat{\Sigma}_n\hat{V}_n'(\hat{W}_p^-)^{-1}$. Here $\hat{W}_f^+$ and $\hat{W}_p^-$ are weighting matrices, which have to be chosen by the user. We will assume throughout that they are nonsingular (a.s.) and that there exist matrices W_f^+ and W_p^- respectively such that $\|\hat{W}_f^+ - W_f^+\| \to 0$ and $\|\hat{W}_p^- - W_p^-\| \to 0$ for $T \to \infty$ and possibly $f, p \to \infty$ as a function of T. $\hat{\Sigma}_n$ contains the n dominant singular values and $\hat{U}_n, \hat{V}_n$ the corresponding left or right respectively singular vectors.
3. Either the estimates $\hat{\mathcal{O}}_f$ and $\hat{\beta}_u$ or the estimate $\hat{\mathcal{K}}_p$ is used to estimate the system.

In the last step, one has two choices: `MOESP` and one variant of `N4SID` use the structure in $\mathcal{O}_f$ and β_u in

order to estimate the system matrices, whereas CCA and another variant of N4SID use $\hat{\mathcal{K}}_p$ to estimate the state as $\hat{x}_t = \hat{\mathcal{K}}_p Z_{t,p}^-$ and then use the system equations (1) in order to obtain estimates of the system.

There are two possible ways to include the preprocessing in this framework:

- Regress y_t and u_t onto p_t and let the residuals be denoted as $\tilde{y}_t$ and $\tilde{u}_t$. This preprocessed data then can be used in the subspace method in order to obtain an estimate of the system (A, B, C, D, K).
- Use y_t and $z_t = [u_t', p_t']'$ as the data for the subspace method to obtain an estimate of the system (A, B, C, D, K) of the transfer function relating u_t to y_t.

The aim of the next section is to show that these two procedures deliver essentially identical estimates.

4 Main results

Note, that in the present setup the regression in the first step to obtain the estimates $\hat{\beta}_z, \hat{\beta}_u$ will have to take into account the singularity of the matrix of the regressors due to the multicollinearity introduced by the process p_t. For $Z_{t,p}^-$ note, that in the actual calculation only the vectors $Z_{t,p}^{-,\Pi}$ are used for the estimate $\hat{\beta}_z$. Here $Z_{t,p}^{-,\Pi}$ denotes the residuals of $Z_{t,p}^-$ regressed onto $U_{t,f}^+$. Using the notation $\langle a_t, b_t \rangle = T^{-1} \sum_{t=p+1}^{T-f} a_t b_t'$ this is equal to

$$Z_{t,p}^- - \langle Z_{t,p}^-, U_{t,f}^+ \rangle \langle U_{t,f}^+, U_{t,f}^+ \rangle^\dagger U_{t,f}^+$$

Here † denotes the Moore Penrose pseudo inverse. Due to the nature of the variables p_t, the residuals in all coordinates corresponding to p_t are zero. Note that this corresponds to using the preprocessed data $\tilde{y}_t, \tilde{u}_t$ up to the difference of at most $f+p$ terms in the calculation of the regressions for the preprocessing. This difference is asymptotically negligible under the usual assumptions on f and p to be $o((\log T)^a)$ for some $a < \infty$. Therefore the columns of the estimate $\hat{\beta}_z$, which correspond to y_t or u_t respectively, are equivalent to the ones obtained from preprocessed data. The remaining columns correspond to p_t and these are essentially zero (i.e. zero up to effects due to initial values in the regressions). Also note that the calculation of the pseudoinverse could be circumvented by using only those coordinates of $U_{t,f}^+$, which are linearily independent. These can be found easily, since the dependence structure of p_t is known. In fact, if the regression is performed according to the formula given above, all coordinates corresponding to $p_{t+j}, j > 0$ simply can be omitted without changing the result.

