ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 32 — 1993

Tests sur le noyau, l'image et le rang de la matrice des coefficients d'un modèle linéaire multivarié

Christian GOURIEROUX, Alain MONFORT, Éric RENAULT*

RÉSUMÉ. – Dans cet article, on s'intéresse aux problèmes d'inférence sur la matrice des coefficients d'un modèle linéaire multivarié ; en particulier on considère des problèmes de test portant sur le noyau, l'image et le rang de cette matrice. On explicite et on compare les diverses procédures de test : rapport des maxima de (pseudo) vraisemblance, Wald (ou Wald généralisé), score. Chaque procédure de test est ramenée à un problème d'analyse canonique empirique.

Tests on the Kernel, the Range and the Rank of the Matrix of Coefficients in a Multivariate Linear Model

ABSTRACT. – In this paper we are interested in inference problems on the matrix of coefficients in a multivariate linear model ; in particular we consider tests on the kernel, the range and the rank of this matrix. Various test procedures are explicited and compared : (pseudo) likelihood ratio, Wald (or generalized Wald), score. Each test procedure is put in a canonical analysis framework.

* C. GOURIEROUX : CREST-CEPREMAP ; A. MONFORT : INSEE-Département de la Recherche ; E. RENAULT : GREMAQ.

1 Introduction

Nous considérons un système de régressions empilées :

$$\mathrm{Y}_t = \mathrm{BX}_t + u_t, \qquad t = 1, \cdots, \mathrm{T},$$

où Y_t, X_t sont des vecteurs de tailles respectives $(n, 1)$, $(\mathrm{K}, 1)$, et où B est une matrice (n, K), de coefficients. Le terme d'erreur u_t, de taille n, est supposé centré, de matrice de variance-covariance $\mathrm{V}u_t = \Omega$ inversible, indépendante du temps et non contrainte en dehors des conditions de symétrie et de positivité ; u_t est de plus supposé non corrélé avec les variables explicatives $\mathrm{X}_t, \mathrm{X}_{t-1}, \cdots$ et temporellement non corrélé.

Nous nous intéressons dans cet article à diverses hypothèses concernant la matrice B des coefficients de régression ; ces hypothèses portent sur le noyau de B, sur son image ou sur son rang. Nous considérons les diverses procédures de tests : rapport des maxima de (pseudo) vraisemblance, Wald (généralisé), score. Ces problèmes de tests se rencontrent fréquemment dans les modèles à facteurs (voir Gourieroux-Monfort-Renault [1991]) et ont des applications pour l'analyse de la causalité (Anderson [1963]b) pour la détermination de relations de codépendance (Gourieroux-Peaucelle [1993], Kugler-Neusser [1993]), pour les écritures structurelles de modèles macroéconomiques (Sargent-Sims [1977]), pour la recherche des ordres autorégressif et moyenne mobile des processus ARMA (Box-Tiao [1977], Tiao-Tsay [1989], Tsay-Tiao [1987]), pour la recherche des facteurs ou des portefeuilles efficients en théorie financière (Chamberlain-Rothschild [1983], Brown [1989], Gourieroux-Monfort-Renault [1991]) ou pour l'étude des conditions d'identification dans les modèles de panels multivariés (Bhargava [1990]). C'est cependant plus ici l'analyse théorique des procédures de tests que leurs simplifications ou interprétations pour chaque type d'application qui nous intéressera et nous nous efforcerons de donner une présentation générale d'un ensemble de résultats apparus de façon éparse et partielle dans la littérature (voir la liste des références).

En particulier, nous établirons divers résultats nouveaux concernant les comparaisons à distance finie et asymptotiques des diverses statistiques de test et leurs propriétés d'invariance vis-à-vis de la forme implicite, explicite ou mixte de l'hypothèse nulle. Nous généralisons ainsi au cadre multivarié des résultats classiques pour les modèles linéaires à une équation.

De plus nous proposons des méthodes d'estimation par maximum de vraisemblance ou moindres carrés asymptotiques des sous-espaces image et noyau associés à la matrice B.

Dans le paragraphe 2, nous considérons des hypothèses dites ex-ante (par analogie avec les applications financières) ; ces hypothèses sont du type : le noyau de B contient-il un ensemble de vecteurs donnés *a priori* ? l'image de B est-elle engendrée par un ensemble de vecteurs donnés *a priori* ? Pour chacun de ces problèmes nous développons les diverses procédures de tests et comparons entre elles les statistiques de test obtenues.

Dans le paragraphe 3, nous considérons des procédures dites ex-post, permettant de déterminer le rang de la matrice B et d'estimer son image et son noyau sans idées *a priori* sur ceux-ci. Nous constatons que les diverses procédures conduisent naturellement à mener des analyses canoniques empiriques. De plus nous vérifions que les procédures d'estimations (maximum de vraisemblance, moindres carrés asymptotiques) associées aux diverses procédures de tests conduisent aux mêmes estimations de l'image et du noyau.

Dans le paragraphe 4, nous généralisons les résultats des paragraphes précédents au cas où il existe d'autres variables Z_t dans le système de régression empilées.

2 Analyse ex-ante

2.1. Tests sur le noyau de B

2.1.1. *Formes de l'hypothèse*

Nous considérons une matrice γ_0 connue, de taille (K, K-r), de rang K-r. L'hypothèse à tester est :

(1) $$\mathrm{H}_0^{\mathrm{N}} = \{\text{le noyau de B contient les vecteurs colonnes de } \gamma_0\} = \{\mathrm{B}\,\gamma_0 = 0\}.$$

Cette hypothèse est écrite sous forme de contraintes implicites, mais admet aussi des écritures sous forme explicite. Ainsi introduisons une matrice α_0 de taille (K, r), de rang r, dont les vecteurs colonnes sont orthogonaux aux vecteurs colonnes de γ_0. Comme γ_0 est connu *a priori*, il en est de même de α_0. L'hypothèse s'écrit alors :

(1′) $$\mathrm{H}_0^{\mathrm{N}} = \{\text{il existe une matrice } \beta, \text{ de taille } (n,\ r) \text{ telle que } \mathrm{B} = \beta\,\alpha_0'\}.$$

2.1.2. *Forme de la statistique du rapport des maxima de (pseudo) vraisemblance*

La statistique de test fondée sur une (pseudo) vraisemblance normale, est de la forme (voir annexe 1).

(2) $$\xi_{\mathrm{RMV}} = \mathrm{T}\,\mathrm{Log}\,\frac{\det \hat{\Omega}_{0\mathrm{T}}}{\det \hat{\Omega}_{\mathrm{T}}},$$

où $\hat{\Omega}_{0T}$ et $\hat{\Omega}_T$ désignent les estimateurs de la matrice Ω calculés sous l'hypothèse nulle et sous l'hypothèse générale respectivement. Cette expression ne dépend ni de la forme (implicite ou explicite) retenue pour écrire l'hypothèse nulle, ni du choix γ_0 ou α_0 de bases du noyau ou de son orthogonal. Utilisons par exemple la forme explicite. Nous voyons que sous l'hypothèse nulle le modèle se réduit à un système de régressions empilées :

$$\text{(3)} \qquad Y_t = \beta\, \alpha_0'\, X_t + u_t, \qquad t = 1, \cdots, T,$$

avec comme nouveaux régresseurs les variables composantes de $\alpha_0'\, X_t$. Désignons par X, Y les matrices de tailles respectives (T, K) et (T, *n*) donnant les observations des diverses variables (ces matrices sont supposées de plein rang colonne), par P_X et $P_{X\alpha_0}$ les projecteurs orthogonaux sur les sous-espaces engendrés par les vecteurs colonnes de X et de $X\alpha_0$. Nous avons :

$$\text{(4)} \qquad \hat{\Omega}_T = \frac{1}{T}\, Y'\,(\mathrm{Id} - P_X)\, Y, \qquad \hat{\Omega}_{0T} = \frac{1}{T}\, Y'\,(\mathrm{Id} - P_{X\alpha_0})\, Y.$$

Comme $\alpha_0'\,\gamma_0 = 0$, nous avons aussi : $(X\alpha_0)'\, X\, (X'X)^{-1}\gamma_0 = 0$, ce qui implique :

$$P_X = P_{X\alpha_0} + P_{X\,(X'X)^{-1}\gamma_0}.$$

Nous pouvons alors expliciter la statistique de test :

$$\xi_{RMV} = T \operatorname{Log} \frac{\det \frac{1}{T}\, Y'\,(\mathrm{Id} - P_{X\alpha_0})\, Y}{\det \frac{1}{T}\, Y'\,(\mathrm{Id} - P_X)\, Y},$$

sous l'une des formes équivalentes ci-dessous :

$$\text{(5)} \begin{cases} \xi_{RMV} = T \operatorname{Log} \det \{\mathrm{Id} + \{Y'\,(\mathrm{Id} - P_X)\, Y\}^{-1}\, Y'\, P_{X\,(X'X)^{-1}\gamma_0}\, Y\} \\ \xi_{RMV} = T \operatorname{Log} \det \{\mathrm{Id} + \{Y'\,(\mathrm{Id} - P_X)\, Y\}^{-1}\, Y'\, X\, (X'X)^{-1}\gamma_0\} \\ \qquad\qquad \{\gamma_0'\,(X'X)^{-1}\gamma_0\}^{-1}\, \gamma_0'\,(X'X)^{-1}\, X'Y\}. \end{cases}$$

2.1.3. *Statistique du rapport des maxima de (pseudo) vraisemblance corrigée*

Sous l'hypothèse nulle, nous savons que $\hat{\Omega}_T$ et $\hat{\Omega}_{0T}$ tendent vers la même limite. La statistique initiale

$$\xi_{RMV} = T \operatorname{Log} \frac{\det \hat{\Omega}_{0T}}{\det \hat{\Omega}_T} = T \operatorname{Log} \det\, [\mathrm{Id} + \hat{\Omega}_T^{-1}\,(\hat{\Omega}_{0T} - \hat{\Omega}_T)],$$

est alors équivalente à :

$$\tilde{\xi}_{RMV} = T \operatorname{Log} \det\, [\mathrm{Id} + \hat{\Omega}_{0T}^{-1}\,(\hat{\Omega}_{0T} - \hat{\Omega}_T)].$$

Cette statistique du rapport des maxima de (pseudo) vraisemblance corrigée est dans notre cas égale à :

$$\tilde{\xi}_{\mathrm{RMV}} = \mathrm{T}\,\mathrm{Log}\det\{\mathrm{Id} + (\mathrm{Y}'\,(\mathrm{Id} - \mathrm{P}_{\mathrm{X}\alpha_0})\,\mathrm{Y})^{-1}\,\mathrm{P}_{\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0}\,\mathrm{Y}\}.$$

Comme $(\mathrm{Id} - \mathrm{P}_{\mathrm{X}\alpha_0})\,\mathrm{X}\,(\mathrm{X}'\,\mathrm{X})^{-1}\gamma_0 = \mathrm{X}\,(\mathrm{X}'\,\mathrm{X})^{-1}\gamma_0$ à cause de la condition d'orthogonalité $\alpha_0'\,\gamma_0 = 0$, nous voyons que la matrice $[\mathrm{Y}'\,(\mathrm{Id} - \mathrm{P}_{\mathrm{X}\alpha_0})\,\mathrm{Y}]^{-1}\,\mathrm{Y}'\,\mathrm{P}_{\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0}\,\mathrm{Y}$ est la matrice de corrélation canonique non centrée empirique entre Y et $\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0$, conditionnellement à $\mathrm{X}\alpha_0$. Cette matrice se réduit à la matrice de corrélation canonique conditionnelle usuelle si les variables X et Y sont centrées ou si les variables $\mathrm{X}\alpha_0$ contiennent la constante. Nous notons : $\mathrm{R}^2\,[\mathrm{Y},\,\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0/\mathrm{X}\alpha_0]$ cette matrice de corrélation conditionnelle.

