Estimation adaptative pour les processus VAR(p)

Moustapha Faizi

Estimation adaptative pour les processus VAR(p)

Moustapha Faizi

Département des sciences économiques,
Faculté des sciences juridiques économiques et sociales,
Université Mohamed I^{er},
BP:724, OUJDA 60000, MAROC
faizi@droit.univ-oujda.ac.ma
Fax +212 36 50 06 00

Résumé

Dans ce travail, nous procédons à l'estimation adaptative pour le paramètre d'un autorégressif vectoriel, on considère pour cela la propriété de normalité asymptotique locale (LAN) de suite centrale basée sur les rangs. L'estimateur $\hat{\theta}_R$ qu'on obtient dépend de la densité f du bruit blanc qu'on va estimer.
Mots-clés : Normalité asymptotique locale, estimateur LAM, estimation adaptative.

Abstract

In this work, we proceed to the adaptive estimation for the parameter of a vectorial autoregressive, we consider for this purpose the property of local asymptotic normality (LAN) of central sequence based on ranks. The estimator $\hat{\theta}_R$ that we obtains depends on the density f of the white noise that we are going to estimate.
Key words : Local asymptotic normality, LAM estimator, adaptive estimation.

1 Introduction

Stein (1956) a donné une condition pour laquelle l'estimation adaptative est possible. Cette condition est vérifiée pour le paramètre de position sous l'hypothèse de symétrie de la distribution. Bickel (1982) a étudié le modèle de régression pour une densité du bruit blanc symétrique et non symétrique. Pour les familles LAN, Fabian et Hannan (1982) ont reformulé les résultats de Stein et ont proposé des conditions pour l'étude de l'adaptation. Kreiss (1987,a) a donné des estimateurs adaptatifs pour le paramètre du modèle ARMA quand la densité du bruit blanc est symétrique, et Kreiss (1987,b) a construit des estimateurs adaptatifs pour le paramètre du modèle AR sans la condition de symétrie.
Dans ce travail on aborde le problème de l'estimation de rangs adaptative pour le paramètre d'un modèle autorégressif vectoriel. L'approche que nous adoptons est essentiellement liée

à la propriété de normalité asymptotique locale basée sur une suite centrale mesurable en les rangs des résidus. Dans le cas univarié, une étude avec la même approche est donnée dans Allal et Kaaouachi (2003) pour l'estimateur du paramètre d'un modèle de régression avec erreurs ARMA.
Considérons $\underline{X} = (X_1, \ldots, X_n)'$ une réalisation finie d'un processus autorégressif vectoriel d'ordre p (VAR(p)) défini par

$$X_t = \sum_{i=1}^{p} A_i X_{t-i} + \varepsilon_t, \qquad t \in \mathbb{Z} \tag{1.1}$$

$(X_t)_{t\in\mathbf{Z}}$ est à valeurs dans $\mathbb{R}^m$, $m \geq 1$, $(\varepsilon_t)_{t\in\mathbf{Z}}$ est un bruit blanc, et $(A_i)_{1\leq i\leq p}$ sont des matrices carrées d'ordre m à coefficients réels.
Posons $\theta = (vec(A_1)', \ldots, vec(A_p)')'$, avec vec l'opérateur défini par $vec(A) = (a'_{.1}, .., a'_{.n})'$, les $a_{.i}$ sont les colonnes de la matrice A. On définit Θ l'ensemble des paramètres $\theta \in \mathbb{R}^{pm^2}$ pour lequels (1.1) est inversible. Le processus $(X_t, t \in \mathbb{Z})$ peut s'écrire : $X_t = \sum_{i=0}^{\infty} \mathcal{G}_i \varepsilon_{t-i},\ t \in \mathbb{Z}$, les $(\mathcal{G}_i)$ sont les matrices de Green.

