

Estimating the System Order by Subspace Methods

Alfredo García-Hiernaux*
José Casals
Miguel Jerez

November 2, 2007

Abstract

This paper discusses how to determine the order of a state-space model. To do so, we start by revising existing approaches and find in them three basic shortcomings: i) some of them have a poor performance in short samples, ii) most of them are not robust and iii) none of them can nowadays accommodate seasonality. We tackle the first two issues by proposing new and refined criteria. The third issue is dealt with by decomposing the system into regular and seasonal sub-systems. The performance of all the procedures considered is analyzed through Monte Carlo simulations.

Keywords: System order, State-space models, subspace methods, information criteria, seasonality

JEL Classification: C32, C51, C52

*Corresponding author. Departamento de Estadística. Universidad Carlos III. Campus de Getafe. 28903 Madrid (SPAIN). email: aghierna@est-econ.uc3m.es, tel: (+34) 91 624 58 52, fax: 91 624 98 49.

1 Introduction

The order of a linear system is the number of dynamic components that must be combined to represent the data dynamics. Determining this value is critical in applied data modelling and, accordingly, there is an extensive literature about this subject which can be broken into two broad categories: i) preliminary data analysis methods and ii) model comparison procedures.

The first group includes very different proposals. Some of them check whether the model dimension is too high using the condition number of the information matrix, see Stoica and Söderström (1982). Other authors, Cooper and Wood (1982), Tsay (1989), Tiao and Tsay (1989), estimate the system order using canonical correlations. In a recent paper, Toscano and Reisen (2000) revise and compare these methods. Aoki (1990) applies the same technique to state space models. Close to this approach, Bauer (2001) presents a subspace-based criterion to determine the system order, while Sorelius (1999) does it by testing the rank of a particular sample covariance matrix, see also Woodside (1971). Finally, other authors apply spectral techniques to fit an empirical transfer function, which points out what system orders are required to describe the data dynamics, see e.g., Bujosa et al. (2007).

The second class of methods typically rely on information criteria. Broadly speaking, they compare several alternative models by balancing their data fit, measured through the log-likelihood, with a penalty that depends on the sample size and a measure of model complexity. The pioneer criterion in this line is due to

Akaike (1976), which seminal work motivated many other proposals such as those of Schwartz (1978) and Hannan and Quinn (1979).

When applied to economic time series, all these procedures share a basic shortcoming: they were not originally designed to cope with seasonality. On the other hand, Monte Carlo experiments reveal two additional weaknesses: (a) many procedures are not robust, meaning that their ability to choose the right system order depends critically of the dynamics of the data; and (b) some methods display a poor performance in short samples. According to this diagnostic, the contribution of this paper is threefold. First, we propose two refinements of the criteria suggested by Bauer (2001), designed to improve its finite-sample performance. Second, we present a simple automatic procedure that has a clear advantage in terms of robustness when compared with other methods. Last, we extend the use of any criterion to determine the order of the seasonal and regular sub-systems in multiplicative seasonal processes.

The paper is structured as follows. Section 2 defines the notation and introduces subspace estimation which will be our framework. The main order-determination procedures are reviewed in Section 3. Section 4 presents the main results of the paper. In Section 5 we discuss the performance of several methods by a simulation study. Finally, Section 6 provides some conclusions and indicates how to obtain a free MATLAB toolbox that implements all the algorithms proposed.

2 Model set and subspace estimation algorithms

Consider a linear fixed-coefficients system that can be described by the state space (SS) model,

$$\mathbf{x}_{t+1} = \mathbf{\Phi}\mathbf{x}_t + \mathbf{E}\boldsymbol{\psi}_t \quad (1a)$$

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \boldsymbol{\psi}_t \quad (1b)$$

where $\mathbf{x}_t \in \mathbb{R}^n$ is a state vector and n denotes the true order of the system. $\mathbf{z}_t \in \mathbb{R}^m$ is an observable output vector, $\boldsymbol{\psi}_t \in \mathbb{R}^m$ is a white noise vector such that $\boldsymbol{\psi}_t \text{ iid}N(\mathbf{0}, \mathbf{Q})$ and $\mathbf{\Phi} \in \mathbb{R}^{n \times n}$, $\mathbf{E} \in \mathbb{R}^{n \times m}$ and $\mathbf{H} \in \mathbb{R}^{m \times n}$ are the parametric matrices of the system. The system is assumed to be stable and strictly minimum-phase, so that all the eigenvalues of $\mathbf{\Phi}$ and $(\mathbf{\Phi} - \mathbf{E}\mathbf{H})$ lie inside the unit circle.

We use the innovations model (1a)-(1b) since is simple and general, in the sense that any fixed-coefficients SS model can be written in this specific form (see e.g., Casals et al., 1999, Theorem 1). Also, no generality is lost by assuming that the model has no observable inputs, since any model with inputs can be decomposed into an input-related and an error-related model (see, Chui and Chen, 1999).

Subspace methods derive from the representation (1a-1b) in matrix form. By recursive substitution in (1a) we obtain,

$$\mathbf{x}_t = \mathbf{\Phi}^t \mathbf{x}_0 + \sum_{j=0}^{t-1} \mathbf{\Phi}^j \mathbf{E} \boldsymbol{\psi}_{t-j-1} \quad (2)$$

and substituting (2) into the observation equation (1b), we get:

$$\mathbf{z}_t = \mathbf{H}\Phi^t\mathbf{x}_0 + \mathbf{H}\sum_{j=0}^{t-1}\Phi^j\mathbf{E}\psi_{t-j-1} + \psi_t \quad (3)$$

so the endogenous variable, \mathbf{z}_t , depends on the initial state vector, \mathbf{x}_0 , and the present and past innovations, ψ_t . Equation (3) can be written in matrix form as,

$$\mathbf{Z}_i = \mathbf{O}_i\mathbf{X}_0 + \mathbf{V}_i\Psi_i \quad (4)$$

where the subscript i is an integer that denotes the dimension of the row space of \mathbf{Z}_i . In this work, we determine it using the rule $i = \max(4, h_t)$, being h_t the nearest integer to $\log T$, which assures the consistency of the estimators for appropriate choice of two weighting matrices in the subspace regression (see Bauer, 2005, Theorem 4).

Now we will define the symbols in equation (4). The following matrices are related to the data:

1) Block-Hankel Matrices (BHM): The dimension of this matrices is defined by the integer i , as:

$$\mathbf{Z}_p = \begin{pmatrix} z_1 & z_2 & \dots & z_{T-2i+1} \\ z_2 & z_3 & \dots & z_{T-2i+2} \\ \vdots & \vdots & & \vdots \\ z_i & z_{i+1} & \dots & z_{T-i} \end{pmatrix}; \quad \mathbf{Z}_f = \begin{pmatrix} z_{i+1} & z_{i+2} & \dots & z_{T-i+1} \\ z_{i+2} & z_{i+3} & \dots & z_{T-i+2} \\ \vdots & \vdots & & \vdots \\ z_{2i} & z_{2i+1} & \dots & z_T \end{pmatrix} \quad (5)$$

where \mathbf{z}_0 is a vector of m components. In (4), Ψ_i is as \mathbf{Z}_i but with ψ_t instead of \mathbf{z}_t .