Si $\hat{\lambda}_{1\mathrm{T}} \geq \cdots \geq \hat{\lambda}_{n\mathrm{T}} \geq 0$ désignent les valeurs propres de cette matrice, nous en déduisons que :

$$\tilde{\xi}_{\mathrm{RMV}} = \mathrm{T}\sum_{i=1}^{n}\mathrm{Log}\,(1 + \hat{\lambda}_{i\mathrm{T}}).$$

Remarquons aussi en appliquant le lemme de l'annexe 2 (avec $\mathrm{U} = \mathrm{Y}$, $\mathrm{P} = \mathrm{P}_{\mathrm{X}}$, $\mathrm{P}_1 = \mathrm{P}_{\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0}$, $\mathrm{P}_2 = \mathrm{P}_{\mathrm{X}\alpha_0}$) que les valeurs propres de la matrice $[\mathrm{Y}'(\mathrm{Id} - \mathrm{P}_{\mathrm{X}})\,\mathrm{Y}]^{-1}\,\mathrm{Y}'\,\mathrm{P}_{\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0}\,\mathrm{Y}$ intervenant dans l'expression de ξ_{RMV} (formule 5) sont égales à $\hat{\lambda}_{i\mathrm{T}}/(1 - \hat{\lambda}_{i\mathrm{T}})$.). Nous en déduisons que la statistique ξ_{RMV} s'écrit :

$$\begin{aligned}\xi_{\mathrm{RMV}} &= \mathrm{T}\sum_{i=1}^{n}\mathrm{Log}\,[1 + \hat{\lambda}_{i\mathrm{T}}/(1 - \hat{\lambda}_{i\mathrm{T}})]\\ &= -\mathrm{T}\sum_{i=1}^{n}\mathrm{Log}\,(1 - \hat{\lambda}_{i\mathrm{T}}).\end{aligned}$$

Bien que la statistique du maximum de vraisemblance corrigée ne soit pas introduite en général parmi les procédures de tests classiques, nous voyons qu'elle apparaît de façon naturelle et avec une interprétation plus simple en terme de valeurs propres que la statistique ξ_{RMV} usuelle.

Propriété 1 : Soient $1 > \hat{\lambda}_{1\mathrm{T}} \geq \cdots \geq \hat{\lambda}_{n\mathrm{T}}$ les carrés des corrélations canoniques entre Y et $\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0$ conditionnellement à $\mathrm{X}\alpha_0$, c'est-à-dire les valeurs propres de la matrice $\mathrm{R}^2\,[\mathrm{Y},\,\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma_0/\mathrm{X}\alpha_0]$, la statistique du rapport des maxima de (pseudo) vraisemblance pour tester $\mathrm{H}_0^{\mathrm{N}}$ est : $\xi_{\mathrm{RMV}} = -\mathrm{T}\sum^{n}\mathrm{Log}\,(1 - \hat{\lambda}_{i\mathrm{T}})$; la statistique corrigée est : $\tilde{\xi}_{\mathrm{RMV}} = -\mathrm{T}\sum_{i=1}^{n}\mathrm{Log}\,(1 - \hat{\lambda}_{i\mathrm{T}})$.

Si n est supérieur à K-r, il y a dans l'expression de la statistique certains termes qui peuvent être omis :

$$\xi_{\mathrm{RMV}} = -\mathrm{T}\sum_{i=1}^{\mathrm{Min}\,(n,\,\mathrm{K}-r)}\mathrm{Log}\,(1 - \hat{\lambda}_{i\mathrm{T}}).$$

Le même type de remarques s'applique d'ailleurs à toutes les expressions des statistiques en terme de valeurs propres.

Remarque : à titre d'exemple, nous pouvons considérer un modèle de régressions empilées avec terme constant :

$$Y_t = b.1 + B^\star X_t^\star + U = (b,\ B^\star) \begin{bmatrix} 1 \\ X_t^\star \end{bmatrix} + u_t.$$

Pour tester l'hypothèse $H_0^N = \{B^\star = 0\}$, la statistique du rapport des maxima de vraisemblance corrigée est :

$$\tilde{\xi}_{RMV} = T \operatorname{Log} \det (\mathrm{Id} + R^2),$$

où : $R^2 = (V_{emp}\, Y)^{-1} \operatorname{Cov}_{emp} (Y,\ X)\, (V_{emp}\, X)^{-1} \operatorname{Cov}_{emp} (X,\ Y)$ est la matrice de corrélation canonique empirique entre les deux familles de variables Y et X. La matrice R^2 est donc l'analogue pour plusieurs équations du coefficient de détermination usuel. L'examen des valeurs propres et des vecteurs propres de cette matrice permet de déterminer les combinaisons linéaires des variables endogènes les plus expliquées par X (celles associées aux grandes valeurs propres) et les combinaisons les moins expliquées (celles associées aux petites valeurs propres).

2.1.4. *Forme de la statistique de Wald généralisée fondée sur la forme explicite*

Contrairement à l'approche du rapport des maxima de (pseudo) vraisemblance, la procédure de type Wald dépend généralement de la forme (explicite, implicite ou mixte) choisie pour exprimer les contraintes (voir Dagenais-Dufour [1991]). Nous allons détailler la procédure pour la forme explicite. La procédure de Wald est construite à partir de l'estimateur non contraint de la matrice B ou de son vectorialisé. Introduisant les vectorialisés : $b = \operatorname{vec}(B')$, $y = \operatorname{vec}(Y)$, $u = \operatorname{vec}(U)$, le modèle initial s'écrit :

$$y = (\mathrm{Id}_n \otimes X)\, b + u, \qquad Vu = \Omega \otimes \mathrm{Id}_T. \tag{6}$$

L'estimateur des m.c.o. non contraint de b est :

$$\hat{b}_T = [\mathrm{Id}_n \otimes (X'X)^{-1} X']\, y. \tag{7}$$

Sa variance est estimée sous l'hypothèse générale par :

$$\hat{V}\, \hat{b}_T = \hat{\Omega}_T \otimes (X'X)^{-1}. \tag{8}$$

L'hypothèse nulle peut aussi être écrite sous une forme explicite vectorialisée :

$$\begin{aligned} H_0^N &= \{\exists \beta : B = \beta \alpha_0'\} \\ &= \{\exists \beta : b = (\mathrm{Id}_n \otimes \alpha_0) \operatorname{vec} \beta'\}. \end{aligned}$$

La statistique de Wald généralisée est alors définie par (voir Szroeter [1983], Gourieroux-Monfort [1989 a]).

$$\xi_W = \underset{\beta}{\mathrm{Min}}[\hat{b}_T - (\mathrm{Id}_n \otimes \alpha_0)\mathrm{vec}\beta']'[\hat{\Omega}_T \otimes (X' X)^{-1}]^{-1}[\hat{b}_T - (\mathrm{Id}_n \otimes \alpha_0)\mathrm{vec}\beta'].$$

Nous établissons dans l'annexe 1 qu'elle est égale à :

$$(9) \qquad \begin{aligned} \xi_W &= \mathrm{T}\, \mathrm{T}r\, [\hat{\Omega}_T^{-1}\, (\hat{\Omega}_{0T} - \hat{\Omega}_T)]. \\ &= \mathrm{T}\, \mathrm{T}r\, \{[Y'\, (\mathrm{Id} - P_X)\, Y]^{-1}\, Y'\, P_{X\,(X'X)^{-1}\gamma_0}\, Y\}. \end{aligned}$$

Introduisons alors les carrés des corrélations canoniques (empiriques non centrées) entre Y et $X\,(X'X)^{-1}\gamma_0$ conditionnellement à $X\alpha_0$, et notons les $\hat{\lambda}_{iT}$. Si on applique le lemme de l'annexe 2 avec les mêmes variables que dans la sous-section 2.1.3., on voit que la statistique de Wald ξ_W s'écrit aussi :

$$(10) \qquad \xi_W = \mathrm{T} \sum_{i=1}^{n} \hat{\lambda}_{iT}/(1 - \hat{\lambda}_{iT}).$$

2.1.5. *Forme de la statistique de Wald généralisée corrigée*

Comme pour la statistique du rapport des maxima de (pseudo) vraisemblance, nous pouvons introduire une statistique de Wald corrigée :

$$(11) \qquad \begin{aligned} \tilde{\xi}_{RMV} &= \mathrm{T}\, \mathrm{T}r\, [\hat{\Omega}_{0T}^{-1}\, (\hat{\Omega}_{0T} - \hat{\Omega}_T)] \\ &= \mathrm{T}\, \mathrm{T}r\, (\mathrm{R}\, [Y,\, X\, (X'\, X)^{-1}\gamma_0/X\alpha_0)] \\ &= \mathrm{T} \sum_{i=1}^{n} \hat{\lambda}_{iT}, \end{aligned}$$

PROPRIÉTÉ 2 : (i) Les statistiques de Wald généralisée et de Wald généralisée corrigée pour tester H_0^N sont respectivement :

$$\xi_W = \mathrm{T} \sum_{i=1}^{n} \hat{\lambda}_{iT}/(1 - \hat{\lambda}_{iT}),$$

et

$$\tilde{\xi}_W = \mathrm{T} \sum_{i=1}^{n} \hat{\lambda}_{iT},$$

où les $\hat{\lambda}_{iT}$ sont les carrés des corrélations canoniques (empiriques et non centrées) entre Y et $X\,(X'X)^{-1}\gamma_0$ conditionnellement à $X\alpha_0$.
(ii) On a les inégalités suivantes :

$$\xi_W \geq \xi_{RMV} \geq \tilde{\xi}_W \geq \tilde{\xi}_{RMV}.$$

(iii) Toutes ces statistiques sont asymptotiquement équivalentes sous l'hypothèse nulle.

Preuve

La partie (i) de la propriété a été montrée.

La partie (ii) résulte de l'inégalité de convexité $x \geq \text{Log}\,(1+x)$, avec $x = \hat{\lambda}_{i\text{T}}/(1-\hat{\lambda}_{i\text{T}})$ pour l'inégalité $\xi_{\text{W}} \geq \xi_{\text{RMV}}$ et $x = \hat{\lambda}_{i\text{T}}$ pour l'inégalité $\tilde{\xi}_{\text{W}} = \xi_{\text{RMV}}$, et de l'inégalité $-\text{Log}\,(1-x) \geq x$ avec $x = \hat{\lambda}_{i\text{T}}$ pour $\xi_{\text{RMV}} = \tilde{\xi}_{\text{W}}$.

La partie (iii) est une conséquence de la convergence de $\hat{\lambda}_{i\text{T}}$ vers zéro sous l'hypothèse nulle ; on a alors les équivalences :

$$\text{Log}\,(1+\hat{\lambda}_{i\text{T}}) \sim -\text{Log}\,(1-\hat{\lambda}_{i\text{T}}) \sim \hat{\lambda}_{i\text{T}} \sim \hat{\lambda}_{i\text{T}}/(1-\hat{\lambda}_{i\text{T}}). \quad \square$$

2.1.6. ***Forme de la statistique de Wald fondée sur la forme implicite***

Dans l'annexe 1, nous établissons le résultat ci-dessous :

PROPRIÉTÉ 3 : La statistique de Wald fondée sur la forme implicite de H_0^{N} coïncide numériquement avec la statistique de Wald généralisée ξ_{W} fondée sur la forme explicite.

La statistique de Wald apparaît ainsi indépendante de la forme retenue pour écrire l'hypothèse. Ceci résulte évidemment de la linéarité des contraintes.

2.1.7. ***Forme de la statistique du score***

Cette statistique est fondée sur la dérivée de la (pseudo) log vraisemblance évaluée en l'estimateur du maximum de (pseudo) vraisemblance contraint.

On montre dans l'annexe 1 que :

$$\xi_{\text{S}} = \text{T}\,\text{T}r\,[\hat{\Omega}_{0\text{T}}^{-1}\,(\hat{\Omega}_{0\text{T}} - \hat{\Omega}_{\text{T}})] = \tilde{\xi}_{\text{W}}.$$

PROPRIÉTÉ 4 : La statistique du score pour tester H_0^{N} est égale à la statistique de Wald corrigée $\xi_{\text{S}} = \tilde{\xi}_{\text{W}}$.

2.1.8. ***Résultats généraux***

Les résultats généraux de l'annexe 1 permettent de trouver le comportement asymptotique sous H_0 des diverses statistiques définies ci-dessus. Il est important de noter que ces résultats ne supposent pas la normalité des u_t.

PROPRIÉTÉ 5 : Les diverses statistiques précédentes ξ_{RMV}, $\tilde{\xi}_{\text{RMV}}$, ξ_{W}, $\tilde{\xi}_{\text{W}} = \xi_{\text{S}}$ sont asymptotiquement équivalentes sous l'hypothèse nulle H_0^{N} et leur loi limite commune est : $\chi^2\,[n\,(\text{K}-r)]$.

Le nombre de degrés de liberté est égal au nombre de contraintes de la forme implicite, c'est-à-dire au nombre d'éléments de la matrice $\text{B}\gamma_0$.

On vérifie qu'il est aussi égal au nombre de paramètres sous l'hypothèse générale nK [nombre d'éléments de B] diminué du nombre de paramètres sous l'hypothèse nulle nr [nombre d'éléments de β].

