

Estimation de l'ordre et des paramètres d'un processus stochastique multivariable

Lardiès Joseph

LMARC, UMR CNRS 6604 - Université de Franche-Comté
UFR Sciences - Rue de l'Épitaphe - 25030 Besançon

RESUME

Une méthode qui permet d'estimer l'ordre et les coefficients d'un processus stochastique multivariable à partir de l'observation d'une de ses réalisations est proposée. Le processus est considéré comme la sortie d'un système linéaire invariant, excité par un bruit blanc vectoriel. Pour estimer l'ordre on effectue un test en utilisant les différentes colonnes de la matrice d'observabilité du processus. L'efficacité de la méthode est comparée à d'autres techniques courantes comme le critère d'Akaike et la décomposition en valeurs singulières de la matrice bloc Hankel. Les paramètres du processus sont ensuite estimés à partir des variables instrumentales issues des états du système.

ABSTRACT

We present a procedure allowing to estimate the order and coefficients of a multivariate stochastic process, from measured output data only. It is assumed that the observation is the output of an invariant linear system driven by white noise. A test for the significance of the coefficients in a particular column of the observability matrix provides the order of the process. The effectiveness of the method is compared with other alternatives like Akaike information criterion, SVD of the bloc Hankel matrix. The parameters are then estimated using the matrix of states as instrumental variables.

1 Introduction du modèle d'état

On considère la représentation d'état d'un système linéaire

$$z_{k+1} = A z_k + B e_k \quad ; \quad y_k = C z_k + e_k \quad (1.1)$$

où y_k est le vecteur ($m \times 1$) des observations ; e_k est le processus d'innovation de y_k , c'est un bruit blanc de matrice de covariance définie positive Q et z_k le vecteur d'état du système d'ordre minimal n . Les matrices A ($n \times n$), B ($n \times m$) et C ($m \times n$) sont les paramètres à estimer en utilisant uniquement les observations y_k [1-3]. Les matrices de covariance R_i des sorties sont obtenues à partir de (1.1)

$$R_i = E(y_{k+i} y_k^T) = C A^{i-1} G \quad (1.2)$$

avec $i > 0$ et $G = E(z_{k+1} y_k^T)$. On définit les vecteurs ($m \times 1$) et ($m \times 1$) du futur et du passé des observations par $y_k^+ = (y_k^T, y_{k+1}^T, \dots, y_{k+f-1}^T)^T$ et par $y_k^- = (y_k^T, y_{k-1}^T, \dots, y_{k-p+1}^T)^T$. La matrice bloc Hankel ($m \times m \times p$) des observations est

$$H = E(y_k^+ y_{k-1}^{-T}) = \begin{bmatrix} R_1 & R_2 & \dots & R_p \\ R_2 & R_3 & \dots & R_{p+1} \\ \dots & \dots & \dots & \dots \\ R_f & \dots & \dots & R_{p+f-1} \end{bmatrix} \\ = \begin{bmatrix} C \\ CA \\ \dots \\ CA^{f-1} \end{bmatrix} [G \quad AG \quad \dots \quad A^{p-1}G] = O K \quad (1.3)$$

avec O et K les matrices d'observabilité et de contrôlabilité associées au triplet (C, A, G) . Il est évident que toutes ces matrices sont en pratique évaluées à partir d'une suite de T échantillons de y_k , et l'estimée de la matrice bloc Hankel sera notée \hat{H} (l'accent circonflexe indiquera une estimée). La décomposition en valeurs singulières de \hat{H} est

$$\hat{H} = \hat{U} \hat{S} \hat{V}^T = \hat{O} \hat{K} \quad (1.4)$$

Lorsqu'on estime un modèle à n états on ne considère que les n valeurs singulières les plus importantes, les autres étant supposées nulles, et on définit \hat{H}_n comme une approximation de \hat{H}

$$\hat{H}_n = \hat{U}_n \hat{S}_n \hat{V}_n^T = \hat{O}_n \hat{K}_n \quad (1.5)$$

où $\hat{O}_n = \hat{U}_n \hat{S}_n^{1/2}$ est la matrice d'observabilité approchée, de pseudo-inverse $\hat{O}_n^\# = (\hat{S}_n)^{-1/2} \hat{U}_n^T$ et $\hat{K}_n = (\hat{S}_n)^{1/2} \hat{V}_n^T$ est la matrice de contrôlabilité approchée, de pseudo-inverse $K_n^\# = \hat{V}_n (\hat{S}_n)^{-1/2}$.