In the MOESP type of procedures this shows that the estimates $\hat{\mathcal{O}}_f$ of the two alternatives are essentially identical. This shows the equivalence of the estimates $\hat{A}$ and $\hat{C}$. For the estimation of B and D consider the estimates corresponding to the MOESP type of procedures: Let $\mathcal{U}_f(\hat{\mathcal{O}}_f)$ denote the estimate of $\mathcal{U}_f$ given B and D, where elements of $\mathcal{O}_f$ are replaced by their respective estimates. Note that $\mathcal{U}_f(\hat{\mathcal{O}}_f)$ is linear in B and D. Then the estimate of B and D can be obtained from minimising

$$\| \hat{\mathcal{O}}_f^\perp \left(\langle Y_{t,f}^+, U_{t,f}^+ \rangle \langle U_{t,f}^+, U_{t,f}^+ \rangle^\dagger - \mathcal{U}_f(\hat{\mathcal{O}}_f) \right) \|_{Fr}$$

Here $\hat{\mathcal{O}}_f^\perp \hat{\mathcal{O}}_f = 0$ and $\hat{\mathcal{O}}_f^\perp$ is of full row rank. Further $\|.\|_{Fr}$ denotes the Frobenius norm. From this it follows using the block matrix inversion, that the columns corresponding to u_t are identical to the respective columns obtained from the preprocessed data $\tilde{y}_t, \tilde{u}_t$. From this it follows that the estimation of the part of B and D, which corresponds to u_t is estimated essentially identical to the estimates using the preprocessed data. This shows, that in this case the estimates of the coordinates of the system (A, B, C, D) corresponding to u_t is essentially the same as obtained from the preprocessed data. Concerning the part due to p_t however the same result does not seem to hold, compare the example given in section 5. This is a matter of further research.

For the Larimore type of approach we have seen that the estimate $\hat{\beta}_z$ is essentially identical to the estimate obtained from the preprocessed data in the coordinates corresponding to the output or the input due to u_t, the other components being zero. Therefore the same holds true for $\hat{\mathcal{K}}_p$, noting that for the essential parts of the usual weighting matrices $\hat{W}_f^+, \hat{W}_p^-$ (note that certain entries of $\hat{\beta}_z$ are zero) also only the data $Z_{t,p}^{-,\Pi}$ and $Y_{t,f}^{+,\Pi}$ is used. Here $Y_{t,f}^{+,\Pi}$ is defined analogously to $Z_{t,f}^{-,\Pi}$. These vectors are essentially identical to the corresponding vectors formed from the preprocessed data. As a consequence the state estimate $\hat{x}_t = \hat{\mathcal{K}}_p Z_{t,p}^-$ is almost identical to the estimate obtained from the preprocessed data except for differences in the space spanned by the components of p_t. Note that the next step in this class of algorithms is to use the state in a regression, where these deterministic components are included. This again shows, that the estimate of the system matrices corresponding to u_t coincide up to initial conditions with the estimate obtained from the preprocessed data. Note, that the nonidentifiability in this setup is circumvented, as the deterministic components in $\hat{x}_{t+1}$ and $\hat{x}_t$ converge to some vectors $\bar{p}_{t+1}^x$ and $\bar{p}_t^x$ respectively. Letting $\bar{p}_t^z$ denote the coefficients of the deterministic components of z_t then it follows from the state equation that

$$\bar{p}_{t+1}^x = A\bar{p}_t^x + B\bar{p}_t^z$$

which determines $B_p p_t$ completely and thus also B_p

is uniquely defined by this. Therefore the algorithm does not have any nonidentifiability problems by construction and the estimates for B_p, the columns due to the deterministic components, are consistent. After the estimation the nonidentifiability of the model structure could be used to obtain a different normalisation as $B_p = 0$ using the structure of the nonidentifiability. In both cases the asymptotic distribution of $(\hat{A}, \hat{B}_m, \hat{C}, \hat{D}_m, \hat{K})$ will be the same, as if the corresponding estimates have been calculated using the preprocessed data, as is straightforward to see. Finally we have obtained the following theorem:

Theorem 1 *Let y_t be generated by a system of the form (1), where the noise is i.i.d. with finite moments up to fourth order. Let the input z_t consist of two different components: $z_t = [u_t', p_t']'$, where u_t is a quasistationary sequence, which is persistently exciting of order $f + p$. Further p_t contains a constant term, a trend component or a periodic component with known frequency and shift as described above. Then the estimates of $G_i(q) = D_m + C(qI_n - A)^{-1}B_m$, the transfer function relating u_t to y_t obtained by the inclusion of p_t in the `MOESP` type of subspace methods and the estimates obtained by using the residuals from a regression of y_t and u_t onto p_t as the data input for the subspace method without p_t are asymptotically equivalent, i.e. the asymptotic distribution of the estimation error are identical.*