2 Hypothèses et notations

Dans ce qui suit, $f(x)$ et $F(x)$, $x \in \mathbb{R}^m$ désignent une fonction densité de probabilité du bruit blanc et la fonction de distribution correspondante. $f_i(x)$ et $F_i(x)$, $x \in \mathbb{R}$, $i = 1, \ldots, m$ sont les fonctions densités de probabilité marginales et les fonctions de distributions marginales correspondantes. Posons $F_i^{-1}(u) = \inf\{x \in \mathbb{R}/F_i(x) \geq u\}$, $u \in (0,1)$ et $i = 1, \ldots, m$, nous définissons $F(x) = (F_1(x_1), \ldots, F_m(x_m))'$, $x = (x_1, \ldots, x_m)'$ et $F^{-1}(u) = (F_1^{-1}(u_1), \ldots, F_m^{-1}(u_m))'$, $u = (u_1, \ldots, u_m)' \in (0,1)^m$.

Les résidus $(Z_t^\theta)_{t\in\mathbf{Z}}$ au point $\theta \in \mathbb{R}^{pm^2}$ sont définis par : $Z_t^\theta = X_t - \sum_{i=1}^{p} A_i X_{t-i}$, avec $\theta = (vec(A_1)', \ldots, vec(A_p)')'$.

On note $\mathbf{H_f}(\theta)$(*resp.* $\mathbf{H}_{.}(\theta)$) l'hypothèse nulle sous laquelle Z_t^θ est un bruit blanc de densité d'innovation f spécifiée (resp. f non spécifiée) et $\mathbf{H_f}(\theta + \dfrac{u}{\sqrt{n}})$ les contres-hypothèse locales, $u = (vec(D_1)', \ldots, vec(D_p)')'$ non nul et
$(D_l)_{l=1,\ldots,p}$ sont des matrices carrées d'ordre m.

$(\mathbf{A_1})$ f est une densité continue par rapport à la mesure de *Lebesgue* sur $\mathbb{R}^m$, non nulle vérifiant: $\int xf(x)dx = 0$ et $\int xx'f(x)dx = \Sigma_f$, où Σ_f est une matrice définie positive. Et pour un certain $\delta > 0$, $\int |x_k|^{2+\delta} dF(x) < \infty$, $\forall k \in \{1, \ldots, m\}$.

$(\mathbf{A_2})$ Il existe un vecteur aléatoire $Df^{\frac{1}{2}}$ de carré intégrable tel que pour tout $0 \neq h \to 0$, on a: $(h'h)^{-1} \int [f^{\frac{1}{2}}(x+h) - f^{\frac{1}{2}}(x) - h' Df^{\frac{1}{2}}(x)]^2 dx \to 0$.

$(\mathbf{A_3})$ Posons $\varphi_f(x) = (\varphi_{f1}(x), \ldots, \varphi_{fm}(x))' = -2\dfrac{Df^{\frac{1}{2}}(x)}{f^{\frac{1}{2}}(x)}$,

$\int[\varphi_{fi}(x)]^4 f(x)dx < \infty$, $i = 1, \ldots, m$.

Remarque 1 $(\mathbf{A_1})$*et*$(\mathbf{A_3})$ *nous garantissent l'existence des moments d'ordre 2 et la matrice d'information de Fisher* $\mathcal{I}(f) = \int \varphi_f(x)\varphi'_f(x)f(x)dx$.

$(\mathbf{A_4})$ La distribution jointe de $(\varepsilon_0, X_{-p+1}, \ldots, X_0)$ admet une densité $f_0(., A)$. Si $A^{(n)}(L) = I - \sum_{i=1}^{p} A_i^{(n)} L^i$, $n \in I\!N$ est tel que : $A_i^{(n)} \to A_i$ pour tout $i = 1, \ldots p$, quand $n \to \infty$, alors
$f_0(\varepsilon_0, X_{-p+1}, \ldots, X_0; A^{(n)}) - f_0(\varepsilon_0, X_{-p+1}, \ldots, X_0; A)$ est un $o_p(1)$, quand $n \to \infty$, pour $(X_t)_{t\in\mathbf{Z}}$ satisfaisant (1.1).

