Estimation des modèles VARMA structurels avec innovations linéaires non corrélées mais non indépendantes

Yacouba Boubacar Mainassara, Christian Francq

Estimation des modèles VARMA structurels avec innovations linéaires non corrélées mais non indépendantes

Yacouba Boubacar Mainassara *, Christian Francq†

Abstract The asymptotic properties of the quasi-maximum likelihood estimator (QMLE) of vector autoregressive moving-average (VARMA) models are derived under the assumption that the errors are uncorrelated but not necessarily independent. Relaxing the independence assumption considerably extends the range of application of the VARMA models, and allows to cover linear representations of general nonlinear processes. Conditions are given for the consistency and asymptotic normality of the QMLE. A particular attention is given to the estimation of the asymptotic variance matrix, which may be very different from that obtained in the standard framework. Modified versions of the Wald, Lagrange Multiplier and Likelihood Ratio tests are proposed for testing linear restrictions on the parameters.

Résumé Dans ce travail, nous étudions les propriétés asymptotiques du quasi-maximum de vraisemblance (QMV) des paramètres d'un modèle VARMA sans faire l'hypothèse d'indépendance sur le bruit, contrairement à ce qui est fait habituellement pour l'inférence de ces modèles. Relâcher cette hypothèse permet aux modèles VARMA faibles de couvrir une large classe de processus non linéaires. Nous faisons des hypothèses d'ergodicité et de mélange afin d'établir la convergence forte et la normalité asymptotique de l'estimateur du QMV. Ensuite, nous accordons une attention particulière à l'estimation de la matrice de variance asymptotique qui a la forme "sandwich" $\Omega := J^{-1}IJ^{-1}$, et qui peut être très différente de la variance asymptotique

*Université Lille III, GREMARS, BP 60 149, 59653 Villeneuve d'Ascq cedex, France. E-mail: yacouba.boubacarmainassara@univ-lille3.fr

†Université Lille III, GREMARS, BP 60 149, 59653 Villeneuve d'Ascq cedex, France. E-mail: christian.francq@univ-lille3.fr

standard dont la forme est $\Omega := 2J^{-1}$. Nous établissons la convergence d'un estimateur de Ω. Enfin, des versions modifiées des tests de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance sont proposées pour tester des restrictions linéaires sur les paramètres libres du modèle.

Mots-clés: Forme échelon, processus non linéaire, QMLE, représentation structurelle des modèles VARMA, test du Multiplicateur de Lagrange, test du rapport de vraisemblance, test de Wald.

1 Introduction

Dans la littérature économétrique et dans l'analyse des séries temporelles (chronologiques), la classe des modèles autorégressifs moyennes mobiles vectoriels (VARMA pour Vector AutoRegressive Moving Average) (voir Lütkepohl, 2005) sont utilisées non seulement pour étudier les propriétés de chacune de ces séries, mais aussi pour décrire de possibles relations croisées entre les différentes séries chronologiques. Ces modèles VARMA occupent une place centrale pour la modélisation des séries temporelles multivariées. Ils sont une extension naturelle des modèles ARMA qui constituent la classe la plus utilisée de modèles de séries temporelles univariées. Cette extension pose néanmoins des problèmes ardus, comme par exemple, l'identification et l'estimation des paramètres du modèle et suscite des axes de recherches spécifiques, comme la cointégration (voir Lütkepohl, 2005). Ces modèles sont généralement utilisés avec des hypothèses fortes sur le bruit qui en limitent la généralité. Ainsi, nous appelons VARMA forts les modèles standard dans lesquels le terme d'erreur est supposé être une suite indépendante et identiquement distribuée (*i.e.* iid), et nous parlons de modèles VARMA faibles quand les hypothèses sur le bruit sont moins restrictives. Le problème qui nous préoccupera sera l'analyse statistique de ces modèles.

2 Modèles et hypothèses

Soit $X_t = (X_{1\,t}, \dots, X_{d\,t})'$ est un processus vectoriel stationnaire au second ordre de dimension d, à valeurs réelles, vérifiant

$$A_{00}X_t - \sum_{i=1}^{p_0} A_{0i}X_{t-i} = B_{00}\epsilon_t - \sum_{j=1}^{q_0} B_{0i}\epsilon_{t-j}, \quad \forall t \in \mathbb{Z} \tag{1}$$