Let the input u_t be an ARMA process generated by i.i.d. white noise having a spectrum $f_u(\lambda)$, such that $0 < c_1 I_{m_i} \leq f_u(\lambda) \leq c_2 I_{m_i}, \lambda \in [-\pi, \pi]$ for some constants $0 < c_1 < c_2 < \infty$. Then for $p = p(T) = o((\log T)^a), p(T) \geq -\frac{d \log T}{2 \log |\rho_0|}$ the same holds for the Larimore type of algorithms. In this case also the estimates of the transfer function relating the noise to the output will be asymptotically equivalent. Furthermore also the estimates of the effects of the deterministic terms are essentially the same.

The asymptotic variances of the estimates of $G_i(q)$ obtained using the preprocessed data are identical to the corresponding variances using data, where the limits of the regression coefficients of the regression of y_t and u_t onto p_t are used to eliminate the deterministic components.

The first two claims have been dealt with above the theorem, the last claim is an easy consequence of the fact, that the subspace methods can be seen as (essentially) implementing a nonlinear function mapping the covariance estimates to the system matrix estimates. For the third statement consider the asymptotic properties of the sample covariance using $\bar{y}_t$ and $\tilde{y}_t$ respectively. The difference is equal to

$$\begin{aligned} \bar{y}t\bar{y}'_{t+j} - \tilde{y}_t\tilde{y}'_{t+j} &= (\bar{y}_t - \tilde{y}_t)\bar{y}'_{t+j} + \tilde{y}_t(\bar{y}_{t+j} - \tilde{y}_{t+j}) \\ &= (\theta_y - \hat{\theta}_y)p_t\bar{y}'_{t+j} + \tilde{y}_t p_t'(\theta_y - \hat{\theta}_y)' \end{aligned}$$

and thus the convergence in distribution to zero follows. The arguments for $\bar{u}_t$ are analogous. Thus the same result holds for the subspace algorithms.

It should be noted, that the assumptions of the theorem are by no means necessary. In fact, much weaker conditions suffice. For a discussion on the necessary conditions in the `MOESP` type of algorithms see (Bauer and Jansson, 2000), for assumptions in a martingale framework for the Larimore type of procedures see (Bauer *et al.*, 1999).

The significance of the theorem is that it gives the user two different but equivalent ways to deal with nonzero means, drifts and periodic components included in many time series. Either one can test for the necessity to preprocess the data beforehand and then accordingly preprocess the data and use the preprocessed data in order to obtain estimates of the dynamical systems, which are the main goal in many applications. The alternative way is to use the deterministic terms as additional inputs in the subspace method and test for the necessity of the inclusion after the estimation. Both procedures have advantages and disadvantages. The big advantage for the preprocessing approach lies in the fact, that the testing procedures, whether the deterministic components are contained in the data at hand, are well established and implemented in many programs. This facilitates the analysis for the user.

On the other hand the inclusion of the terms in the estimation procedure is computationally simple. The original procedures can be used, if the regressions are done in a robust way using the pseudoinverses in the case of multicollinearities. The direct embedding of the preprocessing makes the calculation of the variance straightforward, although at the present time no programs to calculate the asymptotic variance of the subspace procedure exist, that could be used in an industrial context. Tests for more complex hypotheses can be obtained from the asymptotic distribution. One such test could be, whether the deterministic components in the output can be explained only through the components occurring in the input. As is seen in the static regression case, preprocessing in this case leads to higher variances of the estimated parameters. As a simple example take $y_t = Du_t + \varepsilon_t$, where $u_t = c + n_t$. Here ε_t and n_t are i.i.d. and c is a constant. The variance of the coefficient D is proportional to the inverse of $\sum_{t=1}^{T} n_t^2$ in the preprocessed case, whereas without preprocessing the variance is proportional to $(\sum_{t=1}^{T} u_t^2)^{-1}$. The difference can be arbitrarily large depending on c. This justifies the development of a test as suggested above.