2) The state sequence, defined as $\mathbf{X}_t = (\mathbf{x}_t \ \mathbf{x}_{t+1} \ \mathbf{x}_{t+2} \ \dots \ \mathbf{x}_{t+T-2i})$. The past and future state sequences beginning, respectively, at $t = 0$ and $t = i$, can also be written as $\mathbf{X}_p = \mathbf{X}_0$ and $\mathbf{X}_f = \mathbf{X}_i$.

On the other hand, equation (4) includes also the following matrices related to the parameters in model (1a-1b):

3) The Extended Observability Matrix, which is:

$$\mathbf{O}_i = \left(\mathbf{H}' \quad (\mathbf{H}\Phi)' \quad (\mathbf{H}\Phi^2)' \quad \dots \quad (\mathbf{H}\Phi^{i-1})' \right)' \in \mathbb{R}^{im \times n} \quad (6)$$

4) The lower block triangular Toeplitz matrix, defined as:

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\mathbf{E} & \mathbf{I}_m & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Phi\mathbf{E} & \mathbf{H}\mathbf{E} & \mathbf{I}_m & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}\Phi^{i-2}\mathbf{E} & \mathbf{H}\Phi^{i-3}\mathbf{E} & \mathbf{H}\Phi^{i-4}\mathbf{E} & \dots & \mathbf{I}_m \end{pmatrix} \in \mathbb{R}^{im \times im} \quad (7)$$

The future state sequence \mathbf{X}_f can be approximated by $\mathbf{M}\mathbf{Z}_p$, where the rank of \mathbf{M} is n , the true system order. This is because, for proper choices of i , other terms are asymptotically negligible (see, Katayama, 2005). Therefore, shifting

time subscripts in (4) and, substituting \mathbf{X}_f , by $\mathbf{M}\mathbf{Z}_p$, we obtain,

$$\mathbf{Z}_f = \mathbf{O}_i\mathbf{M}\mathbf{Z}_p + \mathbf{V}_i\Psi_f \quad (8)$$

where \mathbf{Z}_f , \mathbf{Z}_p and Ψ_f are as in (5), and \mathbf{O}_i and \mathbf{V}_i are respectively as in (6) and (7).

Subspace methods estimate \mathbf{O}_i , \mathbf{M} and \mathbf{V}_i in (8) by solving a reduced-rank weighted least square problem. This is done by means of the Singular Value Decomposition (SVD, Eckart and Young, 1936) of $\mathbf{W}_1\mathbf{Z}_f\mathbf{W}_2$, being \mathbf{W}_1 and \mathbf{W}_2 two weighting matrices (see, e.g., Favoreel et al., 2000). Then, an estimation of the parameter matrices in (1a-1b) can be straightforwardly obtained from \mathbf{O}_i , \mathbf{M} and \mathbf{V}_i . The literature provides different weighting matrices \mathbf{W}_1 and \mathbf{W}_2 . For instance, Larimore-type methods (Larimore, 1983, 1990) specify $\mathbf{W}_1 = (\mathbf{Z}_f\mathbf{Z}_f')^{-\frac{1}{2}}$ and $\mathbf{W}_2 = \mathbf{Z}_p'(\mathbf{Z}_p\mathbf{Z}_p')^{-1}\mathbf{Z}_p$.

3 An overview of some specification procedures

Estimating the dimension of the state vector, n , is not trivial, in particular when the sample is short. In this section we describe some criteria and test statistics that can be applied to this purpose, distinguishing between two types of approaches. The difference between them is in the need of estimating a model: the latter requires to do so while the former does not.

3.1 Preliminary data analysis methods

Tsay (1989) approaches order determination by testing the null hypothesis that the smallest canonical correlations $\hat{\sigma}_{n+1}, \hat{\sigma}_{n+2}, \dots, \hat{\sigma}_{im}$ are zero. To do so, he suggests the statistic $D(n)$, that can be straightforwardly adapted to subspace methods as:

$$D(n) = -(T - 2i + 1) \sum_{k=n+1}^{im} \log \left(1 - \frac{\hat{\sigma}_k^2}{\hat{d}_k} \right) \quad (9)$$

with

$$\hat{d}_k = 1 + 2 \sum_{l=1}^i \hat{\rho}_{pk}(l) \hat{\rho}_{fk}(l) \quad (10)$$

where $\hat{\rho}_{fk}(l)$ and $\hat{\rho}_{pk}(l)$ are the sample l -order autocorrelations of the canonical variables (see Tsay, 1989). Subscripts p and f refer to the past and future blocks respectively. The term \hat{d}_k is related to the asymptotic variance of the estimated canonical correlation coefficients in a context of serial correlation. Tsay (1989) proves that, under $H_0 : \hat{\sigma}_{n+1} = \hat{\sigma}_{n+2} = \dots = \hat{\sigma}_{im} = 0$, the statistic $D(n)$ follows a χ^2 distribution with $2(im - n)$ degrees of freedom.

In a subspace framework, Bauer (2001) proposes the *SVC* (Singular Values Criterion) which is based on the analysis of the largest neglected singular value in the SVD step. The criterion is defined as:

$$SVC(n) = \hat{\sigma}_{n+1}^2 + C(T)d(n) \quad (11)$$

where $\hat{\sigma}_{n+1}$ is the $n + 1$ singular value of $\mathbf{W}_1 \mathbf{Z}_f \mathbf{W}_2$ and $d(n) = 2nm$ denotes the number of parameters of a (1a-1b) system. The singular values are also denoted

by $\sigma_j \forall j = 1, 2, \dots, im$ as when Larimore-type methods are used, the weighting matrices $\mathbf{W}_1, \mathbf{W}_2$ are such that the singular values of $\mathbf{W}_1 \mathbf{Z}_f \mathbf{W}_2$ coincide with the canonical correlations between \mathbf{Z}_p and \mathbf{Z}_f . In that case, the idea is close to that of Tsay (1989).

The estimated order is obtained as the argument that minimizes (11). Bauer (2001) shows that *SVC* with $C(T) = \log T/T$ outperforms several information criteria when relatively large samples are analyzed, but recognizes that the penalty function is somewhat arbitrary. The author also finds that the choice of \mathbf{W}_1 has a large impact on the performance of *SVC* in finite samples.

3.2 Model comparison methods

Other useful tools to determine the system order are the, so-called, information criteria. The well-known *AIC* (Akaike, 1976), *SBC* (Schwartz, 1978) or *HQ* (Hannan and Quinn, 1979) belong to this family. A general representation of all these criteria is:

$$IC(n) = \frac{-2 \log l(n)}{T} + C(T)d(n) \quad (12)$$

where, $l(n)$ and $d(n)$ denote, respectively, the log-likelihood function value and the number of parameters corresponding to a system of order n . $C(T)$ is again a penalty function that depends on the criterion. In *AIC* $C(T) = 2/T$, in *SBC* $C(T) = \log T/T$ and in *HQ*, $C(T) = 2 \log \log T/T$.

These procedures are well known and widely employed, see recently Stoica et al. (2004) or Bengtsson and Cavanaugh (2006). However, they have an impor-

tant drawback, as they require estimating several models and, therefore, may be computationally expensive.

4 Main results

This section presents three criteria to determine the system order. Two of them are conceived to improve the performance of the *SVC* in small samples, while maintaining its consistency. The third one is devised to be more robust than alternative methods. Finally, we describe a method to determine the regular and seasonal orders in series with seasonal fluctuations.