Par ailleurs, en utilisant les propriétés 2 (ii) et 4 nous pouvons préciser l'ordre entre les diverses statistiques, ordre satisfait à distance finie.

PROPRIÉTÉ 6 : On a les inégalités : $\xi_{\mathrm{W}} \geq \xi_{\mathrm{RMV}} \geq \xi_{\mathrm{S}} = \tilde{\xi}_{\mathrm{W}} \geq \tilde{\xi}_{\mathrm{RMV}}$.

On retrouve ainsi l'ordre entre les statistiques de Wald du score, du rapport de maxima de vraisemblance connu pour les modèles de régression univariés.

2.2. Tests sur l'image de B

De façon analogue nous pouvons développer des procédures de test concernant l'image de la matrice B. Nous n'allons pas détailler toutes les procédures, mais nous allons surtout insister sur le lien avec les procédures concernant le noyau de B.

2.2.1. *Formes de l'hypothèse :*

Une hypothèse sur l'image de B peut, elle aussi, être exprimée sous forme explicite ou sous forme implicite. Désignons par r la dimension de l'image de B, c'est-à-dire le rang de B, et par β_0 une matrice (n, r), de rang r, donnée *a priori*, dont les colonnes engendrent l'image de B. L'hypothèse peut s'écrire, sous forme explicite :

$$\text{(12)} \qquad \begin{aligned} \mathrm{H}_0^{\mathrm{I}} = \{ & \text{il existe une matrice } \alpha, \text{de taille } (\mathrm{K},\ r), \\ & \text{telle que } \mathrm{B} = \beta_0\, \alpha' \}. \end{aligned}$$

Si maintenant nous introduisons une matrice δ_0, de taille $(n, n\text{-}r)$, de rang $n\text{-}r$, dont les colonnes sont orthogonales aux colonnes de β_0 nous voyons que l'hypothèse $\mathrm{H}_0^{\mathrm{I}}$ admet comme forme implicite :

$$\mathrm{H}_0^{\mathrm{I}} = \{\delta_0'\, \mathrm{B} = 0\}.$$

Il est clair qu'il existe une certaine dualité entre les hypothèses concernant le noyau de B et celles concernant l'image de B. En fait cette dualité apparaît directement si nous introduisons la transposée de la matrice B, puisque :

$$\mathrm{H}_0^{\mathrm{I}} = \{\mathrm{B}'\, \delta_0 = 0\} = \{\exists \alpha : \mathrm{B}' = \alpha\, \beta_0'\}.$$

Ainsi, les matrices δ_0 et β_0 apparaissent comme les analogues des matrices γ_0 et α_0 introduites dans le paragraphe précédent.

2.2.2. *Forme de la statistique du rapport des maxima de (pseudo) vraisemblance*

Cette statistique est facile à expliciter à condition de commencer par mettre le modèle initial sous la forme récursive équivalente :

$$\text{(13)} \qquad \begin{cases} \delta_0'\, \mathrm{Y}_t = \mathrm{B}_1\, \mathrm{X}_t + v_{1t}, \\ \beta_0'\, \mathrm{Y}_t = \mathrm{B}_2\, \mathrm{X}_t + \mathrm{C}\delta_0'\, \mathrm{Y}_t + v_{2t}, \end{cases}$$

où B_1, B_2, C sont des matrices de tailles respectives (n-r, K), (r, K), (r, n-r) et où v_{1t}, v_{2t} sont deux vecteurs d'erreurs non corrélés, de matrices de variance-covariance respectives Λ_{11}, Λ_{22}. Avec ce nouveau paramétrage l'hypothèse nulle est simplement $H_0^I = \{B_1 = 0\}$. Comme par ailleurs la forme récursive conduit à une décomposition additive de la (pseudo) log-vraisemblance, nous en déduisons que la statistique de test du rapport des maxima de (pseudo) vraisemblance de H_0^I coïncide avec celle du test de l'hypothèse $(B_1 = 0)$ dans le cadre du seul sous-modèle :

$$\delta_0' Y_t = B_1 X_t + V_{1t}, \; V(V_{1t}) = \Lambda_{11}.$$

Cette statistique est donc donnée par :

$$(14) \quad \begin{cases} \xi_{RMV}^{\star} = T \operatorname{Log} \dfrac{\det \dfrac{1}{T} \delta_0' Y' Y \delta_0}{\det \left(\dfrac{1}{T} \delta_0' Y' (\mathrm{Id} - P_X) Y \delta_0 \right)}, \\ \xi_{RMV}^{\star} = T \operatorname{Log} \det [\mathrm{Id} + (\delta_0' Y' (\mathrm{Id} - P_X) Y \delta_0)^{-1} \delta_0' Y' P_X Y \delta_0]. \end{cases}$$

Elle peut facilement s'exprimer en terme de valeurs propres. Désignons par $\hat{\lambda}_{1T}^{\star} \geq \cdots \geq \hat{\lambda}_{n-r, T}^{\star} \geq 0$ les carrés des corrélations canoniques (empiriques non centrées) entre $Y\delta_0$ et X.

En appliquant le lemme de l'annexe 2, avec $U = Y\delta_0$, $P = P_X$, $P_2 = 0$, $P_1 = P_X$, on obtient une autre forme pour la statistique du rapport des maxima de (pseudo) vraisemblance $\xi_{RMV}^{\star}$:

$$(15) \quad \xi_{RMV}^{\star} = -T \sum_{i=1}^{n-r} \operatorname{Log} (1 - \hat{\lambda}_{iT}^{\star}).$$

2.2.3. *Statistique du rapport des maxima de (pseudo) vraisemblance corrigée*

Comme dans le cas précédent, nous pouvons remplacer dans l'expression finale de la statistique $\xi_{RMV}^{\star}$ l'estimateur non contraint de la variance Λ_{11} par l'estimateur contraint. Nous obtenons la statistique asymptotiquement équivalente sous l'hypothèse nulle :

$$(16) \quad \begin{aligned} \tilde{\xi}_{RMV}^{\star} &= T \operatorname{Log} \det (\mathrm{Id} + (\delta_0' Y' Y \delta_0)^{-1} \delta_0' Y' P_X Y \delta_0) \\ &= T \operatorname{Log} \det (\mathrm{Id} + R^2 (Y\delta_0, X)) \\ &= T \sum_{i=1}^{n-r} \operatorname{Log} (1 + \hat{\lambda}_{iT}^{\star}), \end{aligned}$$

Les résultats sont rassemblés dans la propriété ci-dessous.

PROPRIÉTÉ 7 : Soient $\lambda^*_{1\mathrm{T}} \geq \cdots \geq \lambda^*_{n-r,\,\mathrm{T}} \geq 0$ les carrés des corrélations canoniques (non centrées empiriques) entre $\mathrm{Y}\delta_0$ et X, la statistique du rapport des maxima de (pseudo) vraisemblance est égale à :

$$\tilde{\xi}^*_{\mathrm{RMV}} = -\mathrm{T} \sum_{i=1}^{n-r} \mathrm{Log}\,(1 - \lambda^*_{i\mathrm{T}})\,;$$

la statistique corrigée est : $\tilde{\xi}^*_{\mathrm{RMV}} = \mathrm{T} \sum_{i=1}^{n-r} \mathrm{Log}\,(1 + \lambda^*_{i\mathrm{T}})$.

2.2.4. *Les statistiques de Wald et la statistique du score*

En poursuivant la démarche, nous pourrions aussi expliciter les autres statistiques de test :

PROPRIÉTÉ 8 : (i) La statistique de Wald, indépendante de la forme implicite ou explicite retenue pour l'hypothèse nulle, est :

$$\xi^*_{\mathrm{W}} = -\mathrm{T} \sum_{i=1}^{n-r} \lambda^*_{i\mathrm{T}} / \,(1 - \lambda^*_{i\mathrm{T}}).$$

.

(ii) La statistique de Wald corrigée coïncide avec la statistique du score et est donnée par :

$$\begin{aligned}\tilde{\xi}^*_{\mathrm{W}} = \xi^*_{\mathrm{S}} &= \mathrm{T}\,\mathrm{T}r\,[\mathrm{R}^2\,(\mathrm{Y}\delta_0,\,\mathrm{X})].\\ &= \mathrm{T} \sum_{i=1}^{n-r} \lambda^*_{i\mathrm{T}}.\end{aligned}$$

(iii) Ces diverses statistiques sont asymptotiquement équivalentes sous l'hypothèse nulle et admettent comme loi limite $\chi^2\,(\mathrm{K}\,(n - r))$.
(iv) On a les inégalités :

$$\xi^*_{\mathrm{W}} \geq \xi^*_{\mathrm{RMV}} \geq \xi^*_{\mathrm{S}} = \tilde{\xi}^*_{\mathrm{W}} \geq \tilde{\xi}^*_{\mathrm{RMV}}.$$

3 Procédures ex-post

3.1. Hypothèses et paramètres

Dans ce chapitre nous nous intéressons à la détermination du rang de la matrice B, de son image et de son noyau. Le rang étant un paramètre à valeurs entières, l'estimation de ce paramètre conduit à introduire des procédures analogues à celles utilisées pour l'identification des ordres dans le cas de processus ARMA ; dans la suite nous considérerons des procédures de tests séquentielles concernant ce rang. Les hypothèses successives sont :

(17) $$\mathrm{H}_{0r} = \{\mathrm{Rang\,B} = r\},$$
et

(18) $$\mathrm{H}_r = \{\mathrm{Rang\,B} \geq r\},$$

et nous nous proposons de tester H_{0r} contre $\mathrm{H}_{r+1} = \mathrm{H}_r - \mathrm{H}_{0r}$. Cette suite de tests doit être analysée dans un ordre naturel ; celui-ci est le suivant :

(*) Tester l'hypothèse $\mathrm{H}_{0,\,0} = \{\mathrm{rang\,B} = 0\}$ contre $\mathrm{H}_1 = \{\mathrm{rang\,B} \geq 1\}$.

Si l'hypothèse H_{00} est acceptée, on arrête la procédure et on décide, $\mathrm{Rang\,B} = 0$.

(**) Sinon, on teste, $\mathrm{H}_{0,\,1}$ contre H_2. Si l'hypothèse $\mathrm{H}_{0,\,1}$ est acceptée, la procédure s'arrête et on décide $\mathrm{Rang\,B} = 1$.

Et ainsi de suite...

Une hypothèse sur le rang peut être vue comme une condition sur l'image (car $\dim \mathrm{ImB} = \mathrm{Rang\,B}$) ou sur le noyau (car $\mathrm{Rang\,B} = \mathrm{K} - \dim \mathrm{KerB}$). De plus, les écritures des hypothèses peuvent être données sous des formes explicites ou mixtes. Nous donnons ci-dessous les formes de contraintes caractérisant H_{0r} dans l'hypothèse générale H_r.

Forme explicite :

(19) $\mathrm{H}_{0r} =$ {il existe deux matrices α, β de tailles respectives (K, r) et (n, r) telles que $\mathrm{B} = \beta\alpha'\} \cap \mathrm{H}_r$.
La condition signifie que le rang de B est inférieur ou égal à r et dans le cadre de H_r, nous savons que α et β sont nécessairement de rang r.

Forme mixte à partir du noyau :

(20) $\mathrm{H}_{0r} =$ {il existe une matrice γ, de taille (K, K-r), de rang K-r telle que $\mathrm{B}\gamma = 0\} \cap \mathrm{H}r$.
Dans le cadre de H_r, on ne peut pas trouver une matrice $\tilde{\gamma}$, de taille (K, K$-\tilde{r}$), de rang K$-\tilde{r}$ telle que $\mathrm{B}\tilde{\gamma} = 0$, avec $\tilde{r} < r$.

Forme mixte à partir de l'image :

(21) $\mathrm{H}_{0r} =$ {il existe une matrice δ, de taille (n, n-r), de rang n-r, telle que $\delta'\mathrm{B} = 0\}$.

L'hypothèse H_{0r} peut donc *a priori* être testée à partir de l'une quelconque des formulations précédentes. À ce stade il est important de noter que les procédures de tests vont généralement nécessiter la détermination d'estimateurs contraints par H_{0r} des paramètres auxiliaires α (i.e. de l'image de B' égale à l'orthogonal du noyau de B), β (i.e. de l'image de B), γ (i.e. du noyau de B), et δ (i.e. du noyau de B' égal à l'orthogonal de l'image de B). Les procédures d'estimation seront développées en liaison avec les procédures de test : estimation par maximum de (pseudo) vraisemblance dans le cas du test du rapport des maxima de (pseudo) vraisemblance, estimation par moindres carrés asymptotiques dans le cas du test de Wald.