Method		A	B	D
MOESP	no pre.	0.9183	1.1498	0.0599
	pre.	0.4983	0.9996	0.0013
	incl.	0.4985	1.0001	0.0013
	corr.	0.9983	0.9980	0.9978
CCA	no pre.	0.8886	1.2125	0.0710
	pre.	0.4999	0.9994	0.0011
	incl.	0.5002	0.9998	0.0011
	corr.	0.9970	0.9974	0.9974

Table 1: This table shows the mean values of the estimates of A, B_m and D_m estimated with sample size $T = 1000$ using MOESP and CCA on the original data (no pre.), on the preprocessed data (pre.) or including a constant term (incl.) in the calculations. Also the correlation (corr.) between the estimates (pre.) and (incl.) is given.

5 Numerical Example

In this section we will present some simple examples, which illustrate the theory given in the last section. The first example is a one state SISO system, which is described by the following matrices:

$$A = 0.5, B_m = 1, C = 1, D_m = 1, K = 1$$

The input is Gaussian white noise with mean 1 and variance 1. The noise ε_t is chosen to be zero mean Gaussian white noise of variance 1. The output y_t is equal to

$$y_t = (D_m + C(q-A)^{-1}B_m)u_t + (1 + C(q-A)^{-1}K)\varepsilon_t + 10$$

Here q denotes the forward shift operator. 1000 data sets of dimension $T = 1000$ have been generated and for each data set the system is estimated

- using no preprocessing and no inclusion of a constant,
- mean corrected data and
- using the original data with the addition of a constant input variable.

$f = p = 5$ has been chosen for the MOESP type of procedure and $f = 5, p = 15$ for the Larimore type of procedure, in all the trials. After the estimation the systems have been transformed to echelon canonical form. In both cases the CCA weightings have been used. The mean of the corresponding estimates can be seen in Table 1.

It can be seen clearly, that the estimates without taking the nonzero constants into account is biased. It also can be seen, that the other two approaches give comparable means and are highly correlated, as is expected from the last section. This applies for both types of procedures.

Comparing the mean of y_t according to the estimated model to the actually estimated mean on the other hand shows a different picture: For the CCA procedure a correlation of 0.9989 is obtained, i.e. almost perfect agreement, while for the MOESP procedure the correlation is estimated as 0.0948 showing almost no linear dependence.

As a second example the same system is used, but the input is changed by adding a randomly weighted deterministic component and also the output is contaminated by a similar term:

$$u_t = v_t + \beta x_t \quad , \quad x_t = \begin{bmatrix} t \\ sin(0.5\pi t) \\ sin(0.22\pi t) \\ 1 \\ sin(0.5\pi(t-1)) \\ sin(0.22\pi(t-1)) \end{bmatrix}$$

Here each component of β is drawn from a $[0, 1]$ uniform distribution. Also

$$y_t = G_i(q)u_t + H(q)\varepsilon_t + \alpha x_t$$

where

$$\begin{aligned} G_i(q) &= D_m + C(qI - A)^{-1}B_m \\ H(q) &= I + C(qI - A)^{-1}K \end{aligned}$$

The components of α are uniformely $[0, 1]$ distributed. Finally v_t and ε_t are Gaussian zero mean and unit variance random processes. 1000 time series with these characteristics of sample size $T = 1000$ are constructed. The estimation shows the same picture as the first example: Figure 1 shows the standard deviations of the estimates of the transfer function $D_m + C(qI - A)^{-1}B_m$ at 50 points in the angular frequency range $[-\pi, \pi]$. Hereby $f = p = 5$ was used and only the MOESP type of procedures is considered. The plot shows the standard deviation of the estimates of the transfer function $G_i(q)$ using the MOESP estimates on the preprocessed data and the standard deviation of the difference between the estimates obtained from the preprocessed data and the estimates obtained using the original data and including the additional variables, respectively.

It can be seen clearly, that the difference in between the two approaches is of lower magnitude than the estimation errors themselves. For the CCA case the results are almost identical, therefore the presentation of the results is omitted.