4.1 Estimating the order of non-seasonal systems

To improve the performance of *SVC* we propose two enhancements: refining the choice of the weighting matrix and optimizing its penalty function.

[PLEASE, CHOOSE ONE OF THE TWO FOLLOWING PARAGRAPHS!]

The first improvement consists of substituting $\mathbf{W}_1 = (\mathbf{Z}_f \mathbf{Z}'_f)^{-\frac{1}{2}}$, hereafter $\mathbf{\Omega}_1$, by $\mathbf{W}_1 = (\tilde{\mathbf{Z}}_f \tilde{\mathbf{Z}}'_f)^{-\frac{1}{2}}$, from now on $\mathbf{\Omega}_2$, where $\tilde{\mathbf{Z}}_f$ contains the residuals of regressing \mathbf{Z}_f onto \mathbf{Z}_p in a previous step. This choice is supported by the fact that $\mathbf{\Omega}_2$ is the inverse square root of the prediction error covariance matrix obtained from an autocorrelated noise term and therefore, as in generalized least squares, it should be an adequate weighting matrix. The choice of this weighting matrix will change the finite sample results but will give the same asymptotic distribution (see, e.g. Bauer and Ljung, 2002) of any system invariant.

[ALTERNATIVE]

The first improvement consists of substituting $\mathbf{W}_1 = (\mathbf{Z}_f \mathbf{Z}'_f)^{-\frac{1}{2}}$, hereafter $\mathbf{\Omega}_1$, by $\mathbf{W}_1 = (\tilde{\mathbf{Z}}_f \tilde{\mathbf{Z}}'_f)^{-\frac{1}{2}}$, from now on $\mathbf{\Omega}_2$, where $\tilde{\mathbf{Z}}_f$ contains the residuals of regressing \mathbf{Z}_f onto \mathbf{Z}_p in a previous step. $\mathbf{\Omega}_1$ was chosen because of its asymptotic properties but not for its finite sample behavior. We suggest $\mathbf{\Omega}_2$ for two reasons: 1) because it gives the same asymptotic distribution (see, e.g. Bauer and Ljung, 2002) of any system invariant as $\mathbf{\Omega}_1$, and 2) because it is the inverse square root of the prediction error covariance matrix obtained from an autocorrelated noise term. Therefore, as in feasible generalized least squares, it should show an adequate performance also in finite samples.

The second enhancement consists in using a refined penalty function, denoted by $H(T, i)$, which depends not only on the sample size, but also on the dimension of the row space of the output BHM. Building on this idea, we define the criterion *NIDC*, which stands for “*n* identification criterion”, as:

$$NIDC(n) = \hat{\sigma}_{n+1}^2 + H(T, i)d(n) \tag{13}$$

where, again, $d(n) = 2nm$ denotes the number of parameters. To specify $H(T, i)$ with good performance, both when $T \rightarrow \infty$ and also in finite samples, we build on the idea of Bengtsson and Cavanaugh (2006), who optimize the performance of AIC-type criteria through Monte Carlo simulations.

In the following simulations, we will consider the data generating process (DGP) $z_t = a_t$, with $a_t \sim iidN(0, 1)$ and, accordingly, we will optimize the penalty

criterion to check whether the DGP is white noise. These decisions are motivated by the fact that the resulting criterion will be applied to a sequence of models, with increasing orders, until finding the minimum n which corresponding model filters the data autocorrelation to white noise residuals. On the other hand, choosing a white noise DGP has a secondary advantage, as the properties needed by $NIDC$ to assure a consistent order estimation are straightforward. In this case, $NIDC$ will determine $\hat{n} = 0$ instead of $\hat{n} > 0$, if $NIDC(0) < NIDC(j)$, $j = 1, 2, \dots, i - 1$. Substituting, we should have $\hat{\sigma}_1^2 < \hat{\sigma}_{j+1}^2 + 2jH(T, i)$, since, in this particular case, $d(n) = 2n$. Rearranging terms, we define ν_t as:

$$\nu_t = \max \left\{ \frac{\hat{\sigma}_1^2 - \hat{\sigma}_{j+1}^2}{2j} \right\}$$

Thus the optimal penalty term must be larger than ν_t assuring a correct performance of the criterion for this specific process.

[INSERT FIGURE 1]

Figure 1 shows that $H(T, i) = e^{-2T^{-.9}i^{3/2}}$ is an estimate of the lower bound below which one can find (by simulations) about the 95% of the realizations of ν_t . As we can see, $C(T) = \log T/T$ is systematically above $H(T, i)$ in short samples and the distance between them decreases when $T \rightarrow \infty$. This partially explains the different performance of both criteria in Section 5. Moreover, the following Proposition is obtained.

Proposition 4.1 *The penalty function $H(T, i) = e^{-2T^{-.9}i^{3/2}}$, that fits ν_t , assures the almost sure consistency of the system order estimated by minimizing $NIDC(n)$.*

The proof is given in Appendix A.

By construction, see Figure 1, *NIDC* tends to overestimate n in finite samples when compared to *SVC*, although both criteria lead to consistent estimation of the system order. However, this is not a drawback since, as we will see in the simulations, *SVC* shows a significant underestimate bias in short samples. Moreover, provided that the sample size is large enough, we think that overestimating is better than underestimating n , as further steps in model-building may lead, through the pruning of insignificant parameters, to the correct dimension. In comparison, the standard diagnostics provided by estimation will not reveal the right dimension if the initial value is underestimated.

In any case, the performance of the different methods depends on the DGP and the sample size. This fact explains why there are so many procedures in the literature: none of them dominates systematically the rest and, as a consequence, the ability of each method to choose the right system order depends critically on the dynamics of the data. This lack of robustness motivates the idea of combining several methods, to avoid extreme (sometimes good, sometimes bad) performances. To this end, we suggest the following procedure:

- 1) Compute all (or a selection of) the criteria discussed in Sections 3 and 4: i) SVC_{Ω_2} , which is the Bauer's original *SVC* but computed with our weighting matrix $\tilde{\mathbf{Z}}_f \tilde{\mathbf{Z}}_f'$, ii) the proposed *NICD*, iii) *AIC*, *SBC* and *HQ*, and iv) the χ^2 test due to Tsay (1989). Despite the large number of criteria involved the computational cost is acceptable because, for each n , we only need to run a single least-squares subspace regression.

- 2) Set \hat{n} as the value chosen by most methods, i.e., the sample mode. The mode is not necessarily unique, as different values of \hat{n} can be chosen by the same number of methods. In this case, given our preference for overparametrization, we suggest picking the larger \hat{n} .

As we will see in Section 5, this Mode-based Criterion (*MbC*) is robust and diversifies the risk of error when choosing the system order.

4.2 Estimating the order in seasonal systems

An important limitation of existing methods is that they cannot directly cope with multiplicative seasonal processes. To see this, consider the very common $\text{MA}(1) \times \text{MA}(1)_s$ process, when the seasonal frequency of the data is $s = 12$. In the best case, a standard order-determination method would choose $\hat{n} = 13$. In comparison, a seasonality-tolerant method, such as that of Box and Jenkins (1976), would choose two different orders: one for the regular subsystem and another one for the seasonal subsystem. We will denote these orders by n_r and n_s respectively. Obviously, none of the previously revised methods could directly provide this double choice.