3.2. Estimation et test par maximum de (pseudo) vraisemblance

3.2.1. *Forme de la statistique*

La statistique du rapport des maxima de (pseudo) vraisemblance a une forme indépendante de l'écriture explicite ou mixte retenue pour l'hypothèse nulle. Elle est donnée par :

$$\xi^r_{\mathrm{RMV}} = \mathrm{T\,Log}\,\frac{\det \hat{\Omega}^r_{0\mathrm{T}}}{\det \hat{\Omega}_{\mathrm{T}}}, \tag{22}$$

où $\hat{\Omega}^r_{0\mathrm{T}}$ est la matrice de variance-covariance résiduelle évaluée sous l'hypothèse (RangB $= r$). Cette statistique peut être calculée à partir des statistiques de test analogues portant sur le noyau et sur l'image. Notons :

$$\begin{aligned}\xi_{\mathrm{RMV}}(\gamma) = &\mathrm{T\,Log\,det}\,[\mathrm{Id} + \{\mathrm{Y}'(\mathrm{Id}-\mathrm{P_X})\,\mathrm{Y}\}^{-1}\,\mathrm{Y}'\,\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\gamma\\ &\times[\gamma'\,(\mathrm{X}'\mathrm{X})^{-1}\gamma]^{-1}\,\gamma'\,(\mathrm{X}'\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}],\end{aligned}$$

la statistique obtenue en (5),

$$\xi^{\star}_{\mathrm{RMV}}(\delta) = \mathrm{T\,Log\,det}\,[\mathrm{Id} + \{\delta'\,\mathrm{Y}\,[\mathrm{Id}-\mathrm{P_X})\,\mathrm{Y}\delta\}^{-1}\,\delta'\,\mathrm{Y}'\,\mathrm{P_X}\,\mathrm{Y}\delta],$$

la statistique obtenue en (14). Nous déduisons alors de la définition de la statistique du rapport des maxima de (pseudo) vraisemblance la propriété suivante.

PROPRIÉTÉ 9 : i) $\xi^{r}_{\mathrm{RMV}} = \underset{\gamma}{\mathrm{Min}}\ \xi_{\mathrm{RMV}}(\gamma) = \underset{\delta}{\mathrm{Min}}\ \xi^{\star}_{\mathrm{RMV}}(\delta)$, où la première optimisation est menée sur les matrices γ, de taille (K, K-r) (de rang (K-r)) et la seconde optimisation sur les matrices δ, de taille (n, n-r) (de rang (n-r)).
ii) Toute solution $\hat{\gamma}_0$ [resp $\hat{\delta}_0$] du premier [resp second] problème de minimisation est un estimateur du (pseudo) maximum de vraisemblance contraint d'une base du noyau de B [resp de l'orthogonal de l'image de B].

3.2.2. *La statistique et les estimateurs*

Il suffit donc de résoudre les deux problèmes de minimisation précédents, ce qui est fait dans l'annexe 3. Cette résolution s'appuie sur l'analyse des corrélations canoniques entre X et Y et avant de donner les résultats, il nous faut d'abord expliciter celles-ci.

Considérons les deux matrices de tailles respectives (K, K) et (n, n) :

$$\mathrm{R}^2(\mathrm{X},\mathrm{Y}) = (\mathrm{X}'\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}\,(\mathrm{Y}'\mathrm{Y})^{-1}\,\mathrm{Y}'\,\mathrm{X},$$

et

$$\mathrm{R}^2(\mathrm{Y},\mathrm{X}) = (\mathrm{Y}'\mathrm{Y})^{-1}\,\mathrm{Y}'\,\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}.$$

On sait qu'à des valeurs propres nulles près ces deux matrices ont mêmes valeurs propres et que celles-ci sont réelles positives inférieures à un (voir par exemple Rao (1973]). Nous les notons par ordre décroissant :

$$\hat{\eta}_{1\mathrm{T}} \geq \hat{\eta}_{2\mathrm{T}} \geq \cdots$$

Ces valeurs propres sont les carrés des corrélations canoniques (empiriques et non centrées) entre Y et X.

Par ailleurs, nous pouvons expliciter des bases de vecteurs propres associées. Nous notons $\hat{e}_{1\mathrm{T}}, \cdots, \hat{e}_{\mathrm{KT}}$ une base relative à $\mathrm{R}^2(\mathrm{X}, \mathrm{Y})$, que nous pouvons choisir normalisée par :

$$\hat{e}'_{k\mathrm{T}}\, \mathrm{X}'\, \mathrm{X}\, \hat{e}_{l\mathrm{T}} = \delta_{kl}, \qquad k, l = 1, \cdots, \mathrm{K}.$$

De façon analogue nous notons $\hat{f}_{1\mathrm{T}}, \cdots, \hat{f}_{n\mathrm{T}}$ une base relative à $\mathrm{R}^2(\mathrm{Y}, \mathrm{X})$, que nous pouvons choisir normalisée par :

$$\hat{f}'_{k\mathrm{T}}\, \mathrm{Y}'\, \mathrm{Y}\, \hat{f}_{l\mathrm{T}} = \delta_{kl}, \qquad k, l = 1, \cdots, n.$$

À partir des résultats de l'annexe 3 nous pouvons énoncer la propriété suivante :

PROPRIÉTÉ 10 :

(i)
$$\begin{aligned} \xi^r_{\mathrm{RMV}} &= -\mathrm{T} \sum_{i=r+1}^{n} \mathrm{Log}\,(1 - \hat{\eta}_{i\mathrm{T}}) \\ &= -\mathrm{T} \sum_{i=r+1}^{k} \mathrm{Log}\,(1 - \hat{\eta}_{i\mathrm{T}}). \end{aligned}$$

(ii) Une base α de l'orthogonal du noyau de B peut être estimée par :

$$\hat{\alpha}_{1\mathrm{T}} = \hat{e}_{1\mathrm{T}}, \cdots, \hat{\alpha}_{r\mathrm{T}} = \hat{e}_{r\mathrm{T}}.$$

(iii) Une base δ de l'orthogonal de l'image de B peut être estimée par :

$$\hat{f}_{r+1,\,\mathrm{T}}, \cdots, \hat{f}_{n,\,\mathrm{T}}.$$

(iv) Une base γ du noyau de B peut être estimée par :

$$(\mathrm{X}'\mathrm{X})\, \hat{e}_{r+1,\,\mathrm{T}'}, \cdots, (\mathrm{X}'\mathrm{X})\, \hat{e}_{k,\,\mathrm{T}}.$$

(v) Une base β de l'image de B peut être estimée par :

$$(\mathrm{Y}'\mathrm{Y})\, \hat{f}_{1,\,\mathrm{T}}, \cdots, (\mathrm{Y}'\mathrm{Y})\, \hat{f}_{r,\,\mathrm{T}}.$$

3.2.3. *La symétrie de la procédure*

La propriété précédente 10 montre clairement une symétrie en X et Y de la procédure. En effet il revient au même de tester :

$$\mathrm{H}_{0r} = \{\mathrm{Rang}\,\mathrm{B} = r\},$$

à partir du modèle : $\mathrm{Y}_t = \mathrm{BX}_t + u_t,\ t = 1, \cdots, \mathrm{T}$
ou de tester : $\tilde{\mathrm{H}}_{0r} = \{\mathrm{Rang}\,\mathrm{D} = r\}$,
à partir du modèle de régression empilées :

$$\mathrm{X}_t = \mathrm{DY}_t + \mathrm{W}_t, \qquad t = 1, \cdots, \mathrm{T},$$

où les variables explicatives sont régressées sur les variables expliquées. Ceci se comprend aisément, puisque, dans le cas où $(\mathrm{X}_t, \mathrm{Y}_t)$ est stationnaire :

$$\begin{aligned}\mathrm{D} &= \mathrm{E}\,(\mathrm{XY}')\,(\mathrm{E}\,(\mathrm{YY}'))^{-1} = \mathrm{E}\,(\mathrm{XX}')\,(\mathrm{E}\,(\mathrm{XX}'))^{-1}\,\mathrm{E}\,(\mathrm{XY}')\,(\mathrm{E}\,(\mathrm{YY}')^{-1}\\ &= \mathrm{E}\,(\mathrm{XX}')\,\mathrm{B}'\,[\mathrm{E}\,(\mathrm{YY}')]^{-1},\end{aligned}$$

et donc que $\mathrm{rang D} = \mathrm{rang\ B}'$.

3.3. Test de Wald et estimation par les moindres carrés asymptotiques

(i) Forme du test et des estimateurs.

Utilisant des notations similaires à celles du paragraphe 2.1.4., la statistique de Wald fondée sur la forme mixte est :

$$\text{(23)} \qquad \begin{aligned}\xi_{\mathrm{W}}^{r} = \underset{\alpha,\,\beta}{\mathrm{Min}}\,&(\hat{b}_{\mathrm{T}} - (\mathrm{Id}_n \otimes \alpha)\,\mathrm{vec}\,\beta')'\,[\hat{\Omega}_{\mathrm{T}} \otimes (\mathrm{X}'\mathrm{X})^{-1}]^{-1}\\ &(\hat{b}_{\mathrm{T}} - (\mathrm{Id}_n \otimes \alpha)\,\mathrm{vec}\,\beta'),\end{aligned}$$

où α et β sont des matrices de tailles (K, r) et (n, r).

Cette maximisation diffère de celle du paragraphe 2.1.4., car elle est menée à la fois par rapport à α et β et non pas seulement par rapport à α.

Une solution $\tilde{\alpha}_0, \tilde{\beta}_0$ de ce problème fournit un estimateur par les moindres carrés asymptotiques de bases α et β.

La maximisation en β a déjà été effectuée dans le paragraphe 2.1.4. de sorte que nous avons d'après (9) et en notant $\xi_{\mathrm{W}}(\alpha)$ la statisique concentrée en β :

$$\begin{aligned}\xi_{\mathrm{W}}^{r} = \underset{\alpha}{\mathrm{Min}}\ \xi_{\mathrm{W}}(\alpha) &= \underset{\alpha}{\mathrm{Min}}\ \mathrm{T}\,\mathrm{T}r\,[\hat{\Omega}_{\mathrm{T}}^{-1}\,(\hat{\Omega}_{0\mathrm{T}}(\alpha) - \hat{\Omega}_{\mathrm{T}})]\\ &= \underset{\alpha}{\mathrm{Min}}\,\{\mathrm{T}r\,(\hat{\Omega}_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{P_X}\,\mathrm{Y}) - \mathrm{T}r\,(\hat{\Omega}_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{P}_{\mathrm{X}\alpha}\,\mathrm{Y})\}\\ &= \mathrm{T}r\,\{\hat{\Omega}_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{P_X}\,\mathrm{Y}\} - \underset{\alpha}{\mathrm{Max}}\ \mathrm{T}r\,\{\hat{\Omega}_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{P}_{\mathrm{X}\alpha}\,\mathrm{Y}\}.\end{aligned}$$

Ce problème d'optimisation est résolu dans l'annexe 4 et nous obtenons le résultat suivant :

PROPRIÉTÉ 11 : (i) La statistique de Wald a pour expression :

$$\xi_{\mathrm{W}}^{r} = \mathrm{T} \sum_{i=r+1}^{\mathrm{K}} \frac{\hat{\eta}_{i\mathrm{T}}}{1 - \hat{\eta}_{i\mathrm{T}}}.$$

(ii) Les estimateurs des moindres carrés asymptotiques de α et β coïncident avec les estimateurs du maximum de (pseudo) vraisemblance.

(ii) Comparaison des statistiques de Wald et du rapport des maxima de (pseudo) vraisemblance.