Nous définissons H_l et H_c les matrices formées du premier bloc-ligne (des m premières lignes) et du premier bloc-colonne (des m premières colonnes) de H de sorte que d'après l'équation (1.3) nous obtenons

$$H_l = [R_1 \ R_2 \ \dots \ R_p] = C K \quad (1.6)$$

$$H_c^T = [R_1^T \ R_2^T \ \dots \ R_f^T] = G^T O^T \quad (1.7)$$

Les estimées de C et de G sont obtenues par résolution de (1.6) et (1.7)

$$\hat{C} = \hat{H}_1 \hat{K}_n^\# \quad ; \quad \hat{G} = \hat{O}_n^\# \hat{H}_c \quad (1.8)$$

Pour estimer la matrice de transition A nous définissons la matrice bloc Hankel \bar{H} , translatée de H d'un bloc-colonne vers la gauche, le dernier bloc-colonne à droite étant formé des matrices de covariance R_{p+1}, \dots, R_{p+f}

$$\bar{H} = \begin{bmatrix} R_2 & R_3 & \dots & R_{p+1} \\ R_3 & \dots & \dots & R_{p+2} \\ \dots & \dots & \dots & \dots \\ R_{f+1} & \dots & \dots & R_{p+f} \end{bmatrix} = O A K \quad (1.9)$$

L'estimée de A est obtenue en utilisant à nouveau les pseudo-inverses de \hat{O}_n et de \hat{K}_n . Nous avons

$$\hat{A} = \hat{O}_n^\# \hat{H} \hat{K}_n^\# \quad (1.10)$$

Le gain de Kalman B est obtenu par résolution d'un système de 3 équations matricielles qui contient quatre matrices de covariance : la matrice de covariance de l'état $P = E(z_{k+1} z_{k+1}^T)$, la matrice de covariance de l'innovation $Q = E(e_k e_k^T)$, la matrice de covariance de l'état et de l'observation $G = E(z_{k+1} y_k^T)$ et $R_0 = E(y_k y_k^T)$. En considérant la matrice de covariance de l'état puis la matrice de covariance de l'observation (1.1) et en utilisant la propriété d'orthogonalité de l'état actuel et de l'innovation nous obtenons

$$P = A P A^T + B Q B^T; R_0 = C P C^T + Q; G = A P C^T + B Q \quad (1.11)$$

En combinant ces trois équations nous déduisons l'équation de Ricatti

$$P = A P A^T + (G - A P C^T)(R_0 - C P C^T)^{-1} (G - A P C^T)^T \quad (1.12)$$

qui nous permet de déterminer Q et B

$$Q = R_0 - C P C^T \quad ; \quad B = (G - A P C^T) Q^{-1} \quad (1.13)$$

Il faut noter qu'une variante pour estimer la matrice de transition a été proposée par Zeiger [4] puis par Kung [5], non pas en utilisant la matrice bloc Hankel translatée \hat{H} comme en (1.10) mais les matrices d'observabilité et de contrôlabilité translattées. Pour cela il suffit de considérer les définitions de O et de K qui aboutissent aux expressions suivantes:

$$\hat{A} = \hat{O}_n^\# \hat{O}^\uparrow \quad \text{ou encore} \quad \hat{A} = \hat{K} \hat{K}_n^\# \quad (1.14)$$

où \hat{O}^\uparrow désigne la translatée de \hat{O} d'un bloc ligne vers le haut et \hat{K} la translatée de \hat{K} d'un bloc colonne vers la gauche.

A travers les équations précédentes nous voyons combien il est important de déterminer l'ordre exact n du modèle. Pour cela quatre méthodes vont être exposées et testées en considérant un exemple. Les deux premières méthodes, classiques, utilisent l'une le critère d'Akaike, l'autre la décomposition en valeurs singulières (DVS) de la matrice bloc Hankel. La troisième méthode utilise le rapport des déterminants issus des matrices de covariance des innovations, matrices obtenues en considérant des modèles d'ordre différents. La quatrième méthode, plus efficace que la DVS, teste la valeur des coefficients d'une colonne de la matrice d'observabilité.