To solve this issue, we propose decomposing the process (1a-1b) into a seasonal subsystem:

$$\mathbf{x}_{t+s}^s = \mathbf{\Phi}_s \mathbf{x}_t^s + \mathbf{E}_s \mathbf{r}_t \quad (14a)$$

$$\mathbf{z}_t = \mathbf{H}_s \mathbf{x}_t^s + \mathbf{r}_t \quad (14b)$$

and a regular subsystem:

$$\mathbf{x}_{t+1}^r = \Phi_r \mathbf{x}_t^r + E_r \psi_t \quad (15a)$$

$$\mathbf{r}_t = H_r \mathbf{x}_t^r + \psi_t \quad (15b)$$

and applying the methods previously described to determine the order of each subsystem. These subsystems are defined in different frequencies, as \mathbf{x}_t^s propagates in increments of s periods in (14a), while \mathbf{x}_t^r propagates in increments of one period in (15a); \mathbf{r}_t is an unobserved input of the seasonal subsystem (14a-14b) representing the regular structure of the original process. To estimate sequentially both subsystems, we assume that \mathbf{r}_t is a white noise term in the s frequency, uncorrelated with $\mathbf{x}_t^s \forall t$. To support this, Appendix B shows that, under some assumptions, the expectation of $\mathbf{r}_{t+j} \mathbf{r}_t'$ for all $j = s, 2s, \dots$ converges to zero. This approximation, which is exact when the regular system follows an MA(q) with $q < s$, allows us to estimate both subsystems using subspace methods. However, in (14a-14b) the states propagate in increments of s periods and, as a consequence, we cannot use the standard BHM defined in (5). This issue can be solved by defining the Seasonal Block Hankel Matrices (SBHM) of period s , as:

$$\mathbf{Z}_p^s = \begin{pmatrix} z_1 & \dots & z_{T-s(2i-1)} \\ z_{s+1} & \dots & z_{T-s(2i-2)} \\ \vdots & & \vdots \\ z_{s(i-1)+1} & \dots & z_{T-si} \end{pmatrix}; \quad \mathbf{Z}_f^s = \begin{pmatrix} z_{si+1} & \dots & z_{T-s(i-1)} \\ z_{s(i+1)+1} & \dots & z_{T-s(i-2)} \\ \vdots & & \vdots \\ z_{s(2i-1)+1} & \dots & z_T \end{pmatrix} \quad (16)$$

These matrices are analogous to the standard ones, although the time indices in each row are adapted to the seasonal period. Similarly, we denote both, past and

future-seasonal information blocks by \mathbf{Z}_p^s and \mathbf{Z}_f^s , respectively. Building on these matrices, one can estimate the seasonal parameters and, accordingly, the order of the seasonal subsystem, n_s , by applying any method to the seasonal regression model:

$$\mathbf{Z}_f^s = \mathbf{O}_i^s \mathbf{M}^s \mathbf{Z}_p^s + \mathbf{V}_i^s \mathbf{\Psi}_f^s \quad (17)$$

instead of the standard subspace regression (8). This generalization of conventional subspace methods allows us to apply the procedures discussed in Sections 3 and 4.1 to determine the seasonal and regular orders, by the following structured method:

- 1) Create the SBHM and compute the seasonal order \hat{n}_s by applying the procedures previously described to the seasonal subsystem (14a-14b).
- 2) Estimate the seasonal subsystem, conditional to the choice done in step 1), using e.g., the techniques shown in Section 2. Compute the residuals of this subsystem, \hat{r}_t .
- 3) Determine the order of the regular subsystem, \hat{n}_r , by applying standard methods to \hat{r}_t .

Note that if SVD methods (SVC_{Ω_1} , SVC_{Ω_2} or $NIDC$) are applied to determine the system orders, the seasonal subsystem (14a-14b) does not need to be estimated, as \hat{n}_s can be directly obtained from $\mathbf{W}_1 \mathbf{Z}_f^s \mathbf{W}_2$, where \mathbf{W}_1 and \mathbf{W}_2 are as in Section 2 but replacing \mathbf{Z}_p by \mathbf{Z}_p^s and \mathbf{Z}_f by \mathbf{Z}_f^s .

The order of steps 1) to 3) could be reversed, to determine first \hat{n}_r and, afterwards, \hat{n}_s . However this is not advisable as: i) in series with a short seasonal period, e.g. quarterly data with $s = 4$, it would be difficult to determine the cor-

rect regular order in the first step, and ii) the choice of \hat{n}_r is often more complex than that of \hat{n}_s , increasing the chance to contaminate the second decision with the effects of a previous specification error.

5 Monte Carlo results

This section analyzes the performance of several criteria by simulating univariate and multivariate models. Tables 1 to 6 show the percentage of replications where each criterion chose n and the contiguous dimensions, $n - 1$ and $n + 1$ in non-seasonal processes, while Tables 7 and 8 offer the relative frequency of hits in two seasonal models. In all cases, we discard the first 50 draws of each replication to improve randomization. Each table shows the results obtained using SVC_{Ω_2} , $NICD$, AIC , SBC , HQ , the χ^2 test and MbC . Additionally, SVC_{Ω_1} , as presented in Bauer (2001), is included.

The specific formulations employed in this exercise have been chosen to show how any individual criterion may perform both, very well and badly, depending on the dynamic structure of the DGP and the sample size.

On the other hand, the univariate DGPs assumed in the first four tables show a sequence of common nonseasonal ARMA models, with increasingly complex dynamic structures. Accordingly, the AR(1) DGP in Table 1 is followed by a strong MA(1) structure, and, finally, an AR(2) with complex roots in Table 3. These basic results are supplemented by the three vector processes assumed in tables 4, 5 and 6, being the last two of them taken from the literature. The last two

processes were chosen to illustrate the performance of the proposed decomposition for seasonal processes. The model in Table 7 adds a seasonal MA(1) term to the AR(2) with complex roots that was previously assumed and, finally, the model in Table 8 is a somewhat artificial combination of different seasonal factors, that we specified to show how our decomposition procedure is able to cope with multiple periods.

The first model considered is an autoregressive process with small persistence. In this case: i) *AIC* displays the best performance for almost every sample size, ii) SVC_{Ω_2} and *NIDC* widely dominate SVC_{Ω_1} and iii) assuming that a small overparametrization can lead to an adequate model in next steps, *NIDC* shows remarkable results in very small samples. The DGP M2 has also $n = 1$, but the moving average parameter is large enough to improve the outcomes of all the methods. *SBC* exhibits the best behavior, reaching the asymptotic value with a moderate sample size.

Table 3 displays the results obtained with an autoregressive model with complex roots. This time *NIDC* beats the rest of the methods for almost all the samples analyzed. Surprisingly, *SBC* presents the worst results, with less than 10% of correct estimates even with a large sample.

Table 4 summarizes the results obtained with a bivariate VARMA(1,1) process, with $n = 2$. One of the series has an autoregressive term close to be cancelled out by a moving average root. The performance, which is good in all cases for large samples, is very heterogeneous in small samples. SVC_{Ω_1} tends to underestimate

n , while this is corrected in SVC_{Ω_2} and $NIDC$ that clearly dominate it when $T \leq 100$.