où le terme d'erreur $\epsilon_t = (\epsilon_{1\,t}, \dots, \epsilon_{d\,t})'$ est un bruit blanc faible, c'est-à-dire une suite de variables aléatoires centrées ($E\epsilon_t = 0$), non corrélées, avec une matrice de covariance non singulière Σ_0. Les paramètres A_{0i}, $i \in \{1, \dots, p_0\}$ et B_{0j}, $j \in \{1, \dots, q_0\}$ sont des matrices $d \times d$, et p_0 et q_0 sont des entiers appelés ordres. Nous supposons que ces matrices sont paramétrées suivant le vecteur des vraies valeurs des paramètres noté ϑ_0. Nous notons $A_{0i} = A_i(\vartheta_0)$, $i \in \{1, \dots, p_0\}$, $B_{0j} = B_j(\vartheta_0)$, $j \in \{1, \dots, q_0\}$ et $\Sigma_0 = \Sigma(\vartheta_0)$, où ϑ_0 appartient à l'espace compact des paramètres $\Theta \subset \mathbb{R}^{k_0}$, et k_0 est le nombre de paramètres inconnus qui est inférieur à $(p_0 + q_0 + 3)d^2$. Pour tout $\vartheta \in \Theta$, nous supposons aussi que les applications $\vartheta \mapsto A_i(\vartheta)$ $i = 0, \dots, p_0$, $\vartheta \mapsto B_j(\vartheta)$ $j = 0, \dots, q_0$ et $\vartheta \mapsto \Sigma(\vartheta)$ admettent des dérivées continues d'ordre 3. Pour la suite, nous notons $A_i(\vartheta)$, $B_j(\vartheta)$ et $\Sigma(\vartheta)$ par A_i, B_j et Σ. Pour l'estimation des paramètres, nous utiliserons la méthode du quasi-maximum de vraisemblance (QMV), qui est la méthode du maximum de vraisemblance gaussien lorsque l'hypothèse de bruit blanc gaussien est relâchée.

2.1 Définition de l'estimateur du QMV

Pour des raisons pratiques, nous écrivons le modèle sous sa forme réduite

$$X_t - \sum_{i=1}^{p_0} A_{00}^{-1} A_{0i} X_{t-i} = e_t - \sum_{j=1}^{q_0} A_{00}^{-1} B_{0j} B_{00}^{-1} A_{00} e_{t-j}, \quad e_t = A_{00}^{-1} B_{00} \epsilon_t.$$

On dispose d'une observation de longueur $n, X_1, \dots, X_n$. Pour $0 < t \leq n$ et pour tout $\vartheta \in \Theta$, les variables aléatoires $\tilde{e}_t(\vartheta)$ sont définies récursivement par

$$\tilde{e}_t(\vartheta) = X_t - \sum_{i=1}^{p_0} A_0^{-1} A_i X_{t-i} + \sum_{j=1}^{q_0} A_0^{-1} B_j B_0^{-1} A_0 \tilde{e}_{t-j}(\vartheta),$$

où les valeurs initiales inconnues sont remplacées par zéro: $\tilde{e}_0(\vartheta) = \cdots = \tilde{e}_{1-q_0}(\vartheta) = X_0 = \cdots = X_{1-p_0} = 0$. La quasi-vraisemblance gaussienne s'écrit

$$\tilde{\mathrm{L}}_n(\vartheta) = \prod_{t=1}^{n} \frac{1}{(2\pi)^{d/2}\sqrt{\det \Sigma_e}} \exp\left\{ -\frac{1}{2} \tilde{e}_t'(\vartheta) \Sigma_e^{-1} \tilde{e}_t(\vartheta) \right\}, \quad \Sigma_e = A_0^{-1} B_0 \Sigma B_0' A_0^{-1\prime}.$$

Un estimateur du QMV de ϑ_0 est défini comme toute solution mesurable $\hat{\vartheta}_n$:

$$\hat{\vartheta}_n = \arg\max_{\vartheta \in \Theta} \tilde{\mathrm{L}}_n(\vartheta) = \arg\min_{\vartheta \in \Theta} \tilde{\ell}_n(\vartheta), \quad \tilde{\ell}_n(\vartheta) = \frac{-2}{n} \log \tilde{\mathrm{L}}_n(\vartheta).$$

2.2 Propriétés asymptotiques de l'estimateur du QMV

Pour montrer la convergence l'hypothèse essentielle sera

$\mathbf{H_1}$: Le processus (ϵ_t) est stationnaire et ergodique.

Soit les polynômes $A_\vartheta(z) = A_0 - \sum_{i=1}^{p_0} A_i z^i$ et $B_\vartheta(z) = B_0 - \sum_{i=1}^{q_0} B_i z^i$. Nous ferons également les hypothèses suivantes sur les racines de ces polynômes

$\mathbf{H_2}$: $\det A(z) \det B(z) = 0 \Rightarrow |z| > 1$.