2 Estimation de l'ordre du processus

2.1 Critère d'Akaike

L'approche classique pour la sélection de l'ordre d'un processus est basée sur les principes de l'estimation au sens du maximum de vraisemblance. Ordinairement on choisit le modèle qui fournit la valeur la plus importante de la fonction de vraisemblance, ou, de façon équivalente, dans le cas gaussien la valeur minimale de $\log \det(Q_j)$ avec $j = 1, 2, \dots, n, \dots$ mf. Q_j est la matrice de covariance de l'innovation du modèle considéré à j états. Cependant, si le nombre de paramètres varie cette technique n'est plus adéquate, le critère $\log \det(Q_j)$ étant toujours minimum pour l'ordre maximum considéré (ici mf). Afin de faire face à ce problème Akaike [6] a suggéré d'utiliser le critère

$$AIC(j) = \log \det(Q_j) + 2(\text{nombre de paramètres estimés}) \quad (2.1)$$

Le premier terme décroît avec l'ordre j considéré, le deuxième terme, qui croît avec j, introduit la dispersion issue de l'augmentation du nombre de paramètres estimés. L'ordre n du processus est la valeur de j qui minimise ce critère.

2.2 Décomposition en valeurs singulières de \hat{H}

En supposant que mp et mf sont supérieurs à l'ordre exact du modèle, le rang de la matrice bloc Hankel H est égal au nombre d'états n requis pour caractériser y_i . Cet ordre peut être déterminé directement en utilisant une propriété de la décomposition en valeurs singulières de la matrice bloc Hankel H : le nombre de valeurs singulières non nulles est égal au rang de H. Cependant, lorsqu'on utilise l'estimée \hat{H} , et que l'on effectue sa décomposition en valeurs singulières, d'une part on constate qu'il n'existe pas de valeurs singulières nulles (même si certaines tendent vers zéro lorsque la taille de l'échantillon tend vers l'infini), d'autre part il est difficile de sélectionner les valeurs singulières les plus importantes. Les valeurs singulières étant classées par ordre décroissant, à partir de quel seuil peut on considérer comme nulles les valeurs singulières, la coupure entre les valeurs importantes et faibles n'apparaissant pas ? Cette question est primordiale car l'ordre est estimé en prenant le nombre de valeurs singulières contenues dans \hat{S} , qui sont de façon significative différentes de zéro.

2.3 Test C(j)

La troisième méthode teste le rapport des déterminants des matrices de covariance des innovations, pour des modèles d'ordre différents. Soit λ_j le rapport, à la puissance f, des déterminants des matrices de covariance des innovations pour un modèle avec un maximum d'états mf et pour un modèle à j états

$$\lambda_j = (\det(Q_{mf})/\det(Q_j))^f \quad (2.2)$$

Aoki [2] a montré que

$$(\det(Q_j))^f = \det(I_f \otimes Q_j) = \det(R - H_j R^{-1} H_j^T) \quad (2.3)$$

où R est la matrice bloc de covariance y_{k-1}^- , et H_j , comme précédemment, la matrice bloc Hankel approchée en ne considérant que les j valeurs singulières les plus importantes, les autres étant supposées nulles. En utilisant ces relations λ_j peut être estimée par

$$\hat{\lambda}_j = \det(\hat{R} - \hat{H} \hat{R}^{-1} \hat{H}^T) / \det(\hat{R} - (\hat{U}_j \hat{S}_j \hat{V}_j^T) \hat{R}^{-1} (\hat{V}_j \hat{S}_j \hat{U}_j^T)) \\ = \det(I_{mf-j} - \hat{\Gamma}_2^2) \quad (2.4)$$

où $\hat{U}_j, \hat{S}_j, \hat{V}_j$ ont été définis en (1.5) et

$$\hat{\Gamma} = \text{diag}(\hat{\gamma}_1, \dots, \hat{\gamma}_j, \hat{\Gamma}_2) \quad (2.5)$$