Tables 5 and 6 display the results corresponding to two trivariate VARMA(1,1) models, with $n = 3$. In both cases, we omit the smaller sample sizes due to the complexity of the systems. The first one, M5, is taken from Koreisha and Pukkila (1989). The results obtained are similar to those of previous cases, but the differences in performance are amplified by the increased complexity of the order selection problem. SVC_{Ω_1} shows a clear bias towards underestimation and does not find any $\hat{n} = n - 1$ even when $T \leq 100$. Its enhanced versions produce much better results, as $NIDC$ and SVC_{Ω_2} show a reasonably good behavior. Note, also, that MbC presents remarkably good results. The second process, M6, is taken from Reinsel et al. (1992). The results are better than those of Table 5, because the structure of M5 is closer to a lower-order representation. In this case the χ^2 test beats all the criteria when $T \leq 100$, and all the methods behave well when the sample is large.

Tables 7 and 8 present two seasonal models. Here we are interested in assessing their performance when combined with the method described in Section 4.2. In both models the sample sizes are increased to represent the typical size of a seasonal data sample. The orders are estimated by following the stages stated in Section 4.2, using MbC to obtain the residuals in step 2).

M7 adds to M3 a monthly MA(1) term. In this case all the methods compared behave well, although the χ^2 test has some difficulties when $T = 100$. In the

second step, the results are similar to those in Table 3. This suggests that the method employed to determine the seasonal order does not distort the estimation of the regular order. Model M8 has multiple seasonal factors at frequencies 3, 9 and 36. This could happen, for instance, with financial data observed once each 10 days. Table 8 reveals: i) the remarkable performance of *SBC* in these models, and ii) the outstanding ability of our decomposition to estimate the seasonal order in each frequency, in combination with all the criteria considered. Note that, when using preliminary data analysis methods, we have estimated only 12 models as we have tried $n_s = 0, 1, 2$ for each factor. On the other hand, if we search the dynamic of this data using the same methods over the currently BHM, we should estimate, at least, 50 models as 49 is the order of the original process.

Previous results support the consistency of all the methods compared and also provide some clues about their specific performance.

Criteria based on SVD. In small samples the performance of SVC_{Ω_1} is clearly worst than that of SVC_{Ω_2} and *NIDC*. This difference is even more remarkable in multivariate processes. Note also that *NIDC* is often the best method when the persistence is small.

Information Criteria. The information criteria considered display different performances. *AIC* occasionally overestimates n in univariate process but provides good enough results with multivariate DGPs (see Gonzalo and Pitarakis, 2002). *SBC* is one of the best options in both, univariate and multivariate processes, when the sample is large and/or the parameters persistent enough. Nevertheless,

it shows the worst performance with low-persistence DGPs. *HQ* displays an intermediate performance.

χ^2 test. The χ^2 test shows poor results in small samples, improving substantially as the sample size grows. In univariate process it is not the best option while in multivariate systems its performance improves remarkably.

Mode-based criterion. This procedure is very robust, as it is never the worst option, but sometimes is the best (see Table 5, $T \geq 300$). In our opinion, this should be the method-of-choice for automatic implementations useful, e.g., when the analyst has to work with a large set of series.

6 Conclusions

This paper discusses how can one estimate the order of a SS process using subspace methods. These methods are powerful, flexible and computationally efficient.

Our first contribution consists of proposing two new criteria that improve the small-sample performance of Bauer (2001) *SVC* method. Second, we propose an automatic model-building procedure based on the sample mode of several methods. A simulation analysis confirms that the mode-based procedure is more robust than any specific criterion. Last, we extend the use of any criterion to processes with seasonality. Monte Carlo results show that the resulting criteria display a remarkable performance, even in a complex multiple seasonality framework.

In comparison with the popular VAR approach, which is very close to ours in terms of simplicity and computational efficiency, the methodology illustrated here has two clear advantages, as it: i) supports seasonal time series, and ii) is intrinsically more parsimonious, as it looks for the right system order by increasing the dimension of the state vector one-by-one. To clarify this remark, consider e.g., a vector of m time series and the decision to increase the order of a VAR(p) model to VAR($p + 1$). This implies increasing the dimension of the state vector by adding m new states, while our method would try all the possible choices in the sequence of 1, 2, m additional states.

Finally, the algorithms described in this article are implemented in a MATLAB toolbox for time series modeling called E^4 . The source code of this toolbox is freely provided under the terms of the GNU General Public License and can be downloaded at www.ucm.es/info/icae/e4. This site also provides a complete user manual for the Toolbox and other reference materials.

This research was supported by the Ministerio de Educación y Ciencia, under grant SEJ2005-07388.

References

- Akaike, H. (1976). *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*. Academic Press.
- Aoki, M. (1990). *State Space Modelling of Time Series*. Springer Verlag, New York.
- Bauer, D. (2001). Order estimation for subspace methods. *Automatica*, 37:1561–1573.
- Bauer, D. (2005). Asymptotic properties of subspace estimators. *Automatica*, 41:359–376.
- Bauer, D. and Ljung, L. (2002). Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms. *Automatica*, 38:763–773.
- Bengtsson, T. and Cavanaugh, J. (2006). An improved Akaike information criterion for state-space model selection. *Computational Statistics and Data Analysis*, 50(10):2635–2654.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 2nd ed. edition.
- Bujosa, M., García-Ferrer, A., and Young, P. C. (2007). Linear dynamic harmonic regression. *Computational Statistics and Data Analysis*, 52(2):999–1024.
- Casals, J., Sotoca, S., and Jerez, M. (1999). A fast stable method to compute the likelihood of time invariant state space models. *Economics Letters*, 65(3):329–337.
- Chui, C. K. and Chen, G. (1999). *Kalman Filtering with Real-Time Applications*. Berlin, Springer.
- Cooper, D. M. and Wood, E. F. (1982). Identifying multivariate time series models. *Journal of Time Series Analysis*, 3:277–293.
- Eckart, C. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218.

- Favoreel, W., De Moor, B., and Van Overschee, P. (2000). Subspace state space system identification for industrial processes. *Journal of Process Control*, 10:149–155.
- Gonzalo, J. and Pitarakis, J. Y. (2002). Lag length estimation in large dimensional systems. *Journal or Time Series Analysis*, 23:401–423.
- Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Association, B Series*, 41:713–723.
- Katayama, T. (2005). *Subspace Methods for System Identification*. Springer Verlag, London.
- Koreisha, S. G. and Pukkila, T. H. (1989). Fast linear estimation methods for vector autoregressive moving average models. *Journal of Time Series Analysis*, 10:325–339.
- Larimore, W. E. (1983). System identification, reduced-order filtering and modeling via canonical variate analysis. *Proceedings of the American Control Conference, ACC, San Francisco*, pages 445–451.
- Larimore, W. E. (1990). Canonical variate analysis in identification, filtering and adaptive control. *Proceedings of the 29th Conference on Decision and Control, Hawaii*, pages 596–604.
- Reinsel, G. C., Basu, S., and Yap, S. F. (1992). Maximum likelihood estimators in the multivariate autoregressive moving average model from a generalized least squares viewpoint. *Journal of Time Series Analysis*, 13(2):133–145.
- Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:1–48.
- Sorelius, J. (1999). *Subspace-Based Parameter Estimation Problems in Signal Processing*. PhD thesis, Uppsala University.
- Stoica, P., Selen, Y., and Jian Li (2004). On information criteria and the generalized likelihood ratio test of model order selection. *Signal Processing Letters, IEEE*, 11(10):794–797.
- Stoica, P. and Söderström, T. (1982). On nonsingular information matrices and local identifiability. *International Journal Control*, 36:323–329.
- Tiao, G. C. and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society, B Series*, 51(2):157–213.