$(\mathbf{A_5})$ La fonction score φ_f est lipshitzienne par morceaux.

On note $Z_t^\theta = (Z_{t,1}^\theta, \ldots, Z_{t,m}^\theta)'$ et $R_{t,i}^\theta$ le rang de $Z_{t,i}^\theta$ parmi $\{Z_{1,i}^\theta, \ldots, Z_{n,i}^\theta\}$, $R_t^\theta = (R_{t,1}^\theta, \ldots, R_{t,m}^\theta)'$ et $R^\theta = (R_1^\theta, \ldots, R_n^\theta)$, R^θ est appelée la matrice de rang de la série. Définissons R_*^θ la matrice de rang obtenue à partir de R^θ en réarrangeant ses colonnes de manière que la première ligne ait les éléments 1,2,...,m. Par conséquent, les vecteurs de rangs R_t^θ sont conditionnellement libre par rapport à R_*^θ.
Notons par E_θ l'espérance mathématique sous l'hypothèse $\mathbf{H}_.(\theta)$ et $E_\theta^*(.)$ l'espérance conditionnelle $E_\theta(./R_*^\theta)$.

3 R-estimateur LAM

Soit $T_{n,i}^\theta = S_{n,i}^\theta - E_\theta^*[S_{n,i}^\theta]$ avec $S_{n,i}^\theta = \dfrac{1}{n-i}\sum_{t=i+1}^{n} \varphi_f o F^{-1}(\dfrac{R_t^\theta}{n+1}) F^{-1'}(\dfrac{R_{t-i}^\theta}{n+1})$ pour $i = 1, \ldots, n-1$. On peut donc définir la statistique vectorielle de rangs multivariés sérielle centrée, en posant

$$T_{\mathcal{G},n}(\theta) = \sum_{i=1}^{n-1} \begin{pmatrix} \mathcal{G}_{i-1} \otimes I_m \\ . \\ . \\ \mathcal{G}_{i-p} \otimes I_m \end{pmatrix} vec(T_{n,i}^\theta) \tag{3.1}$$

I_m est la matrice identité d'ordre m et $\otimes$ est le produit de Kronecker de deux matrices.
Posons maintenant $\mathcal{P} = \Sigma_f \otimes \mathcal{I}(f)$, et $W_\mathcal{G}(\theta) = [W_{\mathcal{G},kl}(\theta)]_{k,l=1,\ldots,p}$, avec $W_{\mathcal{G},kl}(\theta) = \sum_{i=1}^{\infty} (\mathcal{G}_{i-k} \otimes I_m)\mathcal{P}(\mathcal{G}_{i-l} \otimes I_m)'$.
Soit $\Lambda_f^{(n)}$ le log du rapport de vraisemblance de $\mathbf{H_f}(\theta + \frac{u}{\sqrt{n}})$ par rapport à $\mathbf{H_f}(\theta)$. On énonce ici la propriété **LAN** de suite centrale basée sur les rangs.

Proposition 1 *Sous les hypothèses* $(\mathbf{A_1})\ldots(\mathbf{A_5})$ *et sous* $\mathbf{H_f}(\theta)$ *on a la propriété LAN de suite centrale* $\sqrt{n}T_{\mathcal{G},n}(\theta)$*: quand* $n \longrightarrow \infty$

$$\Lambda_f^{(n)} = u' \sqrt{n} T_{\mathcal{G},n}(\theta) - \frac{1}{2}\tau^2 + o_P(1) \tag{3.2}$$

et $\sqrt{n}T_{\mathcal{G},n}(\theta)$ est asymptotiquement normale de moyenne 0 et de matrice de variance-covariance $W_{\mathcal{G}}(\theta)$.