avec $\hat{\gamma}_i$ le ième coefficient de corrélation canonique entre y_k^+ et y_{k-1}^- ; $\hat{\Gamma}_2$ est donc la matrice estimée diagonale des coefficients de corrélation canoniques négligés. I_{mf-j} est la matrice identité d'ordre (mf-j). Ce test nous renseigne sur l'importance des coefficients de corrélation canonique omis. Pour obtenir une distribution asymptotique du Chi-deux à $(mf-j)^2$ degrés de liberté on préfère utiliser le test [7]

$$C(j) = -(T - j - (mf + 1)/2) \ln(\hat{\lambda}_j) \quad (2.6)$$

Pour estimer l'ordre du processus on procède de la façon suivante : on commence avec un modèle ayant un nombre d'états mf important ($mf > n$) et on calcule C(j) pour $j = mf - 1$; $mf - 2$; ... jusqu'à ce que la valeur de C(j) dépasse un certain seuil (jusqu'à ce que $C(j) >> C(j+1)$). L'ordre exact du modèle est alors $n = j+1$. En effet, lorsque les (mf - j) plus petits coefficients de corrélation canonique sont exclus du modèle alors le (j+1)ième coefficient devient statistiquement important.

2.4 Test M(j)

Le vecteur du futur se met sous la forme

$$y_k^+ = O z_k + F e_k^+ \quad (2.7)$$

où O est la matrice d'observabilité définie en (1.3), e_k^+ le vecteur des innovations : $e_k^+ = (e_k^T, e_{k+1}^T, \dots, e_{k+f-1}^T)^T$ et F la matrice bloc triangulaire inférieure formée de A, B, C et I_m

$$F = \begin{bmatrix} I_m & 0 & \cdot & 0 \\ CB & I_m & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ CA^{f-2}B & \cdot & CB & I_m \end{bmatrix} \quad (2.8)$$

L'estimée de la matrice d'observabilité à l'ordre n est

$$\hat{O}_n = \hat{H} \hat{V}_n (\hat{S}_n)^{1/2} \quad (2.9)$$

ce qui est une fonction linéaire des vecteurs singuliers. Si un état, ou plusieurs états non nécessaires, sont inclus dans le modèle (modèle supposé d'ordre $j > n$) les coefficients des colonnes de la matrice d'observabilité, qui sont associés à ces états, tendent vers zéro lorsque la taille de l'échantillon tend vers l'infini. Ceci est dû au fait que les valeurs singulières relatives à ces coefficients tendent aussi vers zéro. Ainsi un test sur la valeur significative ou non des coefficients d'une colonne particulière de la matrice d'observabilité nous renseigne sur l'importance de l'état associé à cette colonne. Pour tenir compte de toutes les interactions possibles nous devons calculer la matrice de covariance de la vectorisation de O. Pour cela soient les matrices Y^+ (mf x T); Z (nxT) et E^+ (mf x T) définies par $Y^+ = (y_1^+ \ y_2^+ \ \dots \ y_T^+)$; $Z = (z_1 \ z_2 \ \dots \ z_T)$; $E^+ = (e_1^+ \ e_2^+ \ \dots \ e_T^+)$ (2.10) nous avons alors

$$Y^+ = O Z + F E^+ \quad (2.11)$$

La vectorisation de Y^+ conduit à l'estimée de vec O

$$\text{vec } Y^+ = (Z^T \otimes I_{mf}) \text{vec } O + (I_T \otimes F) \text{vec } E^+$$

$$\text{vec } \hat{O} = [(Z^T \otimes I_{mf})^T (Z^T \otimes I_{mf})]^{-1} (Z^T \otimes I_{mf})^T \text{vec } Y^+ \quad (2.12)$$

et la matrice de covariance de vec \hat{O} est

$$\text{cov vec } \hat{O} = E\{[(Z^T \otimes I_{mf})^T (Z^T \otimes I_{mf})]^{-1} (Z^T \otimes I_{mf})^T (I_T \otimes F) \text{vec } E^+ \\ ((I_T \otimes F) \text{vec } E^+)^T (Z^T \otimes I_{mf}) [(Z^T \otimes I_{mf})^T (Z^T \otimes I_{mf})]^{-1}\} \quad (2.13)$$