PROPRIÉTÉ 12 : (i) La statistique de Wald prend des valeurs plus grandes que la statistique du rapport des maxima de (pseudo) vraisemblance :

$$\xi_{\mathrm{W}}^{r} \geq \xi_{\mathrm{RMV}}^{r}.$$

(ii) Les deux statistiques sont asymptotiquement équivalentes sous l'hypothèse nulle et leur loi asymptotique commune est :

$$\chi^2 \, [(n - r) \, (\mathrm{K} - r)].$$

Preuve : (i) L'inégalité résulte de :

$$-\mathrm{Log}\,(1 - x) = \mathrm{Log}\left(1 + \frac{x}{1 - x}\right) < \frac{x}{1 - x}.$$

(ii) L'équivalence asymptotique se voit immédiatement puisque sous l'hypothèse nulle $\hat{\eta}_{i\mathrm{T}}$ tend vers zéro à la vitesse 1/T, pour $i = r + 1, \cdots, \mathrm{K}$ (voir Anderson-Rubin [1956]). On a alors par développement limité :

$$\xi_{\mathrm{W}}^{r} = \mathrm{T} \sum_{i=r+1}^{\mathrm{K}} \hat{\eta}_{i\mathrm{T}} + 0p\,(1),$$

et

$$\xi_{\mathrm{RMV}}^{r} = \mathrm{T} \sum_{i=r+1}^{\mathrm{K}} \hat{\eta}_{i\mathrm{T}} + 0p\,(1).$$

La loi limite est une conséquence des résultats de l'annexe 3 ; rappelons en particulier que l'hypothèse de normalité des erreurs n'est pas nécessaire. Le nombre de degrés de liberté peut par exemple se calculer à partir de l'écriture explicite : nombre de paramètres initiaux (B)-nombre de paramètres sous H_{0r} (α et β) + nombre de contraintes identifiantes (α et β sont définis à une matrice inversible (r, r) près) soit $n\mathrm{K} - \mathrm{K}r - nr + r^2 = (n - r)\,(\mathrm{K} - r)$ □.

4 Généralisation

Dans cet article nous avons considéré les problèmes de tests concernant le noyau, l'image et le rang de la matrice B intervenant dans le modèle linéaire

$Y_t = BX_t + u_t$. Les résultats obtenus peuvent être facilement généralisés au cas où il y a d'autres variables Z_t. Le modèle s'écrit alors :

$$Y_t = BX_t + CZ_t + u_t.$$

On s'intéresse toujours à la matrice B. Les diverses statistiques de test sur le noyau s'expriment de la même façon en fonction des résidus des m.c.o. dans les modèles :

$$\tilde{y}_i = Xb_i + Z\,c_i + \tilde{u}_i$$
$$\tilde{y}_i = X\alpha_0\,\beta_i + Z\,c_i + \tilde{u}_i, \qquad i = 1, \cdots, n.$$

où $\tilde{y}_i$ et $\tilde{u}_i$ sont les vecteurs de taille T formé par les ièmes composantes des Y_t et u_t et b_i, c_i, β_i sont les ièmes lignes de B, C, β.

Les statistiques de test sur l'image s'expriment aussi de la même façon en fonction des résidus des m.c.o. dans les modèles :

$$Y\delta_{0j} = Xd_j + Z\,c_j + v_j$$

et

$$Y\delta_{0j} = Xc_j + v_j$$

où δ_{0j} est la j-ième colonne de δ_0, $j = 1, \cdots, n - r$ et d_j la j-ième ligne de B_1 (voir 13).

Or, ces résidus peuvent aussi être obtenus à partir des régressions précédentes dans lesquelles les termes en Z sont supprimés, la matrice X est remplacée par $(\mathrm{Id} - P_Z)\,X$ et la matrice Y par $(\mathrm{Id} - P_Z)\,Y$ (les colonnes de Y, i.e. les $\tilde{y}_i$, étant donc remplacées par $(\mathrm{Id} - P_Z)\,\tilde{y}_i$. Ce résultat est une conséquence du théorème de Frisch et Waugh.

Autrement dit toutes les formules du paragraphe 2 restent valables à condition de remplacer Y par $(\mathrm{Id} - P_Z)\,Y$ et X par $(\mathrm{Id} - P_Z)\,X$. Les résultats du paragraphe 3 fondés sur ceux du paragraphe 2 restent aussi valables.

Tests du rapport de (pseudo) vraisemblance, de Wald et du score, dans un modèle de régressions empilées

(i) **Le modèle non contraint**

On considère le modèle suivant :

$$Y_t = BX_t + u_t \qquad t = 1, \cdots, T,$$

où Y_t est un vecteur de taille n, B une matrice de taille nxK, X_t un vecteur de taille K. Les vecteurs u_t, $t = 1, \cdots$, T sont supposés indépendants, de même loi, de matrice de variance-covariance Ω ; on suppose de plus que u_t est indépendant de X_t, $X_{t-1} \cdots$

La pseudo log-vraisemblance (conditionnelle) associée à la normalité des erreurs u_t est :

$$L_T = \sum_{t=1}^{T} l_t,$$

avec :

$$\begin{aligned} l_t &= -\frac{n}{2} \operatorname{Log} 2\pi - \frac{1}{2} \operatorname{Log} \det \Omega \\ &\quad - \frac{1}{2} (Y_t - BX_t)' \, \Omega^{-1} \, (Y_t - BX_t) \\ &= -\frac{n}{2} \operatorname{Log} 2\pi - \frac{1}{2} \operatorname{Log} \det \Omega \\ &\quad - \frac{1}{2} [Y_t - (\mathrm{Id}_n \otimes X_t')\, b]' \, \Omega^{-1} \, [Y_t - (\mathrm{Id}_n \otimes X_t')\, b], \end{aligned}$$

avec $b = \operatorname{vec}(B')$.

Le modèle s'écrit aussi : $y = (\mathrm{Id}_n \otimes X)\, b + u$ avec $y = \operatorname{vec} Y$, Y étant la matrice (T, n) dont les lignes sont les Y_t', $t = 1, \cdots$, T, $u = \operatorname{vec} U$, U étant la matrice (T, n) dont les lignes sont les u_t', X est la matrice (T, K) dont les lignes sont les X_t', $t = 1, \cdots$, T, ces diverses matrices vérifiant $Y' = BX' + U'$.

On a $V(u) = \Omega \otimes \mathrm{Id}_T$ et la pseudo log-vraisemblance L_T peut encore s'écrire :

$$\begin{aligned} L_T = &- \frac{nT}{2} \operatorname{Log} 2\pi - \frac{T}{2} \operatorname{Log} \det \Omega \\ &- \frac{1}{2} \{[y - (\mathrm{Id}_n \otimes X)\, b]' \, (\Omega^{-1} \otimes \mathrm{Id}_T)[y - (\mathrm{Id}_n \otimes X)\, b]\}. \end{aligned}$$

L'estimateur du pseudo maximum de vraisemblance de b, coïncide avec l'estimateur des m.c.o. équation par équation :

$$\hat{b}_T = [\mathrm{Id}_n \otimes (X'X)^{-1} X']\, y = \operatorname{vec} \hat{B}_T',$$

avec :

$$\hat{B}_T' = (X'X)^{-1} X' Y.$$

L'estimateur de Ω est :

$$\hat{\Omega}_T = \frac{1}{T} \hat{U}' \hat{U}, \qquad \text{avec} \qquad \hat{U} = Y' - \hat{B}_T X'.$$

La matrice de variance-covariance asymptotique de $\sqrt{T}(\hat{b}_T - b)$ est estimée de manière convergente par :

$$V_{as} \sqrt{T}(\hat{b}_T - b) = T \hat{\Omega}_T \otimes (X'X)^{-1}.$$

(ii) **Test du rapport de pseudo vraisemblance pour une hypothèse sur B.**

On désire tester une hypothèse nulle quelconque : **linéaire ou non linaire, sous une forme implicite, explicite ou mixte** portant uniquement sur B. Comme Ω est non contrainte sous l'hypothèse générale et sous H_0, la maximisation de L_T peut commencer, sous les deux hypothèses, par une concentration en Ω. On a :

$$\begin{aligned} L_T &= -\frac{nT}{2} \operatorname{Log} 2\pi - \frac{T}{2} \operatorname{Log} \det \Omega \\ &\quad - \frac{1}{2} \sum_{t=1}^{T} (Y_t - BX_t)' \Omega^{-1} (Y_t - BX_t) \\ &= -\frac{nT}{2} \operatorname{Log} 2\pi + \frac{T}{2} \operatorname{Log} \det \Omega^{-1} \\ &\quad - \frac{1}{2} \operatorname{Trace} [\Omega^{-1} \sum_{t=1}^{T} (Y_t - BX_t)(Y_t - BX_t)']. \end{aligned}$$

En dérivant par rapport aux éléments de Ω^{-1}, on obtient :

$$\frac{\partial L_T}{\partial \Omega^{-1}} = \frac{1}{2} \left[T\Omega - \sum_{t=1}^{T} (Y_t - BX_t)(Y_t - BX_t)' \right].$$

En annulant ces dérivées, on en déduit la relation :

$$\Omega = \frac{1}{T} \sum_{t=1}^{T} (Y_t - BX_t)(Y_t - BX_t)',$$

et la pseudo log-vraisemblance concentrée est :

$$L_T^C = -\frac{nT}{2} \operatorname{Log} 2\pi - \frac{T}{2} \operatorname{Log} \det \left[\frac{1}{T} \sum_{t=1}^{T} (Y_t - BX_t)(Y_t - BX_t)' \right] - \frac{1}{2} Tn.$$

Si on note $\hat{B}_{0T}$ l'estimateur du pseudo maximum de vraisemblance sous H_0,

$$\hat{u}_t^0 = Y_t - \hat{B}_0 X_t \text{ et } \hat{\Omega}_{0T} = \frac{1}{T} \sum_{t=1}^{T} (Y_t - \hat{B}_{0T} X_t)(Y_t - \hat{B}_{0T} X_t)',$$

on voit immédiatement que la statistique du rapport de pseudo-vraisemblance est :

$$\xi_{\mathrm{RMV}} = \mathrm{T}\ \mathrm{Log}\ \frac{\det \hat{\Omega}_{0\mathrm{T}}}{\det \hat{\Omega}_{\mathrm{T}}}.$$

Montrons maintenant que cette statistique a pour loi asymptotique sous H_0 une loi du χ^2, dont le nombre de degrés de liberté est égal au nombre de contraintes. Notons $\theta_1 = b$, $\theta_2 = \mathrm{vec}\,\Omega^{-1}$, $\theta' = (\theta_1', \theta_2')$, $\hat{\theta}_{\mathrm{T}}$, $\hat{\theta}_{0\mathrm{T}}$ les estimateurs du PMV de θ sous l'hypothèse générale et sous H_0,

$$\mathrm{J} = -\mathrm{plim}\ \frac{1}{\mathrm{T}} \frac{\partial^2 \mathrm{L}_{\mathrm{T}}}{\partial\theta\partial\theta'} = -\mathrm{E}\ \frac{\partial^2 l_t}{\partial\theta\partial\theta'}$$

et

$$\mathrm{I} = \mathrm{plim}\ \frac{1}{\mathrm{T}} \sum_{t=1}^{\mathrm{T}} \frac{\partial\, l_t}{\partial\theta} \frac{\partial\, l_t}{\partial\theta'} = \mathrm{E}\ \frac{\partial\, l_t}{\partial\theta} \frac{\partial\, l_t}{\partial\theta'}.$$

On sait (Gourieroux-Monfort [1989] proposition 3) que la propriété sera vraie si J est une inverse généralisée de la matrice de variance-covariance asymptotique de $\sqrt{\mathrm{T}}\,(\hat{\theta}_{\mathrm{T}} - \hat{\theta}_{0\mathrm{T}})$. Comme ξ_{RMV} est indépendante de la forme de H_0, on peut supposer que celle-ci a une forme implicite : $g(\theta_1)=0$; alors la matrice de variance-covariance asymptotique de $\sqrt{\mathrm{T}}\,(\hat{\theta}_{\mathrm{T}} - \hat{\theta}_{0\mathrm{T}})$ est :

$$\mathrm{V} = \mathrm{J}^{-1} \frac{\partial g'}{\partial\theta} \left[\frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1} \frac{\partial g'}{\partial\theta}\right]^{-1} \frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1}\, \mathrm{I}\mathrm{J}^{-1} \frac{\partial g'}{\partial\theta} \left[\frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1} \frac{\partial g'}{\partial\theta}\right]^{-1} \frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1}.$$

Dans Gourieroux-Monfort [1989] on montre que J est une inverse généralisée de V, lorsque V s'écrit :

$$\mathrm{J}^{-1} \frac{\partial g'}{\partial\theta} \left[\frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1} \frac{\partial g'}{\partial\theta}\right]^{-1} \frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1},$$

c'est-à-dire lorsque :

$$\frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1}\, \mathrm{I}\, \mathrm{J}^{-1} \frac{\partial g'}{\partial\theta} = \frac{\partial g}{\partial\theta'}\, \mathrm{J}^{-1} \frac{\partial g'}{\partial\theta}.$$

Montrons que c'est bien le cas ici.

