Modelling and forecasting large sets of time series by subspace methods

A. Garcia-Hiernaux[1] M. Jerez[2] J. Casals[2]

[1]Department of Statistics
Universidad Carlos III de Madrid

[2]Department of Quantitative Economics
Universidad Complutense de Madrid

27th International Symposium on Forecasting, 2007

Outline

How to determine the system order?

Definition

Order of linear systems (n): highest degree of the linear differential equation that describes the system

The literature can be broken into:

1. Preliminary data analysis
   - [Tsay, 1989]
   - [Wei, 1990],Garcia-Ferrer (2007)
   - [Bauer, 2001]
2. Model comparison procedures
   - AIC [Akaike, 1976], SBC [Schwartz, 1978], HQ [Hannan and Quinn, 1979]

How to determine the system order?

Definition

Order of linear systems (n): highest degree of the linear differential equation that describes the system

The literature can be broken into:

1. Preliminary data analysis
   - [Tsay, 1989]
   - [Wei, 1990],Garcia-Ferrer (2007)
   - [Bauer, 2001]
2. Model comparison procedures
   - AIC [Akaike, 1976], SBC [Schwartz, 1978], HQ [Hannan and Quinn, 1979]

How to determine the system order?

Definition

Order of linear systems (n): highest degree of the linear differential equation that describes the system

The literature can be broken into:

1. Preliminary data analysis
   - [Tsay, 1989]
   - [Wei, 1990],Garcia-Ferrer (2007)
   - [Bauer, 2001]
2. Model comparison procedures
   - AIC [Akaike, 1976], SBC [Schwartz, 1978], HQ [Hannan and Quinn, 1979]

How to determine the system order?

Definition

Order of linear systems (n): highest degree of the linear differential equation that describes the system

The literature can be broken into:

1. Preliminary data analysis
   - [Tsay, 1989]
   - [Wei, 1990],Garcia-Ferrer (2007)
   - [Bauer, 2001]
2. Model comparison procedures
   - AIC [Akaike, 1976], SBC [Schwartz, 1978], HQ [Hannan and Quinn, 1979]

Basic Shortcoming and Proposals

Shortcomings of the previous methods:

- Some show a poor performance in short samples
- Many are not robust what produces a high risk in an automatic use

Our contribution is threefold:

- 2 refinements of [Bauer, 2001] criterion
- A robust automatic procedure
- Assessment of the procedures in terms of fit and prediction

Basic Shortcoming and Proposals

Shortcomings of the previous methods:

- Some show a poor performance in short samples
- Many are not robust what produces a high risk in an automatic use

Our contribution is threefold:

- 2 refinements of [Bauer, 2001] criterion
- A robust automatic procedure
- Assessment of the procedures in terms of fit and prediction

Brief Introduction to the Subspace Methods

Consider a linear-fixed, stable, strictly minimum-phase system in innovations form:

$$\left.\begin{array}{lcl} \boldsymbol{x}_{t+1} & = & \Phi \boldsymbol{x}_t + \boldsymbol{E}\psi_t \\ \boldsymbol{z}_t & = & \boldsymbol{H}\boldsymbol{x}_t + \psi_t \end{array}\right\} Z_f = O_i \overbrace{M Z_p}^{X_f} + V_i \Psi_f$$

$$Z_p = \begin{pmatrix} z_1 & z_2 & \cdots & z_{T-2i+1} \\ z_2 & z_3 & \cdots & z_{T-2i+2} \\ \vdots & \vdots & & \vdots \\ z_i & z_{i+1} & \cdots & z_{T-i} \end{pmatrix}; \quad Z_f = \begin{pmatrix} z_{i+1} & z_{i+2} & \cdots & z_{T-i+1} \\ z_{i+2} & z_{i+3} & \cdots & z_{T-i+2} \\ \vdots & \vdots & & \vdots \\ z_{2i} & z_{2i+1} & \cdots & z_T \end{pmatrix}$$

Ψ_f is as Z_f but with ψ_t instead of z_t.