In a final experiment $\alpha = 0$ is chosen, making a preprocessing unnecessary. All procedures then lead to consistent estimates, however the asymptotic variances change, as can be seen from Figure 2. In this plot

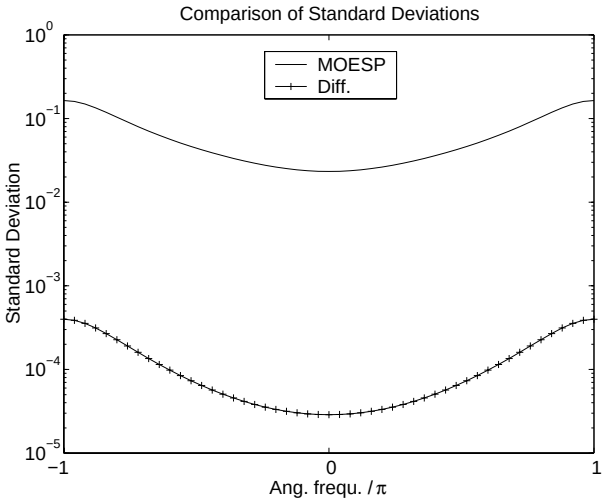

Figure 1: The plots show the standard deviation of the transfer function estimates obtained from the preprocessed data and $f = p = 5$ (—) and the standard deviation of the difference of the two different procedures pre. and incl. according to Table 1 (-+-). The plot has been generated using 1000 time series of sample size $T = 1000$.

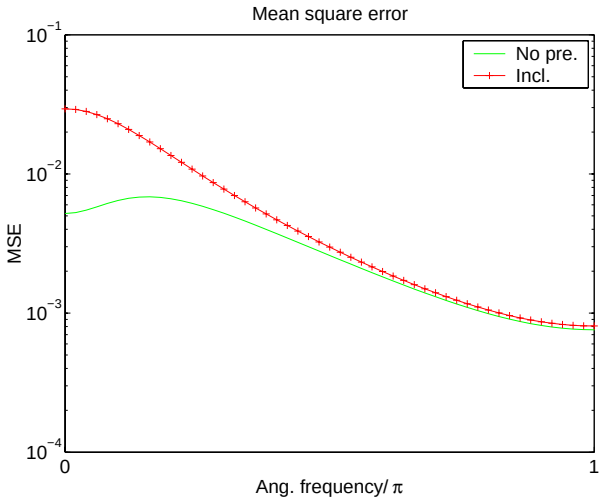

Figure 2: The plots show the mean square error of the transfer function estimates obtained from the preprocessed data and $f = p = 5$ (incl,-+-) and the procedure using the original data (,no pre.,—). The plot has been generated using 1000 time series of sample size $T = 1000$.

the mean square error for the two procedures using the original data without preprocessing (no pre.) and the subspace procedure using x_t as additional inputs are plotted in the angular frequency range $[0, \pi]$. It can be seen clearly, that as expected from the linear regression case the estimates of the transfer function using the preprocessed data are worse than the corresponding mean square errors for the original data case. This is in particular pronounced for the low frequency region, which refers to the excitation introduced by the trend and the mean and also the frequency 0.22π.

6 Conclusions

In this paper we derived the asymptotic properties of subspace procedures, when the data, which is used for the procedure, is preprocessed by removing trends and periodic components. It has been shown, that the preprocessing can alternatively be interpreted as the inclusion of additional input terms. This makes it possible to calculate the asymptotic variance of the estimates obtained with the procedure from the standard subspace algorithms. Also test procedures for more complicated hypotheses can be developed leading to better estimates in certain cases.

Acknowledgements

This work has been done, while the author was holding a post doc position at the Department for Automatic Control, University of Linköping, Sweden. The financial support by the EU TMR project 'SI' and by the Austrian FWF under project No. P14438-INF is gratefully acknowledged.

References

Bauer, D. (1998). Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms. PhD thesis. TU Wien.

Bauer, D. and M. Jansson (2000). Analysis of the asymptotic properties of the MOESP type of subspace algorithms. *Automatica* **36**(4), 497–509.

Bauer, D., M. Deistler and W. Scherrer (1999). Consistency and asymptotic normality of some subspace algorithms for systems without observed inputs. *Automatica* **35**, 1243–1254.

Larimore, W. E. (1983). System identification, reduced order filters and modeling via canonical variate analysis. In: *Proc. 1983 Amer. Control Conference 2.* (H. S. Rao and P. Dorato, Eds.). Piscataway, NJ. pp. 445–451. IEEE Service Center.

Van Overschee, P. and B. DeMoor (1994). N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* **30**, 75–93.

Verhaegen, M. (1994). Identification of the deterministic part of mimo state space models given in innovations form from input-output data. *Automatica* **30**(1), 61–74.