Proposition 2 *(Linéarité asymptotique)*
Sous les hypothèses de la proposition 1 et sous $\mathbf{H_f}(\theta)$, on a : quand $n \to \infty$

$$\sqrt{n}T_{\mathcal{G},n}(\theta + \frac{u}{\sqrt{n}}) - \sqrt{n}T_{\mathcal{G},n}(\theta) + W_{\mathcal{G}}(\theta)\, u = o_P(1) \tag{3.3}$$

Soit la statistique $Q_{\mathcal{G},n}(\tilde{\theta}) = nT'_{\mathcal{G},n}(\tilde{\theta})W_{\mathcal{G}}(\tilde{\theta})^{-1}T_{\mathcal{G},n}(\tilde{\theta})$, nous définissons le R-estimateur $\hat{\theta}_R$ du paramètre θ par

$$\hat{\theta}_R = \underset{\tilde{\theta}}{\text{argmin}}\ Q_{\mathcal{G},n}(\tilde{\theta}) \tag{3.4}$$

Proposition 3 *L'estimateur $\hat{\theta}_R$ est θ-régulier :*

$$\hat{\theta}_R = \theta - W_{\mathcal{G}}(\theta)^{-1}T_{\mathcal{G},n}(\theta) + o_P(n^{-\frac{1}{2}}), \tag{3.5}$$

ce qui entraîne que $\sqrt{n}(\hat{\theta}_R - \theta)$ est asymptotiquement normale $\mathcal{N}(0, W_{\mathcal{G}}(\theta)^{-1})$ sous $\mathbf{H_f}(\theta)$, quand $n \to \infty$.

L'estimateur $\hat{\theta}_R$ correspond ici à l'estimateur localement asymptotiquement minimax au sens de Hájek (1972) $\hat{\theta}_{LAM}$, c'est une conséquence du théorème 6.3 de Fabian et Hannan (1982), voir pour cela Faizi (2002).

Proposition 4 *L'estimateur $\hat{\theta}_R$ vérifie pour toute fonction de perte l sur $I\!R$ et pour chaque $i = 1, \ldots, pm^2$*

$$\lim_{\varepsilon \to 0} \lim_{n \to \infty} \sup_{\|\delta - \theta\| \le \varepsilon} E_\delta\, l \left(\sqrt{n} \frac{e'_i(\hat{\theta}_R - \delta)}{\sqrt{e'_i W_{\mathcal{G}}(\theta)^{-1} e_i}} \right) = E^\phi\, l \tag{3.6}$$

E_δ est l'espérance mathématique sous $\mathbf{H_f}(\delta)$, e_i est le vecteur de $I\!R^{pm^2}$ dont les composantes sont nulles sauf pour la $i^{éme}$ qui est égale à 1 et E^ϕ l'espérance mathèmatique par rapport à la distribution normale centrée réduite.

4 R-estimateur adaptatif

Une construction pratique de l'estimateur $\hat{\theta}_R$ de θ est donné dans l'algorithme suivant :

$$\hat{\theta}^*_R = \hat{\theta}_n + W_{\mathcal{G}}(\hat{\theta}_n)^{-1}T_{\mathcal{G},n}(\hat{\theta}_n) \tag{4.1}$$

avec $\hat{\theta}_n$ un estimateur $\sqrt{n}$-convergeant de θ : $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) = O_P(1)$.
L'estimateur $\hat{\theta}^*_R$ de θ est asymptotiquement équivalent à $\hat{\theta}_R$, donc de même loi, mais cet estimateur dépend toujours de la densité f du bruit blanc, et cela à travers les fonctions scores dans $T_{\mathcal{G},n}(\hat{\theta}_n)$ et les matrices Σ_f et $\mathcal{I}(f)$ dans $W_{\mathcal{G}}(\hat{\theta}_n)$. Dans la pratique la forme de la densité f est inconnue. Néanmoins on voudrait aboutir à des procédures qui évitent ce problème en gardant les propriétés asymptotiques de nos R-estimateurs (en plus l'optimalité du R-estimateur $\hat{\theta}_{LAM}$). Les procédures permettant de telles solutions sont appelées procédures adaptatives. Celle que nous présentont ici est principalement liée à l'estimation de la densité f ou des fonctions scores, qu'on remplacera dans le calcul du R-estimateur. Cette forme d'estimation englobe toutes les densités vérifiant des conditions initiales classiques. En général, ces méthodes convergent moins rapidement que celles où on choisi la densité par un procédé de test parmi une famille de densités spécifique. Pour un compte rendu de ces procédures adaptatives, voir Hogg (1974-1976).