Brief Introduction to the Subspace Methods

Consider a linear-fixed, stable, strictly minimum-phase system in innovations form:

$$\left.\begin{array}{lcl} x_{t+1} & = & \Phi x_t + E\psi_t \\ z_t & = & Hx_t + \psi_t \end{array}\right\} Z_f = O_i \overbrace{M Z_p}^{X_f} + V_i \Psi_f$$

$$Z_p = \begin{pmatrix} z_1 & z_2 & \cdots & z_{T-2i+1} \\ z_2 & z_3 & \cdots & z_{T-2i+2} \\ \vdots & \vdots & & \vdots \\ z_i & z_{i+1} & \cdots & z_{T-i} \end{pmatrix}; \quad Z_f = \begin{pmatrix} z_{i+1} & z_{i+2} & \cdots & z_{T-i+1} \\ z_{i+2} & z_{i+3} & \cdots & z_{T-i+2} \\ \vdots & \vdots & & \vdots \\ z_{2i} & z_{2i+1} & \cdots & z_T \end{pmatrix}$$

Ψ_f is as Z_f but with ψ_t instead of z_t.

Brief Introduction to the Subspace Methods

Consider a linear-fixed, stable, strictly minimum-phase system in innovations form:

$$\left.\begin{array}{lcl} x_{t+1} & = & \Phi x_t + E\psi_t \\ z_t & = & Hx_t + \psi_t \end{array}\right\} Z_f = O_i \overbrace{M Z_p}^{X_f} + V_i \Psi_f$$

$$Z_p = \begin{pmatrix} z_1 & z_2 & \cdots & z_{T-2i+1} \\ z_2 & z_3 & \cdots & z_{T-2i+2} \\ \vdots & \vdots & & \vdots \\ z_i & z_{i+1} & \cdots & z_{T-i} \end{pmatrix}; \quad Z_f = \begin{pmatrix} z_{i+1} & z_{i+2} & \cdots & z_{T-i+1} \\ z_{i+2} & z_{i+3} & \cdots & z_{T-i+2} \\ \vdots & \vdots & & \vdots \\ z_{2i} & z_{2i+1} & \cdots & z_T \end{pmatrix}$$

Ψ_f is as Z_f but with ψ_t instead of z_t.

Brief Introduction to the Subspace Methods

We consider a linear-fixed, stable, strictly minimum-phase system in innovations form:

$$\left.\begin{aligned} x_{t+1} &= \Phi x_t + E\psi_t \\ z_t &= H x_t + \psi_t \end{aligned}\right\} Z_f = O_i \overbrace{M Z_p}^{X_f} + V_i \Psi_f$$

$$O_i = \begin{pmatrix} H' & (H\Phi)' & (H\Phi^2)' & \dots & (H\Phi^{i-1})' \end{pmatrix}'$$

$$V_i = \begin{pmatrix} I_m & 0 & 0 & \dots & 0 \\ HE & I_m & 0 & \dots & 0 \\ H\Phi E & HE & I_m & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H\Phi^{i-2}E & H\Phi^{i-3}E & H\Phi^{i-4}E & \dots & I_m \end{pmatrix}$$

Brief Introduction to the Subspace Methods

We consider a linear-fixed, stable, strictly minimum-phase system in innovations form:

$$\left.\begin{array}{rcl} x_{t+1} & = & \Phi x_t + E\psi_t \\ z_t & = & Hx_t + \psi_t \end{array}\right\} Z_f = O_i \overbrace{M Z_p}^{X_f} + V_i \Psi_f$$

$$O_i = \begin{pmatrix} H' & (H\Phi)' & (H\Phi^2)' & \dots & (H\Phi^{i-1})' \end{pmatrix}'$$

$$V_i = \begin{pmatrix} I_m & 0 & 0 & \dots & 0 \\ HE & I_m & 0 & \dots & 0 \\ H\Phi E & HE & I_m & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H\Phi^{i-2}E & H\Phi^{i-3}E & H\Phi^{i-4}E & \dots & I_m \end{pmatrix}$$