On a :

$$\frac{\partial l_t}{\partial b} = -(\mathrm{Id}_n \otimes \mathrm{X}_t)\, \Omega^{-1}\, [\mathrm{Y}_t - (\mathrm{Id}_n \otimes \mathrm{X}_t')\, b] = (\mathrm{Id}_n \otimes \mathrm{X}_t)\, \Omega^{-1}\, u_t,$$

$$\frac{\partial^2 l_t}{\partial b \partial b'} = (\mathrm{Id}_n \otimes \mathrm{X}_t)\, \Omega^{-1}\, (\mathrm{Id}_n \otimes \mathrm{X}_t'),$$

et donc

$$\begin{aligned} \mathrm{I}_{11} &= \mathrm{E}\ \frac{\partial\, l_t}{\partial b} \frac{\partial\, l_t}{\partial b'} \\ &= \mathrm{E}\, [(\mathrm{Id}_n \otimes \mathrm{X})\, \Omega^{-1}\, (\mathrm{Id}_n \otimes \mathrm{X}') \\ &= -\mathrm{E}\ \frac{\partial^2 l_t}{\partial b \partial b'} \\ &= \mathrm{J}_{11}. \end{aligned}$$

De plus, comme $\mathrm{E}u_t = 0$, on a $\mathrm{J}_{12} = \mathrm{E}\,\dfrac{\partial^2 l_t}{\partial\theta_1\,\partial\theta_2} = 0$.

Les propriétés $\mathrm{I}_{11} = \mathrm{J}_{11}$, $\mathrm{J}_{12} = 0$ jointes à l'égalité $\dfrac{\partial g}{\partial\theta'} = \left(\dfrac{\partial g}{\partial\theta'_1}, 0\right)$, impliquent bien : $\dfrac{\partial g}{\partial\theta'}\,\mathrm{J}^{-1}\mathrm{I}\,\mathrm{J}^{-1}\,\dfrac{\partial g'}{\partial\theta} = \dfrac{\partial g}{\partial\theta'}\,\mathrm{J}^{-1}\,\dfrac{\partial g'}{\partial\theta}$.

(iii) **Test du score**

Compte-tenu des égalités $\mathrm{I}_{11} = \mathrm{J}_{11}$, $\mathrm{J}_{12} = 0$, la statistique de test du score pour une hypothèse nulle quelconque portant sur b, s'écrit :

$$\xi_\mathrm{S} = -\left(\frac{\partial \mathrm{L_T}}{\partial b'}\right)_{\mathrm{H}_0}\left(\frac{\partial^2 \mathrm{L_T}}{\partial b\,\partial b'}\right)_{\mathrm{H}_0}^{-1}\left(\frac{\partial \mathrm{L_T}}{\partial b}\right)_{\mathrm{H}_0},$$

avec

$$\left(\frac{\partial \mathrm{L_T}}{\partial b}\right)_{\mathrm{H}_0} = (\mathrm{Id}_n \otimes \mathrm{X}')\,(\hat{\Omega}_{0\mathrm{T}}^{-1} \otimes \mathrm{Id_T})\,[y - (\mathrm{Id}_n \otimes \mathrm{X})\,\hat{b}_{0\mathrm{T}}],$$

$$\left(\frac{\partial^2 \mathrm{L_T}}{\partial b\partial b'}\right)_{\mathrm{H}_0} = -(\mathrm{Id}_n \otimes \mathrm{X}')\,(\hat{\Omega}_{0\mathrm{T}}^{-1} \otimes \mathrm{Id_T})\,(\mathrm{Id}_n \otimes \mathrm{X}) = -(\hat{\Omega}_{0\mathrm{T}}^{-1} \otimes \mathrm{X}'\mathrm{X}).$$

On a donc :

$$\xi_\mathrm{S} = [y - (\mathrm{Id}_n \otimes \mathrm{X})\,\hat{b}_{0\mathrm{T}}]'\,(\hat{\Omega}_{0\mathrm{T}}^{-1} \otimes \mathrm{P_X})\,[y - (\mathrm{Id}_n \otimes \mathrm{X})\hat{\mathrm{b}}_{0\mathrm{T}}],$$

avec $\mathrm{P_X} = \mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\,\mathrm{X}'$.

On en déduit :

$$\xi_\mathrm{S} = \mathrm{Tr}\,[\hat{\Omega}_{0\mathrm{T}}^{-1}\,\hat{\mathrm{U}}'_0\,\mathrm{P_X}\,\hat{\mathrm{U}}_0],$$

où $\hat{\mathrm{U}}_0$ est la matrice (T, n) dont les lignes sont $(\mathrm{Y}_t - \hat{\mathrm{B}}_{0\mathrm{T}}\,\mathrm{X}_t)'$ et les colonnes sont $\tilde{u}_{i0} = \tilde{y}_i - \mathrm{X}\hat{b}_{i0}$ ($\hat{y}_i$ étant le vecteur de taille T correspondant aux observations de la ième variable endogène et $\hat{b}_{i0}$ l'estimateur sous H_0 des paramètres de la ième équation).

Notons que la matrice $\hat{\mathrm{U}}'_0\,\mathrm{P_X}\,\hat{\mathrm{U}}_0$ s'écrit aussi :

$$\begin{aligned}\hat{\mathrm{U}}'_0\,\mathrm{P_X}\,\hat{\mathrm{U}}_0 &= \hat{\mathrm{U}}'_0\,\hat{\mathrm{U}}_0 - \hat{\mathrm{U}}'_0\,(\mathrm{Id_T} - \mathrm{P_X})\,\hat{\mathrm{U}}_0 \\ &= \hat{\mathrm{U}}_0\,\hat{\mathrm{U}}_0 - \hat{\mathrm{U}}'\,\hat{\mathrm{U}} = \mathrm{T}\,(\hat{\Omega}_{0\mathrm{T}} - \hat{\Omega}_\mathrm{T}),\end{aligned}$$

car les vecteurs colonnes de $\hat{\mathrm{U}} - \hat{\mathrm{U}}_0$ sont dans le sous-espace engendré par les colonnes de X.

Donc ξ_S s'écrit : $\xi_\mathrm{S} = \mathrm{T}\ \mathrm{Trace}\,[\hat{\Omega}_{0\mathrm{T}}^{-1}\,(\hat{\Omega}_{0\mathrm{T}} - \hat{\Omega}_\mathrm{T})]$.

(iv) **Test de Wald généralisé**

La statistique du test de Wald généralisé pour tester une hypothèse nulle quelconque portant sur b s'écrit :

$$\xi_\mathrm{W} = \underset{b \in \mathrm{H}_0}{\mathrm{Min}}\,(\hat{b}_\mathrm{T} - b)'\,[\hat{\Omega}_\mathrm{T} \otimes (\mathrm{X}'\mathrm{X})^{-1}]^{-1}\,(\hat{b}_\mathrm{T} - b).$$

Soit $\tilde{b}_{0\mathrm{T}}$ l'estimateur de b ainsi obtenu ; cet estimateur est asymptotiquement équivalent à l'estimateur du PMV contraint (compte tenu des égalités ci-dessous $I_{11}=J_{11}$, $J_{12}=0$, voir Gourieroux-Monfort [1989-b], propriété 10-64).

On a :

$$\begin{aligned}\xi_W &= (\hat{b}_T - \tilde{b}_{0T})' [\hat{\Omega}_T^{-1} \otimes X' X] (\hat{b}_T - \tilde{b}_{0T}) \\ &= \mathrm{T}r [\hat{\Omega}_T^{-1} (\hat{B}_T - \tilde{B}_{0T}) X' X (\hat{B}'_T - \tilde{B}'_{0T})] \\ &= \mathrm{T}r [\hat{\Omega}_T^{-1} (\tilde{U}_0 - \hat{U})' (\tilde{U}_0 - \hat{U})] \\ &= \mathrm{T}r[\hat{\Omega}_T^{-1} (\tilde{U}'_0 \tilde{U}_0 - \hat{U}' \hat{U})],\end{aligned}$$

car : $\hat{U}' \tilde{U}_0 = \hat{U}' (\hat{U} + \tilde{U}_0 - \hat{U}) = \hat{U}' [\hat{U} + X (\tilde{B}'_{0T} - \hat{B}'_T)] = \hat{U}' \hat{U}$.

Donc : $\xi_W = T \,\mathrm{Trace}\, [\hat{\Omega}_T^{-1} (\tilde{\Omega}_{0T} - \hat{\Omega}_T)]$,

avec : $\tilde{\Omega}_{0T} = \dfrac{1}{T} \tilde{U}'_0 \tilde{U}_0$.

(v) **Tests de Wald dans le cas d'une hypothèse linéaire sur Ker B**

Supposons que l'hypothèse nulle soit : $H_0^N = \{\exists \beta : B = \beta \alpha'_0\}$, où α_0 est une matrice (K, r) donnée. Cette contrainte peut s'écrire aussi : $H_0^N = \{\exists \beta : b = (\mathrm{Id}_n \otimes \alpha_0) \,\mathrm{vec}\, \beta'\}$. La statistique du test de Wald généralisée est alors :

$$\begin{aligned}\xi_W = \underset{\beta}{\mathrm{Min}}\, &[\hat{b}_T - (\mathrm{Id}_n \otimes \alpha_0) \,\mathrm{vec}\, \beta']' [\hat{\Omega}_T^{-1} \otimes X' X] \\ &[\hat{b}_T - (\mathrm{Id}_n \otimes \alpha_0) \,\mathrm{vec}\, \beta'].\end{aligned}$$

Le minimum en β est atteint pour :

$$\begin{aligned}\hat{\mathrm{vec}}\, \beta &= \{[\mathrm{Id}_n \otimes \alpha'_0] [\hat{\Omega}_T^{-1} \otimes (X' X)] [\mathrm{Id}_n \otimes \alpha_0]\}^{-1} \\ &\quad \times [\mathrm{Id}_n \otimes \alpha'_0] [\hat{\Omega}_T^{-1} \otimes (X' X) \hat{b}_T \\ &= \{[\mathrm{Id}_n \otimes (\alpha'_0 X' X \alpha_0)^{-1} \alpha'_0 X'\} \{\mathrm{Id}_n \otimes X\} \hat{b}_T \\ &= \{[\mathrm{Id}_n \otimes (\alpha'_0 X' X \alpha_0)^{-1} \alpha'_0 X'\} \tilde{y}.\end{aligned}$$

Donc $\hat{\mathrm{vec}}\beta$ est l'estimateur des m.c.o. équation par équation

$$y_i = X \alpha_0 \beta'_i + \tilde{u}_i,$$

où β_i est la ième ligne de β.

Il y a ici égalité numérique avec l'estimateur du PMV contraint et donc d'après le paragraphe précédent :

$$\xi_W = T \,\mathrm{Trace}\, [\hat{\Omega}_T^{-1} (\hat{\Omega}_{0T} - \hat{\Omega}_T)],$$

où $\hat{\Omega}_{0T}$ est l'estimateur de Ω obtenu à partir des résidus des m.c.o. sous H_0.

Il est aussi possible de fonder la statistique de Wald sur la forme implicite de l'hypothèse nulle :

$$B \gamma_0 = 0 \quad \text{ou} \quad (\mathrm{Id}_n \otimes \gamma'_0) b = 0,$$

où γ_0 est une matrice (K, K$-r$) dont les vecteurs colonnes sont orthogonaux à ceux de α_0. On montre que cette version de la statistique de Wald est identique à la précédente. En effet :

$$\begin{aligned}
\xi_{\mathrm{W}} &= [(\mathrm{Id}_n \otimes \gamma_0')\,\hat{b}_{\mathrm{T}}]'\,[\hat{\mathrm{V}}\,[\mathrm{Id}_n \otimes \gamma_0']\,\hat{b}_{\mathrm{T}}]^{-1}\,[(\mathrm{Id}_n \otimes \gamma_0')\,\hat{b}_{\mathrm{T}}] \\
&= y'\,[\mathrm{Id}_n \otimes \mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\,\gamma_0]\,[\hat{\Omega}_{\mathrm{T}} \otimes \gamma_0'\,(\mathrm{X}'\mathrm{X})^{-1}\,\gamma_0]^{-1} \\
&\quad \times [\mathrm{Id}_n \otimes \gamma_0'(\mathrm{X}'\mathrm{X})^{-1}\,\mathrm{X}']\,y \\
&= y'\,[\hat{\Omega}_{\mathrm{T}}^{-1} \otimes \mathrm{P}_{\mathrm{X}\,(\mathrm{X}'\mathrm{X})^{-1}\,\gamma_0}]\,y \\
&= y'\,[\hat{\Omega}_{\mathrm{T}}^{-1} \otimes (\mathrm{P}_{\mathrm{X}} - \mathrm{P}_{\mathrm{X}\,\alpha_0})]\,y \\
&= \mathrm{T}\ \mathrm{Trace}\,[\hat{\Omega}_{\mathrm{T}}^{-1}\,(\hat{\Omega}_{0\mathrm{T}} - \hat{\Omega}_{\mathrm{T}})].
\end{aligned}$$

LEMME 13 : Soit U une matrice de taille (n, p) et P un projecteur orthogonal de $\mathbb{R}^n$ sur un sous-espace vectoriel E. On suppose que la matrice $U'(\mathrm{Id} - P)U$ est inversible et que E est la somme de deux sous-espaces E_1, E_2 orthogonaux admettant respectivement comme bases les vecteurs colonnes des matrices V_1 et V_2. On note P_1 et P_2 les projecteurs orthogonaux sur E_1 et E_2.