$\mathbf{H_3}$: pour tout $\vartheta \in \Theta$ tel que $\vartheta \neq \vartheta_0$, soit les fonctions de transfert

$$A_0^{-1} B_0 B_\vartheta^{-1}(z) A_\vartheta(z) \neq A_{00}^{-1} B_{00} B_{\vartheta_0}^{-1}(z) A_{\vartheta_0}(z) \text{ pour un } z \in \mathbb{C}$$

$$\text{ou alors} \quad A_0^{-1} B_0 \Sigma B_0' A_0^{-1'} \neq A_{00}^{-1} B_{00} \Sigma_0 B_{00}' A_{00}^{-1'}.$$

Nous pouvons maintenant énoncer le théorème de convergence forte suivant.

Théorème 2.1 (Convergence forte) *Sous les hypothèses* $\mathbf{H_1}$, $\mathbf{H_2}$ *et* $\mathbf{H_3}$, *nous avons presque sûrement* $\hat{\vartheta}_n \to \vartheta_0$ *quand* $n \to \infty$.

Pour la normalité asymptotique l'estimateur du QMV, nous ferons les hypothèses supplémentaires suivantes

$\mathbf{H_4}$: Nous avons $\vartheta_0 \in \overset{\circ}{\Theta}$. où $\overset{\circ}{\Theta}$ est l'intérieur de Θ,

$\mathbf{H_5}$: Il existe un réel $\nu > 0$ tel que $E\|\epsilon_t\|^{4+2\nu} < \infty$ et les coefficients de mélange du processus (ϵ_t) vérifient $\sum_{k=0}^{\infty} \{\alpha_\epsilon(k)\}^{\frac{\nu}{2+\nu}} < \infty$.

Nous définissons la matrice des coefficients de la forme réduite du modèle par

$$M_{\vartheta_0} = [A_{00}^{-1} A_{01} : \cdots : A_{00}^{-1} A_{0p} : A_{00}^{-1} B_{01} B_{00}^{-1} A_{00} : \cdots : A_{00}^{-1} B_{0q} B_{00}^{-1} A_{00} : \Sigma_{e0}].$$

Nous avons besoin d'une hypothèse qui spécifie comment cette matrice dépend du paramètre ϑ_0. Soit $\dot{M}_{\vartheta_0} := \partial \mathrm{vec}(M_\vartheta)/\partial \vartheta'$ appliquée en ϑ_0.

$\mathbf{H_6}$: La matrice $\dot{M}_{\vartheta_0}$ est de plein rang k_0.

Nous pouvons maintenant énoncer le théorème suivant qui nous donne le comportement asymptotique de l'estimateur du QMV.

Théorème 2.2 *Sous les hypothèses* $\mathbf{H_1}$–$\mathbf{H_6}$*, nous avons*

$$\sqrt{n}\left(\hat{\vartheta}_n - \vartheta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega := J^{-1}IJ^{-1}), \quad \textit{où} \quad J = J(\vartheta_0) \quad \textit{et} \quad I = I(\vartheta_0),$$

$$\textit{avec} \quad J(\vartheta) = \lim_{n\to\infty} \frac{\partial^2}{\partial\vartheta\partial\vartheta'}\tilde{\ell}_n(\vartheta) \quad \textit{p.s.}, \quad I(\vartheta) = \lim_{n\to\infty} \operatorname{Var} \frac{\partial}{\partial\vartheta}\tilde{\ell}_n(\vartheta).$$

Notons que, sous les hypothèses précédentes, la consistance et la normalité asymptotique de l'estimateur des moindres carrés ordinaires des modèles ARMA faibles univariés ont été montrés par Francq et Zakoïan (1998). Pour les modèles VARMA sous la forme réduite, nous faisons l'hypothèse:

$\mathbf{H_7}$: Le paramètre ϑ se décompose en $\vartheta^{(1)}$ et $\vartheta^{(2)}$, où $\vartheta^{(1)} \in \mathbb{R}^{k_1}$ dépend des coefficients $A_0, \dots, A_{p_0}$ et $B_0, \dots, B_{q_0}$, et où $\vartheta^{(2)} = D \operatorname{vec} \Sigma_e \in \mathbb{R}^{k_2}$ dépend de Σ_e, pour une matrice D de taille $k_2 \times d^2$, avec $k_1 + k_2 = k_0$.

Le théorème suivant montre que pour des modèles VARMA sous la forme réduite, les estimateurs du QML et du LS coincident.