Soit un estimateur $\hat{f}_n$ de f qui satisfait la condition suivante :

($\mathbf{A_6}$) On a sous $\mathbf{H_f}(\theta)$, pour tous $k, l \in \{1, \ldots, m\}$:

$$E^*_\theta[(\varphi_{f,k} o F^{-1}(\frac{R^\theta_t}{n+1}) F_l^{-1}(\frac{R^\theta_{t-i,l}}{n+1}) - \varphi_{\hat{f}_n,k} o \hat{F}_n^{-1}(\frac{R^\theta_t}{n+1}) \hat{F}_{n,l}^{-1}(\frac{R^\theta_{t-i,l}}{n+1}))^2 | \hat{f}_n] = o_P(1). \quad (4.2)$$

Cette condition est la même donnée pour le cas univarié dans Hallin et Werker (1997).
Sous la condition ($\mathbf{A_6}$) on va adapter l'algorithme (4.1) en estimant les fonctions scores $\varphi_f o F^{-1}$ et F^{-1} par $\varphi_{\hat{f}_n} o \hat{F}_n^{-1}$ et $\hat{F}_n^{-1}$, avec $\hat{F}_n$ la fonction de distribution généralisé.
On peut déduire facilement de la proposition 2.4. de Hallin et al. (1989) que: $E^*_\theta[S^\theta_{n,i}]$ converge en probabilité vers 0.
Posons $\widehat{T}^\theta_{n,i} = \dfrac{1}{n-i} \sum_{t=i+1}^{n} \varphi_{\hat{f}_n} o \hat{F}_n^{-1}(\dfrac{R^\theta_t}{n+1}) \hat{F}_n^{-1\prime}(\dfrac{R^\theta_{t-i}}{n+1})$, on obtient une statistique $\widehat{T}_{\mathcal{G},n}(\theta)$ en remplaçant $T^\theta_{n,i}$ par $\widehat{T}^\theta_{n,i}$ dans la statistique $T_{\mathcal{G},n}(\theta)$, on a le résultat suivant :

Proposition 5 *Sous les hypothèses* ($\mathbf{A_1}$) ... ($\mathbf{A_6}$) *et sous* $\mathbf{H_f}(\theta)$, *on a :*

i. $vec(\widehat{T}^\theta_{n,i}) = vec(T^\theta_{n,i}) + o_P(n^{-\frac{1}{2}})$, *quand* $n \to \infty$,

ii. $\widehat{T}_{\mathcal{G},n}(\theta) = T_{\mathcal{G},n}(\theta) + o_P(n^{-\frac{1}{2}})$, *quand* $n \to \infty$.

Maintenant, soit $\hat{\theta}_n$ un estimateur $\sqrt{n}$-convergeant de θ. On définit l'estimateur adaptatif $\hat{\theta}^{ad}_R$ par

$$\hat{\theta}^{ad}_R = \hat{\theta}_n + \widehat{W}_{\mathcal{G}}(\hat{\theta}_n)^{-1} \widehat{T}_{\mathcal{G},n}(\hat{\theta}_n) \quad (4.3)$$

Ici $\widehat{W}_{\mathcal{G}}(\theta)$ est un estimateur de $W_{\mathcal{G}}(\theta)$ obtenu en remplaçant Σ_f et $\mathcal{I}(f)$ dans $\mathcal{P}$ par des estimateurs $\Sigma_{\hat{f}_n}$ et $\widehat{\mathcal{I}}(\hat{f}_n)$ qui convergent respectivement en probabilité vers Σ_f et $\mathcal{I}(f)$. On prendra par exemple sous $\mathbf{H_f}(\theta)$ les estimateurs $\Sigma_{\hat{f}_n} = \frac{1}{n} Z^\theta_t Z^{\theta\prime}_t$ et $\widehat{\mathcal{I}}(\hat{f}_n) =$