Les valeurs propres de $A_1 = [U'(\mathrm{Id} - P)U]^{-1} U' P_1 U$ sont égales à $\nu_i/(1 - \nu_i)$, $i = 1, \cdots, n$, où les ν_i sont les valeurs propres de $A_2 = [U'(\mathrm{Id} - P_2)U]^{-1} U' P_1 U$, matrice de corrélation canonique empirique entre U et V_1 conditionnellement à V_2. En outre les vecteurs propres de A_2 et A_1 associés à ν_i et $\nu_i/(1 - \nu_i)$ sont les mêmes.

Preuve

Soit e_i un vecteur propre de $A_1 = [U'(\mathrm{Id} - P)U]^{-1} U' P_1 U$ associé à la valeur μ_i, on a :

$$[U'(\mathrm{Id} - P)U]^{-1} U' P_1 U e_i = \mu_i e_i,$$

ou :

$$U' P_1 e_i = \mu_i U'(\mathrm{Id} - P) U e_i.$$

En remplaçant P par $P_1 + P_2$, on obtient une écriture équivalente :

$$U' P_1 U e_i = \mu_i [U'(\mathrm{Id} - P_2)U - U' P_1 U] e_i,$$

ou :

$$A_2 e_i = \mu_i (\mathrm{Id} - A_2) e_i,$$

ou :

$$A_2 e_i = \frac{\mu_i}{1 + \mu_i} e_i.$$

Donc si on note ν_i les valeurs propres (positives et inférieures à 1) de A_2, on a $\mu_i = \nu_i/(1 - \nu_i)$. Notons aussi que comme la fonction $\nu \to \nu/(1 - \nu)$ est croissante, si les ν_i sont classées par ordre décroissant, il en est de même des μ_i. Par ailleurs, e_i est vecteur propre de A_1 pour la valeur propre $\nu_i/(1 - \nu_i)$ si et seulement s'il est vecteur propre de A_2 associé à ν_i. □

Résolution des optimisations par maximum de vraisemblance

(i) Deux lemmes préliminaires

Nous utiliserons pour la démonstration deux lemmes déjà utilisés dans des cas analogues par Anderson (1984), et Johansen (1988). Nous considérons une matrice symétrique positive admettant la décomposition par blocs :

$$\begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix},$$

où S_{00}, S_{01}, S_{11} sont respectivement de tailles (n, n), (n, K), (K, K). Nous introduisons les valeurs propres de la matrice $S_{11}^{-1} S_{10} S_{00}^{-1} S_{01}$ qui sont réelles positives inférieures à 1 et peuvent être rangées par ordre décroissant ; elles sont notées

$$1 \geq \eta_1 \geq \cdots \geq \eta_K \geq 0.$$

Nous introduisons aussi une base de vecteurs propres associés $e_1, \cdots, e_K$,, normalisée par :

$$e'_k S_{11} e_l = \delta_{kl} \quad k, l = 1, \cdots, K.$$

Nous nous intéressons alors au problème d'optimisation :

$$\operatorname*{Min}_{\alpha} \frac{\det(S_{00} - S_{01} \alpha(\alpha' S_{11} \alpha)^{-1} \alpha' S_{10})}{\det S_{00}},$$

par rapport aux matrices α, de taille (K, r) donnée. Notons que la fonction objectif n'est pas modifiée, si on remplace α par αQ où Q est une matrice (r, r) inversible.

LEMME 14 :

$$\frac{\det(S_{00} - S_{01} \alpha(\alpha' S_{11} \alpha)^{-1} \alpha' S_{10})}{\det S_{00}} = \frac{\det(\alpha' S_{11} \alpha - \alpha' S_{10} S_{00}^{-1} S_{01} \alpha)}{\det(\alpha' S_{11} \alpha)}$$

LEMME 15 : (i) une solution en α du problème d'optimisation est :

$$\alpha_1 = e_1, \cdots, \alpha_r = e_r.$$

(ii) La valeur à l'optimum de la fonction objectif est :

$$\prod_{i=1}^{r} (1 - \eta_i)$$

Le problème peut être résolu de façon symétrique, si on s'intéresse à une maximisation en α de la fonction objectif.

(ii) **Résolution du problème** $\xi^r_{\mathrm{RMV}} = \underset{\gamma}{\mathrm{Min}}\ \xi_{\mathrm{RMV}}(\gamma)$.

Il faut minimiser la quantité :

$$\begin{aligned}
\xi_{\mathrm{RMV}}(\gamma) &= \mathrm{T\,Log}\,\frac{\det \frac{1}{\mathrm{T}} \mathrm{Y}'(\mathrm{Id} - \mathrm{P}_{\mathrm{X}\alpha})\mathrm{Y}}{\det \frac{1}{\mathrm{T}} \mathrm{Y}'(\mathrm{Id} - \mathrm{P}_{\mathrm{X}})\mathrm{Y}} \\
&= \mathrm{T\,Log}\,\frac{\det \frac{1}{\mathrm{T}} \mathrm{Y}'\,\mathrm{Y}}{\det \frac{1}{\mathrm{T}} \mathrm{Y}'(\mathrm{Id} - \mathrm{P}_{\mathrm{X}})\mathrm{Y}} \\
&\quad + \mathrm{T\,Log}\,\frac{\det \frac{1}{\mathrm{T}} \mathrm{Y}'(\mathrm{Id} - \mathrm{P}_{\mathrm{X}\alpha})\mathrm{Y}}{\det \frac{1}{\mathrm{T}} \mathrm{Y}'\,\mathrm{Y}} \\
&= -\mathrm{T\,Log}\,\det(\mathrm{Id} - (\mathrm{Y}'\mathrm{Y})^{-1}\mathrm{Y}'\mathrm{P}_{\mathrm{X}}\mathrm{Y}) \\
&\quad + \mathrm{T\,Log}\,\det(\mathrm{Id} - (\mathrm{Y}'\mathrm{Y})^{-1}\mathrm{Y}'\mathrm{P}_{\mathrm{X}\alpha}\mathrm{Y}) \\
&= -\mathrm{T\,Log}\,\det(\mathrm{Id} - \mathrm{R}^2\,[\mathrm{Y}, \mathrm{X}]) \\
&\quad + \mathrm{T\,Log}\,\det(\mathrm{Id} - \mathrm{R}^2\,(\mathrm{Y}, \mathrm{X}\alpha)).
\end{aligned}$$

Ainsi nous sommes conduits à optimiser par rapport à une matrice α dont la connaissance de l'image équivaut à celle de γ.

La quantité à minimiser est :

$$\det[\mathrm{Id} - \mathrm{R}^2(\mathrm{Y}, \mathrm{X}\alpha)] = \det[\mathrm{Id} - (\mathrm{Y}'\mathrm{Y})^{-1}\mathrm{Y}'\mathrm{X}\alpha(\alpha'\mathrm{X}'\mathrm{X}\alpha)^{-1}\alpha'\mathrm{X}'\mathrm{Y}].$$

Appliquant le lemme 15, nous sommes conduits à introduire la matrice $\mathrm{R}^2(\mathrm{X}, \mathrm{Y}) = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y}(\mathrm{Y}'\mathrm{Y})^{-1}\mathrm{Y}'\mathrm{X}$, à déterminer ses valeurs propres $\hat{\eta}_{1\mathrm{T}} \geq \hat{\eta}_{2\mathrm{T}} \cdots \geq \hat{\eta}_{\mathrm{KT}}$ et des vecteurs propres associés $\hat{\mathrm{e}}_{1\mathrm{T}} \cdots \hat{\mathrm{e}}_{\mathrm{KT}}$ normalisés par :

$$\hat{e}'_{\mathrm{KT}}\,\mathrm{X}'\mathrm{X}\,\hat{e}_{l\,\mathrm{T}} = \delta_{kl}.$$

Une solution en α du problème est alors :

$$\hat{\alpha}_{1\mathrm{T}} = \hat{e}_{1\mathrm{T}} \cdots,\ \hat{\alpha}_{r\,\mathrm{T}} = \hat{e}_{r\,\mathrm{T}}.$$

Cette solution est donc obtenue en prenant les r vecteurs propres associés aux r plus grandes valeurs propres. La statistique de test est alors donnée par :

$$\begin{aligned}
\xi^r_{\mathrm{RMV}} &= -\mathrm{T}\sum_{i=1}^{\mathrm{K}} \mathrm{Log}(1 - \hat{\eta}_{i\,\mathrm{T}}) + \mathrm{T}\sum_{i=1}^{r} \mathrm{Log}(1 - \hat{\eta}_{i\,\mathrm{T}}) \\
&= -\mathrm{T}\sum_{i=r+1}^{\mathrm{K}} \mathrm{Log}(1 - \hat{\eta}_{i\,\mathrm{T}}).
\end{aligned}$$

(iii) **Résolution du problème** $\xi^r_{\mathrm{RMV}} = \underset{\gamma}{\mathrm{Min}}\ \xi^*_{\mathrm{RMV}}(\delta)$.

La quantité à minimiser est :

$$\xi^{*}_{\mathrm{RMV}}(\delta) = \mathrm{T}\,\mathrm{Log}\,\frac{\det(\delta'\,\mathrm{Y}'\,\mathrm{Y}\,\delta)}{\det(\delta'\,\mathrm{Y}'\,(\mathrm{Id}-\mathrm{P_X})\,\mathrm{Y}\,\delta)}$$

$$= \mathrm{T}\,\mathrm{Log}\,\frac{\det(\delta'\,\mathrm{Y}'\,\mathrm{Y}\,\delta)}{\det[\delta'\,\mathrm{Y}'\,\mathrm{Y}\,\delta - \delta'\,\mathrm{Y}'\,\mathrm{X}\,(\mathrm{X}'\,\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}\,\delta]}.$$

Appliquant les lemmes 14-15, nous sommes amenés à introduire la matrice :

$$\mathrm{R}^2(\mathrm{Y},\mathrm{X}) = (\mathrm{Y}'\,\mathrm{Y})^{-1}\,\mathrm{Y}'\,\mathrm{X}\,(\mathrm{X}'\,\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y},$$

et à retenir comme estimateurs des vecteurs δ_i, $i = 1, \cdots, n-r$ par exemple les vecteurs : $\hat{f}_{r+1,\mathrm{T}}, \cdots, \hat{f}_{n,\mathrm{T}}$ associés aux $n-r$ plus petites valeurs propres. La valeur associée de la fonction objectif est la même que précédemment :

$$\xi^{r}_{\mathrm{RMV}} = \mathrm{T}\,\mathrm{Log}\,\frac{1}{\sum\limits_{i=r+1}^{n}(1-\hat{\eta}_{i\mathrm{T}})} = -\mathrm{T}\sum_{i=r+1}^{n}\mathrm{Log}\,(1-\hat{\eta}_{i\mathrm{T}})$$

$$= -\mathrm{T}\sum_{i=r+1}^{\mathrm{K}}\mathrm{Log}\,(1-\hat{\eta}_{i\mathrm{T}}).$$

Détermination de la statistique de Wald et des estimateurs des moindres carrés asymptotiques

(i) Il faut résoudre le problème :

$$\begin{aligned} & \operatorname*{Max}_{\alpha} \operatorname{T}r\,(\hat\Omega_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{P}_{\mathrm{X}\alpha}\,\mathrm{Y}) \\ \Leftrightarrow\quad & \operatorname*{Max}_{\alpha} \operatorname{T}r\,(\hat\Omega_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{X}\,\alpha(\alpha'\,\mathrm{X}'\,\mathrm{X}\,\alpha)^{-1}\,\alpha'\,\mathrm{X}'\,\mathrm{Y}) \\ \Leftrightarrow\quad & \operatorname*{Max}_{\alpha} \operatorname{T}r\,[\alpha'\,\mathrm{X}'\,\mathrm{Y}\hat\Omega_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{X}\,\alpha(\alpha'\,\mathrm{X}'\,\mathrm{X}\,\alpha)^{-1}]. \end{aligned}$$