Théorème 2.3 *Sous les hypothèses* $\mathbf{H_1}$–$\mathbf{H_7}$*, l'estimateur du QMV* $\hat{\vartheta}_n = (\hat{\vartheta}_n^{(1)'}, \hat{\vartheta}_n^{(2)'})'$ *peut être obtenu par*

$$\hat{\vartheta}_n^{(2)} = D \operatorname{vec} \hat{\Sigma}_e, \; \hat{\Sigma}_e = \frac{1}{n}\sum_{t=1}^{n} \tilde{e}_t(\hat{\vartheta}_n^{(1)})\tilde{e}'_t(\hat{\vartheta}_n^{(1)}), \; \textit{et}\; \hat{\vartheta}_n^{(1)} = \arg\min_{\vartheta^{(1)}} \left| \sum_{t=1}^{n} \tilde{e}_t(\vartheta^{(1)})\tilde{e}'_t(\vartheta^{(1)}) \right|.$$

De plus, nous avons

$$J = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{22} \end{pmatrix}, \quad \textit{avec} \;\; J_{11} = 2E\left\{\frac{\partial}{\partial\vartheta^{(1)}} e'_t(\vartheta_0^{(1)})\right\} \Sigma_{e0}^{-1} \left\{\frac{\partial}{\partial\vartheta^{(1)'}} e_t(\vartheta_0^{(1)})\right\}$$

et $J_{22} = D(\Sigma_{e0}^{-1} \otimes \Sigma_{e0}^{-1})D'$.

3 Estimation de la matrice Ω

Afin d'obtenir des intervalles de confiance ou de tester la significativité des coefficients VARMA faibles, il sera nécessaire de disposer d'un estimateur au moins faiblement consistant de la matrice de variance asymptotique Ω. La matrice bloc diagonale J peut facilement être estimée empiriquement par $\hat{J}$ avec

$$\hat{J}_{11} = \frac{2}{n}\sum_{t=1}^{n} \left\{\frac{\partial}{\partial\vartheta^{(1)}} \tilde{e}'_t(\hat{\vartheta}_n^{(1)})\right\} \hat{\Sigma}_{e0}^{-1} \left\{\frac{\partial}{\partial\vartheta^{(1)'}} \tilde{e}_t(\hat{\vartheta}_n^{(1)})\right\} \quad \text{et} \quad \hat{J}_{22} = D(\hat{\Sigma}_{e0}^{-1} \otimes \hat{\Sigma}_{e0}^{-1})D'.$$

Nous avons besoin d'un estimateur convergent de la matrice I. Notons que

$$I = \mathrm{var}_{as} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \Upsilon_t, \quad \text{où} \quad \Upsilon_t = \frac{\partial}{\partial \vartheta} \left\{ \log \det \Sigma_e + e_t'(\vartheta^{(1)}) \Sigma_e^{-1} e_t(\vartheta^{(1)}) \right\}_{\vartheta = \vartheta_0}.$$

Nous utiliserons la méthode d'estimation paramètrique de la densité spectrale. On note $\hat{\mathbf{\Phi}}_r(z) = \mathrm{I}_{k_0 - s} + \sum_{i=1}^{r} \hat{\Phi}_{r,i} z^i$, où $\hat{\Phi}_{r,1}, \ldots, \hat{\Phi}_{r,r}$ sont les coefficients de la régression des moindres carrés de $\hat{\Upsilon}_t$ sur $\hat{\Upsilon}_{t-1}, \ldots, \hat{\Upsilon}_{t-r}$ et par $\hat{\Sigma}_{\hat{u}_r}$ la variance empirique de ces résidus. Nous énonçons maintenant le théorème suivant qui est une extension de celui de Francq, Roy et Zakoïan (2005).

Théorème 3.1 *Sous les hypothèses* $\mathbf{H_1}$–$\mathbf{H_7}$ *et certaines conditions de régularité, nous avons*

$$\hat{I}^{\mathrm{SP}} := \hat{\mathbf{\Phi}}_r^{-1}(1) \hat{\Sigma}_{\hat{u}_r} \hat{\mathbf{\Phi}}_r'^{-1}(1) \to I$$

en probabilité quand $r = r(n) \to \infty$ *et* $r^3/n \to 0$ *quand* $n \to \infty$.

Enfin, nous proposerons des tests modifiés de Wald, du multiplicateur de Lagrange et du rapport de vraisemblance afin de tester s_0 contraintes linéaires sur les coefficients, sous l'hypothèse nulle $H_0 : R_0 \vartheta_0 = \mathfrak{r}_0$, où R_0 est une matrice $s_0 \times k_0$ connue de rang s_0 et $\mathfrak{r}_0$ un vecteur connu de dimension s_0.

References

[1] Francq, C., Roy, R. and Zakoïan, J-M. (2005) Diagnostic checking in ARMA Models with Uncorrelated Errors, *Journal of the American Statistical Association* 100, 532–544.

[2] Francq, C. and Zakoïan, J-M. (1998) Estimating linear representations of nonlinear processes, *Journal of Statistical Planning and Inference* 68, 145–165.

[3] Francq, and Zakoïan, J-M. (2005) Recent results for linear time series models with non independent innovations. In *Statistical Modeling and Analysis for Complex Data Problems,* Chap. 12 (eds P. Duchesne and B. Rémillard). New York: Springer Verlag, 137–161.

[4] Lütkepohl, H. (2005) *New introduction to multiple time series analysis.* Springer Verlag, Berlin.