Brief Introduction to the Subspace Methods

- Subspace methods face the reduced-rank weighted least square problem:

$$Z_f = O_i M Z_p + V_i \Psi_f \tag{1}$$

efficiently solved applying the SVD to $W_1 Z_f W_2$, being

- $$W_1 = (Z_f Z_f')^{-\frac{1}{2}} \tag{2}$$

- $$W_2 = Z_p'(Z_p Z_p')^{-1} Z_p \tag{3}$$

Determining the system order from the SVs

[Bauer, 2001]:

- $rank(\boldsymbol{W_1 Z_f W_2}) =$ number of SVs statistically nonzero $= n$
- $SVC(n) = \hat{\sigma}^2_{n+1} + C(T)d(n)$
- SVC is consistent with $C(T) = \log T/T$

However:

- SVC exhibits an underestimation bias in small samples,
- $\boldsymbol{W_1}$ has an important impact on the performance of SVC,
- "the penalty term is somewhat arbitrary"

Determining the system order from the SVs

[Bauer, 2001]:

- $rank(\boldsymbol{W_1 Z_f W_2})$ = number of SVs statistically nonzero $= n$
- $SVC(n) = \hat{\sigma}^2_{n+1} + C(T)d(n)$
- SVC is consistent with $C(T) = \log T/T$

However:

- SVC exhibits an underestimation bias in small samples,
- $\boldsymbol{W_1}$ has an important impact on the performance of SVC,
- "the penalty term is somewhat arbitrary"

Refining the SVC

How to improve the small sample performance of SVC?

- Substituting $\boldsymbol{W_1} = (\boldsymbol{Z_f Z_f'})^{-\frac{1}{2}}$ by $\tilde{\boldsymbol{W}}_1 = (\tilde{\boldsymbol{Z}}_f \tilde{\boldsymbol{Z}}_f')^{-\frac{1}{2}}$, being $\tilde{\boldsymbol{Z}}_f$ the residuals of regressing $\boldsymbol{Z_f}$ onto $\boldsymbol{Z_p}$. [$SVC_2(n)$]
- Optimizing the penalty function through simulation:

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

Proposition

The penalty function $H(T,i) = e^{-2}T^{-.9}i^{1.6}$ assures the almost sure consistency of the system order estimated by minimizing NIDC(n).

Refining the SVC

How to improve the small sample performance of SVC?

- Substituting $\boldsymbol{W_1} = (\boldsymbol{Z_f}\boldsymbol{Z_f'})^{-\frac{1}{2}}$ by $\tilde{\boldsymbol{W}}_1 = (\tilde{\boldsymbol{Z}}_f\tilde{\boldsymbol{Z}}_f')^{-\frac{1}{2}}$, being $\tilde{\boldsymbol{Z}}_f$ the residuals of regressing $\boldsymbol{Z_f}$ onto $\boldsymbol{Z_p}$. [$SVC_2(n)$]
- Optimizing the penalty function through simulation:

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

Proposition

The penalty function $H(T,i) = e^{-2}T^{-.9}i^{1.6}$ assures the almost sure consistency of the system order estimated by minimizing NIDC(n).

Looking for the robustness

On the other hand,

- Many works show that the performance of the different methods depends on the DGP and T.
- To avoid extreme performance (sometimes very good/bad) we devise a criterion (*MbC*) which combines several methods:
  1. Compute a selection of criteria (SVC_2, $NIDC$, AIC, SBC, HQ and χ^2 due to [Tsay, 1989]).
  2. Select $\hat{n}$ as the value chosen by most methods, i.e., the sample mode.

System order estimates: Univariate case

Relative frequency of hits. 2000 replications.