$\frac{1}{n}\varphi_{\hat{f}_n}(Z_t^\theta)\varphi_{\hat{f}_n}(Z_t^\theta)'$ et en ajoutant la condition suivante

$$E_\theta^*[(\varphi_{f,k}(Z_t^\theta) - \varphi_{\hat{f}_n,k}(Z_t^\theta))^2|\hat{f}_n] = o_P(1),\ k = 1, \ldots, m, \tag{4.4}$$

on a la convergence de $\widehat{\mathcal{I}}(\hat{f}_n)$ en probabilité vers $\mathcal{I}(f)$.
L'estimateur adaptatif $\hat{\theta}_R^{ad}$ hérite des propriétés asymptotiques de $\hat{\theta}_R$.

Proposition 6 *Sous les hypothèses de la proposition 5, on a :*

$$\sqrt{n}\|\hat{\theta}_R^{ad} - \hat{\theta}_R\| = o_P(1) \tag{4.5}$$

et $\sqrt{n}(\hat{\theta}_R^{ad} - \theta)$ *est asymptotiquement normale* $\mathcal{N}(0, W_{\mathcal{G}}(\theta)^{-1})$*, quand* $n \to \infty$.

Remarque 2 *On peut toujours construire des estimateurs* $\hat{f}_n$ *de* f *satisfaisant la condition* $(\mathbf{A_6})$ *ou encore la condition (4.4), voir par exemple Bickel et al. (1993).*

Bibliographie

[1] Allal, J. & Kaaouachi, A. (2004) *Adaptive R-estimation in a linear regression model with ARMA errors.* Statistics. 37, 271-286.

[2] Bickel, P. J. (1982) *On adaptive estimation.* Ann. Statist., 10, 647-671.

[3] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. & Wellner, J.A. (1993) *Efficient and adaptive estimation for semiparametric models.* Johns Hopkins University Press, Baltimore.

[4] Fabian, V. & Hannan, J. (1982) *On Estimation and Adaptive Estimation for Locally Asymptotically Normal Families.* Wahrscheinlichkeitstheorie verw. Gebiete. Springer Verlag.

[5] Faizi, M. (2002) *Estimation basée sur les méthodes de rangs pour les modèles autorégressifs vectoriels d'ordre p.* Thèse de doctorat, Université Sidi Mohamed Ben Abdellah, Faculté des Sciences Dhar Mehraz, Fès.

[6] Hájek, J. (1972) *Local asymptotic minimax and admissibility in estimation.* Proc. Sixth Berkeley Sympos. Math. Statist. Probab. I, 175-194. Univ. Calif. Press.

[7] Hallin, M., J. F. Ingenbleek & M. L. Puri (1989) *Asymptotically Most Powerful Rank Tests for Multivariate Randomness against Serial Dependence.* Academic Press, Inc.

[8] Hallin, M. & Werker B.J.M. (1997) *Efficient rank-based semi-parametric inference.* Institut de Statistique. ULB, Brussels.

[9] Hogg, R. V. (1974) *Adaptive robust procedures: partial review on some suggestions for futur applications and theory.* J. Amer. Statist. Assoc., 69, 909-923.

[10] Hogg, R. V. (1976) *A new dimension to nonparametric tests.* Comm. Statist. th. Methods A 5, 1313-1325.

[11] Kreiss, J. P. (1987a) *On Adaptive Estimation In Stationary ARMA Processes.* The Annals of Statistics, Vol 15, $N°$1, 112-133.

[12] Kreiss, J. P. (1987b) *On adaptive estimation in autoregressive models when there are nuisance functions.* Statistics and Decisions, Vol 5, 59-76.

[13] Stein, C. (1956) *Efficient nonparametric testing and estimation.* Proc. Third Berkeley Sympos. Math. Statist. and Probab. I, Univ. Calif. Press : 187-195.