On peut vérifier que

$$Z = K R^{-1} Y_{-1}^- \quad \text{et} \quad R = T^{-1} E(Y_{-1}^- Y_{-1}^{-T}) \quad (2.14)$$

de sorte que

$$\text{cov vec } \hat{O} = T^{-1} [(K R^{-1} K^T)^{-1} \otimes (F (I_f \otimes Q) F^T)] \quad (2.15)$$

La matrice bloc R sera estimée en considérant les matrices de covariance \hat{R}_i et la matrice de contrôlabilité K sera estimée d'après (1.4) par

$$\hat{K} = \hat{O}^\# \hat{H} = (\hat{S})^{-1/2} \hat{U}^T \hat{H} \quad (2.16)$$

les matrices F et Q seront estimées après résolution de (1.8), (1.10) et (1.13). On définit \hat{o}_j la jème colonne de \hat{O} ; le test qui nous permet de savoir si les coefficients de cette jème colonne tendent vers zéro peut se mettre sous la forme

$$M(j) = \hat{o}_j^T [T(\hat{K} \hat{R}^{-1} \hat{K}^T)]_{jj} (\hat{F} (I_f \otimes \hat{Q}) \hat{F}^T)^{-1} \hat{o}_j \quad (2.17)$$

asymptotiquement ce test suit une distribution du Chi-deux, l'expression entre crochets étant reliée à l'inverse de la matrice de covariance de \hat{o}_j . La procédure pour estimer l'ordre du processus est la suivante:

- on considère un modèle ayant un nombre d'états important ($j = mf > n$) et on calcule l'estimée de la matrice d'observabilité \hat{O} , qui nous fournira les diverses colonnes \hat{o}_j , ainsi que la matrice cov vec \hat{O}

- on calcule M(j) pour $j = mf$ et on note H_0 l'hypothèse selon laquelle les coefficients de la jème colonne de la matrice d'observabilité sont nuls

- si H_0 est retenue, alors on considère $j = mf - 1$ et on calcule la nouvelle valeur de M(j)

- on répète la procédure jusqu'à ce que l'hypothèse H_0 soit rejetée. Dans ces conditions M(j) dépasse un certain seuil. La

première valeur de j pour laquelle M(j) prend une valeur importante (dépassement d'un seuil) correspond à l'ordre du processus. On a alors j = n.

Il faut noter que la procédure est effectuée avec les mêmes estimées de la matrice d'observabilité \hat{O} et de la matrice cov vec \hat{O} , le modèle n'est pas réestimé après chaque test.

2.5 Exemple

On applique les quatre méthodes précédentes à un exemple extrait de [8] où les coefficients A, B et C sont déterminés à partir d'un système en vibration, système excité par un bruit blanc. L'ordre exact du processus est n = 6. Le nombre d'états maximum considéré est 9 (mf = mp = 9) et le nombre d'échantillons est T = 1000. Une seule réalisation a été effectuée. Le tableau ci-dessous nous fournit les valeurs des différents tests pour diverses valeurs de j

j	AIC(j)	DVS(j)	C(j)	M(j)
1	-7,34	3,56	1,1x10 ⁴	5,1x10 ⁴
2	-8,72	3,03	6,9x10 ³	2,7x10 ⁴
3	-9,39	1,23	4,8x10 ³	2,8x10 ⁴
4	-10,37	0,87	1,9x10 ³	3,3x10 ⁴
5	-10,72	0,70	8x10 ²	1,1x10 ⁴
6 *	-10,97	0,016	-2,15	6,1x10 ²
7	-10,95	0,0015	-9,58	14,38
8	-10,93	0,0013	-8,45	10,41
9	-10,91	0,0002	0	0,16

Comme l'indique cet exemple le test M(j), le test C(j) (avec n=j+1), ainsi que le critère AIC(j) donnent des résultats meilleurs que la DVS(j)