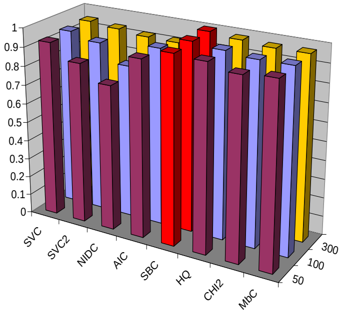


Figure: MA(1): $z_t = (1 - .8B)a_t$
$a_t \sim iidN(0,1) \rightarrow n = 1$

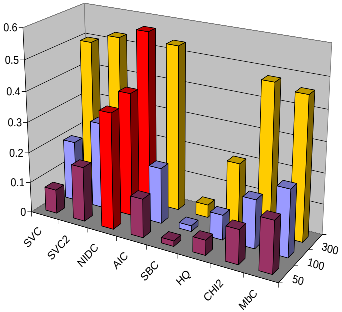
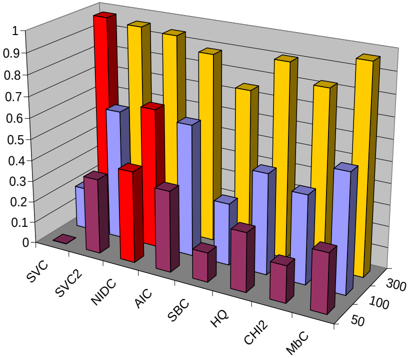


Figure: AR(2) $z_t = (1 - .4B + .3B^2)a_t$
$a_t \sim iidN(0,1) \rightarrow n = 2$

System order estimates: multivariate case I



$$(I + \Phi B)Z_t = (I + \Theta B)a_t$$

$$\Phi = \begin{pmatrix} -0.4 & 0 \\ 0 & -0.8 \end{pmatrix}$$

$$\Theta = \begin{pmatrix} -0.7 & 0.5 \\ -0.3 & -0.7 \end{pmatrix}$$

$$\Sigma_a = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 2.0 \end{pmatrix}$$

Figure: VARMA(1,1) with $n = 2$

System order estimates: multivariate case II

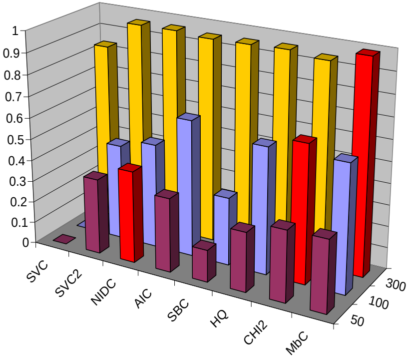


Figure: VARMA(1,1) with $n = 3$

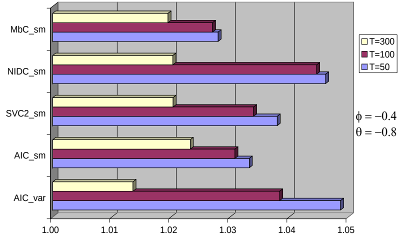
$$(I + \Phi B)Z_t = (I + \Theta B)a_t$$

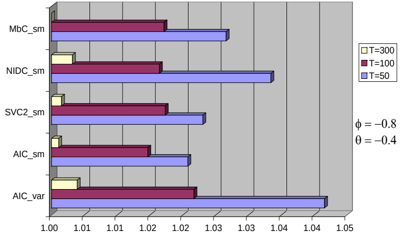
$$\Phi = \begin{pmatrix} -0.4 & -0.3 & 0.6 \\ 0 & -0.8 & -0.4 \\ -0.3 & 0 & 0 \end{pmatrix}$$

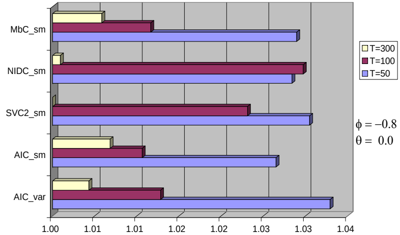
$$\Theta = \begin{pmatrix} -0.7 & 0 & 0 \\ -0.1 & -0.2 & 0 \\ 0.4 & -0.5 & 0.1 \end{pmatrix}$$