- Toscano, E. M. M. and Reisen, V. A. (2000). The use of canonical correlation analysis to identify the order of multivariate ARMA models: Simulation and application. *Journal of Forecasting*, 19:441–455.
- Tsay, R. S. (1989). Identifying multivariate time series models. *Journal of Time Series Analysis*, 10(4):357–372.
- Woodside, C. M. (1971). Estimation of the order of linear systems. *Automatica*, 7:727–733.

Appendix A

Proof. The conditions on the penalty function that are sufficient for the almost sure consistency of the order estimated by minimizing (13) are: 1) $H(T, i)T > 0$, 2) $H(T, i) \rightarrow 0$ and, 3) $H(T, i)T/(i^2 \log \log T) \rightarrow \infty$, when $T \rightarrow \infty$ (see Bauer, 2001, Theorem 3).

Although in practice we implement $i = \max(4, h_t)$, being h_t the integer closer to $\log T$, in sections 2) and 3) of this proof we consider, that $i = \log T$. Let $H(T, i) = aT^{-b}i^c$ with $(a, b, c) = (e^{-2}, .9, 3/2)$, under these conditions:

- 1) It is straightforward to see that,

$$H(T, i)T = aT^{-b+1}i^c > 0, \quad \forall T > 0$$

- 2) The limit,

$$\lim_{T \rightarrow \infty} aT^{-b}(\log T)^c = \frac{a(\log T)^c}{T^b}$$

converges to the indeterminate $\frac{\infty}{\infty}$. Applying twice L'Hôpital rule we get that:

$$\lim_{T \rightarrow \infty} H(T, i) = 0$$

- 3) Finally,

$$\frac{H(T, i)T}{i^2 \log \log T} = \frac{aT^{-b+1}(\log T)^{c-2}}{\log \log T}$$

Again applying L'Hôpital we can obtain:

$$\lim_{T \rightarrow \infty} \frac{aT^{-b+1}(\log T)^{c-2}}{\log \log T} = \lim_{T \rightarrow \infty} aT^{1-b}[(1-b)(\log T)^{c-1} + (c-2)(\log T)^{c-2}] = \infty$$

□

Appendix B

By recursive substitution in (15a) and shifting time subscripts, the regular subsystem (15a-15b) can be expressed as:

$$\mathbf{x}_{t+s}^r = \Phi_r^{s-1} \mathbf{x}_{t+1}^r + \sum_{j=0}^{s-1} \Phi_r^j \mathbf{E}_r \psi_{t+s-j-1} \quad (\text{A.1a})$$

$$\mathbf{r}_{t+s} = \mathbf{H}_r \mathbf{x}_{t+s}^r + \psi_{t+s} \quad (\text{A.1b})$$

From these two equations, it is easy to see that:

$$\mathbb{E}[\mathbf{r}_{t+s} \mathbf{r}'_t] = \mathbb{E}[(\mathbf{H}_r \mathbf{x}_{t+s}^r + \psi_{t+s}) \mathbf{r}'_t] = \mathbf{H}_r \Phi_r^{s-1} \mathbb{E}[\mathbf{x}_{t+1}^r \mathbf{r}'_t] \quad (\text{A.2})$$

where $\mathbb{E}[\cdot]$ is the mathematical expectation. Since \mathbf{z}_t is stationary, and as a consequence all the eigenvalues of Φ_r are inside the unit circle, then, provided that s is large enough, we will assume that:

$$\mathbb{E}[\mathbf{r}_{t+j} \mathbf{r}'_t] \simeq \mathbf{0}, \quad \forall j = s, 2s, \dots \quad (\text{A.3})$$

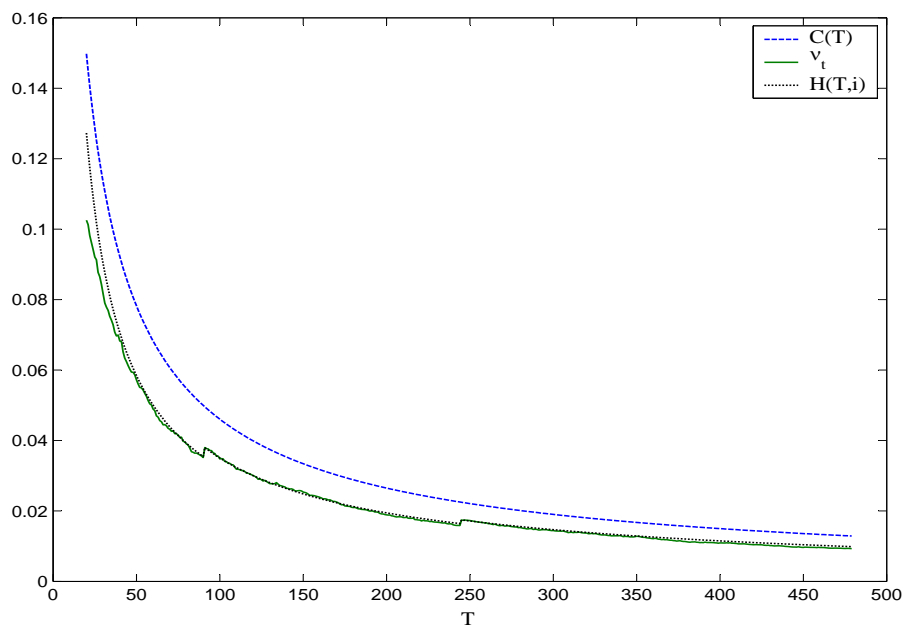


Figure 1: Estimated penalty function of NIDC. ν_t is computed using the singular values obtained as the 95 percentile of 2000 simulations of the univariate model $z_t = a_t$. $H(T,i)$ is the lost function fitted to ν_t , proposed in NIDC. $C(T)$ is the penalty term suggested in Bauer (2001).

Table 1: System order estimates, M1 model*.

M1($n = 1$): $(1 - 0.2B)z_t = a_t, \quad a_t \sim iidN(0, 1)$.

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n} = n - 1$								
30	0.913	0.801	0.389	0.680	0.866	0.682	0.878	0.719
50	0.861	0.728	0.436	0.551	0.818	0.672	0.841	0.660
100	0.782	0.676	0.489	0.314	0.697	0.521	0.729	0.537
300	0.486	0.415	0.380	0.037	0.273	0.130	0.246	0.219
500	0.243	0.204	0.171	0.007	0.081	0.027	0.051	0.073
$\hat{n} = n$								
30	0.086	0.185	<u>0.389</u>	0.315	0.134	0.305	0.098	0.270
50	0.133	0.244	0.392	<u>0.435</u>	0.182	0.311	0.127	0.323
100	0.204	0.290	0.402	<u>0.649</u>	0.301	0.470	0.238	0.448
300	0.490	0.548	0.574	<u>0.916</u>	0.724	0.849	0.713	0.775
500	0.724	0.757	0.780	<u>0.967</u>	0.915	0.961	0.906	0.919
$\hat{n} = n + 1$								
30	0.001	0.014	0.211	0.005	0.0	0.013	0.019	0.016
50	0.006	0.027	0.111	0.013	0.0	0.017	0.026	0.016
100	0.013	0.033	0.086	0.036	0.002	0.009	0.023	0.015
300	0.020	0.036	0.039	0.043	0.003	0.021	0.027	0.013
500	0.032	0.039	0.044	0.024	0.004	0.012	0.029	0.021

* The table shows the percentage cases where each criteria chose n , and the contiguous choices $n - 1$ and $n + 1$. The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{Z}_f \tilde{Z}'_f)^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.