Les vecteurs colonnes de α étant définis à un changement de base près nous pouvons toujours imposer la contrainte de normalisation :

$$\alpha'\,\mathrm{X}'\,\mathrm{X}\,\alpha = \mathrm{Id}_r$$

Le problème devient alors :

$$\begin{cases} \operatorname*{Max}_{\alpha} \operatorname{T}r\,(\alpha'\,\mathrm{X}'\,\mathrm{Y}\,\hat\Omega_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{X}\,\alpha) \\ \text{sous } \alpha'\,\mathrm{X}'\,\mathrm{X}\,\alpha = \mathrm{Id}_r. \end{cases}$$

Il s'agit d'un problème dont la solution est obtenue en cherchant les valeurs propres $\hat\mu_{1\mathrm{T}} \geq \cdots \geq \hat\mu_{\mathrm{KT}}$ de la matrice $\mathrm{Q} = \frac{1}{\mathrm{T}}\,(\mathrm{X}'\,\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}\,\hat\Omega_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{X}$ et en explicitant une base de vecteurs propres associés : $\tilde e_{1\mathrm{T}}, \cdots, \tilde e_{\mathrm{KT}}$, qui peuvent être normalisés par :

$$\tilde e_{k\,\mathrm{T}}\,\mathrm{X}'\,\mathrm{X}\,\tilde e_{l\,\mathrm{T}} = \delta_{kl},\ k,\ l = 1, \cdots, \mathrm{K}.$$

La solution est obtenue en prenant :

$$\hat\alpha_{1\mathrm{T}} = \tilde e_{1\mathrm{T}}, \cdots, \hat\alpha_{r\,\mathrm{T}} = \tilde e_{r\,\mathrm{T}},$$

et la valeur à l'optimum de la fonction objectif est :

$$\mathrm{T}\sum_{j=1}^{r} \hat\mu_{j\,\mathrm{T}}.$$

Par conséquent la statistique ξ_{W}^{r} est égale à $\mathrm{T}\sum_{j=r+1}^{\mathrm{K}} \hat\mu_{j\,\mathrm{T}}$.

(ii) Il reste à lier le résultat précédent à la procédure d'analyse canonique. Pour cela nous pouvons remarquer que la matrice Q a mêmes valeurs propres que la matrice : $\mathrm{Q}^* = \frac{1}{\mathrm{T}}\,\hat\Omega_{\mathrm{T}}^{-1}\,\mathrm{Y}'\,\mathrm{X}\,(\mathrm{X}'\,\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}$, et, si on note $e_{j\,\mathrm{T}}^{*}$ des vecteurs propres de Q*, correspondant aux valeurs propres $\hat\mu_{j\,\mathrm{T}} > 0$ et normalisés par : $e_{j\,\mathrm{T}}^{*\prime}\,\mathrm{Y}'\,\mathrm{X}\,(\mathrm{X}'\,\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}\,e_{k\,\mathrm{T}}^{*} = \delta_{jk}$, on peut prendre : $\tilde e_{j\,\mathrm{T}} = (\mathrm{X}'\,\mathrm{X})^{-1}\,\mathrm{X}'\,\mathrm{Y}\,e_{j\,\mathrm{T}}^{*}$.

En appliquant le lemme 13 de l'annexe 2 (avec $\mathrm{U} = \mathrm{Y}$, $\mathrm{P} = \mathrm{P}_{\mathrm{X}}$, $\mathrm{P}_1 = \mathrm{P}_{\mathrm{X}}$, $\mathrm{P}_2 = 0$) on obtient que les valeurs propres $\hat\mu_{j\,\mathrm{T}}$ sont égales à $\hat\eta_{j\,\mathrm{T}}/(1-\hat\eta_{j\,\mathrm{T}})$, où les $\hat\eta_{j\,\mathrm{T}}$ sont les corrélations canoniques entre Y et X.

Par ailleurs, le lemme montre aussi que les $e^*_{j\,\mathrm{T}}$ sont les vecteurs propres de R^2 (Y, X) ; on en déduit que les $\tilde{e}_{j\,\mathrm{T}}$ sont vecteurs propres de R^2 (X, Y).

La statistique ξ^r_{W} est donc égale à $\mathrm{T} \sum_{j=r+1}^{\mathrm{K}} \hat{\eta}_{j\,\mathrm{T}}/(1 - \hat{\eta}_{j\,\mathrm{T}})$, tandis que les colonnes de la matrice α sont estimées par les vecteurs propres $\tilde{e}_{1\mathrm{T}}, \cdots, \tilde{e}_{r\,\mathrm{T}}$ de R^2 (X, Y), satisfaisant les conditions de normalisation $\tilde{e}_{k\,\mathrm{T}}\,\mathrm{X}'\,\mathrm{X}\,\tilde{e}_{l\,\mathrm{T}} = \delta_{kl}$.

(iii) Cherchons maintenant l'estimateur des moindres carrés asymptotiques de $\boldsymbol{\beta}$. Nous savons (voir annexe 1.v) que :

$$
\begin{aligned}
\mathrm{v\hat{e}c}\,\beta' &= \{\mathrm{Id}_n \otimes (\hat{\alpha}'\,\mathrm{X}'\,\mathrm{X}\,\hat{\alpha})^{-1}\,\hat{\alpha}'\,\mathrm{X}'\}\,y \\
&= \{\mathrm{Id}_n \otimes (\hat{\alpha}'\,\mathrm{X}'\}\,y \quad \text{(à cause de la condition de normalisation)} \\
&\Leftrightarrow \quad \hat{\beta} = \mathrm{Y}'\,\mathrm{X}\,\hat{\alpha}. \\
&(\mathrm{Y}'\,\mathrm{Y})^{-1}\,\hat{\beta} = (\mathrm{Y}'\,\mathrm{Y})^{-1}\,\mathrm{Y}'\,\mathrm{X}\,\hat{\alpha}.
\end{aligned}
$$

Or $\hat{\alpha}_{j\,\mathrm{T}}$ est vecteur propre de R^2 (X, Y), ce qui entraîne que $(\mathrm{Y}'\,\mathrm{Y})^{-1}\,\mathrm{Y}'\,\mathrm{X}\,\hat{\alpha}_{j\mathrm{T}}$ est vecteur propre de R^2 (Y, X) associé à la même valeur propre. Comme de plus la condition de normalisation est satisfaite, nous en déduisons que : $(\mathrm{Y}'\,\mathrm{Y})^{-1}\,\hat{\beta}_{j\,\mathrm{T}} = \hat{f}_{j\,\mathrm{T}} \quad \Leftrightarrow \quad \hat{\beta}_{j\,\mathrm{T}} = \mathrm{Y}'\,\mathrm{Y}\,\hat{f}_{j\,\mathrm{T}}$

• Références bibliographiques

AHN, S., REINSEL, G. (1988). – "Nested Reduced Rank Autoregressive Models for Multiple Time Series", *JASA 83*, pp. 849-856.

ANDERSON, T. W., RUBIN, H. (1949). – "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, 20, pp. 46-63.

ANDERSON, T. W. (1951). – "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions", *Annals of Mathematical Statistics*, 22, pp. 327-351.

ANDERSON, T. W., RUBIN H. (1956). – "Statistical Inference in Factor Analysis", in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Ed. J. Neyman, University of California Press, 5, pp. 111-150.

ANDERSON, T. W. (1963a). – "Asymptotic Theory for Principal Component Analysis", *Annals of Mathematical Statistics*, 34, pp. 122-148.

ANDERSON, T. W. (1963b). – "The Use of Factor Analysis in the Statistical Analysis of Multiple Time Series", *Psychometrika*, 28, pp. 1-25.

ANDERSON, T. W. (1984a). – "An Introduction to Multivariate Statistical Analysis", New York, Wiley.

ANDERSON, T. W. (1984b). – "Estimating Linear Statistical Relationship", *The Annals of Statistics*, 12, pp. 1-45.

BHARGAVA, A. (1991). – "Identification and Panel Data Models with Endogenous Regressions", *Review of Economic Studies*, 58, pp. 129-140.

BOX, G., TIAO, G. (1977). – "A Canonical Analysis of Multiple Time Series", *Biometrika*, 64. pp. 355-365.

BROWN, S. (1989). – "The Number of Factors in Security Returns", *Journal of Finance*, 44, pp. 1247-1262.

CHAMBERLAIN, G., ROTHSCHILD, M. (1983). – "Arbitrage and Mean Variance Analysis in Large Asset Markets", *Econometrica*, 51, pp. 1281-1304.

DAGENAIS, M., DUFOUR, J. M. (1991). – "Invariance, Nonlinear Models and Asymptotic Tests", *Econometrica*, forthcoming.

DAVIDSON, J., HENDRY, D., SRBA, F., YEO, S. (1978). – "Econometric Modelling of the Aggregate Time Series Relationship Between Consumer's Expenditure and Income in the United Kingdom", *Economic Journal*, 88, pp. 661-692.

ENGLE, R., GRANGER, C. (1987). – "Cointegration and Error Correction : Representation, Estimation and Testing", *Econometrica*, 55, pp. 251-276.

GEARY, R. C. (1949). – "Determination of Linear Relations Between Economic Time Series", *Econometrica*, 17, pp. 30-58.

GILL, L., LEWBEL, A. (1991). – "Testing the Rank and Definiteness of Estimated Matrices with Applications to Factors, State Space and ARMA Models", *Manchester University D.P.*

GOURIEROUX, C., MONFORT, A. (1989a). – "A General Framework for Testing a Null Hypothesis in a Mixed Form", *Econometric Theory*, 5, pp. 63-82.

GOURIEROUX, C., MONFORT, A. (1989b). – "Statistique et modèles économétriques", *Economica* (2 volumes).

GOURIEROUX, C., PEAUCELLE, I. (1993). – "Séries codépendantes : Application à l'Hypothèse de Parité du Pouvoir d'Achat", dans Macroeconomie Developpements Récents, *Economica*, pp. 285-306.

GOURIEROUX, C., MONFORT, A., RENAULT, E. (1991). – "A General Framework for Factor Models", Document de Travail ENSAE/INSEE n° 9107.

GREGOIR, S., LAROQUE, G. (1990). – "Multivariate Integrated Time Series : A General Error Correction Representation with Associated Estimation and Test Procedures", *mimeo*.

HSU, P. (1941). – "Canonical Reduction of the General Regression Problem", *Ann. Eugen.*, 11, pp. 42-46.

JOHANSEN, S. (1988). – "Statistical Analysis of Cointegration Vector", *Journal of Economics Dynamics and Control*, 12, 2/3, pp. 231-254.

KUGLER, P., NEUSSER, K. (1993). – "International Real Interest Rate Equalization", *Journal of Applied Econometrics*, 8, pp. 163-174.

LAWLEY, D., MAXWELL, A. (1971). – "Factor Analysis as a Statistical Method", Londres, Butterworth.

MUIRHEAD, R. (1982). – "Aspects of Multivariate Statistical Theory", Wiley.

RAO, C. R. (1973). – "Linear Statistical Inference and Its Applications", Wiley, New York.

REINSEL, G. C. (1983). – "Some Results on Multivariate Autoregressive Index Models", *Biometrika*, pp. 145-156.

SARGENT, T., SIMS, C. (1977). – "Business Cycle Modeling Without Pretending to Have Too Much a Priori Economic Theory", dans "New Methods in Business Cycle Research", Federal Reserve Bank of Minneapolis.

SZROETER, J. (1983). – "Generalized Wald and Methods for Testing Nonlinear Implicit and Overidentifying Restrictions", *Econometrica*, 51, pp. 335-353.

TIAO, G., TSAY, R. (1989). – "Model Specification in Multivariate Time Series", JRSS B, 51, pp. 157-213.

TINTNER, G (1950). – "A Test for Linear Relations Between Weighted Regression Coefficients", JRSS B, 12, pp. 273-277.

TSAY, R., TIAO, G. (1985). – "Use of Canonical Analysis in Time Series Model Identification", *Biometrika*, 72, pp. 299-315.

VELU, R., REINSEL, G., WICHERN, D. (1986). – "Reduced Rank Models for Multiple Time Series", *Biometrika*, 73, pp. 105-118.

WILLIAMS, E. J. (1955). – "Significance Tests for Discriminant Functions and Linear Functional Relationships", *Biometrika*, 42, pp. 360-381.