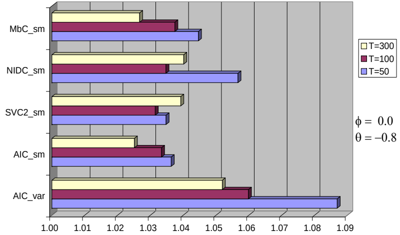
$$\Sigma_a = \begin{pmatrix} 1.0 & 0.5 & 0.4 \\ 0.5 & 1.0 & 0.7 \\ 0.4 & 0.7 & 1.0 \end{pmatrix}$$

RMSE ratios for ARMA models with $n = 1$

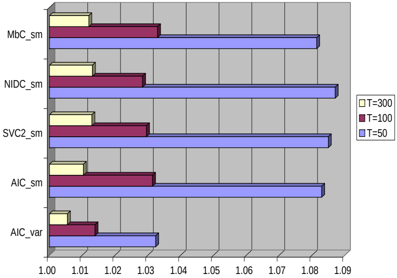


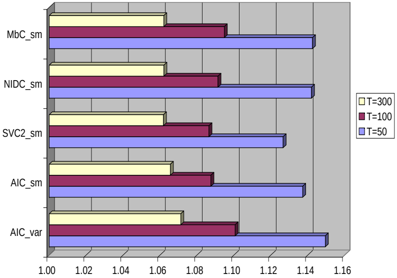






RMSE ratios for VAR and VMA models with $n = 2$





$$(I + \Phi B)Z_t = a_t,\ a_t \sim iidN(0, I)$$

$$\Phi = \begin{pmatrix} 0.4 & 0.0 \\ -0.6 & 0.5 \end{pmatrix}$$

$$Z_t = (I + \Theta B)a_t,\ a_t \sim iidN(0, I)$$

$$\Theta = \begin{pmatrix} -0.8 & 0.5 \\ 0.0 & -0.7 \end{pmatrix}$$

RMSE ratios for VARMA model with $n = 2$

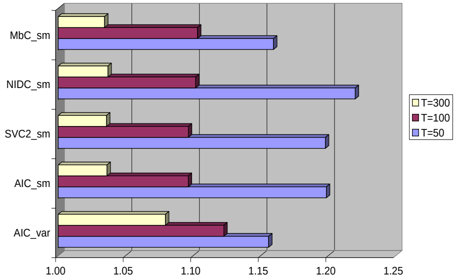


Figure: VARMA(1,1) with $n = 2$

$$(I+\Phi B)Z_t = (I+\Theta B)a_t$$

$$\Phi = \begin{pmatrix} 0.4 & 0.0 \\ -0.6 & 0.5 \end{pmatrix}$$

$$\Theta = \begin{pmatrix} -0.8 & 0.5 \\ 0.0 & -0.7 \end{pmatrix}$$

$$\Sigma_a = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{pmatrix}$$

Summary

We suggest using the SM to automatically model and forecast time series.

- We propose 3 criteria to select the system order using SM.
- In terms of forecasting:
  - Univariate: SM beats the common autoregressive model
  - Multivariate: It is not clear, but it seems that SM perform VAR when VMA or VARMA.
- What's next?
  - Extend the forecast analysis to other than one-step-ahead.
  - Adapt the methodology to seasonal models and compare with other automatic procedures which can accommodate seasonality, e.g., Tramo-Seats or DHR.

For Further Reading I

- Akaike, H. (1976).
  Canonical Correlation Analysis of Time Series and the Use of an Information Criterion.
  Academic Press.
- Bauer, D. (2001).
  Order estimation for subspace methods.
  Automatica, 37:1561–1573.
- 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.

For Further Reading II

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.

Tsay, R. S. (1989).
Identifying multivariate time series models.
Journal of Time Series Analysis, 10(4):357–372.

For Further Reading III

Wei, W. W. S. (1990).
Time Series Analysis.
Addison Wesley.