Table 2: System order estimates, M2 model*.

M2($n = 1$): $z_t = (1 - 0.8B)a_t$, $a_t \sim iidN(0, 1)$.

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n} = n - 1$								
30	0.179	0.032	0.019	0.021	0.057	0.025	0.284	0.019
50	0.019	0.001	0.0	0.0	0.006	0.0	0.018	0.0
100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
300	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\hat{n} = n$								
30	0.793	0.819	0.771	0.932	<u>0.934</u>	0.929	0.676	0.919
50	0.929	0.850	0.774	0.932	<u>0.988</u>	0.978	0.946	0.962
100	0.941	0.907	0.810	0.935	<u>0.998</u>	0.981	0.965	0.967
300	0.949	0.934	0.922	0.915	<u>1.0</u>	0.984	0.971	0.971
500	0.954	0.944	0.935	0.909	<u>1.0</u>	0.981	0.966	0.980
$\hat{n} = n + 1$								
30	0.027	0.124	0.162	0.045	0.009	0.041	0.03	0.055
50	0.048	0.120	0.174	0.067	0.006	0.022	0.017	0.028
100	0.052	0.076	0.128	0.063	0.002	0.019	0.012	0.033
300	0.044	0.053	0.065	0.082	0.0	0.016	0.013	0.019
500	0.040	0.047	0.055	0.090	0.0	0.019	0.01	0.014

* The table shows the percentage cases where each criteria chose n , and the contiguous choices $n - 1$ and $n + 1$. The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{Z}_f \tilde{Z}'_f)^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.

Table 3: System order estimates, M3 model*.

M3($n = 2$): $(1 - 0.4B + 0.3B^2)z_t = a_t$, $a_t \sim iidN(0, 1)$.

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n} = n - 1$								
30	0.247	0.419	0.417	0.652	0.441	0.628	0.257	0.563
50	0.439	0.555	0.482	0.800	0.673	0.803	0.520	0.664
100	0.653	0.639	0.525	0.814	0.939	0.909	0.767	0.756
300	0.487	0.445	0.398	0.460	0.958	0.796	0.464	0.527
500	0.290	0.263	0.235	0.229	0.908	0.605	0.204	0.250
$\hat{n} = n$								
30	0.026	0.147	<u>0.329</u>	0.085	0.022	0.062	0.082	0.110
50	0.079	0.179	<u>0.374</u>	0.126	0.018	0.049	0.111	0.167
100	0.199	0.282	<u>0.406</u>	0.182	0.019	0.081	0.157	0.216
300	0.500	0.532	<u>0.561</u>	0.539	0.042	0.204	0.481	0.463
500	0.689	0.710	<u>0.736</u>	<u>0.771</u>	0.092	0.395	0.738	0.740
$\hat{n} = n + 1$								
30	0.0	0.006	0.039	0.002	0.0	0.0	0.012	0.009
50	0.001	0.005	0.050	0.001	0.0	0.0	0.021	0.006
100	0.004	0.015	0.042	0.003	0.0	0.0	0.023	0.004
300	0.013	0.023	0.035	0.001	0.0	0.0	0.055	0.008
500	0.021	0.027	0.040	0.0	0.0	0.0	0.058	0.014

* The table shows the percentage cases where each criteria chose n , and the contiguous choices $n - 1$ and $n + 1$. The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{Z}_f \tilde{Z}'_f)^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.

Table 4: System order estimates, $M4$ model*.

$$M4(n = 2): (\mathbf{I} + \Phi B)\mathbf{Z}_t = (\mathbf{I} + \Theta B)\mathbf{a}_t, \quad \mathbf{a}_t \sim iidN(\mathbf{0}, \Sigma_a)$$

$$\Phi = \begin{pmatrix} -0.4 & 0 \\ 0 & -0.8 \end{pmatrix}; \Theta = \begin{pmatrix} -0.7 & 0.5 \\ -0.3 & -0.7 \end{pmatrix}; \Sigma_a = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n} = n - 1$								
30	0.002	0.355	0.285	0.519	0.536	0.581	0.540	0.614
50	0.068	0.493	0.361	0.517	0.702	0.632	0.699	0.665
100	0.432	0.323	0.201	0.302	0.701	0.501	0.554	0.410
300	0.007	0.002	0.001	0.005	0.249	0.078	0.054	0.011
500	0.0	0.0	0.0	0.0	0.035	0.0	0.005	0.001
$\hat{n} = n$								
30	0.0	0.311	<u>0.315</u>	0.270	0.080	0.203	0.097	0.200
50	0.001	0.354	<u>0.425</u>	0.379	0.137	0.275	0.171	0.275
100	0.199	0.611	<u>0.652</u>	0.617	0.290	0.470	0.414	0.555
300	<u>0.975</u>	0.957	0.938	0.885	0.751	0.909	0.825	0.969
500	0.982	0.972	0.954	0.896	0.965	<u>0.992</u>	0.851	0.982
$\hat{n} = n + 1$								
30	0.0	0.183	0.227	0.081	0.008	0.031	0.007	0.032
50	0.0	0.072	0.145	0.084	0.003	0.026	0.011	0.022
100	0.001	0.047	0.115	0.071	0.001	0.027	0.024	0.036
300	0.018	0.040	0.054	0.106	0.0	0.014	0.119	0.018
500	0.018	0.028	0.040	0.098	0.0	0.002	0.148	0.015

* The table shows the percentage cases where each criteria chose n , and the contiguous choices $n - 1$ and $n + 1$. The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{\mathbf{Z}}_f \tilde{\mathbf{Z}}_f')^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.

Table 5: System order estimates, M5 model*.

$$M5(n = 3): (\mathbf{I} + \Phi B)\mathbf{Z}_t = (\mathbf{I} + \Theta B)\mathbf{a}_t, \quad \mathbf{a}_t \sim iidN(\mathbf{0}, \Sigma_a)$$

$$\Phi = \begin{pmatrix} -0.7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & -0.4 & 0 \end{pmatrix}; \Theta = \begin{pmatrix} 0 & 1.1 & 0 \\ 0 & -0.6 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}; \Sigma_a = \begin{pmatrix} 1 & -0.7 & 0.4 \\ -0.7 & 1 & 0 \\ 0.4 & 0 & 1 \end{pmatrix}$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n} = n - 1$								
50	0.0	0.232	0.272	0.362	0.644	0.604	0.344	0.536
100	0.0	0.485	0.481	0.215	0.678	0.399	0.298	0.382
300	0.164	0.026	0.029	0.0	0.043	0.004	0.003	0.008
500	0.001	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\hat{n} = n$								
50	0.0	0.353	<u>0.426</u>	0.345	0.149	0.276	0.330	0.333
100	0.0	0.448	0.444	0.638	0.317	0.590	<u>0.641</u>	0.592
300	0.836	0.965	0.966	0.954	0.957	0.962	0.943	<u>0.992</u>
500	0.997	0.995	0.993	0.968	<u>1.0</u>	0.999	0.945	<u>1.0</u>
$\hat{n} = n + 1$								
50	0.0	0.270	0.248	0.178	0.013	0.052	0.023	0.099
100	0.0	0.060	0.065	0.125	0.0	0.010	0.031	0.025
300	0.0	0.009	0.017	0.044	0.0	0.004	0.047	0.003
500	0.002	0.005	0.009	0.031	0.0	0.001	0.051	0.0

* The table shows the percentage cases where each criteria chose n , and the contiguous choices $n - 1$ and $n + 1$. The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{\mathbf{Z}}_f \tilde{\mathbf{Z}}_f')^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.