3 Estimation des paramètres par les variables instrumentales

Une fois que l'ordre a été déterminé nous allons estimer les paramètres A, B et C de façon itérative à partir des variables instrumentales issues des états du processus. Pour cela on met l'équation des observations sous forme matricielle, en considérant les T échantillons du signal

$$Y = CZ + E \tag{3.1}$$

avec Y et E les matrices (mxT) définies par Y = (y₁ y₂ . . . y_T) et E = (e₁ e₂ . . . e_T) et Z la matrice (nxT) définie par Z = (z₁ z₂ . . . z_T). On vectorise (3.1)

$$\text{vec } Y = (Z^T \otimes I_m) \text{vec } C + \text{vec } E \tag{3.2}$$

Soit Z_{VI} = (z_{VI1} z_{VI2} . . . z_{VI T}) la matrice (nxT) des instruments. En multipliant les 2 membres de l'équation (3.2) par (Z_{VI} ⊗ I_m) et choisissant l'estimée de C telle que les résidus soient orthogonaux à (Z_{VI} ⊗ I_m), on obtient

$$(Z_{VI} \otimes I_m) \text{vec } Y = (Z_{VI} \otimes I_m) (Z^T \otimes I_m) \text{vec } \tilde{C} \tag{3.3}$$

La forme variable instrumentale de vec C est donc

$$\text{vec } \tilde{C} = [(Z_{VI} Z^T)^{-1} Z_{VI} \otimes I_m] \text{vec } Y \tag{3.4}$$

on utilise l'inverse de l'opérateur vec pour obtenir

$$\tilde{C} = Y Z_{VI}^T (Z Z_{VI}^T)^{-1} \tag{3.5}$$

Une démarche identique est utilisée pour estimer la forme variable instrumentale de la matrice de transition A

$$\tilde{A} = Z_{+1} Z_{VI}^T (Z Z_{VI}^T)^{-1} \tag{3.6}$$

avec Z₊₁ = (z₂ z₃ . . . z_{T+1}). Il faut noter que les états (les matrices Z et Z₊₁) ne sont pas observés. Cependant, ils peuvent être remplacés par leurs estimés à partir de l'équation (2.14). Le problème consiste à déterminer la matrice des instruments. Toute matrice du passé Y_{-j} peut être utilisée comme matrice instrumentale. Cependant, d'après [2], la matrice Z = KR⁻¹Y_{-j} est asymptotiquement la plus efficace et sera prise comme matrice des instruments. Pour estimer le paramètre B il suffit de considérer l'équation (1.13). L'efficacité de la technique par rapport à la méthode conventionnelle qui utilise les équations (1.8) et (1.10) ou (1.14) est en voie de développement.

4 Conclusion

On a développé une méthode qui permet d'estimer l'ordre d'un processus stochastique multivariable à partir d'une de ses réalisations. Cette méthode a été comparée à d'autres techniques plus courantes et donne des résultats satisfaisants. L'estimation des paramètres a été effectuée par la méthode des variables instrumentales, en utilisant la matrice des états comme matrice des instruments. Une technique itérative permettant d'estimer efficacement les paramètres est en cours d'étude. Ainsi la matrice de transition peut être estimée en utilisant $\tilde{A}^{(i+1)} = Z_{+1}^{(i)T} (Z^{(i)} Z^{(i)T})^{-1}$, la valeur initiale étant la valeur conventionnelle.

5 Références

[1] FAURRE P.; CLERGET M. et GERMAIN F. : " Opérateurs rationnels positifs ", M.M.I., Paris, 1978
 [2] AOKI M. : " State space modeling of time series ", Springer Verlag, Berlin, 1990
 [3] HAVENNER A. : " A guide to state space modeling of multiple time series ", Lectures Notes in Statistics n° 119, pp.15-72, 1997
 [4] ZEIGER H.P. and Mc EWEN A.J. : " Approximate linear realization of given dimension via Ho's algorithm " IEEE Trans. on Aut. Control , vol. 19, pp.153-155, 1974
 [5] KUNG S.Y. : " A new identification and model reduction algorithm via singular value decomposition " Proc. 12th Asilomar Conference on Circuits, Systems and Computers, 1978, pp. 705-714
 [6] AKAIKE H. : "Fitting autoregressive models for prediction" Ann. Inst. Stat. Math., Vol.21, pp.243-247, 1969
 [7] DORFMAN J. and HAVENNER A.: "Model specification tests for balanced representation " Commu.Statist. - Theory Methods; Vol. 24 , pp. 97-119; 1995
 [8] LARDIES J. : " Frequency response functions of vibrating systems from output only measurements " à paraître dans Mechanics Research Communications .