Table 6: System order estimates, M6 model*.

$$M6(n = 3): (\mathbf{I} + \Phi B)\mathbf{Z}_t = (\mathbf{I} + \Theta B)\mathbf{a}_t, \quad \mathbf{a}_t \sim iidN(\mathbf{0}, \Sigma_a)$$

$$\Phi = \begin{pmatrix} -0.4 & -0.3 & 0.6 \\ 0 & -0.8 & -0.4 \\ -0.3 & 0 & 0 \end{pmatrix}; \Theta = \begin{pmatrix} -0.7 & 0 & 0 \\ -0.1 & -0.2 & 0 \\ 0.4 & -0.5 & 0.1 \end{pmatrix}; \Sigma_a = \begin{pmatrix} 1 & 0.5 & 0.4 \\ 0.5 & 1 & 0.7 \\ 0.4 & 0.7 & 1 \end{pmatrix}$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n} = n - 1$								
50	0.0	0.199	0.231	0.257	0.658	0.440	0.284	0.345
100	0.002	0.314	0.317	0.055	0.295	0.110	0.057	0.115
300	0.025	0.002	0.003	0.0	0.0	0.0	0.0	0.0
500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$\hat{n} = n$								
50	0.0	0.380	0.425	0.428	0.264	0.490	<u>0.602</u>	0.545
100	0.0	0.607	0.603	0.789	0.700	0.869	<u>0.908</u>	0.861
300	0.971	0.978	0.971	0.979	<u>0.999</u>	0.998	0.966	0.991
500	0.998	0.987	0.982	0.987	<u>1.0</u>	0.999	0.961	0.997
$\hat{n} = n + 1$								
50	0.0	0.286	0.271	0.226	0.026	0.068	0.047	0.102
100	0.0	0.071	0.075	0.123	0.005	0.021	0.033	0.023
300	0.004	0.020	0.026	0.021	0.001	0.002	0.034	0.007
500	0.002	0.013	0.016	0.013	0.0	0.001	0.038	0.003

* The table shows the percentage cases where each criteria chose n , and the contiguous choices $n - 1$ and $n + 1$. The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{\mathbf{Z}}_f \tilde{\mathbf{Z}}_f')^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.

Table 7: System orders estimates, M7 model*.

M7($n_r = 2, n_s = 1$): $(1 - 0.4B + 0.3B^2)z_t = (1 - 0.6B^s)a_t, a_t \sim iidN(0, 1)$.

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n}_s = n_s, s = 4$								
30	0.453	0.634	0.736	<u>0.816</u>	0.569	0.747	0.470	0.690
50	0.771	0.842	0.789	<u>0.954</u>	0.887	0.943	0.664	0.905
100	0.925	0.897	0.815	0.952	<u>0.996</u>	0.985	0.888	0.960
300	0.916	0.896	0.879	0.943	<u>1.0</u>	0.991	0.884	0.960
$\hat{n}_r = n_r, s = 4$								
30	0.020	0.121	<u>0.283</u>	0.065	0.011	0.037	0.071	0.060
50	0.041	0.117	<u>0.311</u>	0.068	0.009	0.031	0.104	0.071
100	0.133	0.197	<u>0.329</u>	0.122	0.009	0.043	0.142	0.119
300	0.415	0.457	<u>0.501</u>	0.443	0.021	0.152	0.467	0.400
$\hat{n}_s = n_s, s = 12$								
100	0.969	0.972	0.950	0.984	0.992	<u>0.996</u>	0.460	0.988
300	0.944	0.935	0.941	0.971	<u>1.0</u>	0.996	0.878	0.973
500	0.946	0.938	0.934	0.942	<u>1.0</u>	0.998	0.894	0.976
$\hat{n}_r = n_r, s = 12$								
100	0.181	0.264	<u>0.399</u>	0.162	0.023	0.071	0.205	0.186
300	0.495	0.529	<u>0.558</u>	0.487	0.040	0.192	0.536	0.476
500	0.712	0.733	0.766	0.759	0.089	0.396	<u>0.802</u>	0.754

* The table shows the percentage cases where each criteria chose n_s and n_r . The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{Z}_f \tilde{Z}'_f)^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.

Table 8: System orders estimates, M8 model*.

$$\text{M8}(n_{36} = 1, n_9 = 1, n_3 = 1, n_1 = 1):$$

$$z_t = (1 - 0.6B)(1 - 0.5B^3)(1 - 0.4B^9)(1 - 0.3B^{36})a_t, \quad a_t \sim iidN(0, 1).$$

T	SVC_{Ω_1}	SVC_{Ω_2}	$NIDC$	AIC	SBC	HQ	χ^2	MbC
$\hat{n}_{36} = n_{36}$								
400	0.890	0.887	0.881	0.947	0.981	<u>0.984</u>	0.573	0.940
600	0.902	0.888	0.868	0.955	<u>0.997</u>	0.992	0.764	0.952
1000	0.846	0.832	0.870	0.949	<u>0.998</u>	0.991	0.760	0.931
$\hat{n}_9 = n_9$								
400	0.937	0.917	0.898	0.967	<u>0.999</u>	0.994	0.922	0.980
600	0.931	0.920	0.896	0.972	<u>0.999</u>	0.995	0.989	0.996
1000	0.929	0.918	0.944	0.980	<u>1.0</u>	0.997	0.989	0.997
$\hat{n}_3 = n_3$								
400	0.958	0.950	0.937	0.983	<u>1.0</u>	0.998	0.989	0.996
600	0.950	0.946	0.935	0.987	<u>1.0</u>	0.998	0.988	0.996
1000	0.932	0.928	0.946	0.985	<u>1.0</u>	<u>1.0</u>	0.988	0.992
$\hat{n}_1 = n_1$								
400	0.974	0.969	0.964	0.994	<u>1.0</u>	0.999	0.995	0.997
600	0.981	0.979	0.975	0.996	<u>1.0</u>	<u>1.0</u>	0.993	0.999
1000	0.963	0.959	0.974	0.986	<u>1.0</u>	<u>1.0</u>	0.989	0.998

* The table shows the percentage cases where each criteria chose n_s , $s = 3, 6, 9$ and n_r . The number of replications is 2000. The methods analyzed are the following: SVC_{Ω_1} from Bauer (2001), SVC_{Ω_2} that has the same structure but computed with the proposed weighting matrix $\Omega_2 = (\tilde{Z}_f \tilde{Z}'_f)^{-\frac{1}{2}}$, $NIDC$ which is as SVC_{Ω_2} but replacing the penalty term $C(T) = \log T/T$ by $H(T, i) = e^{-2T^{-.9}i^{3/2}}$. AIC , SBC and HQ are the information criteria due to Akaike (1976), Schwartz (1978) and Hannan and Quinn (1979) respectively. χ^2 is the test proposed by Tsay (1989) with conventional significance level 5% and MbC is the \hat{n} value most often for each realization in the criteria collection, except SVC_{Ω_1} . For each sample size, the largest relative frequency of hits is underlined.